

The U.S. House Committee Votes Dataset*

Josh M. Ryan[†]
Department of Political Science
Utah State University

Ryan Bell[‡]
Department of Politics
Princeton University

Abstract

The U.S. House Committee Votes Dataset is a new dataset of roll call votes taken within House standing committees. The data begin with the 104th Congress and extend through the 114th Congress, though future updates are planned. We provide a brief history of committee votes since 1970, discuss the data collection process, the structure of the data, and limitations. Using the dataset, we conduct a brief examination of majority party agenda control within committees. We find majority and minority party roll rates are, on average, similar to those observed on the floor, but with significant heterogeneity across committees and across time. Our analyses examine the effects of majority party size and legislator preferences on roll rates within committees. The results show that roll rates may not necessarily be due to majority party control of the committee's agenda.

Word Count: 4,330

*Working Paper, comments welcome. Please cite this paper when using the dataset. This project was funded by the Social Science Research Council's "Negotiating Agreement in Congress Research Grants."

[†]0725 Old Main Hill, Logan, UT 84322. josh.ryan@usu.edu.

[‡]001 Fisher Hall, Princeton, NJ 08540. rwbell@princeton.edu.

Standing committees remain the center of legislative activity in Congress by engaging in oversight, developing legislation, and processing introduced bills prior to floor consideration (Curry 2019). Despite their importance, committee activity has been less amenable to empirical analysis because, among other reasons, record-keeping has been inconsistent and less accessible to congressional scholars. As a result, a wide variety of theoretical questions about the operation and effects of congressional committees, and about how individual members behave within them, remain unanswered.

The U.S. House Committee Votes Dataset, available online at ANONYMIZED FOR REVIEW, illuminates one aspect of committee activity: roll call votes taken during the markup of legislation. These data will allow researchers to test a variety of theories that heretofore have not been subject to quantitative analysis. For example, scholars may be interested in the extent to which the majority party controls the committee's agenda (Bussing and Treul Forthcoming), the ideological extremity of individual committee members, the partisanship of committees, or whether legislators are consistent in their voting behavior from the committee to the floor (Hamm 1982; Maltzman 1995; Unekis 1978).

We provide a brief history of recorded committee votes, explain the data collection strategy and the structure of the data. As an application, we aggregate individual member votes within committees to the vote-level and explore how successful the majority party is in controlling committee agendas. We find that roll rates of the majority and minority parties vary across both committees and congresses, suggesting the power of the majority party to control the agenda varies based on party size and the distribution of preferences.

Background on Recorded Committee Votes

The Legislative Reorganization Act (LRA) of 1970 introduced a number of reforms intended to make Congress more transparent, including the introduction of electronic voting (implemented in 1973), a requirement that committee meetings be open to the public¹, and that committee votes be "made available". Specifically, the LRA requires that committee votes be recorded, open for public inspection, that information on each vote include a description of the issue, and that members' individual roll call vote choice be listed (see Appendix A for more details). The law also requires that a vote on the motion to report the bill be printed in the committee report. Even after enactment of the LRA, House Rules only required votes to be kept in committee offices, and it remained very difficult to systematically collect committee vote data. Unekis (1978) was able to access some votes from 1971

¹Unless a majority of committee members vote to close the meeting.

to 1974 kept in offices by staff, and that *most* committees had separated the votes from committee minutes, but he was required to transcribe—by hand—the votes for the committees he examined (see footnote 8 in Unekis 1978). Using this approach, he was able to collect data on 128 total votes across 21 committees from 1971-1974.

The Republican leadership in the 104th Congress called for “accountability for committee votes” and adopted rules that required committee reports to report roll call votes taken during the markup process. This rule was likely adopted to give the party leadership better oversight of committee activities, and increase party discipline within committees; every subsequent Congress has adopted a similar rule (House Rule XIII, Clause 3(b), see Appendix A). While some committees were reporting roll call votes within committee reports prior to the 104th, no House rule required it, nor did the LRA, with the exception of motions to report; even this requirement seems to have been unenforced.²

Data Collection Process

In recent congresses a report typically has a table of contents listing the page on which committee votes are recorded and a separate section listing all votes taken in committee. The committee votes section in the report has a brief introduction, a description of each vote, and the results. Figure 1 shows an example of how the votes are indicated in the report’s table of contents (left panel), and then in the report language itself (right panel).

Python was used to scrape committee reports from `Congress.gov`, then parse the text from the report itself into a usable format. The Python script first uses regular expressions to separate a report into its component sections. The script then identifies a report’s section—if any—that contains committee votes by using regular expressions on the section headings to test for variations on phrases such as “votes of the committee” or “committee roll call votes”. If found, the script then uses additional regular expressions to further separate the section containing committee votes into a list of the different votes taken by the committee.

For each vote, the script then captures the information describing the vote (e.g., a motion, amendment, final passage). The report also states whether the vote was a roll call vote or non-roll call, such as a voice vote or adoption by unanimous consent. If the vote is a roll call vote, the script then parses the lists or tables of committee member votes. While the text formatting for votes can vary across reports, committees, and Congress, there is a limited set of formats used to provide votes in the reports.

²As a result, the dataset begins with the 104th Congress.

Figure 1: Committee Report 114-660, Accompanying H.R. 4854, “Supporting America’s Innovators Act of 2016”

Mr. UPTON, from the Committee on Energy and Commerce,
submitted the following

R E P O R T
together with
DISSENTING VIEWS
[To accompany H.R. 1770]

[Including cost estimate of the Congressional Budget Office]

The Committee on Energy and Commerce, to whom was referred the bill (H.R. 1770) to require certain entities who collect and maintain personal information of individuals to secure such information and to provide notice to such individuals in the case of a breach of security involving such information, and for other purposes, having considered the same, report favorably thereon with an amendment and recommend that the bill as amended do pass.

CONTENTS

| | Page |
|---|------|
| Purpose and Summary | 9 |
| Background and Need for Legislation | 9 |
| Hearings | 10 |
| Committee Consideration | 10 |
| Committee Votes | 10 |
| Committee Oversight Findings | 17 |
| Statement of General Performance Goals and Objectives | 17 |
| New Budget Authority, Entitlement Authority, and Tax Expenditures | 17 |
| Earmark, Limited Tax Benefits, and Limited Tariff Benefits | 17 |
| Committee Cost Estimate | 17 |
| Congressional Budget Office Estimate | 17 |

COMMITTEE VOTES

Clause 3(b) of rule XIII of the Rules of the House of Representatives requires the Committee to list the record votes on the motion to report legislation and amendments thereto. The sole record vote in committee was a motion by Chairman Hensarling to report the bill favorably to the House as amended. That motion was agreed to by a recorded vote of 57 yeas to 2 nays (Record vote no. FC-110), a quorum being present.

Note: Left panel shows how committee votes are usually displayed in table of contents, right panel shows how committee votes section is usually displayed in body of report. Highlighting in left panel by authors.

Each of these roll call formats have distinctive features that allow the script to identify how to parse a given vote. Once the format of the vote is identified, the script identifies members voting yea, nay, and present. The yea and nay counts are then compared against the totals listed in the vote description to ensure accuracy.

In some cases, the vote table itself is contained within the committee report as an image document. The script identifies these roll calls and saves the image files for these votes. We were not able to develop an effective way of parsing these images, so they were coded by hand by research assistants. There are approximately 510 votes contained in image files that were hand coded out of more than 9,800 total votes. Some other votes were hand-coded for idiosyncratic reasons including poor text quality that prevented an accurate OCR process, or a table format that was not easily parsed by the Python script. Random sampling of these hand-coded votes was performed to check for accuracy. Figure 2 shows how votes are contained within committee reports, with the left panel displaying the table style, while the right panel shows the text style.

Figure 2: Example Committee Roll Call Vote Formats

Committee on Natural Resources
U.S. House of Representatives
114th Congress

Date: 09.08.16

Recorded Vote: #1

FC Mark Up on 4 bills: **Grijalva_037 Amendment to H.R. 3764 (Rep. Rob Bishop)**, To provide that an Indian group may receive Federal acknowledgment as an Indian tribe only by an Act of Congress, and for other purposes. "Tribal Recognition Act of 2015"

| MEMBERS | | | Yes | No | Pres | MEMBERS | | | Yes | No | Pres |
|----------------------------------|---|---|-----|----|------|--------------------|--|--|-----|----|------|
| Mr. Bishop, UT, Chairman | | | | X | | Mr. LaMalfa, CA | | | | X | |
| Mr. Grijalva, AZ, Ranking Member | X | | | | | Mrs. Dingell, MI | | | | | |
| Mr. Young, AK | | | | | | Mr. Denham, CA | | | | X | |
| Mrs. Napolitano, CA | X | | | | | Mr. Gallego, AZ | | | X | | |
| Mr. Gohmert, TX | | X | | | | Mr. Cook, CA | | | | X | |
| Mrs. Bordallo, Guam | X | | | | | Mrs. Capps, CA | | | X | | |
| Mr. Lamborn, CO | | X | | | | Mr. Westerman, AR | | | | | |
| Mr. Costin, CA | X | | | | | Mr. Polis, CO | | | X | | |
| Mr. Wittman, VA | | X | | | | Mr. Graves, LA | | | | X | |
| Mr. Sablan, CNMI | X | | | | | Mr. Clay, MO | | | X | | |
| Mr. Fleming, LA | | X | | | | Mr. Newhouse, WA | | | | X | |
| Mrs. Tsongas, MA | X | | | | | Mr. Zinke, MT | | | | | |
| Mr. McClintock, CA | | X | | | | Mr. Hice, GA | | | | X | |
| Mr. Pinarviti, Puerto Rico | | | | | | Mrs. Radevagen, AS | | | | X | |
| Mr. Thompson, PA | | | | | | Mr. MacArthur, NJ | | | | X | |
| Mr. Huffman, CA | | | | | | Mr. Mooney, WV | | | | X | |
| Mrs. Lummis, WY | | | | | | Mr. Hardy, NV | | | | X | |
| Mr. Ruiz, CA | X | | | | | Mr. LaHood, IL | | | | X | |
| Mr. Benishel, MI | | X | | | | | | | | | |
| Mr. Loeventhal, CA | X | | | | | | | | | | |
| Mr. Duncan, SC | | X | | | | | | | | | |
| Mr. Corvino, PA | X | | | | | | | | | | |
| Mr. Gosar, AZ | | X | | | | | | | | | |
| Mr. Beyer, VA | | | | | | | | | | | |
| Mr. Labrador, ID | | | | | | | | | | | |
| Mrs. Torres, CA | X | | | | | TOTALS | | | 14 | 20 | |

dismissing the election contest against loretta sanchez

The Committee on House Oversight, having had under consideration the resolution H. Res. 355, dismissing the election contest against Loretta Sanchez, reports the same to the House with the recommendation that the resolution be agreed to.

committee action

On February 4, 1998, by a vote of 8-1, a quorum being present, the Committee agreed to a motion to report the resolution favorably to the House. Yeas: Mr. Thomas, Mr. Ney, Mr. Ehlert, Mr. Boehner, Ms. Granger, Mr. Gejdenson, Mr. Hoyer, Ms. Kilpatrick. Nay: Mr. Mica.

Note: Left panel shows table style committee vote from Committee Report 114-847, Accompanying H.R. 3764, "Tribal Recognition Act of 2015". Right panel shows text style committee vote from Committee Report 105-416, Accompanying H.Res. 355, "Dismissing the election contest against Loretta Sanchez".

Committee reports are numbered sequentially within a Congress and not all reports are relevant to the dataset (e.g., end-of-the-year summaries such as committee activity reports, or conference committee reports.) Even the vast majority of reports related to bills contain no roll call votes or only voice votes. Overall, about 30% of all committee reports issued within a congressional term contain roll call votes, with more recent congresses having a higher percentage of reports with recorded votes. Appendix B provides additional technical details on the scraping and parsing of roll call votes and voice votes from committee reports.

There are 2,638 legislative items in the dataset across the 11 congresses in the sample. The vast majority are House bills (H.R.), with only eight Senate bills for which a committee reported the bill with roll call votes. We speculate that important or controversial Senate legislation referred to committee is instead reported through a House legislative vehicle as either a companion H.R. bill or through the inclusion of its text in another bill (Kirkland and Kroeger 2018; Wilkerson, Smith and Stramp 2015). Most House Resolutions (H. Res.) in the dataset are the rules for consideration attached to another piece of legislation as reported by the Rules Committee. A variable in the dataset captures the bill to which the resolution is attached.

Structure of the Dataset

The completed dataset is composed of legislator-vote choices, nested within a vote number by committee report. Because the committee reports are the means by which roll call votes were collected, all votes are nested within a committee report. And, because all committee reports are associated with a bill, votes are also nested within bills, though a bill may have more than one committee report if, for example, the bill was referred to multiple committees. Because of this, we recommend researchers view the report number, not the bill number, as the appropriate nesting unit. Each vote within a committee report-congress is sequentially numbered, though these numbers may not match the vote number printed in the committee report. Some committees include voice votes in the vote number, number votes within the committee rather than within the report, or include vote results from other committees.

Votes are sometimes taken in a different committee than issues the report. Most notably, the Budget Committee, which compiles the budget resolution, sometimes reports reconciliation bills and issues a report in which votes from other committees are included. For example, the Sequester Replacement

Act of 2012 (H.R. 5652) required six committees, “to achieve specified amounts of deficit reduction from programs within their jurisdictions (House Report 112-740, 3).” These reconciliation bills are, in essence, a combination of text developed by other standing committees. In these cases, the vote is assigned to the committee which took it, not the committee which produced the report. Users of the data would not be aware that the vote was included in a committee report issued by a different committee unless they examined the reporting committee for the bill (which is not included in the data). This also sometimes occurs in rules votes embedded in Rules Committee reports, and in a few other idiosyncratic cases.

Figure 3: Structure of the Dataset

| Congress | icpsr | Last Name | Report Number | Bill Type | Bill Number | Vote Number | Vote as Scraped | Vote |
|----------|-------|-----------|---------------|-----------|-------------|-------------|-----------------|------|
| 114 | 10713 | CONYERS | 114-235 | HR | 9 | 1 | yea | 1 |
| 114 | 10713 | CONYERS | 114-235 | HR | 9 | 2 | nay | 0 |
| 114 | 10713 | CONYERS | 114-235 | HR | 9 | 3 | yea | 1 |
| 114 | 10713 | CONYERS | 114-235 | HR | 9 | 4 | nay | 0 |
| 114 | 10713 | CONYERS | 114-8 | HR | 285 | 1 | yea | 1 |
| 114 | 10713 | CONYERS | 114-228 | HR | 348 | 1 | yea | 1 |
| 114 | 10713 | CONYERS | 114-228 | HR | 348 | 2 | yea | 1 |
| 114 | 10713 | CONYERS | 114-228 | HR | 348 | 3 | yea | 1 |
| 114 | 10713 | CONYERS | 114-228 | HR | 348 | 4 | yea | 1 |
| 114 | 10713 | CONYERS | 114-228 | HR | 348 | 5 | yea | 1 |

Note: Ten observations from the dataset, sorted by legislator, showing how votes are nested within vote numbers and reports. “Vote as Scraped” shows the text in the report, while the “Vote” column indicates the vote as coded, where one equals “yea” and 0 equals “nay”. See Appendix C for more details.

The current dataset, for the 104th through 114th Congresses, contains more than 295,000 individual roll call vote choices across approximately 2,700 total committee reports with about 9,800 committee roll call votes, or approximately 30 individual member votes per roll call. See Appendix D for the number of votes by committee-congress. Figure 3 shows a ten row sample of the dataset to demonstrate its structure, with individual legislator-votes identified by vote number within a report.

Integrating the Committee Votes Data with Committee Assignments

Parsed individual votes typically identifies legislators by last name, and in some cases, states are listed for members. These individuals are identified using known committee members from Stewart’s committee data (Stewart and Woon 2016), which contain a list of legislators with their ICPSR identification codes, along with other variables, by committee-congress. Each legislator-vote whose parsed last name matches the last name of a legislator in the same committee-congress in the Stewart data

is then identified with their unique ICPSR number, which is taken from the NOMINATE data (Lewis et al. 2022).

In some cases, there are two legislators with the same last name in the same committee-congress, and either there is no state identification by legislator, or the report does have a state identification but there are two legislators with the same last name from the same state. For example, both Ann Davis and Jim Davis served on the International Relations committee in the 107th Congress. There are approximately 200 vote choices cast by members with the same last name on the same committee. Attempts were made to identify these members by state, but in many cases that information was not listed and we were not able to match these legislators with an ICPSR number from the Stewart data.

In other instances, last names were sometimes difficult to merge due to differences between the Stewart data and the committee report. This occurs because of suffixes (e.g., Sr.), errors in the OCR/parsing process, or because the names differed across the two datasets. For example, Enid Greene Waldholtz (R-UT) is listed as “Waldholtz” in the Stewart committee data, but her votes are recorded as “Greene” in the committee reports. After the initial matching process, the non-matched names were “fuzzy” matched in R. This was very effective at capturing problematic last names and fixed the vast majority of issues. After using this technique, any remaining issues were hand-corrected, which was usually a straightforward process (e.g., FROBES becomes FORBES). For example, in the 104th Congress, there were 637 total legislator-vote choices that did not match with the Stewart committee data. 570 were successfully fuzzy matched, while 66 were corrected by hand. One scraped last name could not be matched to the Stewart committee data using these techniques.

Legislator-vote choices which could not be matched to a corresponding legislator in the Stewart committee data were coded as having a missing ICPSR number and vote choice within the dataset. These include members whose last name did not match a committee member in the Stewart data, legislators on the same committee-congress with the same last name and no other identifying information (i.e., state), or legislators whose vote was not scraped correctly and could not be coded. That is, a missing ICPSR number indicates a legislator-vote choice was cast for a given committee roll call vote, but we could not determine who cast the vote, or occasionally, what the vote choice was. There are approximately 2,400 legislator-vote choices missing a vote choice, about .8% of the dataset. We are not aware of any component of the data collection process that would produce non-random missingness for these legislator-vote choices so their exclusion should not bias inferences drawn from the data.

The dataset also reports the total number of yeas and nays and the total number of votes cast for a

given vote. These values are given in the vote text within the committee report, and are matched with the scraped votes data. Rarely, these values may not match what is given in the committee report due to missing votes. Legislators who cast a vote other than yea or nay (e.g., present) or who did not vote despite serving on the committee at the time are not included in the dataset. There were approximately ten committee reports that contain votes reported in images but were not printed with the report in `Congress.gov` and are missing from the data.

Variables Included in the Dataset

The dataset contains a legislator's last name, their ICPSR number, and the committee identification number (as coded by the Stewart data) in which the vote was taken. It also includes the committee's name, though as these change across time, committee names as of the 114th Congress are used. As noted, vote choices are nested within vote numbers, committee reports, bills, and congresses. The text of the vote as scraped from the committee report is included, though this varies across committees and across congresses (e.g., yea/nay, yes/no, aye/nay, Y/N). A separate vote column classifies votes cast as a one for an affirmative vote (yea/aye/yes/y), and a zero for a negative vote (nay/no/n). The vote description, as scraped from the committee report, is also available in the dataset.

The dataset includes aggregated columns for total number of votes cast for a given vote, total number of affirmatives and negatives, and the percentage of the committee voting for or against. There are separate columns for total number of affirmative and negative votes for each party as well, along with the percentages of each party voting in favor or against. These values do not vary within votes and do not include votes with missing ICPSR numbers (and missing votes). Other variables included are taken directly from the Stewart committee data and include party, majority or minority party status, and the legislator's state, if it is available from the Stewart data. See Appendix C for a detailed description of all variables included in the dataset.

Application: Agenda Control within Committees

Majority roll rates on the House floor in a given term are less than 5% (and often less than 1%) (Cox and McCubbins 2002, 2005). Similar results have been shown in other institutional settings, and vary based on the extent to which institutional rules empower the majority party to control the agenda (Anzia and Jackman 2012; Cox, Kousser and McCubbins 2010). The low majority party roll

rate, in which a majority of the majority party votes on the losing side, is taken as evidence that the majority party successfully prevents divisive issues from reaching the floor, and more broadly, solves the collective action problems associated with heterogeneous preferences within the party.

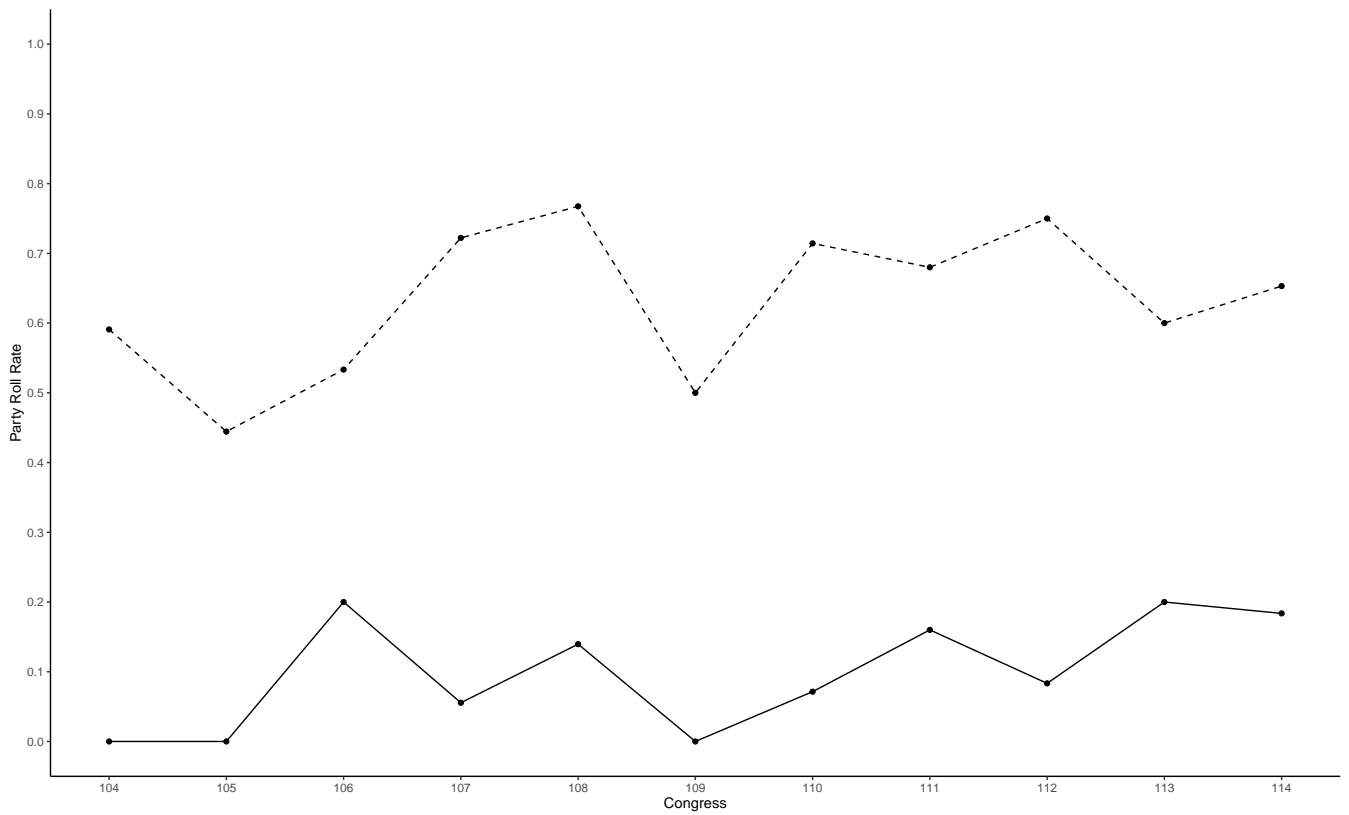
Standing committees pose an interesting test of cartel theory because votes directly affect the content of legislation, and while there is a norm of majority party agenda control (e.g., the chairman's mark, Curry 2015), there is also substantial ambiguity in the formal rules which govern the agenda. For example, in the 117th Congress, the agenda rules for Energy and Commerce require only that, "items of business...shall be provided to each member of the Committee at least 36 hours in advance of such meeting" (Rule 2(d), see Appendix D.) The lack of formal rules may empower the majority party, through the chairperson or median party member, to cartelize the agenda. Alternatively, the markup process may be more bipartisan than is commonly assumed if the minority has proposal rights. And, if this is the case, the minority may propose moderate policies which roll the majority party more frequently than on the floor.³

Majority and minority party roll rates within committees are found by aggregating individual committee votes and generating roll rates for both parties in each committee-congress. We exclude committee-congresses in which the committee took fewer than 10 votes, which removes 54 committee-congress observations for a total of 146 from the 104th through 114th Congresses (see Appendix D). Across all committees and congresses, the majority roll rate is about 6% and the minority roll rate is about 76% for a ratio of nearly 13 to 1. There is significant variation across committees, and within committees across time, however. For example, the average majority party roll rate in Appropriations is about 8%, but in the 107th and 114th Congresses the rates were 13% and 14%, respectively. The data also show that the majority party in the Rules committee was *never* rolled, while the minority party roll rate was 97%, consistent with expectations about the role of the Rules Committee in the modern House (Finocchiaro and Rohde 2008). Figure 4 shows roll rates for the majority and minority parties in a sample committee, Armed Services. Appendix Figure F1 shows roll rates for all committees within the sample.

Krehbiel (2007) suggests roll rates themselves are not dispositive evidence for majority party negative agenda control because there exists a baseline rate solely due to majority party size and the distribution of preferences across the parties. Similarly Patterson Jr. and Schwartz (2020) show that low roll rates similar to those observed in the House could occur by chance alone, and that sometimes

³Committee rules allow for a recorded roll call vote on the request of 1/5 of members.

Figure 4: Majority and Minority Party Roll Rates by Congress in the Armed Services Committee



Note: Dashed line shows roll rate for minority party, solid line shows roll rate for majority party.

the majority party has an incentive to allow bills which roll the party.⁴

Party theories suggests that as intraparty heterogeneity increases, the number of party rolls should increase (for both the majority and minority). Conversely, in a non-partisan legislature, Krehbiel (2007) expects no relationship between intraparty heterogeneity and roll rate, but does expect that greater interparty distance will increase minority party rolls and decrease majority party rolls “marginally.”⁵ Finally, a larger sized majority party will increase the roll rate for the minority and decrease the rate for the majority (also see Jackman 2013 for evidence of this claim). Table 1 summarizes the theoretical expectations and measures for each of the concepts, along with the empirical results.

Table 1: Differing Theoretical and Empirical Expectations With and Without a Majority Party Cartel on Committees

| Theoretical Concept | Theorized Effect on on Roll Rates w/ Maj. Party Cartel | Theorized Effect on on Roll Rates w/o Maj. Party Cartel | Measure | Empirical Results |
|---------------------------|--|---|--|----------------------------------|
| Intraparty heterogeneity | Minority and majority rolls increase | None | Sum of proportion of each party overlap | Support for no maj. party cartel |
| Interparty distance | None | Minority rolls increase, majority rolls increase marginally | Absolute ideal point distance between party medians on committee | Mixed |
| Proportion Majority Party | None | Minority rolls increase, majority rolls decrease | Percentage of majority party seats on committee | Support for no maj. party cartel |

Three independent variables are created from the committee votes data and follow the measures used by Krehbiel (2007). To find these ideal points, we use the “oc” package in R to create an optimal classification score for each legislator, which are similar to W-NOMINATE scores and more robust to small sample sizes (see Lo (2020) and ANONYMIZED Forthcoming for more details on creating OC scores from committee votes).

Intraparty heterogeneity is measured within a committee-congress through the sum of the proportion of Republican committee members more liberal than the most conservative Democrat and the proportion of Democratic committee members more conservative than the most liberal Republican.

⁴These situations may occur as part of voting on an overall set of bills in which the majority benefits from allowing legislators “free” votes.

⁵According to Krehbiel, the expectation for the majority is marginal because, “the rate is bounded below at zero and, under most parameter settings, low to begin with [19].”

The variable ranges from zero to 1.9 with a mean of .46 and a median of zero. Interparty heterogeneity is the absolute ideal point distance of the committee’s Republican and Democratic medians. The variable ranges from zero to 1.52 with a mean of .73 and a median of .78. Finally, the proportion of committee seats controlled by the majority party is taken from the Stewart committee data and is the total number of majority party members at the start of the term divided by the total number of majority and minority party members on the committee.⁶ The value ranges from .52 to .67, with a mean of .56 and a median of .56.

The dependent variables are majority and minority party roll rates within committee-congresses. The models are OLS regressions with clustered standard errors for committees. Committee fixed effects hold the policy agenda and other committee-level factors constant, while confounders across time within Congress, such as the party leadership or polarization, are controlled for with congressional term fixed effects.

Table 2: Predicting Minority Party Roll Rates in Committees, 104th-114th Congresses

| | Minority Party Roll Rate | | | | | |
|----------------------------------|--------------------------|-----------------|-----------------|-----------------|-----------------|-----------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Intraparty Heterogeneity | -0.02 (0.02) | | 0.01 (0.02) | -0.01 (0.02) | | 0.02 (0.02) |
| Interparty Distance | | 0.07 (0.06) | 0.07 (0.07) | | 0.08 (0.05) | 0.10 [#] (0.06) |
| Proportion Majority Party | | 2.02* (0.69) | 2.06* (0.71) | | 1.29 (0.95) | 1.33 (0.91) |
| Constant | 0.71* (0.01) | -0.49 (0.41) | -0.52 (0.43) | 0.62* (0.03) | -0.16 (0.50) | -0.22 (0.48) |
| Committee Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Congressional Term Fixed Effects | | | | Yes | Yes | Yes |
| N | 124 | 124 | 124 | 124 | 124 | 124 |
| R ² | 0.49 | 0.53 | 0.53 | 0.62 | 0.64 | 0.65 |

Note: [#]p<0.1; *p<0.05. Dependent variable is minority party roll rate, only committees with more than ten votes in a congressional term are included. Members’ OC votes used to generate intraparty heterogeneity and proportion overlap. A member will not have an OC score generated if they cast fewer than ten votes within a committee-congress.

The results for minority party roll rates are shown in Table 2. The first two columns include each of the preference measurements separately with committee fixed effects, while column three includes both preference measurements and the party size measure. Consistent with baseline expectations for a non-partisan legislature, the intraparty heterogeneity variable is not a statistically significant predictor

⁶Bernie Sanders is coded as a Democrat

of minority party rolls. In models two and three, majority party size is positive and statistically significant, also consistent with the non-partisan legislature predictions, though the result does not hold with congress fixed effects. This variable changes across committees but depends largely on the overall size of the majority within a term. Thus, including fixed effects for congress controls for majority size within the chamber, and may mute the effects of the committee-level variable. Finally, the effect of interparty distance is positive, as expected, but not statistically significant except in model six, where it is significant at the .1 level.

Table 3 shows the same models, but predicts majority party roll rates within committee-congresses. As with minority party roll rates, intraparty heterogeneity has no effect on majority party roll rates, inconsistent with predictions from partisan theories. Interparty distance, which is predicted to have a “marginal” positive effect is not statistically significant in the models, though the direction is positive. Finally, the proportion of the majority party on the committee is negative and statistically significant (at the .1 level in models 2 and 3), the opposite sign from the coefficient predicting minority party roll rates. This is evidence in support of the claim that the observed roll rates are not the result of partisan agenda control. However, as with the previous results, including congress fixed effects makes these results insignificant, though the sign remains negative.

To summarize, the null hypothesis in a non-partisan legislature that intraparty heterogeneity should affect roll rates is borne out by the data. There is weak support for hypotheses about the effect of interparty distance; with greater distance comes a higher minority party roll rate when fixed effects for congress and committee are included. The effect on majority roll rates is positive, but small and insignificant. The hypotheses about the effect of majority party size within the committee are supported. The effect is positive and significant for minority party roll rates, and negative and statistically suggestive for majority party roll rates if congress fixed effects are not included.

The overall evidence does not provide strong evidence of majority party control of committee agendas as measured by roll rates. However, the distribution of preferences and the size of the majority party on committees are themselves functions of majority party power within the chamber. It also seems likely that the majority exercises more control over the agenda in some committees than others. The party leadership stacks certain committees with certain types of members to produce more favorable policy outcomes, and perhaps that is where party power manifests, rather than through overt control of the committee’s agenda. And, perhaps some committees are more important to the majority party’s brand, as parties likely seek to balance the inherent trade-off between strong control over some

committees and weaker control over others. Indeed, the significant variation in majority and minority party roll rates across committees and congresses, despite the same institutional rules, suggests that majority party agenda control is conditional on other factors. Future research should investigate this variation in order to further inform theories of legislative organization.

Table 3: Predicting Majority Party Roll Rates in Committees, 104th-114th Congresses

| | Majority Party Roll Rate | | | | | |
|----------------------------------|--------------------------|------------------------------|------------------------------|-----------------|-----------------|-----------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Intraparty Heterogeneity | -0.01 (0.02) | | -0.02 (0.02) | -0.02 (0.02) | | -0.01 (0.02) |
| Interparty Distance | | 0.02 (0.03) | 0.002 (0.03) | | 0.02 (0.03) | 0.01 (0.03) |
| Proportion Majority Party | | -1.24 [#] (0.60) | -1.30 [#] (0.62) | | -0.39 (0.86) | -0.41 (0.88) |
| Constant | 0.18* (0.01) | 0.87* (0.35) | 0.92* (0.37) | 0.20* (0.02) | 0.39 (0.46) | 0.42 (0.48) |
| Committee Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Congressional Term Fixed Effects | | | | Yes | Yes | Yes |
| N | 124 | 124 | 124 | 124 | 124 | 124 |
| R ² | 0.34 | 0.37 | 0.38 | 0.48 | 0.48 | 0.48 |

Note: [#]p<0.1; *p<0.05. Dependent variable is majority party roll rate, only committees with more than ten votes in a congressional term are included. Members' OC votes used to generate intraparty heterogeneity and proportion overlap. A member will not have an OC score generated if they cast fewer than ten votes within a committee-congress.

Discussion

The House Committee Votes dataset offers new opportunities to better understand legislative action inside committees, allowing researchers to examine agendas, voting behavior, representation, and interest groups. We offer an initial examination of roll rates within committees, though the dataset will allow researchers to undertake more complex and detailed analyses of this question. The data currently extend from the 104th Congress to the 114th, but we plan on updating the dataset with additional committee votes at the end of each congressional term.

The House recently created a Modernization Committee to study ways of improving the operation of Congress. One of its key suggestions was to make committee votes more easily available to the public through posting vote results online, similar to how floor votes are currently posted. This would allow researchers to access voting data in future congresses, and would connect with this dataset to

allow for a time-series of committee votes. Unfortunately, the House has taken few steps to implement the suggestions made by the Modernization Committees.

References

- Anzia, Sarah F. and Molly C. Jackman. 2012. "Legislative Organization and the Second Face of Power: Evidence from U.S. State Legislatures." *The Journal of Politics* 75(1):210–224.
- Bussing, Austin and Sarah A. Treul. Forthcoming. "Majority Party Agenda Setting: Picking Fights or Avoiding Them?" *Legislative Studies Quarterly* .
- Cox, Gary W. and Mathew D. McCubbins. 2002. Agenda Power in the U.S. House Of Representatives, 1877-1986. In *Party, Process, and Political Change in Congress*. Stanford, CA: Stanford University Press.
- Cox, Gary W. and Matthew D. McCubbins. 2005. *Setting the Agenda: Responsible Party Government in the US House of Representatives*. Cambridge, UK: University of Cambridge Press.
- Cox, Gary W., Thad Kousser and Mathew D. McCubbins. 2010. "Party Power or Preferences? Quasi-Experimental Evidence from American State Legislatures." *Journal of Politics* 72(3):799–811.
- Curry, James M. 2015. *Legislating the Dark: Information and Power in the House of Representatives*. Chicago, IL: University of Chicago Press.
- Curry, James M. 2019. "Knowledge, Expertise, and Committee Power in the Contemporary Congress." *Legislative Studies Quarterly* 44(2):203–237.
- Finocchiaro, Charles J. and David W. Rohde. 2008. "War for the Floor: Partisan Theory and Agenda Control in the U.S. House of Representatives." *Legislative Studies Quarterly* 33(1):35–61.
- Hamm, Keith E. 1982. "Consistency Between Committee and Floor Voting in U.S. State Legislatures." *Legislative Studies Quarterly* 7(4):473–490.
- Jackman, Molly C. 2013. "Parties, Median Legislators, and Agenda Setting: How Legislative Institutions Matter." *The Journal of Politics* 76(1):259–272.
- Kirkland, Justin H. and Mary A. Kroeger. 2018. "Companion Bills and Cross-Chamber Collaboration in the U.S. Congress." *American Politics Research* 46(4):629–670.
- Krehbiel, Keith. 2007. "Partisan roll rates in a nonpartisan legislature." *The Journal of Law, Economics, & Organization* 23(1):1–23.

- Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin and Luke Sonnet. 2022. "Voteview: Congressional Roll-Call Votes Database." Voteview.com.
- Lo, James. 2020. "Using Poole's Optimal Classification in R." <https://cran.r-project.org/web/packages/oc/vignettes/oc.pdf>.
- Maltzman, Forrest. 1995. "Meeting Competing Demands: Committee Performance in the Postreform House." *American Journal of Political Science* 39(3):653–682.
- Patterson Jr., Shawn and Thomas Schwartz. 2020. "Parties, Agendas, and Roll Rates." *Journal of Theoretical Politics* 32(2):348–359.
- Ryan, Josh M. Forthcoming. "The Partisanship of House Committees and Member Self-Selection." *Legislative Studies Quarterly* . <https://doi.org/10.1111/lsg.12313>.
- Stewart, III, Charles and Jonathan Woon. 2016. "Congressional Committee Assignments, 103rd to 114th Congresses, 1993–2017: House of Representatives." http://web.mit.edu/17.251/www/data_page.html.
- Unekis, Joseph K. 1978. "From Committee to the Floor: Consistency in Congressional Voting." *The Journal of Politics* 40(3):761–769.
- Wilkerson, John, David Smith and Nicholas Stramp. 2015. "Tracing the flow of policy ideas in legislatures: A text reuse approach." *American Journal of Political Science* 59(4):943–956.

Appendix

Appendix A: Committee Vote Requirements in Law and House Rules

Legislative Reorganization Act of 1970 P.L. 91-510, Section 104(b), Requiring Committee Votes to Be Recorded and Made Available

(b) Clause 27(b) of Rule XI of the Rules of the House of Representatives is amended by adding at the end thereof the following: “The result of each rollcall vote in any meeting of any committee shall be made available by that committee for inspection by the public at reasonable times in the offices of that committee. Information so available for public inspection shall include a description of the amendment, motion, order, or other proposition and the name of each Member voting for and each Member voting against such amendment, motion, order, or proposition, and whether by proxy or in person, and the names of those Members present but not voting. With respect to each record vote by any committee on each motion to report any bill or resolution of a public character, the total number of votes cast for, and the total number of votes cast against, the reporting of such bill or resolution shall be included in the committee report.”

H.Res.6, 104th Congress, Requiring Committee Votes to be Reported in Committee Reports

Accountability for Committee Votes

Sec. 209. In clause 2(1)(2) of rule XI amend subdivision (B) to read as follows: “(B) With respect to each rollcall vote on a motion to report any measure or matter of a public character, and on any amendment offered to the measure or matter, the total number of votes cast for and against, and the names of those members voting for and against, shall be included in the committee report on the measure or matter.”.

House Rule XIII, Clause 3(b), Committee Vote Reporting Requirement as of the 117th Congress (2021-2023)

(b) With respect to each record vote on a motion to report a measure or matter of a public nature, and on any amendment offered to the measure or matter, the total number of votes cast for and against, and the names of members voting for and against, shall be included in the committee report. The preceding sentence does not apply to votes taken in executive session by the Committee on Ethics, and applies only to the maximum extent practicable to a report by the Committee on Rules on a rule, joint rule, or

the order of business.

(e)(1) Each committee shall keep a complete record of all committee action which shall include a record of the votes on any question on which a roll call vote is demanded. The result of each such roll call vote shall be made available by the committee for inspection by the public at reasonable times in the offices of the committee. Information so available for public inspection shall include a description of the amendment, motion, order, or other proposition and the name of each Member voting for and each Member voting against such amend. meant, motion, order, or proposition, and whether by proxy or in person, and the names of those Members present but not voting.

Appendix B: Scraping and Parsing Details

Scraping

Reports from the 104th through the 114th Congress are scraped from Congress.gov. If scraping by Congress, the url is:

```
'https://www.congress.gov/search?q="report-chamber":"House+Report","congress":"%s"'
```

where %s is replaced with the specified Congress. If scraping all congresses at once, the url is:

```
'https://www.congress.gov/search?q="report-chamber":"House+Report"'
```

The text from each report, along with other information gathered from the report's webpage, is saved and stored in JSON Lines format.

Parsing Scraped Reports

The *ReportData* class serves as a vehicle to store the data from each report. Using load, each entry in the JSON Lines files—one line per report for each Congress—is loaded into Python.

Each report loaded is then passed to the *Report* class to be parsed. The *Report* class serves as a vehicle to parse the report's information and locates the roll call votes within the report. The information from each JSON Lines entry are added as attributes to the *Report* class instance, and then additional information not scraped is parsed and added to the report. The text of the report is divide into its sections and the section containing the votes is identified.

The section containing the vote information is divided into individual votes, and each vote is passed to *Vote*, with each *Vote* instance stored post-parsing in a list attribute for the *Report* instance. The text of the vote is parsed to identify whether the vote is a roll call or non-roll call vote.

Roll Call Votes

Appended votes—votes which are not contained in the text of the report, but as a scanned image in the pdf—are identified for later processing. For in-text votes, information is collected from the text and the members' votes collected.

The pdfs for reports that contain appended vote images are downloaded using 'download_report_url'.

Scanned images from the reports are extracted using 'extract_images'. All images from the pdf are extracted, so images that do not contain a vote are removed. Using Adobe, OCR is then performed on the appended vote images. The text from the newly-created appended image pdfs is extracted using 'extract_image_text'. The text files are cleaned as necessary.

The appended votes are parsed similar to non-appended or in-text-votes, except that the text of the vote is loaded from the vote's stored .txt. file. After completing this process, the next time the data is loaded, the appended votes will be parsed as well.

Voice Votes

Voice votes are sometimes identified in the text of the report, but this appears to be the case for only some committees, some of the time. Further, if the report text mentions phrases such as "vote" in certain sections, it may be counted as a non-roll call vote. Thus, while we collected data on voice votes, these data are excluded from the dataset because of the inconsistency with their reporting and collection. Please contact the authors for access to this data, but be aware that voice votes are likely under-counted.

Appendix C: Codebook

Variable Names and Descriptions

| Column Name | Description |
|------------------|--|
| cong | Congress of vote. |
| billtype | one of: House Concurrent Resolution (HCONRES), House Joint Resolution (HJRES), House of Representatives Bill (HR), House Resolution (HRES), Senate Bill (S). |
| billnumber | Bill Number of report and vote. |
| Bill.Name | Bill title taken from committee report. |
| Vote.Description | Text of vote description as scraped from committee report. |
| Report.Number | Unique within bills; report numbers with a number after a second hyphen indicates multiple parts, treated separately (e.g., 104-280-1 indicates part 1 of report 104-280). |
| Vote.Number | Unique vote identifier within a report; not necessarily in same order or match roll call vote number reported in report. |
| stewartcommid | Stewart committee data identification number of committee which took the vote; not necessarily the same committee which wrote the report or reported the bill. |
| icpsr | ICPSR number of voting member; merged in from Stewart committee assignment data. |
| lastname | formatted last name of voting member; used to merge with Stewart committee assignment data. |
| Member | In most cases, name as scraped from report; in some cases (hand-coded image file) taken directly from Stewart committee assignment data. |
| Vote | Member's vote as coded, 1=yes, 0=no. Other votes (e.g., present) not coded. |
| Vote.Scrape | Text of member's vote as scraped from committee report. |
| Party | Member's party, 100=Democrat, 200=Republican, other parties included; taken from Stewart committee assignment data. |
| Maj.Min | Party status code, collapsed (i.e., majority=1 or minority=2 party); taken from Stewart committee assignment data, see Stewart codebook for more details. |
| State.Name | State abbreviation for member; taken from Stewart committee assignment data; not available for all members. |
| Total.Yea | Total voting in favor on vote; vote-level variable. |
| Total.Nay | Total voting against on vote; vote-level variable. |

Table 4

| Column Name | Description |
|------------------|---|
| percentyea | percentage voting in favor on vote; vote-level variable. |
| percentnay | percentage voting against on vote; vote-level variable. |
| totalvotes | Sum of Total.Yea and Total.Nay. |
| demyesvotes | Number of Democrats voting in favor on vote; vote-level variable. |
| demnovotes | Number of Democrats voting against on vote; vote-level variable. |
| demtotalvotes | Sum of demyesvotes and demnovotes; vote-level variable. |
| demperyea | demyesvotes divided by demtotalvotes; vote-level variable. |
| gopyesvotes | Number of Republicans voting in favor on vote; vote-level variable. |
| gopnovotes | Number of Republicans voting against on vote; vote-level variable. |
| goptotalvotes | Sum of gopyesvotes and gopnovotes; vote-level variable. |
| gopperyea | gopyesvotes divided by goptotalvotes; vote-level variable. |
| Consideration.of | Rules Committee vote on resolution considering bill; not all consideration votes coded as such. |

Appendix D: Frequency Table of Votes Per Committee-Congress

Number of Committee Votes by Committee-Congress

| Committee | Congress | | | | | | | | | | | Total |
|-----------------------------------|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------|
| | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | |
| Agriculture | 10 | 2 | 3 | 2 | 0 | 1 | 11 | 7 | 20 | 12 | 1 | 69 |
| Appropriations | 21 | 8 | 13 | 10 | 6 | 16 | 2 | 6 | 17 | 7 | 11 | 117 |
| Armed Services | 10 | 8 | 9 | 9 | 28 | 8 | 11 | 14 | 12 | 18 | 28 | 155 |
| Budget | 39 | 20 | 22 | 19 | 26 | 8 | 0 | 0 | 8 | 4 | 3 | 149 |
| Education and the Workforce | 23 | 8 | 43 | 12 | 12 | 18 | 14 | 21 | 16 | 5 | 20 | 192 |
| Energy and Commerce | 41 | 5 | 11 | 17 | 31 | 65 | 7 | 38 | 21 | 6 | 17 | 259 |
| Financial Services | 18 | 17 | 15 | 13 | 4 | 5 | 20 | 12 | 13 | 5 | 3 | 125 |
| Homeland Security | NA | NA | NA | 19 | 6 | 25 | 10 | 11 | 25 | 5 | 7 | 108 |
| House Administration | 6 | 8 | 4 | 5 | 2 | 5 | 12 | 7 | 2 | 0 | 0 | 51 |
| Intelligence (Select) | 1 | 1 | 1 | 0 | 11 | 10 | 13 | 16 | 4 | 6 | 1 | 64 |
| International Relations | 21 | 12 | 4 | 7 | 7 | 8 | 2 | 1 | 32 | 0 | 2 | 96 |
| Judiciary | 41 | 19 | 29 | 18 | 14 | 32 | 17 | 17 | 21 | 12 | 9 | 229 |
| Natural Resources | 20 | 7 | 10 | 19 | 3 | 19 | 23 | 12 | 11 | 6 | 7 | 137 |
| Rules | 30 | 9 | 17 | 26 | 22 | 16 | 16 | 65 | 87 | 12 | 7 | 307 |
| Oversight and Government Reform | 21 | 3 | 2 | 0 | 7 | 2 | 3 | 5 | 12 | 7 | 3 | 65 |
| Science, Space, And Technology | 24 | 2 | 5 | 1 | 11 | 2 | 4 | 4 | 5 | 4 | 32 | 94 |
| Small Business | 0 | 0 | 1 | 0 | 0 | 0 | 4 | 2 | 10 | 7 | 4 | 28 |
| Transportation and Infrastructure | 21 | 4 | 3 | 4 | 0 | 1 | 3 | 0 | 17 | 5 | 2 | 60 |
| Veterans Affairs | 2 | 0 | 1 | 1 | 1 | 0 | 0 | 2 | 1 | 0 | 4 | 12 |
| Ways and Means | 35 | 18 | 7 | 9 | 10 | 15 | 21 | 26 | 9 | 5 | 17 | 172 |

Note: Homeland Security was a select committee in the 107th and 108th Congresses, became a standing committee in the 109th Congress.

Appendix E: Rules of the House Energy and Commerce Committee—Agenda

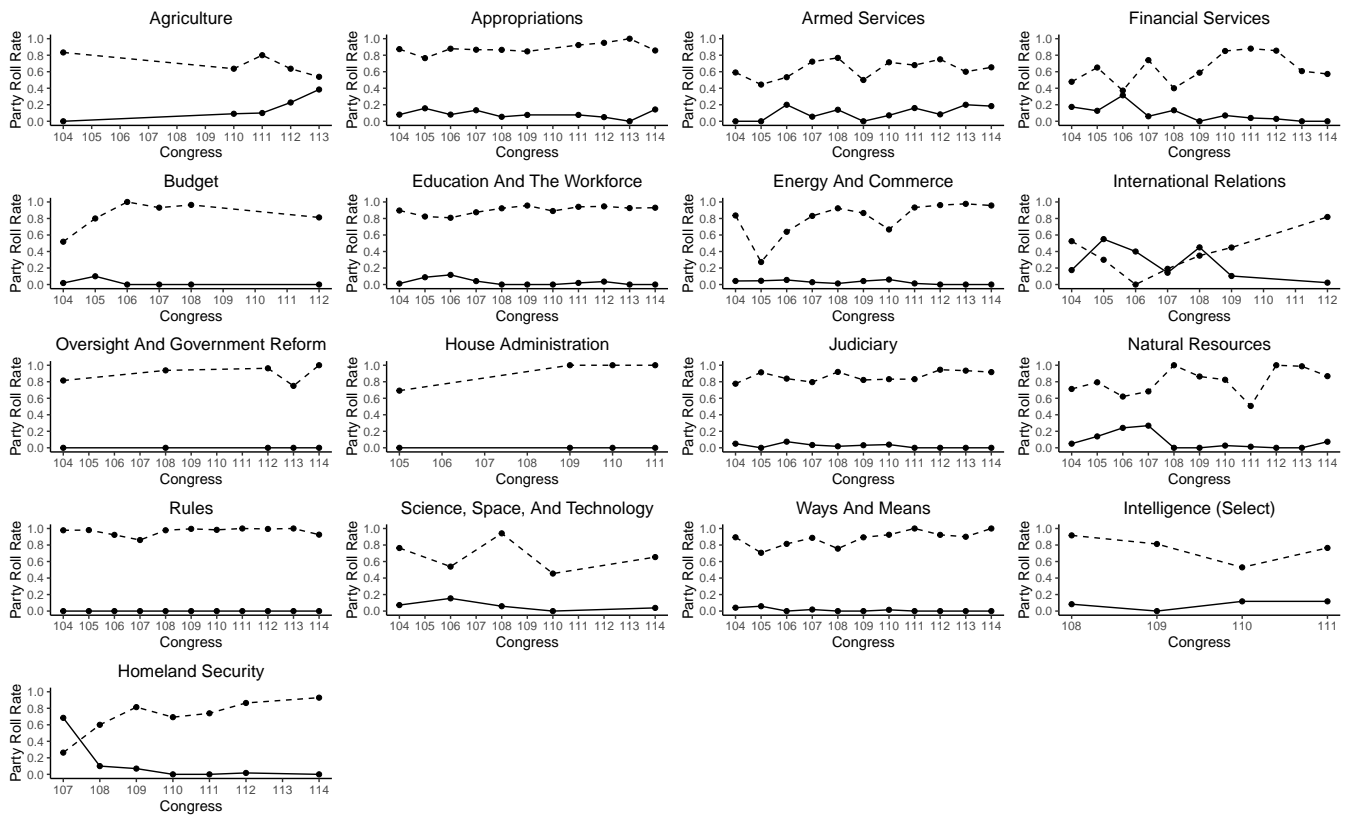
RULE 2. BUSINESS MEETINGS/MARKUPS

(d) Agenda. The agenda for each Committee meeting, setting out all items of business to be considered, shall be provided to each member of the Committee at least 36 hours in advance of such meeting.

Accessed at: <https://energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/117th%20Congress%20Committee%20on%20Energy%20and%20Commerce%20Resolution%20Rules%20Package.pdf>

Appendix F: Party Roll Rates Within Committees Across Congresses

Figure F1: Majority and Minority Party Roll Rates by Committee-Congress



Note: Dashed line shows roll rate for minority party, solid line shows roll rate for majority party. Committee-congresses in which there are fewer than ten committee votes are dropped from the sample. In the graph, those congresses without a scatter point indicate the observation is missing from the sample.