

Supplement to Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech

Matthew Gentzkow, *Stanford University and NBER**

Jesse M. Shapiro, *Brown University and NBER*

Matt Taddy, *Amazon*

March 2019

Online Appendix

Contents

A	Audit of Congressional Record Data	2
B	Filtering Phrases	3
C	Discussion of Partisan Phrases	5

List of Tables

1	Summary Statistics by Session	9
2	Details on Topic Classification	12
3	Most Partisan Phrases by Session	14
4	Manually Selected Words Used to Exclude Phrases as Procedural	20

List of Figures

1	Average Partisanship of Speech Estimated on Alternative Samples	21
2	Average Verbosity per Speaker-Session	22
3	Different Target Measures of Party Differences in Speech	23
4	Average Partisanship of Speech with Different Covariate Designs	24
5	Parametric Bootstrap of Leave-out and Penalized Estimators	25

*E-mail: gentzkow@stanford.edu, jesse_shapiro_1@brown.edu, mataddy@amazon.com.

6	Average Partisanship of Speech Using Alternative Estimation Settings	26
7	Average Partisanship of Speech, Bias-Corrected Penalized Estimates	27
8	Average Partisanship of Speech, Alternative Subsampling Designs	28
9	Sample-splitting Inference for Penalized Estimator	29
10	Estimated vs. True Average Partisanship, Alternative Estimators	30
11	Isolating Changes in Average Partisanship Due to Changes in Verbosity	31
12	Approximate Bias in Maximum Likelihood Estimate of Partisanship	32
13	Maximum Likelihood Estimate from Data in Jensen et al. (2012)	33
14	Maximum Likelihood Estimates with Frequently Occurring Phrases	34
15	Average Partisanship of Speech, Minimal Penalty ($\lambda \approx 0$)	35
16	Average Partisanship of Speech Estimated on Alternative Time Frames	36
17	Partisanship over Time for Phrases within Topics	37
18	Discontinuity in Democratic Vote Share from Lee et al. (2004)	39
19	Average Partisanship of Speech between and within Parties	40
20	Distribution of Phrase Partisanship	41
21	Evidence on the Role of Neologisms	42
22	Plug-in Estimates of Average Partisanship for Different Choice Dimensions	43

A Audit of Congressional Record Data

We conduct an audit to check the accuracy of our algorithms for parsing the *Record* and matching speeches to speakers. We manually parse text selected by a multistage sampling procedure that chooses 2539 (319) speeches from the bound (daily) edition. The procedure randomly selects a two-year period among the first 10 (5) periods of the bound (daily) edition and randomly samples four days for inspection. It then samples four days from every subsequent tenth (fifth) period, until no more remain. On each sampled day, we select 1,000 contiguous lines of text by randomly choosing a start line and the following 999 lines of text. We require start lines to be outside the last 1,000 lines of text if there are more than 1,000 lines in a day, and we choose all the lines in a day if there are 1,000 or fewer.

We manually record each speaker’s name, chamber (including a flag for Extensions of Remarks), and whether the speaker is referred to by a title, as well as the starting and ending lines of each speech. We reduce the number of human errors in the manual parsing by double-checking the raw text whenever there is a discrepancy between our work and the automated parser and correcting any obvious errors of our own. We compare a speaker’s name between manual and automated parsings after removing all whitespace and punctuation and coercing all characters to lower case.

The automated parser recognizes the beginning of a speech in the same location as our manual parsing in 94 (99.7) percent of speeches in the respective editions. Of speeches with the same starting location not made by special speakers, the parsings agree on the exact length of a speech in 88 (77) percent of cases and differ by two lines or fewer in 94 (92) percent of cases. Speeches with the same starting location also agree on the chamber in 99 (98) percent of cases and on the speaker’s name in 98 (92) percent.

Our audit of the speaker mapping covers 138 (36) speeches from the bound (daily) edition of the *Record*. The speeches are chosen by a stratified sampling procedure that randomly selects two speeches from the automated parsing at the session-edition level. We treat the two audits sequentially: conditional on the set of automatically parsed speeches from the *Record*, we measure how well our algorithm matches speeches to members of Congress.

The automated and manual mappings agree in all but three cases for the bound edition and in all cases for the daily edition. The three disagreements arose when the manual mapping was made on contextual information in the *Record* that our parsing algorithm was not designed to recognize.

B Filtering Phrases

Procedural or parliamentary language appears frequently in Congress, and may be used asymmetrically by parties (e.g., depending on whether they are in the majority or the minority). Because this language is distinct from the substantive partisanship of language we seek to identify, we filter out procedural language to the extent possible prior to our analysis.

We start our identification of procedural phrases by obtaining an electronic copy of *Robert’s Rules of Order* (Robert 1876), a widely accepted manual that explains the procedures of assemblies.¹ We also obtain an electronic copy of the appendix of *Riddick’s Senate Procedure* (Riddick 1992) for the 101st session (1989–1991), a glossary-style document detailing the rules, practices, and customs of the United States Senate’s operations and meetings.² We consider each bigram appearing in the manuals to be procedural, calling them Robert and Riddick procedural phrases as appropriate. If a speech contains many procedural phrases, it is likely to be a procedural speech. Guided by this understanding, we identify additional procedural phrases

¹We downloaded a text version from Project Gutenberg <http://www.gutenberg.org/etext/9097> in early August 2009. It is the original 1876 version of the document.

²We downloaded a PDF version from <http://www.gpoaccess.gov/riddick/1441-1608.pdf> on August 11, 2010 and converted its contents to text using Optical Character Recognition with metadata removal.

using co-occurrence rules. First, we define a highly Robert (Riddick) speech as one for which Robert (Riddick) phrases account for at least 30 percent of all bigrams. We similarly define a procedural speech as one for which phrases appearing in either manual constitute at least 30 percent of bigrams.

We use two separate rules to identify procedural phrases not appearing in the manuals. A phrase qualifies as procedural by our first rule if one of the following sets of conditions applies:

- The phrase appears in at least 5 procedural speeches in more than 5 sessions and one of: 1) it appears in more than 5,200 highly Robert speeches, and at least 1.75 percent of speeches it appears in are highly Robert; 2) it appears in more than 100 highly Robert speeches, and at least 7.5 percent of speeches it appears in are highly Robert; or 3) it appears in more than 50 highly Robert speeches, and more than 30 percent of speeches it appears in are highly Robert.
- The phrase appears in at least 5 highly Robert speeches in more than 10 sessions and one of: 1) it appears in more than 2,000 highly Robert speeches, and at least 1 percent of speeches it appears in are highly Robert; 2) it appears in more than 100 highly Robert speeches, and at least 5 percent of speeches it appears in are highly Robert; or 3) it appears in more than 50 highly Robert speeches, and at least 20 percent of speeches it appears in are highly Robert.
- The phrase appears in at least 5 highly Riddick speeches in more than 10 sessions and one of: 1) it appears in at least 3,000 highly Riddick speeches, and at least 1.75 percent of speeches it appears in are highly Riddick; 2) it appears in at least 100 highly Riddick speeches, and at least 7 percent of speeches it appears in are highly Riddick; or 3) it appears in at least 50 highly Riddick speeches, and at least 20 percent of speeches it appears in are highly Riddick.

We compute, for every phrase, the average percentage of Robert's and of Riddick's procedural phrases across speeches containing the phrase. Of the phrases not identified by our first rule, a phrase qualifies as procedural by our second rule if one of the following sets of conditions applies:

- 1) It is mentioned over 500 times; and 2) it appears in more than 5 sessions; and 3) over 5 percent of bigrams in speeches in which it occurs, on average, are Robert phrases.

- 1) It is mentioned over 20,000 times; and 2) it appears in more than 10 sessions; and 3) over 7.5 percent of bigrams in speeches in which it occurs, on average, are Riddick phrases.
- 1) It is mentioned over 500 times; and 2) it appears in more than 10 sessions; and 3) over 9.6 percent of bigrams in speeches in which it occurs, on average, are Riddick phrases.

The cut-off points above are chosen to maximize the share of excluded phrases, and minimize the share of non-excluded phrases, that we judge subjectively to be procedural.

We also remove all phrases that include (i) a congressperson’s surname; (ii) the name of a state or month; (iii) numbers, or symbols; (iv) fewer than five characters, including the space; (v) a one-letter word; or (vi) any of the words from Online Appendix Table 4.

C Discussion of Partisan Phrases

Table 1 in the main paper lists the 10 most partisan phrases in every tenth session (along with the 114th session) according to our definition of phrase partisanship. Here we discuss the historical context of these phrases, and show that in each session they align closely with the policy positions and narrative strategies of the parties.

Our discussion draws mainly on the original congressional text and on the national party platforms (from Peters and Woolley 2016). We cite proceedings in Congress using the format “CR Date,” with a hyperlink to ProQuest Congressional, a gated service to which many universities subscribe.

Some phrases from the 50th session (1887-88) are discussed in the main text. In addition to those discussed there, the highly Republican phrase “sugar trust” is indicative of the Republican opposition “to all combinations of capital organized in trusts” and corresponds with the formation of the sugar trust in 1887 (Zerbe 1969).³ Controversies over the appropriation of land grants, particularly to railroad companies, are reflected in the partisanship of terms like “public domain” (Democratic) and “railroad compani” (Republican).⁴

The Republican Party platform of 1908 devotes a section to the need for “generous provision” for veterans, which is reflected in the highly Republican phrase “pay pension” in the 60th

³CR July 9, 1888 includes debate of a proposition by Nelson Dingley Jr. (R-ME) that would reduce the sugar tariff in an attempt to “strike down the sugar trusts.”

⁴Both party platforms mention the issue of the public domain and decry the behavior of the other party, but only the Republican platform explicitly focuses on “railroad land grants.”

session (1907-08).⁵ The highly Republican phrase “postal save” reflects the Republican Party platform’s support for a “postal savings bank system,” while the Democratic Party platform preferred the establishment of a national “guaranteed bank” that would provide “prompt payment of the depositors of any insolvent national bank” in response to the financial panic of 1907.⁶ The Democratic Party platform of that same year has as its key theme “to free the Government from the grip of those who have made it a business asset of the favor-seeking corporations.” Several Democratic phrases of the 60th session relate to this theme, for example “bureau corpor,” a reference to the Bureau of Corporations, the predecessor to the Federal Trade Commission, and “standard oil.” The Democratic platform also emphasizes trade and shipping issues, declaring support for the Panama Canal (“canal zone,” “panama canal”), and “demanding” the repeal of tariffs on several commodities (“revis tariff”).

In the 70th session (1927-1928), both party platforms discuss flood control and waterways with the Democratic platform additionally emphasizing waterpower; this emphasis reveals itself in highly partisan, water-related phrases for both Republicans (“cove creek,” “creek dam,” “muscl shoal,” “steam plant”)⁷ and Democrats (“flowag right,” “imperi valley”).⁸ Both parties also devote attention in their platforms to the nascent radio market and associated regulation; Democratic partisan phrases like “radio commiss” and “wave length” reflect discussions about regulating this market. Both party platforms include a section on “merchant marine,” and the highly Republican phrase “american ship” reflects the passage of the Republican-supported Merchant Marine Act of 1928.⁹

Phrases from the 80th session (1947-1948) are discussed in the main text. In addition to what is mentioned there, we note that both party platforms discuss support for agriculture, reflected in phrases like “depart agricultur” (Republican). Both party platforms also mention the public debt, which aligns with the partisan phrases “nation debt” (Republican) and “budget estim” (Democratic).

In the 90th session (1967-68), Congress discussed the UN Conventions on Forced Labor

⁵CR May 12, 1908 includes an extensive amendment specifying pension benefits to accrue to individuals, for example: “The name of Annie A. Robbins, late nurse, Medical Department, United States Army, war with Spain, and pay her a pension at the rate of \$12 per month.”

⁶In CR January 6, 1909, the Senate debated a bill to “establish postal savings banks for depositing savings at interest with the security of the Government for repayment thereof.”

⁷CR May 16, 1928 contains debate on a bill “for the construction of Cove Creek Dam” and CR March 12, 1928 contains discussion of joint resolution (S.J.Res.46) “providing for the completion of Dam No. 2 and the steam plant at nitrate plant No. 2 in the vicinity of Muscle Shoals for the manufacture and distribution of fertilizer, and for other purposes.”

⁸CR May 22, 1928 includes discussion on building “a canal from the Laguna Dam to Imperial Valley.”

⁹See <http://documents.law.yale.edu/jones-white-act-see-merchant-marine-act-1928> accessed on April 18, 2017.

and the Political Rights of Women (e.g., CR February 8, 1968), both of which connect to highly Democratic terms (“human right,” “unit nation,” “men women”). The Republicans, on the other hand, are marked by support for the Human Investment Act (“invest act,” “tax credit”), which provided tax credits to employers for training and hiring certain kinds of workers. Although both parties’ 1968 platforms emphasize transportation issues, the only transportation-related phrases in the top ten partisan phrases for either party are Republican (“federalaid highway,” “highway program”). The highly Democratic phrase “gun control” reflects the passage of the Gun Control Act of 1968; only the Democratic Party platform mentions gun control. Both party platforms mention pollution-related issues, but the only pollution-related phrase on the partisan phrase list (“air pollut”) is Democratic.

Partisan language in the 100th session (1987-88) centers around familiar Cold War themes. Both parties focus on the Iran-Contra scandal and the related conflict in Nicaragua. Democrats refer to the insurgents as “Contras” (“contra aid,” “aid contra”). Republicans instead call the insurgents “freedom fighter[s]” (e.g., CR March 17, 1988). Democrats criticize the Strategic Defense Initiative (“star war”) and have more foreign policy phrases on their list (“persian gulf,” “central american”). The Republican and Democratic phrase lists also reflect different domestic priorities. Republicans emphasize labor relations (“minimum wage”), tax policy (“tax increas”), and business conditions (“plant close”), while the Democrats focus on macroeconomic policies (“feder reserve,” “interest rate,” “presid budget”). The most Republican phrase (“judg bork”) reflects the nomination of Robert Bork to the Supreme Court by Ronald Reagan; Bork failed to be confirmed by the Senate.¹⁰ Republicans in this session are also associated with some procedural language (“demand second,” “reserv object”).¹¹

Language in the 110th (2007-2008) and 114th (2015-2016) sessions are discussed in the main text.

References

Peters, G., and J. T. Woolley (2016): Political party platforms of parties receiving electoral votes. *American Presidency Project*. Accessed at <http://www.presidency.ucsb.edu/platforms.php> on June 16, 2016.

¹⁰See <https://constitutioncenter.org/blog/on-this-day-senate-rejects-robert-bork-for-the-supreme-court/> accessed on April 18, 2017.

¹¹For “I demand a second,” see CR October 3, 1988. For “I withdraw my reservation of objection,” see CR October 21, 1988.

- ProQuest Congressional (2016): Legislative and executive publications. Accessed at <<http://congressional.proquest.com/congressional/search/basic/basicsearch>> on June 16, 2016.
- Riddick, F. (1992): Riddick's Senate procedure: Precedents and practices. Accessed at <<http://www.gpoaccess.gov/riddick/1441-1608.pdf>> on August 11, 2010.
- Robert, H. M. (1876): *Robert's rules of order*. Chicago, IL: S. C. Griggs & Company.
- Taddy, M. (2016): One-step estimator paths for concave regularization. *Journal of Computational and Graphical Statistics*.
- Zerbe, R. (1969): The American sugar refinery company, 1887–1914: The story of a monopoly. *The Journal of Law & Economics* 12(2): 339–375.

Online Appendix Table 1: Summary Statistics by Session

Source	Session	Unique speakers	Unique phrases	Total phrases		Total speeches	Matched speeches	Match rate
				Republicans	Democrats			
Bound	43	359	210289	751168	279903	101060	96115	0.95
Bound	44	363	212858	567837	512794	96614	93653	0.97
Bound	45	371	222919	585143	585001	110628	107866	0.98
Bound	46	358	232531	616686	754944	141771	139464	0.98
Bound	47	366	254328	822175	890470	169496	163358	0.96
Bound	48	398	230857	566459	670154	108973	107765	0.99
Bound	49	404	250714	809991	740501	131328	127992	0.97
Bound	50	392	254988	796176	816842	151984	149698	0.98
Bound	51	423	277726	1059871	1035847	191412	187416	0.98
Bound	52	406	235578	484210	749566	125801	124243	0.99
Bound	53	436	274824	899227	1042184	191132	184629	0.97
Bound	54	420	231630	693232	476673	126808	118936	0.94
Bound	55	410	260356	789793	778349	166690	164452	0.99
Bound	56	427	243893	691601	611922	129679	126652	0.98
Bound	57	420	246685	722309	571837	111773	110448	0.99
Bound	58	453	234035	579238	523295	102828	101016	0.98
Bound	59	463	267737	935032	678318	154410	149825	0.97
Bound	60	467	247820	730390	521558	129937	127586	0.98
Bound	61	482	299505	1453545	877656	238139	234526	0.98
Bound	62	501	311931	1295609	1281470	279163	265977	0.95
Bound	63	527	342249	1843013	1856349	385605	375854	0.97
Bound	64	510	309136	1199470	1191241	257456	250601	0.97
Bound	65	534	335626	1686025	1632614	352095	342109	0.97
Bound	66	527	332005	1509100	1362454	290273	281876	0.97
Bound	67	531	351091	1937730	1778356	352614	339663	0.96
Bound	68	521	306109	888733	1087263	183248	176898	0.97
Bound	69	519	304198	937943	949982	208505	199448	0.96
Bound	70	520	284685	721871	792831	158665	152196	0.96
Bound	71	545	334545	1384811	1307605	269087	264510	0.98
Bound	72	527	325314	1079677	1323340	200428	197112	0.98
Bound	73	523	315747	756013	1275693	165336	160779	0.97
Bound	74	521	344586	759319	1815933	207655	203782	0.98
Bound	75	505	335226	671779	1681788	211598	207385	0.98

Bound	76	535	361900	1126855	1987257	251706	247074	0.98
Bound	77	536	350917	1094207	1744980	210657	207537	0.99
Bound	78	534	361352	1344364	1625019	214005	209633	0.98
Bound	79	528	374779	1532390	1784567	217654	212038	0.97
Bound	80	535	368380	1662533	1203244	171374	168026	0.98
Bound	81	536	398938	1995102	2156839	286584	279844	0.98
Bound	82	536	377166	1451767	1667322	220472	216926	0.98
Bound	83	541	389175	1605444	1851275	225715	216909	0.96
Bound	84	531	381498	1179239	1890519	176799	170403	0.96
Bound	85	540	413675	1703018	2661917	224588	218286	0.97
Bound	86	546	421819	1502357	3133507	229241	224321	0.98
Bound	87	554	430791	2031538	3109717	232008	225273	0.97
Bound	88	544	439876	2049780	3595093	233111	227557	0.98
Bound	89	539	449360	2125174	4331838	236912	227843	0.96
Bound	90	539	455258	2536756	4367251	254592	249141	0.98
Bound	91	547	460108	2837072	4520494	257164	249367	0.97
Bound	92	544	457552	2548526	4217206	233250	226014	0.97
Bound	93	539	461410	2863368	4655539	265523	259026	0.98
Bound	94	534	460186	2734980	4654275	287431	281977	0.98
Bound	95	542	462486	3085221	4519141	308217	304013	0.99
Bound	96	536	460426	3109963	4196188	264032	260831	0.99
Bound	97	537	454785	3004650	3486414	208286	205670	0.99
Bound	98	538	457923	3214728	3528536	208522	205402	0.99
Bound	99	538	462756	3462978	3866423	211479	209903	0.99
Bound	100	540	460503	3285198	3953632	194583	192943	0.99
Bound	101	545	457885	3021336	3977673	171317	169517	0.99
Bound	102	541	458716	3057798	4396474	166219	164631	0.99
Bound	103	542	453180	3451296	3797938	164526	162739	0.99
Bound	104	540	456166	4232893	4382993	197222	195800	0.99
Bound	105	542	446289	3500502	3302132	143345	140057	0.98
Bound	106	534	446256	3621940	3753529	142350	141407	0.99
Bound	107	535	436586	2775103	3497856	116994	116415	1.00
Bound	108	534	436508	2905906	3857291	126302	125664	0.99
Bound	109	537	431219	2751444	3552897	120437	119506	0.99
Bound	110	548	430383	3008329	3967795	135148	133182	0.99
Bound	111	549	416574	2755426	3251817	114016	112550	0.99
Daily	112	538	397883	2236407	2660227	91705	91060	0.99

Daily	113	545	383762	1997823	2232865	78143	77621	0.99
Daily	114	533	369185	1782203	1758981	68322	67971	0.99

Notes: Table shows data from the bound and daily editions of the *Congressional Record*. We show for each session the number of unique speakers, the number of unique phrases, and the the total phrase counts for Democratic and Republican speakers used in the estimation sample. We also show for each session the number of speeches that we identify in the *Record* and flag as neither appearing in the Extensions of Remarks nor as being delivered by speakers identified by office rather than name (e.g., “The Clerk”), the number of these speeches that we match to members of Congress, and the match rate of these speeches to members of Congress.

Online Appendix Table 2: Details on Topic Classification

Topic	Threshold	Keywords	Phrases in topic	Phrases excluded as false matches
alcohol	50	antisaloon, distil, alcohol, liquor, spirit, malt, saloon	alcohol beverag, alcohol content, alcohol tobacco, drug alcohol, abus alcohol, bureau alcohol, cent alcohol, alcohol drug, intox liquor, antisaloon leagu (17)	letter spirit, bipartisan spirit, spirit constitut, spirit compromis, spirit bipartisanhip, public spirit, spirit freedom, american spirit, human spirit, spiritu leader (13)
budget	125	budget, debt, defici, expenditur, spend, balanc, treasuri, expens, revenu	budget reconcili, budget plan, level spend, unifi budget, spend limit, rais debt, reduc spend, loss revenu, support budget, spend less (204)	tax expenditur, famili budget, revenu increas, balanc need, billion revenu, vote balanc, unexpend balanc, continu spend, unoblig balanc, reserv balanc (34)
business	75	bankruptci, agricultur, coal, owner, capit, stock, commod, canal, oil, credit, electr, farm, loan, financ, leather, railroad, patent, plant, potato, privat, wool, refin, sugar, salt, textil, profit, enterpri, commerc, fishermen, busi, petroleum, industri, chief execut, compani, gas, banker, manufactur, mine	gas pump, credit countri, interst commerc, build canal, american farmer, loan author, agricultur commod, insur loan, timber industri, farmer want (891)	american soil, capitol build, like mine, financi assist, privat concern, provid financi, presid nomine, financi aid, privat social, home ownership (85)
crime	50	crime, traffick, opium, rehabilit, crim, drug, gun, narcot, shoot, sentenc, prison	state crimin, drug countri, violent crimin, heinous crime, increas drug, drug import, drug war, crime occur, drug one, crimin enterpris (211)	institut drug, prison popul, racial discrimin, anim drug, state prison, problem drug, one sentenc, prohibit discrimin, discrimin duti, bureau prison (65)
defense	75	arm, navi, armi, atom, defens, war, guard, missil, homeland, fighter, naval, navyyard, soldier, prison, veteran, intellig, enlist, bomb, weapon, combat, militari, torpedo	weapon capabl, privat soldier, intellig oper, warsaw pact, war think, intellig surveil, armi corp, war labor, veteran educ, honor veteran (584)	freight forward, individu farmer, farm subsidi, support farm, warn us, farmer produc, famili farmer, releas prison, put forward, farmer rancher (252)
economy	50	growth, recoveri, econom, unemploy, poverti, stimulus	economi job, longterm unemploy, economi area, econom disloc, recoveri program, assist econom, econom life, economi put, make economi, econom success (424)	conserv recoveri, growth countri, growth defens, growth develop, growth feder, growth govern, growth polici, old growth, percent growth, popul growth (10)
education	50	campus, educ, school, lunch, student, teacher, academ	student faculti, educ unit, american school, import educ, school dropout, educ relat, build school, school superintend, econom educ, lowincom student (477)	reduc deficit, reduc feder, debt reduct, reduc nuclear, actual reduc, reduc paperwork, reduc time, reduc oil, go reduct, state reduct (219)
elections	50	campaign, elect, vote, ballot	local elect, vote poll, secret ballot, influenc elect, bipartisan vote, vote reduct, vote precinct, popular elect, vote feder, vote unit (201)	educ campaign, electron equip, state elect, electr generat, media campaign, vote tariff, electr co, time elect, vote dont, director select (56)
environment	50	pollut, garden, climat, dam, forest, wast, hurrican, natur, plant, conserv, eros, contam, soil, canyon, environ, shale, lake	understand natur, environment assess, natur thing, mother natur, environment degrad, lowlevel wast, develop environment, dam across, wast dump, dam reservoir (192)	reserv forc, militari reserv, conserv liber, strateg reserv, war plant, repair damag, grain reserv, indian reserv, immigr natur, damag economi (183)
federalism	125	regul, state, southern, local, union, agenc, confeder, secede, execut, feder, interst	state air, appropri state, agenc must, feder respons, feder constitut, feder deficit, state plan, state line, feder effort, local offic (538)	farmer union, japanes govern, percent state, particular state, arab state, unit statessoviet, chief execut, trade union, german govern, accord statement (66)
foreign	125	duti, qaeda, canada, chines, china, aid, salvador, foreign, britain, persia, philipin, porto, puerto, saddam hussein, spanishamerican, communism, troop, cold war, latin, khmer, independ, postcold war, spanish, german, spain, iran, asia, bolivia, vietnam, afghanistan, treati, communist, contra, cuba, genocid, haiti, iraq, mexic, nicaragua, panama, soviet, kingdom	vietnam era, foreign languag, depend foreign, state soviet, asian develop, foreign aid, chines communist, veteran foreign, mexican govern, freedom independ (184)	said look, year said, direct toward, defens budget, economi nation, nation labor, medicar medicaid, nation academi, repeat said, balanc budget (368)

Details on Topic Classification (*continued*)

Topic	Threshold	Keywords	Phrases in topic	Phrases excluded as false matches
government	75	govern, depart, highway, ineitem, medicaid, regulatori, medicar, stimulus, welfar, privati, benefit, bureau, municip, agenc	manag agenc, welfar payment, industri govern, right govern, govern creat, author govern, concern welfar, govern also, spend govern, state highway (697)	benefit colleagu, oper depart, guarante benefit, survivor benefit, go benefit, percent benefit, direct benefit, econom benefit, iraqi govern, benefit packag (101)
health	50	care, care reform, health, medicar, mental, nurs, coverag, nutrit, hiv, hospit, medicaid, gyn, medic, cell	use medic, overal health, militari health, care studi, insur coverag, includ health, prevent health, health care, primari care, bill medicar (603)	cell phone, public career, year career, profession career, care unemploy, term care, care also, examin care, care much, like health (100)
immigration	50	alien, illeg, immigr, foreign, citizen, undoc, border, detain	foreign born, legal illeg, countri illeg, involv foreign, nation border, come foreign, allow foreign, subcommitte immigr, side border, reform immigr (80)	vulner citizen, concern citizen, foreign mail, depend foreign, issu foreign, patriot citizen, foreign countri, citizen whose, competit foreign, paid foreign (230)
justice	50	court, justic, freedman, right, trial, lawyer, habea, attorney, clerk, judici	right particip, right front, justic burger, justic scienc, right record, preserv right, provid court, right selfdefens, court order, tri right (485)	left right, clinic trial, right away, go right, thing right, done right, right approach, way right, right answer, given right (38)
labor	75	compens, worker, employ, salari, job, longshoremen, labor, pension, pay, unemploy, vocat, wage, strike, retir, union	labor practic, cheap labor, reduc salari, job peopl, lost job, employe govern, provid job, postal worker, rise unemploy, govern employ (624)	depart commerc, pay offic, program payment, payment lieu, intern union, expens taxpay, commerc depart, cash payment, billion taxpay, requir pay (179)
mail	50	mail, freedeliveri, posta, postoffic, messeng, postmast	mail pay, rate postag, receiv mail, postal receipt, postal system, ocean mail, deliv mail, mail rate, postal card, postmast general (61)	n/a
minorities	50	african, choctaw, color, jew, equal, hate, right, women, philippin, negro, discrimin, woman, segreg, right act, filipino, muslim, hawaian, cherokee, indian, tribe, tribal	women voter, bureau indian, indian agent, leagu women, ensur women, indian educ, tribal govern, elimin discrimin, women militari, indian communiti (230)	upon right, bargain right, right justic, right measur, right fix, system right, right talk, right concern, right away, right just (371)
money	50	dollar, treasuri, gold, silver, tender, coin, payment, note, cash, monet	dollar valu, dollar limit, payabl coin, secretari treasuri, tender payment, dollar just, coinag silver, silver gold, cash payment, treasuri bill (178)	import note, presid note, one dollar, goldman sach, payment lieu, payment limit, payment adjust, payment hospit, disabl payment, pay silver (144)
religion	50	islam, religi, prayer, jew, mormon, jehovah, muslim, buddhi, christian, cathol, hindu, protest, judai, sectarian	american jewish, religi leader, exercis religion, million jew, religi belief, religion sex, christian scienc, roman cathol, soviet jewish, prayer school (47)	administr depart, administr cost, state administr, administr veteran, administr develop, administr polici, electrif administr, administr nation, within administr, enter protest (290)
tax	50	tax, estat, excessprofit, internalrevenu	increas taxat, general tax, state tax, pay estat, extend tax, stamp tax, tax appeal, dollar tax, huge tax, tax legisl (433)	billion sinc, billion debt, break cycl, quarter billion, ten billion, billion spent, billion gallon, billion time, billion addit, one billion (114)
trade	50	valorem, trade, tariff, paynealdrich, duti, exportimport	pay duti, eastwest trade, tariff import, dutiabi list, tariff reform, trade china, nontariff barrier, duti manufacturer, duti increas, trade compani (194)	upon duti, duti bill, high duti, imposit duti, duti unit, duti high, duti militari, constitut duti, duti see, higher duti (97)

Notes: For each topic, this table shows the minimum number of times a phrase needs to be used (in at least 5 sessions) to be included in the topic, all the keywords we associated with the topic, a random sample of 10 phrases included in the topic, and a random sample of 10 phrases (if any) that we manually exclude from the topic. The numbers in parentheses represent the total number of phrases in each category.

Session 55 (1897-1898)

Republican	#R	#D	Democratic	#R	#D
philippin island	78	52	stand armi	20	53
line offic	18	2	peopl countri	57	85
coastwis trade	19	3	ad valorem	51	75
elector board	13	0	sugar trust	9	33
acquir territori	29	17	cent ad	28	48
merit system	12	2	tax upon	34	53
servic commiss	21	11	silver dollar	31	49
dalla counti	10	1	compar statement	4	22
secretari navi	67	58	gold silver	23	40
san francisco	31	22	duti upon	26	42

Session 56 (1899-1900)

Republican	#R	#D	Democratic	#R	#D
poll tax	33	4	secretari treasuri	84	117
geodet survey	28	6	great britain	43	76
secretari navi	82	60	american peopl	75	103
naval offic	41	19	retir list	10	34
coast geodet	28	6	war tax	5	28
coast survey	25	6	stand armi	15	39
pneumaticub servic	24	8	old soldier	16	38
puerto rico	248	232	secondclass matter	8	30
rule regul	35	20	treasuri note	14	36
color peopl	17	2	philippin island	84	105

Session 57 (1901-1902)

Republican	#R	#D	Democratic	#R	#D
pension also	42	8	philippin island	215	283
panama canal	62	33	great britain	37	89
secretari agricultur	54	28	volunt infantri	1	22
railroad compani	73	51	anthracit coal	7	28
said court	23	4	philippin commiss	95	111
secretari interior	57	40	american peopl	19	37
secretari navi	50	33	capit stock	7	20
panama rout	26	10	wrn alden	28	40
built navyyard	16	2	old soldier	13	24
depart agricultur	29	15	present system	9	19

Session 58 (1903-1904)

Republican	#R	#D	Democratic	#R	#D
judg swayn	360	49	pension also	25	78
lake bluff	69	0	lackawanna counti	0	50
submarin boat	37	12	american peopl	64	88
norfolk navyyard	24	0	peopl countri	38	59
secretari agricultur	63	42	commerc commiss	25	44
gate receipt	18	4	citi chicago	9	24
secretari treasuri	91	78	traction compani	0	14
naval train	16	4	armor plate	17	30
train station	20	7	sugar trust	5	18
servic commiss	28	17	postal save	1	13

Session 59 (1905-1906)

Republican	#R	#D	Democratic	#R	#D
pension also	132	22	leather compani	0	87
pipe line	73	13	secondclass matter	3	32
oil compani	64	8	subig bay	22	48
basingpoint system	36	0	watch movement	0	24
dri dock	46	14	watch compani	5	27
claim also	34	7	waltham watch	2	17
nation cemeteri	27	1	secondclass mail	1	15
geolog survey	43	19	construct canal	15	28
circuit court	46	23	canal zone	7	20
philippin island	103	80	railway compani	27	39

Session 60 (1907-1908)

Republican	#R	#D	Democratic	#R	#D
postal save	39	3	canal zone	18	66
census offic	31	2	also petit	0	47
reserv balanc	36	12	standard oil	4	25
war depart	62	39	indirect contempt	0	19
secretari navi	62	39	bureau corpor	5	24
secretari agricultur	58	36	panama canal	23	41
pay pension	20	2	nation govern	12	30
boat compani	24	8	coal mine	9	27
twelfth census	14	0	revis tariff	8	26
forestri servic	20	7	feet lake	0	17

Session 61 (1909-1910)

Republican	#R	#D	Democratic	#R	#D
commerc commiss	123	91	sugar trust	7	54
manufactur jute	30	0	porto rico	66	108
railroad compani	84	55	postal save	25	58
circuit court	74	47	necessari life	4	29
great mani	70	43	revis tariff	26	47
rate fare	23	1	natur resourc	10	29
san francisco	46	25	special interest	5	21
mexican ore	20	0	increas salari	35	51
volunt infantri	20	0	old soldier	10	25
fare charg	18	1	refin compani	7	21

Session 62 (1911-1912)

Republican	#R	#D	Democratic	#R	#D
pig iron	32	10	american peopl	85	111
judg archbald	44	23	high protect	5	21
armi reserv	18	0	panama canal	41	55
foreign countri	51	34	direct tax	5	18
regular armi	21	5	nation govern	16	27
organ militia	16	1	pension legis	9	20
porto rico	23	9	depart agricultur	23	34
know whether	66	51	raw wool	15	26
great britain	67	53	payncaldrich bill	7	17
war depart	63	50	trial juri	4	14

Session 63 (1913-1914)

Republican	#R	#D	Democratic	#R	#D
great britain	96	69	feder reserv	72	114
lock dam	33	9	volunt infantri	0	32
foreign countri	51	36	peopl countri	50	67
know whether	61	49	special privileg	10	24
ship board	14	4	soldier sailor	3	17
year ago	149	140	potato starch	2	16
botan garden	19	9	reserv board	31	44
industri corpor	10	1	armor plate	5	17
vocat educ	14	6	feder court	10	21
rock creek	16	7	navi yard	34	45

Session 64 (1915-1916)

Republican	#R	#D	Democratic	#R	#D
porto rico	68	26	navi yard	56	91
favor trade	33	0	feder reserv	52	72
secretari war	130	97	manufactur powder	2	15
nation guard	164	136	unavoid absenc	5	17
trade balanc	25	1	announc unavoid	5	17
american citizen	66	47	american peopl	75	84
armi engin	26	10	feder treasuri	11	20
regular armi	82	66	peopl countri	47	56
pension also	16	1	war depart	73	82
agricultur subject	16	2	navi leagu	0	9

Session 65 (1917-1918)

Republican	#R	#D	Democratic	#R	#D
colleagu senior	105	2	feder reserv	70	120
announc colleagu	72	3	ship board	38	57
unmark ballot	69	2	standard return	3	18
account ill	72	10	time war	50	64
secretari treasuri	112	67	nation defenc	35	46
absenc colleagu	41	7	rule regul	21	30
war profit	43	17	advisor commiss	5	13
govern unit	55	33	reserv note	5	13
american peopl	98	78	command chief	23	30
announc stand	23	3	interest commerc	67	74

Session 66 (1919-1920)

Republican	#R	#D	Democratic	#R	#D
secretari treasuri	96	43	leagu nation	75	129
take care	116	81	distinguishedservic medal	0	28
war depart	187	158	militari train	18	40
muscl shoal	41	12	univers militari	7	24
regular armi	52	34	nation world	19	36
loan board	27	8	american peopl	100	113
public health	47	29	feder reserv	69	80
foreign countri	42	24	men women	17	28
great mani	64	47	postal employ	8	19
farm loan	38	22	also petit	1	12

Session 67 (1921-1922)

Republican	#R	#D	Democratic	#R	#D
war depart	111	67	american peopl	99	159
nation guard	54	13	great britain	77	123
american valuat	41	7	tax upon	17	43
take care	102	70	feder reserv	109	134
attorney general	79	53	excessprofit tax	27	48
foreign countri	62	38	proptert custodian	15	35
ad valorem	71	49	net incom	11	29
secretari agricultur	52	31	increas rate	14	30
muscil shoal	31	12	cash regist	3	17

Session 68 (1923-1924)

Republican	#R	#D	Democratic	#R	#D
nitrat plant	71	9	adjust compens	32	80
war depart	104	50	postal employe	36	75
secretari war	86	55	american peopl	63	100
muscil shoal	87	61	exservic men	16	46
regular armi	27	6	navi yard	30	58
nation guard	28	7	net incom	21	42
steam plant	19	2	tax upon	17	37
take care	81	65	social mellon	1	20
chief engin	22	7	soldier bonus	5	23
pension also	14	0	secretari navi	51	64

Session 69 (1925-1926)

Republican	#R	#D	Democratic	#R	#D
war depart	86	52	american peopl	52	113
nation defens	54	20	feder reserv	62	101
regular armi	38	6	veteran bureau	38	59
muscil shoal	62	35	american citizen	29	47
nation guard	34	7	milk cream	5	21
reserv offic	26	2	peopl countri	30	45
secretari war	66	44	antisaloon leagu	5	20
corp area	19	0	exservic men	5	16
leagu nation	39	21	game relig	1	11
alien proptert	32	14	high protect	3	12

Session 70 (1927-1928)

Republican	#R	#D	Democratic	#R	#D
war depart	97	63	pension also	0	163
take care	105	72	american peopl	51	91
foreign countri	54	28	radio commiss	8	44
muscil shoal	97	71	spoken drama	0	30
steam plant	25	3	civil war	27	54
nation guard	39	18	trade commiss	19	46
air corp	32	12	feder trade	19	45
creek dam	25	6	wave length	6	25
cove creek	30	13	imperit valley	12	28
american ship	29	12	flowag right	5	20

Session 71 (1929-1930)

Republican	#R	#D	Democratic	#R	#D
feder reserv	74	36	american peopl	65	108
tariff commiss	137	104	stabil corpor	17	53
muscil shoal	77	46	increas cost	21	47
war depart	78	53	men women	33	55
reserv object	32	9	high tariff	6	26
communis parti	23	1	dollar worth	6	26
flexibl tariff	26	4	industri alcohol	10	26
nitrat plant	24	4	interst commerc	59	76
cost product	67	49	commerc commiss	37	52
hydraul laboratori	19	3	farm relief	40	54

Session 72 (1931-1932)

Republican	#R	#D	Democratic	#R	#D
cove creek	58	1	pension also	21	106
war depart	99	62	farm board	75	117
muscil shoal	46	8	american peopl	91	126
poll tax	34	1	reconstruct financ	100	132
feder reserv	193	161	financ corpor	117	148
nation defens	66	35	affair also	11	39
flowag right	23	1	sale tax	99	122
pay pool	20	0	feder employe	18	39
air mail	45	29	foreign coal	0	18
reserv offic	26	11	claim also	18	36

Session 73 (1933-1934)

Republican	#R	#D	Democratic	#R	#D
feder reserv	254	165	home owner	28	79
rule regul	69	25	dir tugwel	2	37
veteran bureau	54	15	financ corpor	75	103
econom act	74	40	reconstruct financ	73	99
war veteran	102	70	oldag positon	4	24
veteran administr	65	35	men women	32	50
cove creek	32	3	american peopl	101	118
servic connect	44	16	philippin island	18	33
spanishamerican war	62	35	loan corpor	15	30
tariff commiss	47	23	reserv object	14	26

Session 74 (1935-1936)

Republican	#R	#D	Democratic	#R	#D
feder reserv	160	105	home owner	33	72
valley author	59	23	world war	55	79
reserv board	51	25	presid director	2	22
militari train	36	11	commerc commiss	29	43
counti agent	26	1	loan corpor	21	34
trade agreement	35	14	reconstruct financ	46	59
lengn nation	34	14	regul valu	1	13
present administr	41	22	air corp	7	17
interst commerc	123	105	announc absenc	1	11
flowag right	21	3	ship board	3	13

Session 75 (1937-1938)

Republican	#R	#D	Democratic	#R	#D
benefit payment	53	7	feder reserv	24	56
trade agreement	66	20	labor relat	18	38
undistributprofit tax	41	10	war depart	37	58
nation defens	70	43	puerto rico	8	27
nation debt	29	9	import public	1	18
secretari agricultur	67	49	wage hour	45	60
loan corpor	31	13	coast guard	4	17
sale tax	26	10	nation labor	16	29
soil conserv	50	35	navi yard	6	19
valley author	32	18	public health	7	17

Session 76 (1939-1940)

Republican	#R	#D	Democratic	#R	#D
nation defens	256	131	servic commiss	12	41
nation debt	45	18	war depart	78	103
world war	137	110	polit activ	8	31
arm embargo	48	25	import public	1	21
foreign polici	40	17	interst commerc	49	65
great britain	80	58	wage hour	33	48
foreign countri	61	41	panama canal	24	38
nation guard	76	58	labor relat	27	39
trade treat	21	4	commerc commiss	33	43
billion dollar	64	48	us fate	0	10

Session 77 (1941-1942)

Republican	#R	#D	Democratic	#R	#D
nation defens	263	159	percent parti	17	65
world war	134	95	navi depart	51	78
great britain	104	66	commod credit	25	47
win war	104	75	organ labor	6	28
defens program	88	61	agricultur commod	15	36
san francisco	27	6	credit corpor	26	45
nation debt	28	9	reconstruct financ	16	35
tax bill	53	37	financ corpor	18	35
billion dollar	60	43	war depart	125	140
merchant ship	30	14	depart justic	19	31

Session 78 (1943-1944)

Republican	#R	#D	Democratic	#R	#D
pay roll	61	28	unit nation	34	56
san francisco	35	8	product board	34	53
servic court	22	0	war depart	93	111
reserv object	19	4	war product	62	80
smoke opium	14	0	poll tax	13	29
pearl harbor	55	41	offic war	16	29
win war	80	66	silver purchas	3	16
execut agreement	16	3	ruml plan	17	30
consum subsid	17	6	side aisl	18	29
exercis jurisdict	12	0	war program	7	18

Session 103 (1993-1994)				Session 104 (1995-1996)				Session 105 (1997-1998)						
Republican	#R	#D	Democratic	Republican	#R	#D	Democratic	Republican	#R	#D	Democratic			
american peopl	263	194	fas fac	307	188	minimum wage	50	140	american peopl	245	149	campaign financ	44	95
tax increas	84	33	deficit reduct	69	20	tax break	12	77	tax relief	94	24	financ reform	30	72
unit nation	82	58	feder reserv	142	94	nurs home	7	44	tax increas	56	8	public school	44	79
rais tax	40	18	interest rate	76	35	deficit reduct	22	53	feder debt	47	4	minimum wage	9	35
illeg alien	32	10	univers coverage	63	27	head start	9	30	tax code	54	24	bill right	14	38
men women	72	51	care reform	38	6	student loan	24	44	rais tax	37	7	tobacco compani	12	33
term limit	25	5	head start	50	18	medicard medicaid	17	37	side aisl	99	75	tax break	9	29
justic depart	31	12	crime bill	97	68	inreaz minimum	10	26	govern spend	26	2	insur compani	7	24
employ mandat	22	5	peopl countri	31	3	welfar reform	72	87	big govern	25	2	credit union	23	40
american taxpay	38	21	human right	42	14	budget resolut	41	56	save account	32	9	budget resolut	23	39

Session 106 (1999-2000)				Session 107 (2001-2002)				Session 108 (2003-2004)						
Republican	#R	#D	Democratic	Republican	#R	#D	Democratic	Republican	#R	#D	Democratic			
tax relief	87	30	prescript drng	65	3	prescript drng	90	162	reserv balanc	132	67	head start	19	73
death tax	55	2	insur compani	80	24	trust fund	46	80	side aisl	133	79	homeland secur	99	140
secur surplus	59	20	bill right	95	66	homeland secur	74	105	tax relief	71	22	first respond	18	58
feder debt	43	4	minimum wage	67	41	secur medicar	16	46	saddam hussein	106	60	war iraq	19	56
american peopl	172	134	juvenil justic	50	28	insur compani	17	42	al qaeda	56	18	american peopl	152	183
tax increas	38	4	credit card	23	2	african american	13	37	war terror	84	49	prescript drng	142	173
trust fund	89	56	african american	30	9	drug benefit	25	48	hate crime	38	4	nation debt	7	35
side aisl	109	79	tax break	44	25	nation debt	10	33	tax increas	37	7	tax break	6	34
teacher empower	29	1	estat tax	27	9	estat tax	20	42	econom growth	44	19	insur compani	12	38
tax code	44	16	gun violenc	84	68	american peopl	139	159	rais tax	31	6	guard reserv	14	39

Session 109 (2005-2006)				Session 110 (2007-2008)				Session 111 (2009-2010)						
Republican	#R	#D	Democratic	Republican	#R	#D	Democratic	Republican	#R	#D	Democratic			
reserv balanc	138	75	american peopl	87	20	dog coalit	0	90	american peopl	331	218	insur compani	41	107
war terror	106	44	minimum wage	77	20	war iraq	18	78	general kegan	62	1	african american	7	56
natur gas	85	26	tax break	147	105	african american	6	62	pleas bless	59	0	care reform	70	116
side aisl	128	75	credit card	44	10	american peopl	230	278	tax increas	64	9	middl class	14	54
al qaeda	62	15	oil compani	34	3	oil compani	20	65	god pleas	54	0	colleagu support	65	102
death tax	46	2	oil compani	34	7	civil war	17	45	stimulus bill	56	14	unemploy benefit	14	45
hate crime	35	1	hurricane katrina	132	106	troop iraq	11	39	care bill	62	21	unemploy insur	5	36
tax relief	42	10	african american	33	8	children health	17	42	rais tax	45	7	progress caucus	1	31
illeg immigr	46	16	nation debt	32	8	noblid contract	0	24	reserv balanc	150	115	recoveri act	5	35
stem cell	96	67	nation guard	26	4	middl class	15	39	govern takeov	33	4	colleagu join	24	53

Session 112 (2011-2012)				Session 113 (2013-2014)				Session 114 (2015-2016)						
Republican	#R	#D	Democratic	Republican	#R	#D	Democratic	Republican	#R	#D	Democratic			
fornia madam	195	0	middl class	364	223	minimum wage	14	99	american peopl	327	205	homeland secur	96	205
american peopl	322	252	tax break	87	6	immigr reform	23	94	al qaeda	50	7	climat chang	23	94
rais tax	85	22	care act	42	1	climat chang	11	82	men women	123	83	gun violenc	3	74
job creator	71	12	afford care	173	143	student loan	27	89	side aisl	133	93	african american	11	71
tax increas	79	26	million american	35	4	care act	94	149	human traffick	60	26	vote right	2	62
job creation	94	55	insur compani	27	4	afford care	95	149	colleagu support	123	89	public health	24	83
govern spend	46	10	student loan	27	5	middl class	42	90	religi freedom	34	4	depart homeland	48	93
bust owner	47	17	oil compani	20	0	farm bill	36	79	taxpay dollar	47	19	plan parenthood	66	104
nation debt	57	27	clean air	20	0	interest rate	19	61	mental health	59	32	afford care	40	77
american energi	30	2	public health	25	6	comprehens immigr	6	40	radio islam	22	0	puerto rico	42	79

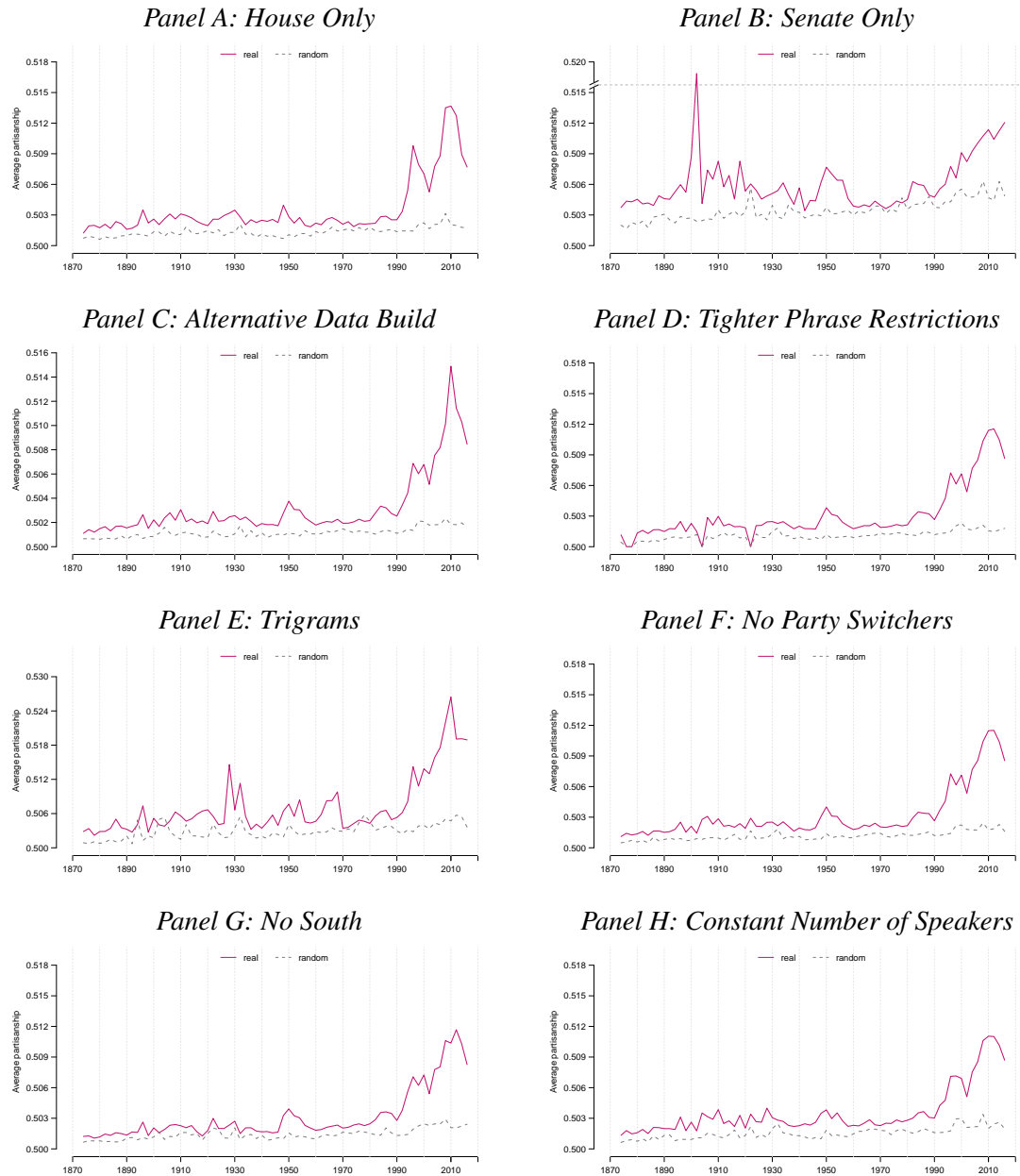
Notes: Calculations are based on our preferred specification in Panel B of Figure 2. The table shows the Republican and Democratic phrases with the greatest magnitude of estimated partisanship ζ_{jt} alongside the predicted number of occurrences of each phrase per 100,000 phrases spoken by Republicans or Democrats. Phrases with positive values of ζ_{jt} are listed as Republican and those with negative values are listed as Democratic. See the body of the paper for the definition of partisanship ζ_{jt} .

Online Appendix Table 4: Manually Selected Words Used to Exclude Phrases as Procedural

absent	adjourn	ask	can	chairman
committee	con	democrat	etc	gentleladies
gentlelady	gentleman	gentlemen	gentlewoman	gentlewomen
hereabout	hereafter	hereat	hereby	herein
hereinafter	hereinbefore	hereinto	hereof	hereon
hereto	heretofore	hereunder	hereunto	hereupon
herewith	month	mr	mrs	nai
nay	none	now	part	per
pro	republican	say	senator	shall
sir	speak	speaker	tell	thank
thereabout	thereafter	thereagainst	thereat	therebefore
therebeforn	thereby	therefor	therefore	therefrom
therein	thereinafter	thereof	thereon	thereto
theretofore	thereunder	thereunto	thereupon	therewith
therewithal	today	whereabouts	whereafter	whereas
whereat	whereby	wherefore	wherefrom	wherein
whereinto	whereof	whereon	whereto	whereunder
whereupon	wherever	wherewith	wherewithal	will
yea	yes	yield		

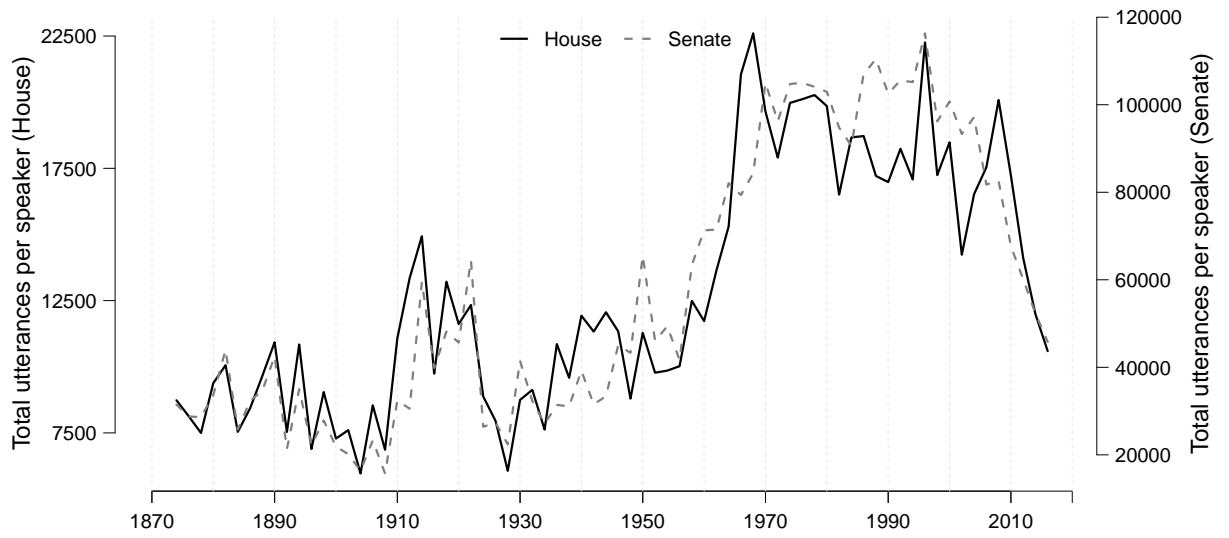
Notes: Table presents words we use to identify phrases for removal. Any phrase containing the stem of at least one of these words is removed from our final vocabulary.

Online Appendix Figure 1: Average Partisanship of Speech Estimated on Alternative Samples



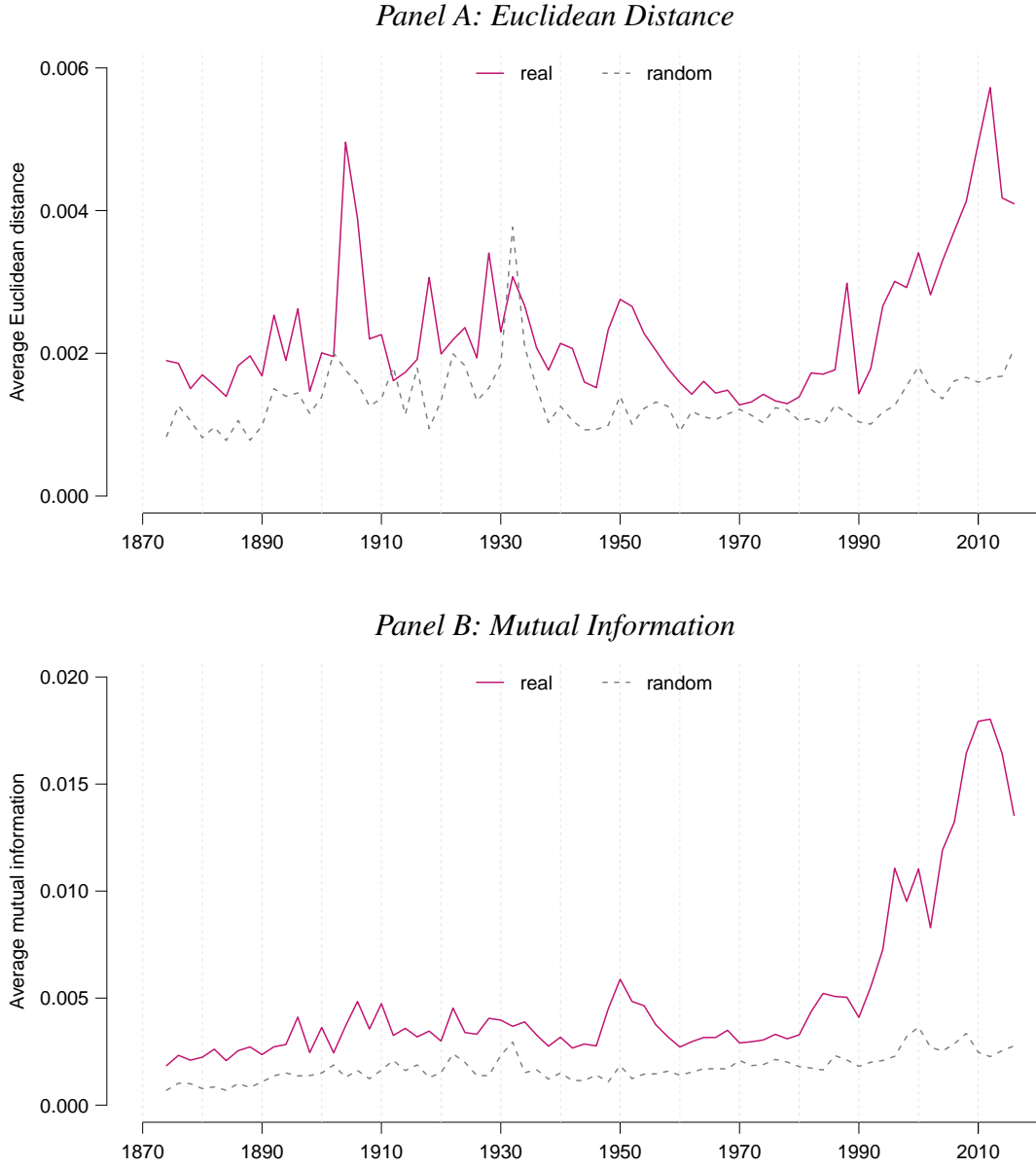
Notes: All estimates are based on our preferred penalized estimator $\hat{\pi}_t^*$ defined in section 4.3. “Real” series is from actual data; “random” series is from hypothetical data in which each speaker’s party is randomly assigned with the probability that the speaker is Republican equal to the average share of speakers who are Republican in the sessions in which the speaker is active. Panel A’s specification restricts the sample for estimation to speaker-sessions observed in the House, whereas Panel B’s applies the same restriction for speaker-sessions observed in the Senate. Note that the vertical axis of Panel B contains a break to facilitate comparison with the other panels. Panel C’s specification uses data from the *Congressional Record* bound edition for sessions 43-96 and from its daily edition for sessions 97-114 and the same vocabulary as our baseline estimates. Panel D’s specification decreases the number of phrases used in estimation by roughly 10 percent. We implement this restriction by requiring that phrases in our vocabulary are spoken at least 11 times in at least one session, at least 110 times across all sessions, and in at least 11 speaker sessions. (We use the minimums of 10, 100, and 10 to produce the baseline specification’s vocabulary.) Panel E’s specification uses alternate counts matrices \mathbf{C}_t whose rows correspond to speakers and whose columns correspond to distinct three-word phrases (trigrams), produced under the same rules as our vocabulary of bigrams. Panel F’s specification excludes speakers who ever switched parties from estimation. Panel G’s specification excludes speakers representing states in the South census region from estimation. Lastly, Panel H’s specification uses an estimation sample with a constant number of speakers in each session of Congress. We construct this sample by finding the number of speakers in the session of Congress with the fewest number of speakers, and then randomly sampling that many speakers without replacement from each session of Congress. We alter our covariate design when estimating partisanship using an alternative sample, omitting covariates corresponding to features universally possessed or lacked by speaker-sessions in this sample.

Online Appendix Figure 2: Average Verbosity per Speaker-Session



Notes: Plot shows, separately for each chamber and session, the average number of phrase (bigram) utterances across speakers in our estimation sample before making any vocabulary restrictions.

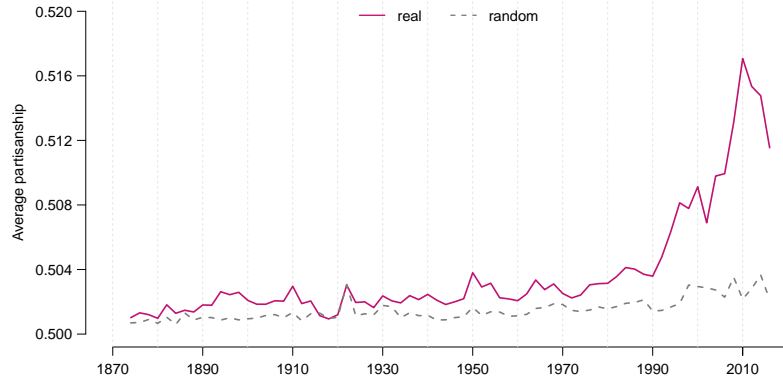
Online Appendix Figure 3: Different Target Measures of Party Differences in Speech



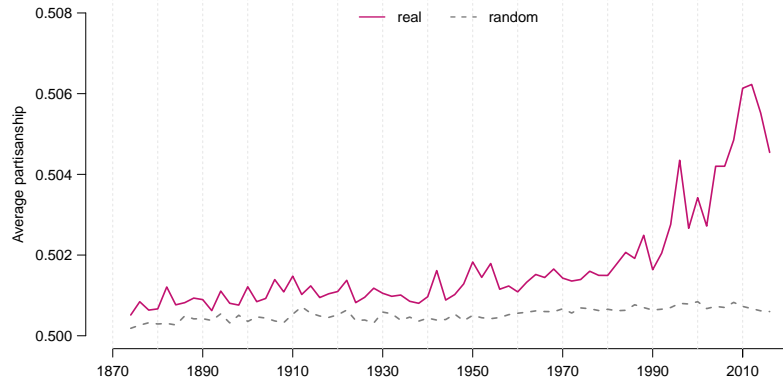
Notes: Estimates are from our preferred specification in Panel B of Figure 2 in the paper. “Real” series is from actual data; “random” series is from hypothetical data in which each speaker’s party is randomly assigned with the probability that the speaker is Republican equal to the average share of speakers who are Republican in the sessions in which the speaker is active. Panel A reports the average of Euclidean distance between $\mathbf{q}_t^R(\mathbf{x}_{it})$ and $\mathbf{q}_t^D(\mathbf{x}_{it})$ across speakers in session t for each session in our sample, defined as $\frac{1}{|R_t \cup D_t|} \sum_{i \in R_t \cup D_t} \sqrt{\sum_{j \in J} (q_{jt}^R(\mathbf{x}_{it}) - q_{jt}^D(\mathbf{x}_{it}))^2}$. Panel B reports average mutual information between party and phrase choice, given a neutral prior. This is defined as $\frac{1}{|R_t \cup D_t|} \sum_{i \in R_t \cup D_t} I(\mathbf{x}_{it})$ where $I(\mathbf{x}_{it}) = 1 - \sum_j \frac{q_{jt}^R(\mathbf{x}_{it})}{2} \log_2\left(\frac{1}{\rho_{jt}(\mathbf{x}_{it})}\right) + \frac{q_{jt}^D(\mathbf{x}_{it})}{2} \log_2\left(\frac{1}{1 - \rho_{jt}(\mathbf{x}_{it})}\right)$.

Online Appendix Figure 4: Average Partisanship of Speech with Different Covariate Designs

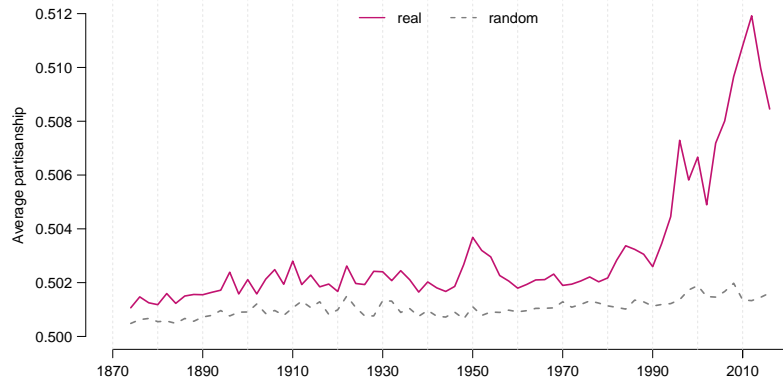
Panel A: No Covariates



Panel B: Speaker Effects



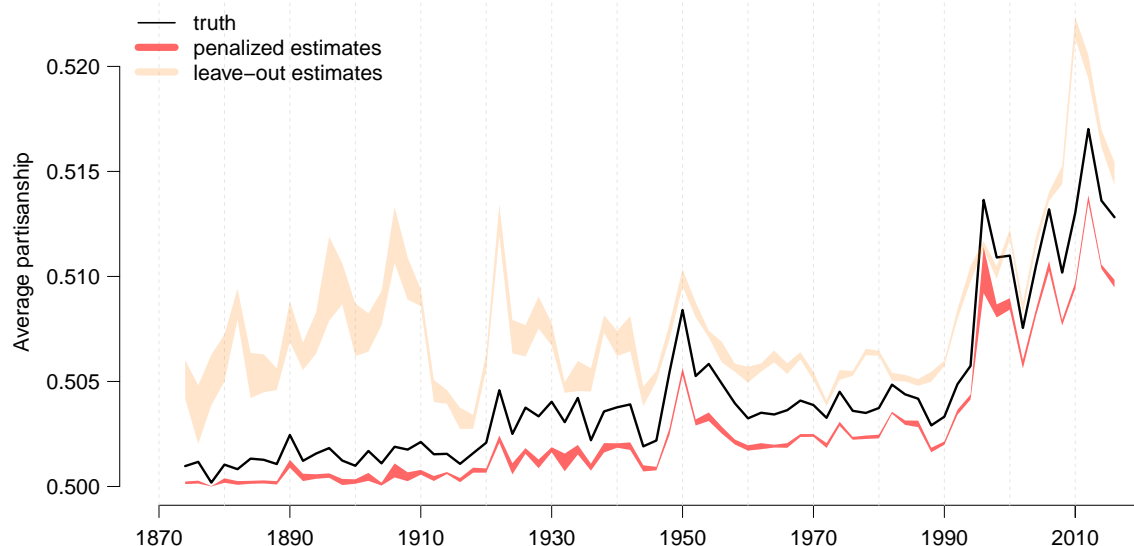
Panel C: Additional Covariates



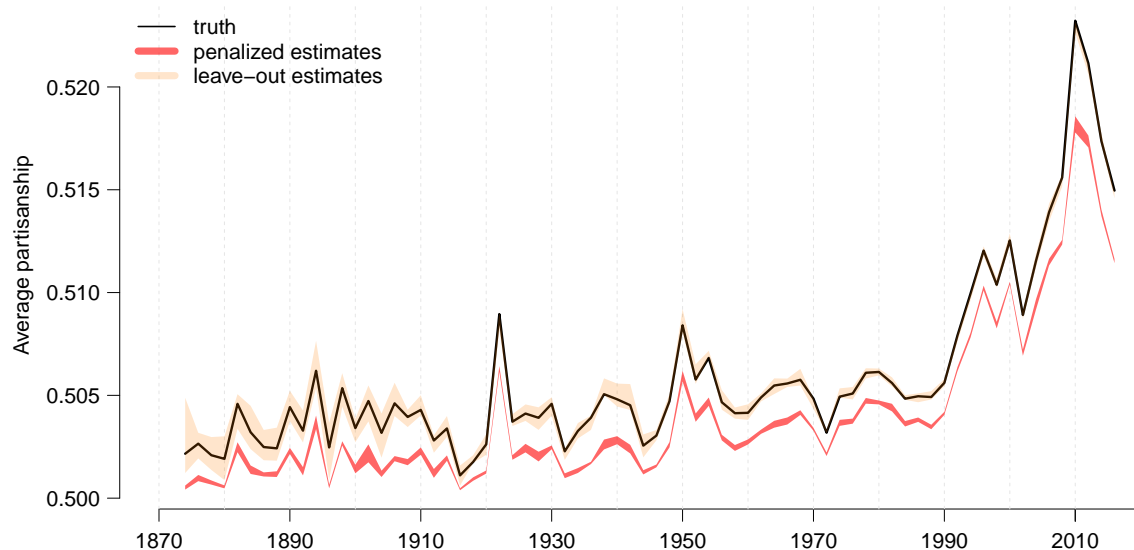
Notes: “Real” series is from actual data; “random” series is from hypothetical data in which each speaker’s party is randomly assigned with the probability that the speaker is Republican equal to the average share of speakers who are Republican in the sessions in which the speaker is active. Each panel modifies the covariate design relative to our preferred specification in Panel B of Figure 2 in the paper. The specification in Panel A imposes that $\gamma_{jt} = 0$. The specification in Panel B includes in \mathbf{x}_{it} a speaker random effect v_{ijt} that is independent of covariates and is distributed as *Laplace* $(0, \kappa)$ so that its standard deviation is $\sqrt{2}/\kappa$. We set $\hat{\kappa} = \sqrt{2}/\text{sd}(\hat{e}_{ijt})$ where $\hat{e}_{ijt} = \log(c_{ijt}/\exp[\hat{\mu}_i + u_{ijt}])$ are the observed Poisson residuals from our baseline model when $c_{ijt} > 0$. We then estimate the random effects v_{ijt} and the remaining parameters of the model by exploiting the fact that posterior maximization under the Laplace assumption is equivalent to L_1 -penalized deviance minimization with cost κ/n , where n is the number of speaker-sessions. We impose a minimal penalty of $\psi = 4 \times 10^{-5}$ on the speaker random effects v_{ijt} and choose $\hat{\lambda}_j^G$ with $\iota = 10^{-3}$. The specification in Panel C includes in \mathbf{x}_{it} indicators for having had family members in Congress, having attended college, military service, prior public service, being a lawyer, and whether any of the aforementioned variables are missing.

Online Appendix Figure 5: Parametric Bootstrap of Leave-out and Penalized Estimators

Panel A: Baseline Specification

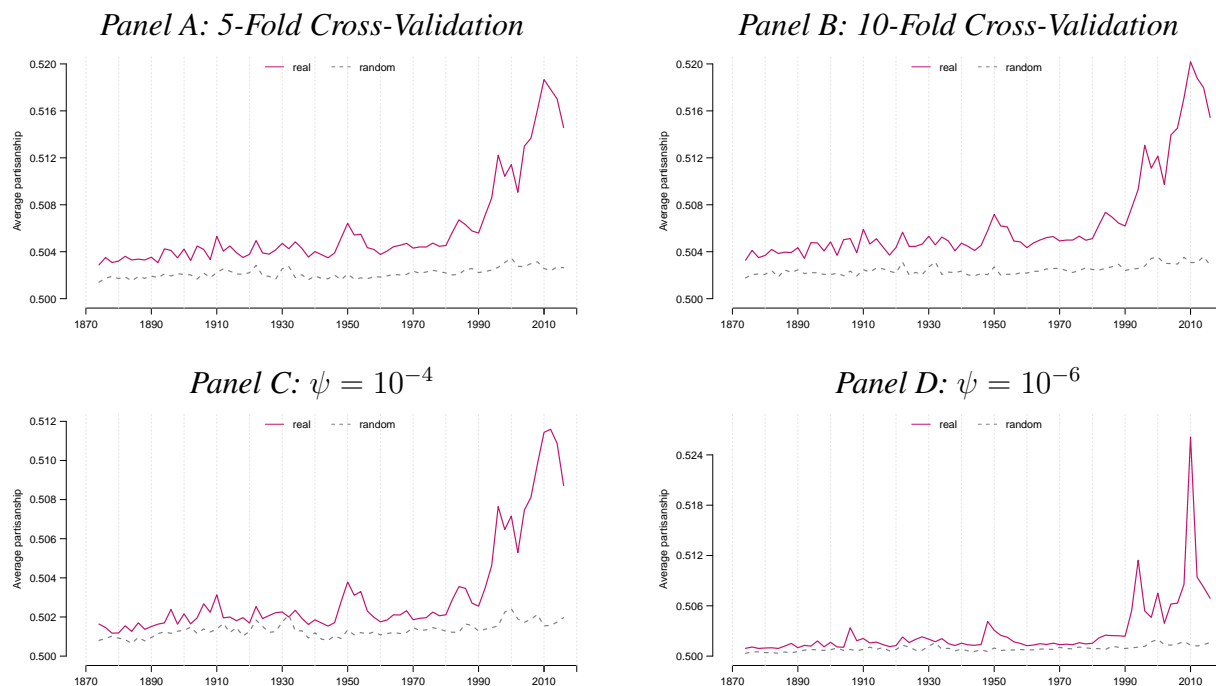


Panel B: No Covariates Specification



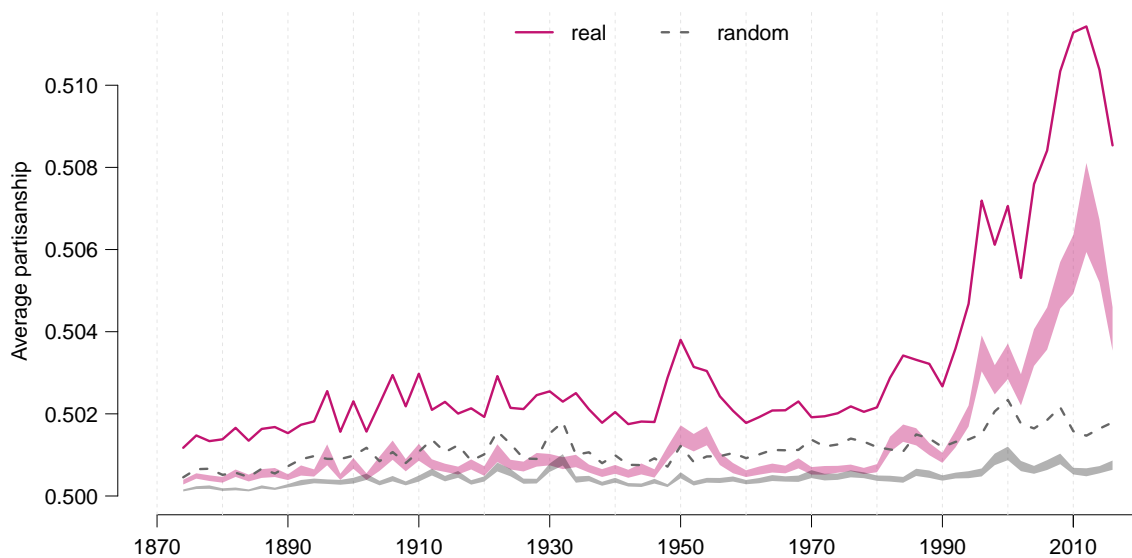
Notes: Figure shows parametric bootstrap results for leave-out and penalized estimates of average partisanship $\bar{\pi}_t$. We restrict attention to the data from the most frequently spoken 1,000 phrases for computational reasons. We begin by computing our penalized estimates on the restricted data. We produce 10 bootstrap replicates by generating data for all speaker-sessions using the estimated model. We then use the 10 generated datasets to compute the leave-out estimate and our preferred penalized estimate of average partisanship. Each panel shows the 10th–90th percentile range of the replicates for each of the estimates. Panel A shows results where the true series is based on our preferred specification with covariates. Panel B shows results where the true series is based on our preferred specification without covariates.

Online Appendix Figure 6: Average Partisanship of Speech Using Alternative Estimation Settings



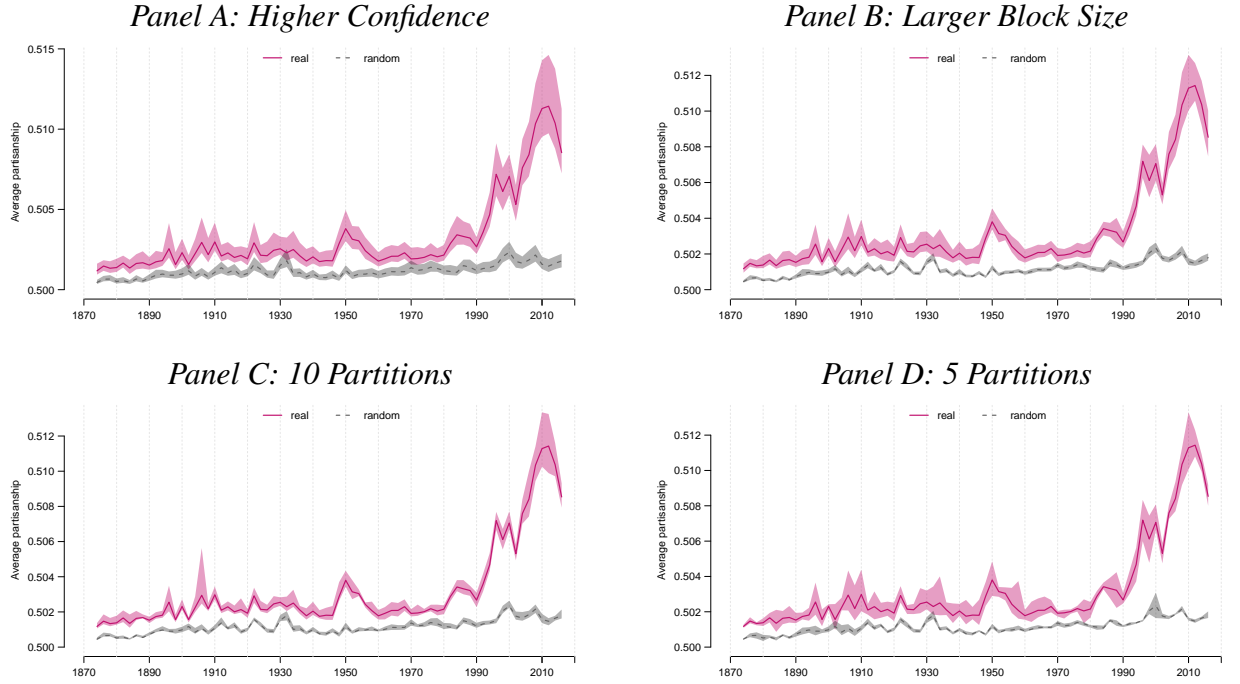
Notes: All panels show average partisanship from variants of our preferred specification in Panel B of Figure 2 in the paper. “Real” series is from actual data; “random” series is from hypothetical data in which each speaker’s party is randomly assigned with the probability that the speaker is Republican equal to the average share of speakers who are Republican in the sessions in which the speaker is active. The specifications in Panel A and B use 5-fold and 10-fold cross-validation to select the λ_j that minimizes average out-of-sample deviance for each phrase j . The specifications in Panels C and D change the fixed cost penalty ψ to the values indicated by their titles.

Online Appendix Figure 7: Average Partisanship of Speech, Bias-Corrected Penalized Estimates



Notes: Estimates are from our preferred specification in Panel B of Figure 2 in the paper. “Real” series is from actual data; “random” series is from hypothetical data in which each speaker’s party is randomly assigned with the probability that the speaker is Republican equal to the average share of speakers who are Republican in the sessions in which the speaker is active. The shaded region corresponding to each series represents a bias-corrected pointwise confidence interval obtained by subsampling (Politis et al. 1999). Specifically, we randomly draw speakers without replacement to create 100 equal-sized subsamples each containing (up to integer restrictions) one-tenth of all speakers and, for each subsample k , we compute the penalized estimate $\hat{\pi}_t^k$. Let τ_k be the number of speakers in the k th subsample and let τ be the number of speakers in the full sample. Then our confidence interval is $\frac{1}{2} + (\exp[\log(\hat{\pi}_t - \frac{1}{2}) - (Q_t^k)_{(90)}/\sqrt{\tau}], \exp[\log(\hat{\pi}_t - \frac{1}{2}) - (Q_t^k)_{(11)}/\sqrt{\tau}])$, where $(Q_t^k)_{(b)}$ is the b th order statistic of $Q_t^k = \sqrt{\tau_k} (\log(\hat{\pi}_t^k - \frac{1}{2}) - \log(\hat{\pi}_t - \frac{1}{2}))$.

Online Appendix Figure 8: Average Partisanship of Speech, Alternative Subsampling Designs



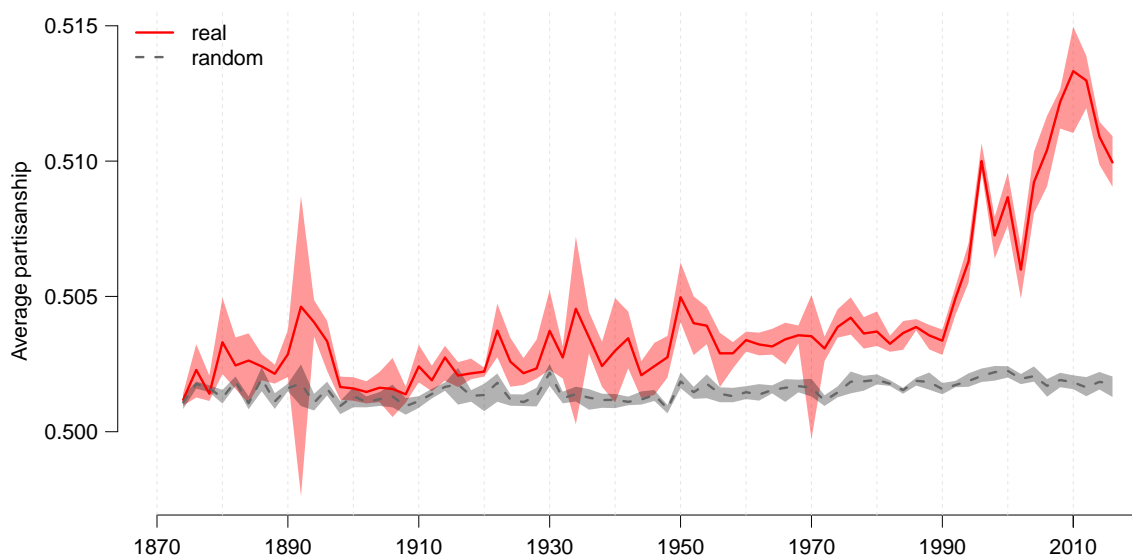
Notes: All panels show the results from our preferred specification in Panel B of Figure 2 in the paper, with inference from alternative methods of subsampling. “Real” series is from actual data; “random” series is from hypothetical data in which each speaker’s party is randomly assigned with the probability that the speaker is Republican equal to the average share of speakers who are Republican in the sessions in which the speaker is active. The shaded region around each series represents a pointwise confidence interval obtained by subsampling (Politis et al. 1999). Specifically, we randomly draw speakers without replacement to create K equal-sized subsamples and, for each subsample k , we compute the penalized estimate $\hat{\pi}_t^k$. Let τ_k be the number of speakers in the k th subsample and let τ be the number of speakers in the full sample. Our confidence interval is $\frac{1}{2} + (\exp[\log(\hat{\pi}_t - \frac{1}{2}) - (Q_t^k)_{(b_h)}/\sqrt{\tau}], \exp[\log(\hat{\pi}_t - \frac{1}{2}) - (Q_t^k)_{(b_l)}/\sqrt{\tau}])$ where $(Q_t^k)_{(b)}$ is the b th order statistic of $Q_t^k = \sqrt{\tau_k} \left(\log(\hat{\pi}_t^k - \frac{1}{2}) - \log([\frac{1}{K} \sum_{l=1}^K \hat{\pi}_t^l] - \frac{1}{2}) \right)$. Panel A corresponds to our subsampling procedure in the main paper with $K = 100$, τ_k is one-tenth of all speakers (up to integer restrictions), and order statistics $b_h = 98$ and $b_l = 3$ chosen so that our confidence interval has nominal 96 percent coverage. Panel B doubles the number of speakers in each subsample so that $K = 100$, τ_k is one-fifth of all speakers (up to integer restrictions), and uses order statistics $b_h = 90$ and $b_l = 11$. Panel C uses $K = 10$ subsamples of size one-tenth of all speakers (up to integer restrictions) with no overlap in the sets of speakers and order statistics $b_h = 9$ and $b_l = 2$. Panel D similarly uses $K = 5$ subsamples of size one-fifth of all speakers (up to integer restrictions) with no overlap in the sets of speakers and order statistics $b_h = 5$ and $b_l = 1$.

Online Appendix Figure 9: Sample-splitting Inference for Penalized Estimator

Panel A: Full Set of Phrases

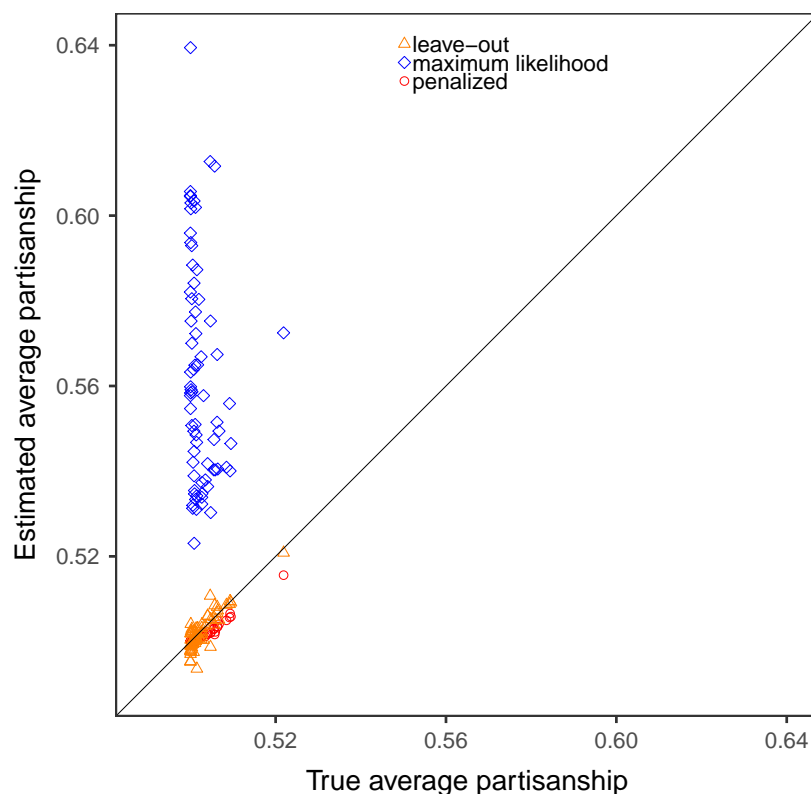


Panel B: Excluding Outlier Phrases



Notes: Figure shows confidence intervals for penalized estimates of average partisanship $\bar{\pi}_t$ based on a sample-splitting procedure. We randomly split the set of speakers into two equal-sized partitions and restrict the vocabulary to be the set of phrases spoken by at least one speaker from both partitions. We begin by computing estimates using our preferred specification on the first partition. We then estimate our minimal penalty specification (Online Appendix Figure 15) on the second partition, constraining to zero any party loading φ_{jt} that is penalized to zero in the first partition. To form confidence intervals, we perform a nonparametric bootstrap with 10 replicates. For each replicate, we draw speakers with replacement from the second partition, and re-estimate the minimal penalty specification. In each panel we plot the estimated average partisanship on the second partition, as well as the 10th–90th percentile range of the estimated average partisanship from the replicates. Panel A shows results from the full set of phrases. Panel B shows results where we exclude two outlier phrases, “judg swayn” and “creek dam,” from the calculation of average partisanship.

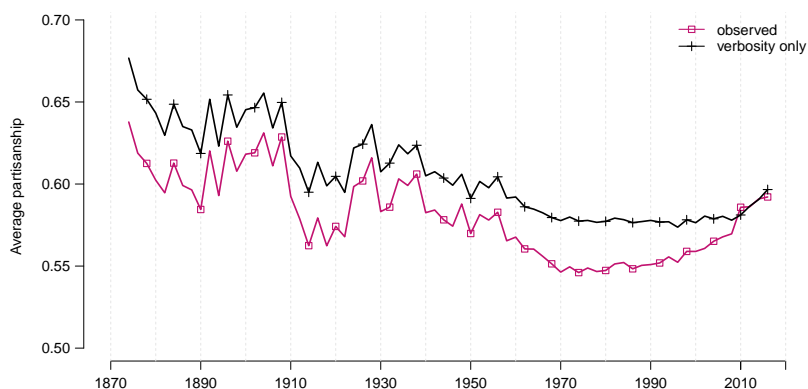
Online Appendix Figure 10: Estimated vs. True Average Partisanship, Alternative Estimators



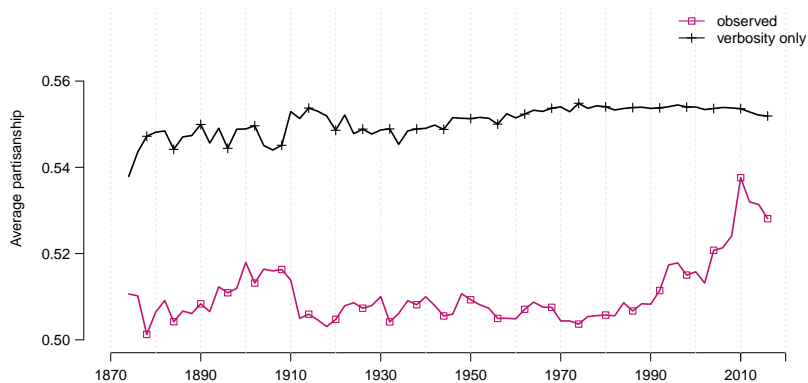
Notes: Plot shows the estimated and true average partisanship, for three alternative estimators, across sessions, based on a Monte Carlo exercise. To perform the Monte Carlo exercise, we first estimate our model with no covariates (specification in Online Appendix Figure 4A) on a subset of 1000 randomly drawn phrases using our preferred penalized estimator. We then simulated 10 datasets from the estimated model and, on each simulated dataset, we estimate average partisanship using our preferred penalized estimator, the leave-out estimator, and the maximum likelihood estimator. For each estimator and each session, the plot depicts the mean of the estimated average partisanship (y-axis) against the true average partisanship in the data-generating process used to generate the simulated dataset (x-axis). The solid line depicts the 45-degree line.

Online Appendix Figure 11: Isolating Changes in Average Partisanship Due to Changes in Verbosity

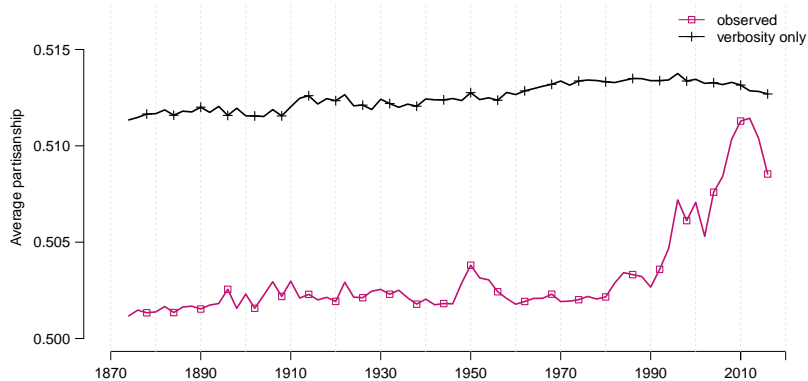
Panel A: Maximum Likelihood Estimator



Panel B: Leave-out Estimator

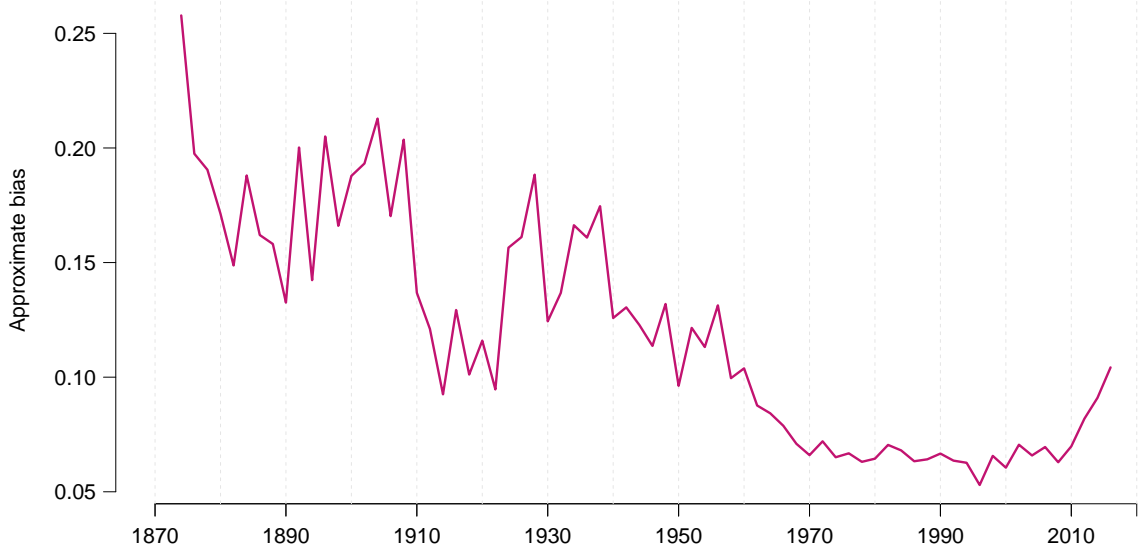


Panel C: Penalized Estimator



Notes: For each estimator, the figure shows estimates on real data (“observed”) and on data simulated from the distribution $MN(m_{it}, \hat{\mathbf{q}}_{t^*}^{P(i)})$ where t^* is the 103rd session (“verbosity only”). Thus, in the “verbosity only” series, verbosity follows its empirical distribution but phrase frequencies are held constant at their empirical values for the 103rd session. Panel A is for the maximum likelihood estimator $\hat{\pi}^{MLE}$. Panel B is for the leave-out estimator $\hat{\pi}^{LO}$. Panel C is for our preferred penalized estimator $\hat{\pi}_t^*$. These estimators are defined in Section 4 of the paper.

Online Appendix Figure 12: Approximate Bias in Maximum Likelihood Estimate of Partisanship



Notes: Plot shows the approximate bias in $\hat{\pi}_t^{MLE}$ when covariates do not matter ($\mathbf{x}_{it} := \mathbf{x}_t$) and true partisanship is $\frac{1}{2}$ (i.e., $\mathbf{q}_t^R(\mathbf{x}) = \mathbf{q}_t^D(\mathbf{x})$). We can rewrite $\hat{\pi}_t^{MLE}$ as

$$\hat{\pi}_t^{MLE} = \frac{1}{2} \sum_{j \in J_t} \frac{(\hat{q}_{jt}^R)^2 + (\hat{q}_{jt}^D)^2}{\hat{q}_{jt}^R + \hat{q}_{jt}^D}, \quad (1)$$

where J_t is the set of unique phrases spoken in session t . To approximate the bias in $\hat{\pi}_t^{MLE}$, let $c_{jt}^R = \sum_{i \in R_t} c_{ijt}$, $c_{jt}^D = \sum_{i \in D_t} c_{ijt}$ and suppose that $c_{jt}^R \sim \text{Pois}(\eta_{jt}^R)$ and $c_{jt}^D \sim \text{Pois}(\eta_{jt}^D)$. Each summand in equation (1) is independent. Taking the expectation of a first-order Taylor expansion of the right-hand-side of equation (1) around its true value gives

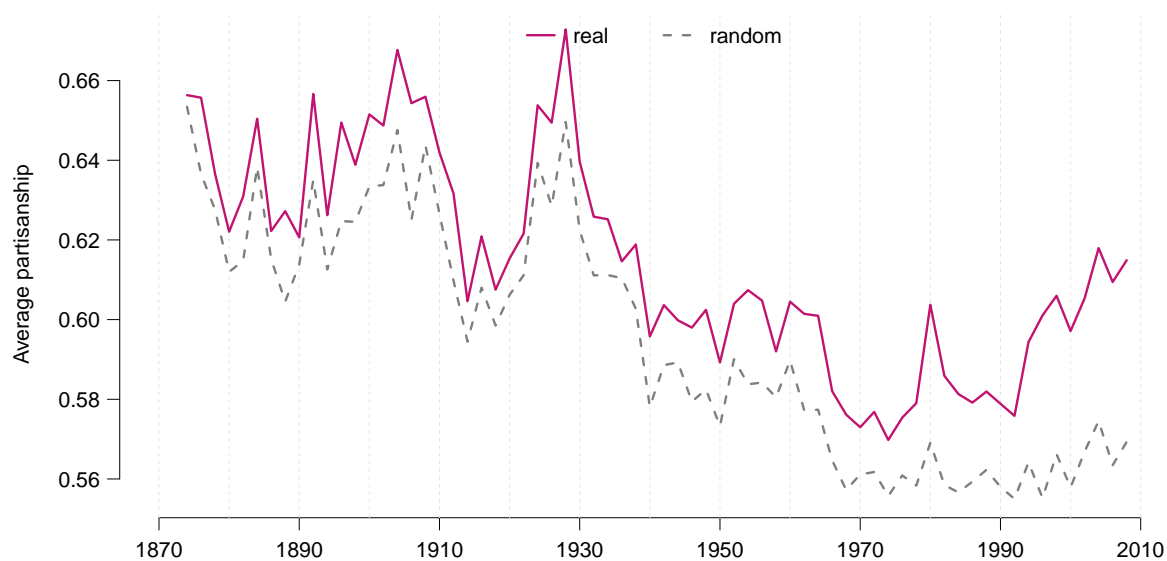
$$\mathbb{E}(\hat{\pi}_t^{MLE}) \approx \frac{1}{2} \sum_{j \in J_t} \frac{(\eta_{jt}^R + (\eta_{jt}^R)^2) / (m_t^R)^2 + (\eta_{jt}^D + (\eta_{jt}^D)^2) / (m_t^R)^2}{\eta_{jt}^R / m_t^R + \eta_{jt}^D / m_t^D},$$

where here we treat $m_t^R = \sum_j c_{jt}^R$ and $m_t^D = \sum_j c_{jt}^D$ as fixed and known. Imposing that $\eta_{jt}^R = \eta_{jt}^D$ for all j , we have that

$$\mathbb{E}(\hat{\pi}_t^{MLE}) - \frac{1}{2} \approx \frac{|J_t|}{4} \left(\frac{m_t^R + m_t^D}{m_t^R m_t^D} \right), \quad (2)$$

where the term on the right-hand side of equation (2) is the approximate bias plotted above.

Online Appendix Figure 13: Maximum Likelihood Estimate from Data in Jensen et al. (2012)



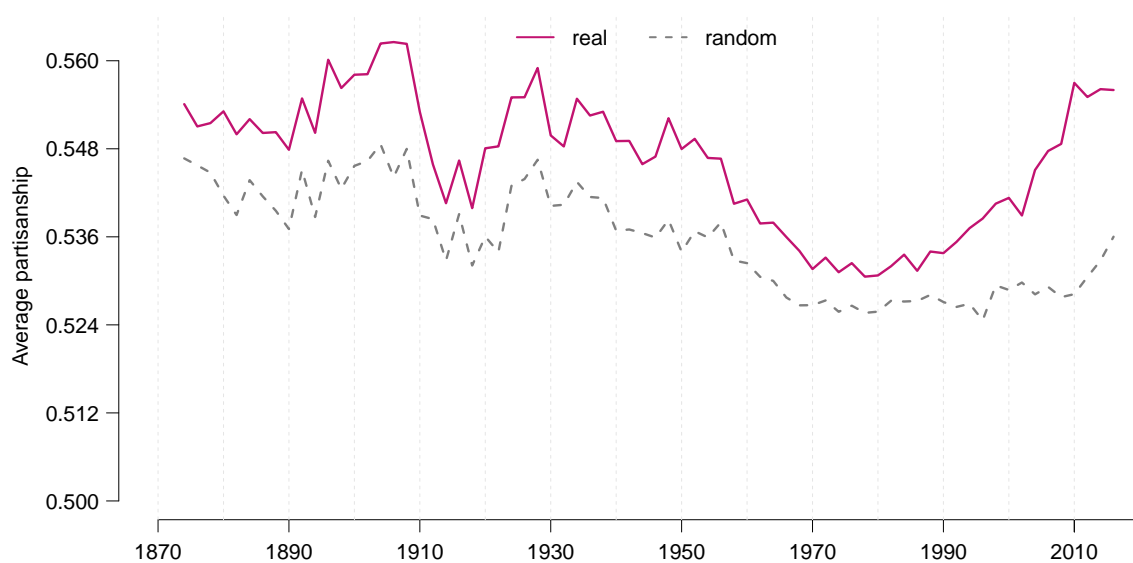
Notes: Plot shows $\hat{\pi}^{MLE}$ computed using replication data for Jensen et al. (2012). “Real” series is from actual data; “random” series is from hypothetical data in which each speaker’s party is randomly assigned with the probability that the speaker is Republican equal to the average share of speakers who are Republican in the sessions in which the speaker is active.

Online Appendix Figure 14: Maximum Likelihood Estimates with Frequently Occurring Phrases



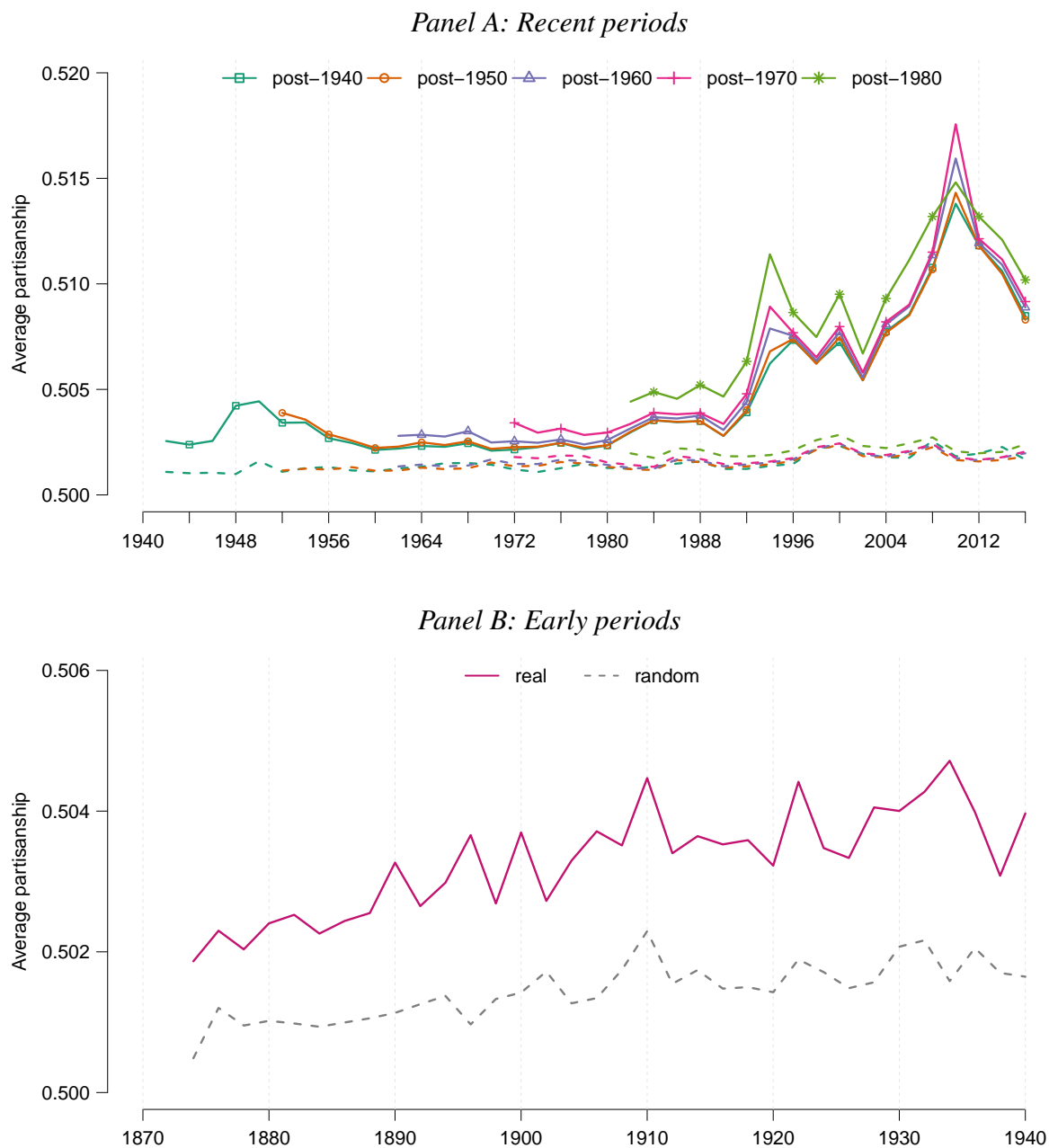
Notes: Each plot shows $\hat{\pi}_t^{MLE}$ computed on the most frequently occurring phrases. Specifically, let $m_{jt} = \sum_i c_{ijt}$. The plots on the left show $\hat{\pi}_t^{MLE}$ computed on the set of phrases with $m_{jt} > g_t$ where g_t is chosen to be the $(100 - P)$ th percentile of m_{jt} across all phrases j spoken in a given session t , for each $P \in \{90, 50, 10, 1\}$. The plots on the right show $\hat{\pi}_t^{MLE}$ computed on the set of phrases with $m_{jt} > g$ for $g \in \{5, 20, 100, 500\}$. “Real” series is from actual data; “random” series is from hypothetical data in which each speaker’s party is randomly assigned with the probability that the speaker is Republican equal to the average share of speakers who are Republican in the sessions in which the speaker is active.

Online Appendix Figure 15: Average Partisanship of Speech, Minimal Penalty ($\lambda \approx 0$)



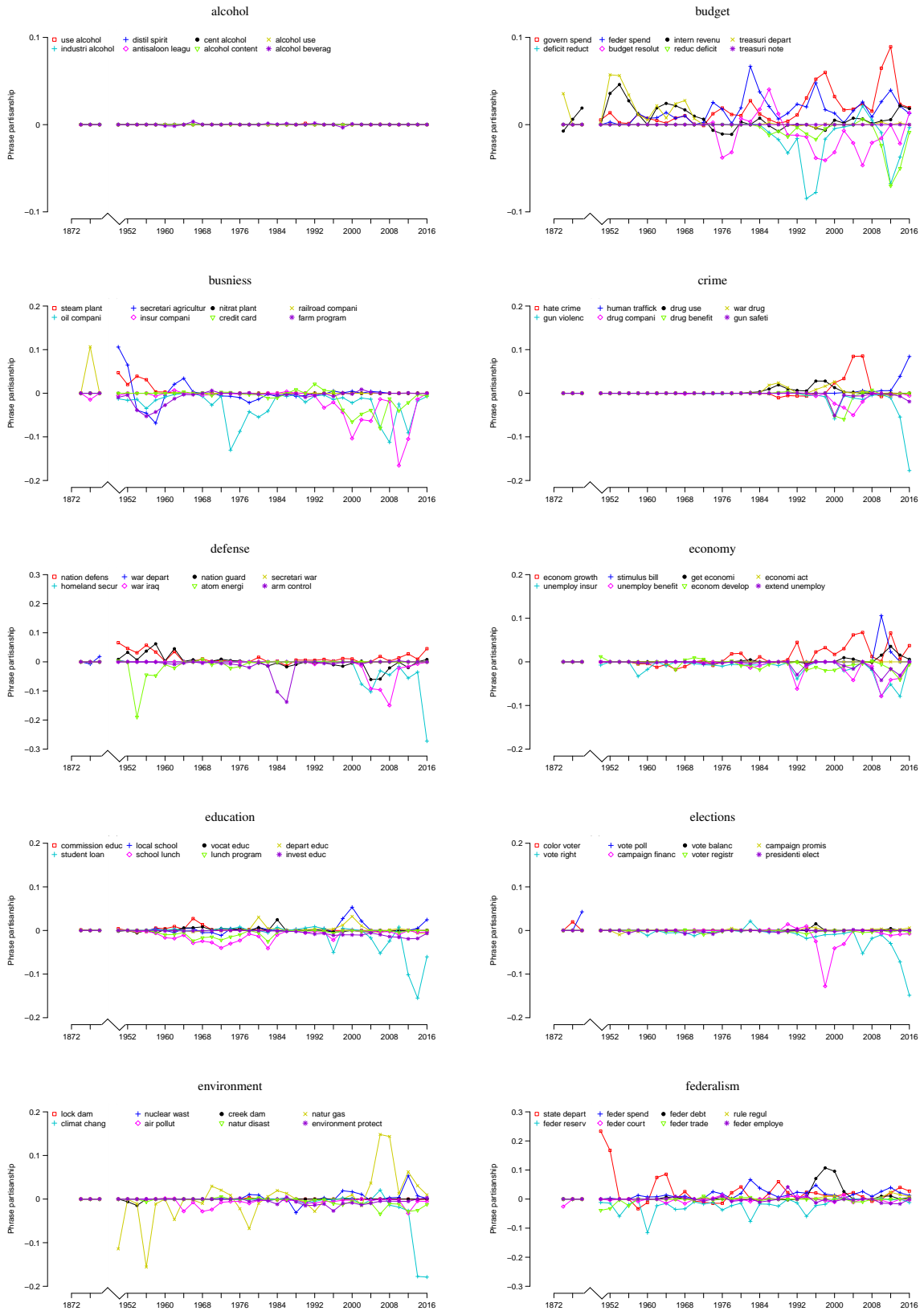
Notes: Plot shows estimated average partisanship when we modify our preferred penalized estimator to set λ_j for all j equal to its final value λ_j^G in the regularization path, which is approximately zero. “Real” series is from actual data; “random” series is from hypothetical data in which each speaker’s party is randomly assigned with the probability that the speaker is Republican equal to the average share of speakers who are Republican in the sessions in which the speaker is active.

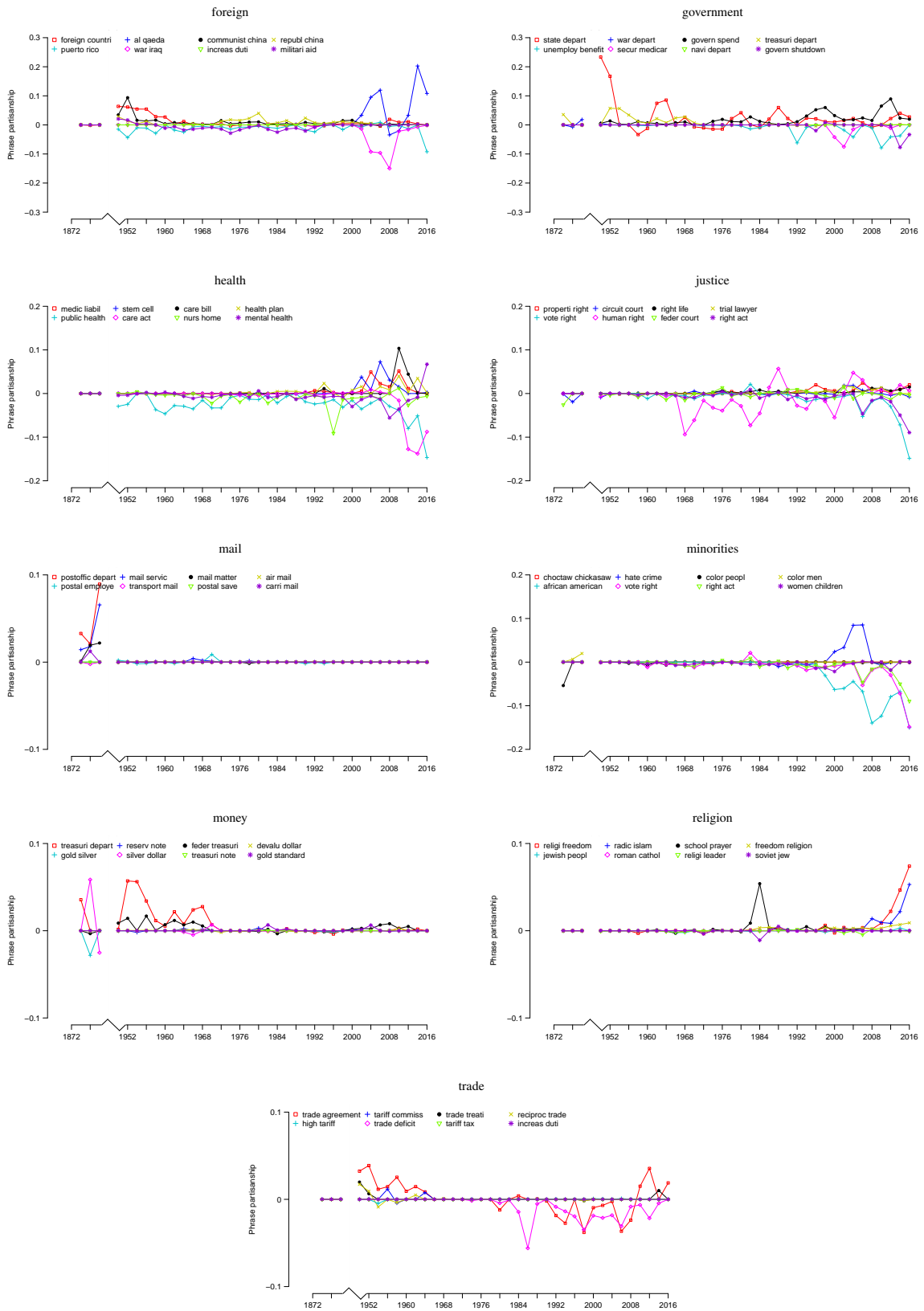
Online Appendix Figure 16: Average Partisanship of Speech Estimated on Alternative Time Frames



Notes: For each group of speaker-sessions, the solid line shows estimated average partisanship for the real data. The corresponding dashed line shows estimated average partisanship for hypothetical data in which each speaker’s party is randomly assigned with the probability that the speaker is Republican equal to the average share of speakers who are Republican in the sessions in which the speaker is active. Panel A restricts our full estimation sample to the years indicated in the legend. Panel B restricts our input data to Congresses starting before 1940 and produces a new estimation sample for those years following the steps in Section 2. The estimation sample in Panel B is a subset of Panel A’s. Both panels are based on our preferred specification in Panel B of Figure 2 in the paper.

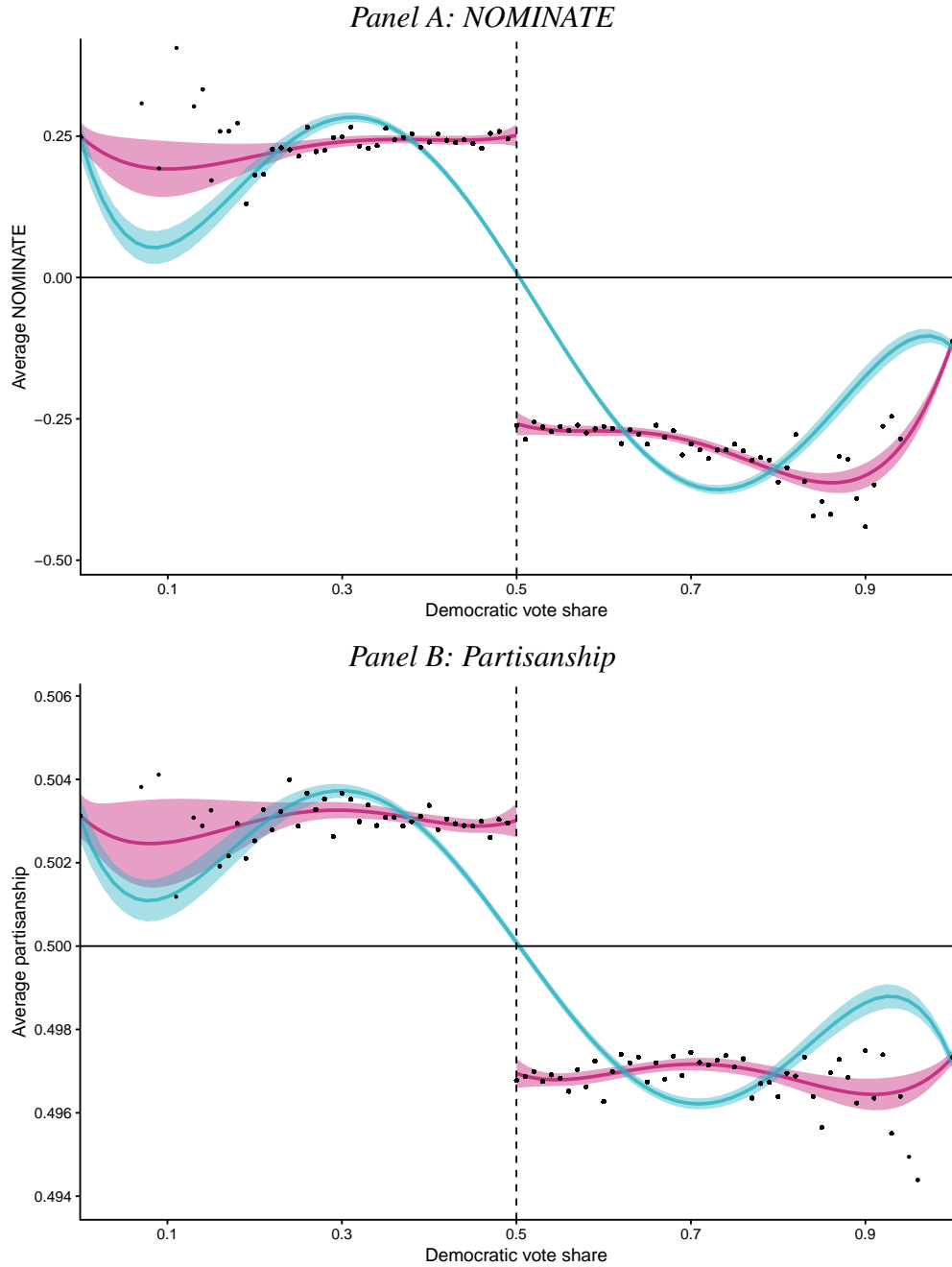
Online Appendix Figure 17: Partisanship over Time for Phrases within Topics





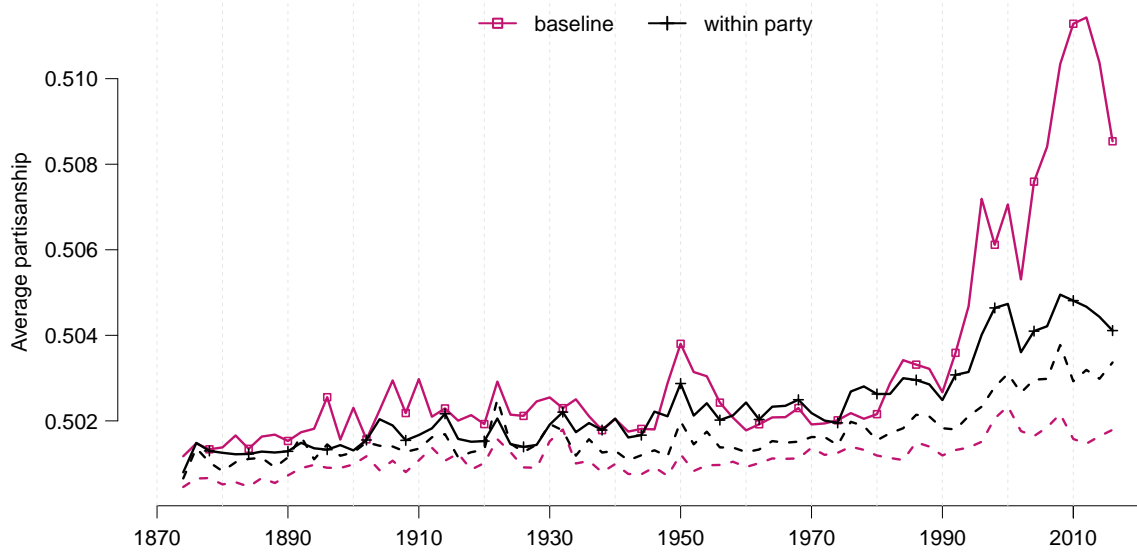
Notes: Calculations are based on our preferred specification in Panel B of Figure 2 in the paper. Each plot shows the mean estimated value of 1,000 times the partisanship ζ_{jt} for the four Republican (Democratic) phrases in a given topic that have the highest (lowest) average partisanship across all sessions. See the body of the paper for the definition of partisanship ζ_{jt} .

Online Appendix Figure 18: Discontinuity in Democratic Vote Share from Lee et al. (2004)



Notes: Panel A follows the first panel of Figure VI from Lee et al. (2004). It plots the non-Common-Space DW-NOMINATE scores (Carroll et al. 2015b) for the House representatives in a session against the Democratic vote shares in the corresponding elections. Panel B reproduces Panel A but plots each representative's posterior probability $\hat{\rho}_{it}$ of being Republican based on speech on the vertical axis. As in the main paper, we define $\hat{\rho}_{it} = \hat{\mathbf{q}}_{it} \cdot \hat{\boldsymbol{\rho}}_t^*(\mathbf{x}_{it})$, where we recall that $\hat{\mathbf{q}}_{it} = \mathbf{c}_{it}/m_{it}$ are the empirical phrase frequencies for speaker i in session t and where we define $\hat{\boldsymbol{\rho}}_t^*(\mathbf{x}_{it})$ as the estimated value of $\boldsymbol{\rho}_t(\mathbf{x}_{it})$ from our baseline penalized estimates. Both panels are restricted to representatives in our main speech sample who serve between 1946 and 1992, have a non-missing DW-NOMINATE score in session t , and which we are able to match to election outcomes (King 1995). As in Lee et al. (2004), each dot is the average of the variable on the vertical axis within 0.01 bins of Democratic vote share. Fourth-order polynomials and their associated 95 percent pointwise confidence intervals are separately fit on either side of the discontinuity at 0.5. A fifth-order polynomial and its associated 95 percent pointwise confidence interval is fit over the entire sample. The horizontal line marks the neutral value of the variable on the vertical axis.

Online Appendix Figure 19: Average Partisanship of Speech between and within Parties



Notes: Figure provides our preferred estimate of average partisanship (from Panel B of Figure 2 in the paper) and our estimate of “within-party partisanship.” In producing the latter, we define the set of extreme congresspeople in session t , denoted by E_t , as the set containing each speaker in session t with a first-dimension Common Space DW-NOMINATE score from Carroll et al. (2015a) that is more extreme than the median score of that speaker’s party in session t ; we define the moderates M_t as the set of all other speaker-sessions in t . We estimate within-party partisanship using estimated coefficients from the adjusted model:

$$u_{ijt} = \alpha_{jt} + \mathbf{x}'_{it} \boldsymbol{\gamma}_{jt} + \varphi_{jt}^{RE} \mathbf{1}_{i \in R_t \cap E_t} + \varphi_{jt}^{RM} \mathbf{1}_{i \in R_t \cap M_t} + \varphi_{jt}^{DE} \mathbf{1}_{i \in D_t \cap E_t} + \varphi_{jt}^{DM} \mathbf{1}_{i \in D_t \cap M_t}.$$

where φ_{jt}^{RE} , φ_{jt}^{RM} , φ_{jt}^{DE} , and φ_{jt}^{DM} are phrase-time-specific group loadings. Here, the groups are given by the two-way interaction of E_t and R_t in the natural way. We estimate the model’s coefficients using our baseline penalized estimator and use the results to compute our phrase-frequency vector estimates $\hat{\mathbf{q}}_t^{RE}(\mathbf{x}_{it})$, $\hat{\mathbf{q}}_t^{RM}(\mathbf{x}_{it})$, $\hat{\mathbf{q}}_t^{DE}(\mathbf{x}_{it})$, and $\hat{\mathbf{q}}_t^{DM}(\mathbf{x}_{it})$. We use these, in turn, to estimate in-party analogues of $\boldsymbol{\rho}_t(\mathbf{x}_{it})$. For each Republican $i \in R_t$, for instance, we compute

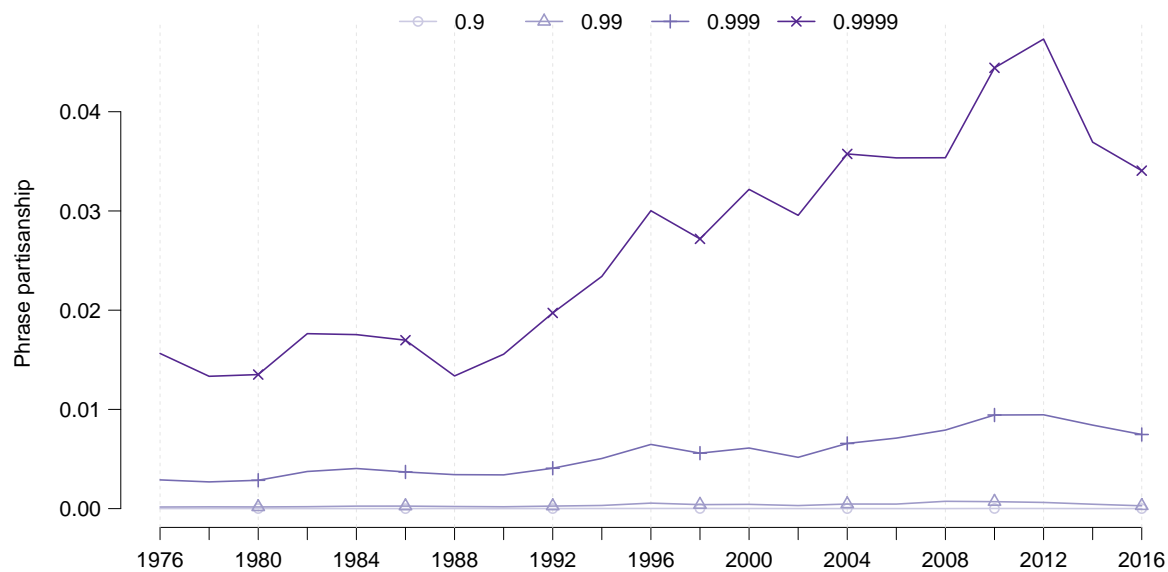
$$\hat{\boldsymbol{\rho}}_{jt}^R(\mathbf{x}_{it}) = \frac{\hat{\mathbf{q}}_{jt}^{RE}(\mathbf{x}_{it})}{\hat{\mathbf{q}}_{jt}^{RE}(\mathbf{x}_{it}) + \hat{\mathbf{q}}_{jt}^{RM}(\mathbf{x}_{it})}$$

for each phrase to estimate $\boldsymbol{\rho}_t^R(\mathbf{x}_{it})$. We define average Republican within-party partisanship for session t as

$$\bar{\pi}_t^R = \frac{1}{|R_t|} \sum_{i \in R_t} [\mathbf{q}_t^{RE}(\mathbf{x}_{it}) \cdot \boldsymbol{\rho}_t^R(\mathbf{x}_{it}) + \mathbf{q}_t^{RM}(\mathbf{x}_{it}) \cdot (1 - \boldsymbol{\rho}_t^R(\mathbf{x}_{it}))]$$

and estimate this quantity by replacing objects with their empirical analogues. We proceed analogously for the Democrats, and we estimate the model after randomly assigning each speaker-session to either the extreme or moderate wing of the speaker’s own party to produce the random series. Last, we form the within-party partisanship series shown in the figure as the average of the Republican and Democratic series.

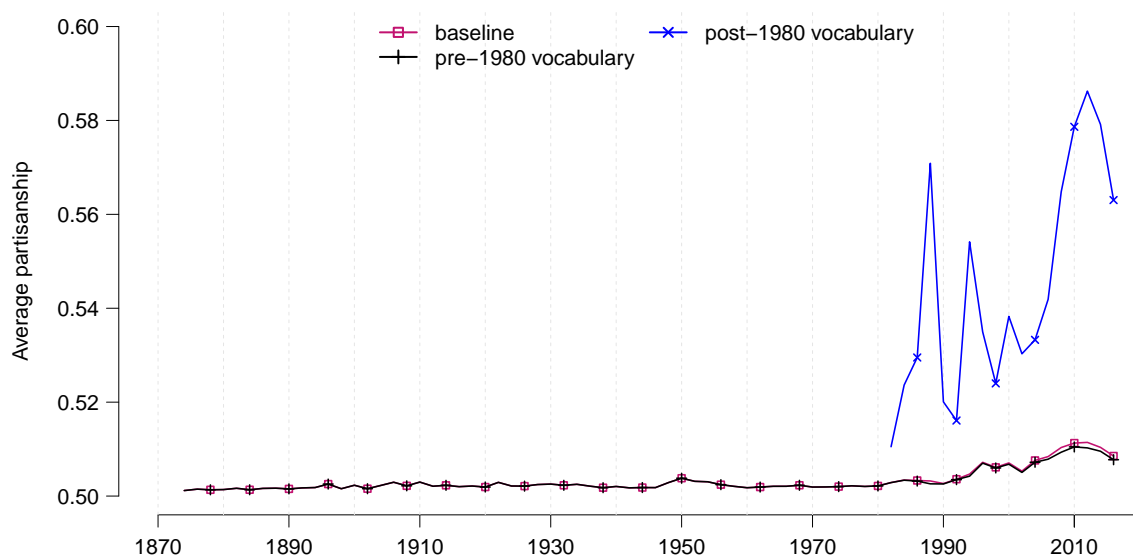
Online Appendix Figure 20: Distribution of Phrase Partisanship



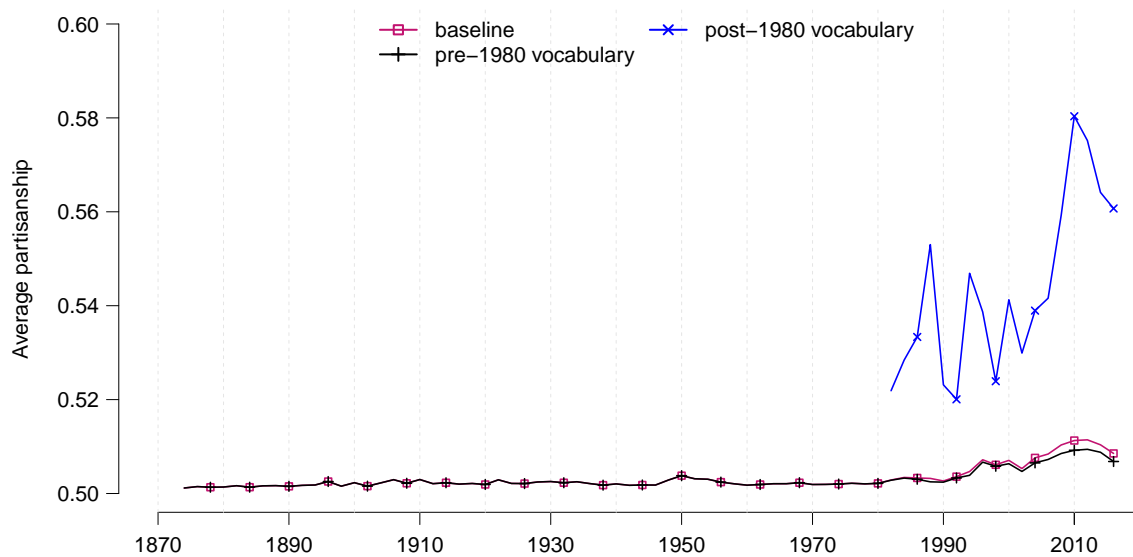
Notes: Calculations are based on our preferred specification in Panel B of Figure 2 in the paper. The solid lines denote the Q th quantile of 1,000 times the absolute value of estimated phrase partisanship in a session ζ_{jt} , as defined in the paper.

Online Appendix Figure 21: Evidence on the Role of Neologisms

Panel A: First occurrence after 1980



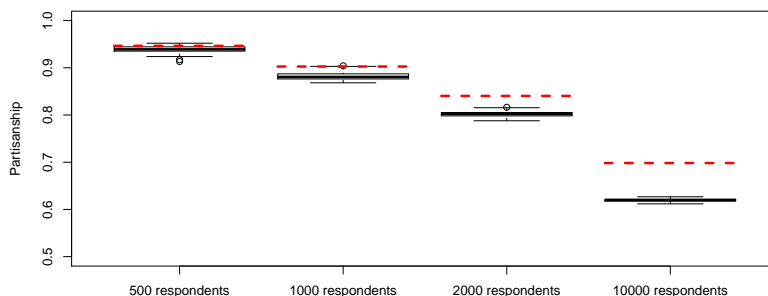
Panel B: 99 percent of occurrences after 1980



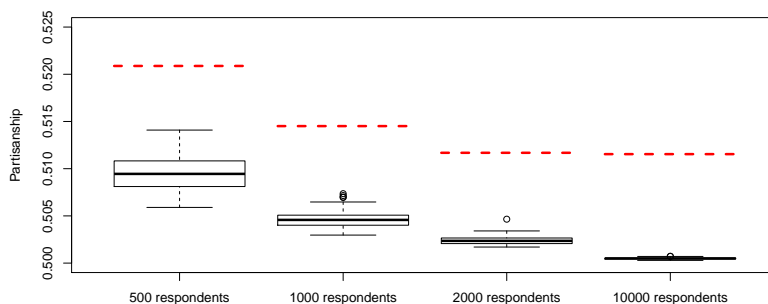
Notes: Panel A presents the partisanship of phrases first occurring after 1980. The “baseline” series is our preferred estimate of average partisanship from Panel B of Figure 2. The other two series are based on the same parameter estimates. The “pre-1980 vocabulary” series recomputes average partisanship exclusively on phrases spoken at least once during or prior to 1980 (the 96th session), while the “post-1980 vocabulary” does so for phrases only spoken after 1980. Panel B is analogous to Panel A. The “post-1980 vocabulary” series recomputes average partisanship exclusively on the set of phrases such that 99 percent of the occurrences of the phrase are after 1980 (the 96th session), while the “pre-1980 vocabulary” does so for the rest of phrases.

Online Appendix Figure 22: Plug-in Estimates of Average Partisanship for Different Choice Dimensions

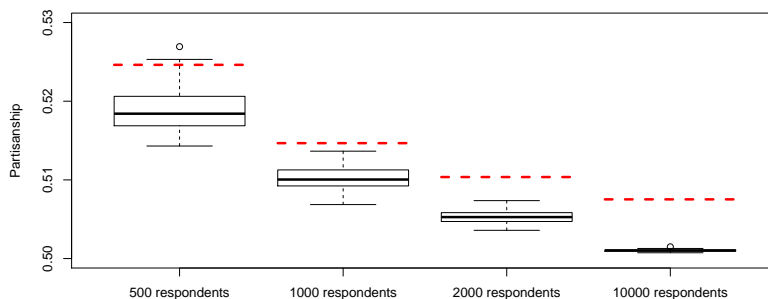
Panel A: Residential Location



Panel B: Websites



Panel C: TV Programs



Notes: We use GfK Mediamark Research & Intelligence (MRI) 2016 survey data to obtain plug-in estimates of average partisanship for different choice dimensions including residential location, websites visited, and TV programs watched. In Panel A, the dashed, red horizontal lines indicate the plug-in estimates of average partisanship in residential location choice for a sample of 500, 1000, 2000, 10000 respondents. For a given size of respondents, the boxplots show the distribution of the plug-in estimates for the simulated data in which each respondent's party is randomly assigned with the probability that the respondent is Republican equal to the average share of respondents who are Republican among respondents who identify themselves as either Republican or Democratic. Panels B and C show similar plots with the choice dimensions of websites visited and TV programs watched, respectively.

© GfK 2016. The data herein derives from a confidential, proprietary syndicated product owned by GfK US, LLC, the GfK MRI division ("GfK").