Jason MacDonald
Department of Political Science
West Virginia University
Morgantown, WV 26506
Jason.MacDonald@mail.wvu.edu

Overview of Data and Syntax Files for Replication of Figures and Tables in "The Rise and Fall. . ."

This file describes all of the data files and syntax files necessary to replicate the graphs and tables in "The Rise and Fall of Congressional Oversight: The House Committee on Energy and Commerce, 1969-2018," in the *Journal of Political Institutions and Political Economy*.

Please direct any questions to the author at the email address above or at jmacdon37@gmail.com . Thanks for your interest in this research!

(1)

"figures for revision manuscript.do" is the Stata syntax that is used to for figures 2, 3a, and 3b in the article.

(2)

"energy_and_commerce_91_115_transformed.dta" is the Stata dataset from which figures 2, 3a, and 3b are created. The data is "transformed" in the sense that I took data from individual hearings across 24 congressional sessions and aggregated it to the congressional session level.

Within the file:

"congress" is the variable running from 91 to 115 that is the session of the Congress during which the hearings occurred.

"sum_days" is the sum of the number of oversight hearing days that occurred in the House Committee on Energy and Commerce (or whatever its name was, e.g., Interstate and Foreign Commerce) during that session.

"sum_nagency" is the sum of the number of witnesses from the executive branch testifying before the Committee during that session (as the article notes, this count includes civil service witnesses, senior executive service witnesses, political appointees (anyone from an agency) and it includes personnel from the Executive Office of the President.

“pro_nagency” is the proportion of agency personnel who testified at oversight hearings, out of all witnesses who testified at oversight hearings, during the relevant session.

“sum_zero” is the count of hearings at which zero agency personnel testified during the session.

“pro_zero” is the proportion of oversight hearings during the relevant session at which zero agency personnel testified.

(3)

“figures4aAnd4b.do” is the Stata syntax that is used to created figures 4a and 4b in the article.

(4)

“DoeBudgetOversight.dta” is the Stata dataset from which Figure 4a is created.

“Congress” is the session during which hearings occurred.

“hearing_no” is the number of the hearing assigned by the Committee.

“pages” is the number of pages spanned by the hearing transcript.

“over_days” is the number of calendar days over which the hearing occurred.

“agency_witnesses” is the number of executive branch witnesses who testified at the hearing (includes executive office of the president, political appointees, SES, civil service).

(5)

“CleanAirActOversight.dta” is the Stata dataset from which Figure 4b is created.

Same as variable names and definitions “DoeBudgetOversight.dta”

(6)

“da100.dta” is the Stata dataset from which Table 1, columns 1 and 2 are estimated and from which Table 1, column 1 is created.

“Agency_Scrutiny1” is a categorical variable coded 0 if the purpose of the hearing involved no oversight, 1 if the primary purpose was not oversight but involved some oversight, 2 if the primary purpose of the hearing was oversight, 3 if the primary purpose of the hearing was to engage in advocacy for programs.

“d_oversight1” is a dummy variable coded 0 if Agency_Scrutiny1 equaled 0, 1, or 3; coded 1 if Agency_Scrutiny1 equaled 2 (so it’s a dummy variable for whether the primary purpose of a hearing is oversight.

“d_oversight2” is a dummy variable coded 0 if Agency_Scrutiny1 equaled 0, 1, or 3; coded 1 if Agency_Scrutiny1 equaled 2 (so it’s a dummy variable for whether the primary purpose of a hearing is oversight.

“nagency” is variable that is the count of the number of executive branch personnel testifying in the hearing (no distinction between civil service, SES, political appointees, or individuals from the Executive Office of the President—it includes them all in one count).

“nwitnesses” is a variable that is the count of the total number of witnesses at the hearing.

“days” is the number of calendar days over which the hearing occurred.

“pages” is the number of pages (in hundreds) spanned by the hearing transcript.

“dum_sc1” is a dummy variable indicating that the subcommittee that held the hearing was "Energy and Power".

“dum_sc2” is a dummy variable indicating that the subcommittee that held the hearing was "Health and the Environment".

“dum_sc3” is a dummy variable indicating that the subcommittee that held the hearing was "Telecommunications and Finance".

“dum_sc4” is a dummy variable indicating that the subcommittee that held the hearing was "Transportation, Tourism, and Hazardous Materials".

“dum_sc5” is a dummy variable indicating that the subcommittee that held the hearing was "Commerce, Consumer Protection, and Competitiveness".

(The sixth subcommittee was “Oversight and Investigations” and it is the reference category in the analysis).

(7)

“da91_115” is the sample of hearings from the 91st to the 115th Congress, as described in the article.

“Agency_Scrutiny1” is a categorical variable coded 0 if the purpose of the hearing involved no oversight, 1 if the primary purpose was not oversight but involved some oversight, 2 if the primary purpose of the hearing was oversight, 3 if the primary purpose of the hearing was to engage in advocacy for programs.

"d_oversight1" is a dummy variable coded 0 if Agency_Scrutiny1 equaled 0, 1, or 3; coded 1 if Agency_Scrutiny1 equaled 2 (so it's a dummy variable for whether the primary purpose of a hearing is oversight.

"d_oversight2" is a dummy variable coded 0 if Agency_Scrutiny1 equaled 0, 1, or 3; coded 1 if Agency_Scrutiny1 equaled 2 (so it's a dummy variable for whether the primary purpose of a hearing is oversight.

"nagency" is variable that is the count of the number of executive branch personnel testifying in the hearing (no distinction between civil service, SES, political appointees, or individuals from the Executive Office of the President—it includes them all in one count).

"nwitnesses" is a variable that is the count of the total number of witnesses at the hearing.

"days" is the number of calendar days over which the hearing occurred.

"pages" is the number of pages (in hundreds) spanned by the hearing transcript.

"c91" through "c115" are dummy variables coded 1 indicating that the observation/hearing occurred during the session designated by the number; 0 otherwise.

"p0" through "p6" are dummy variables (p1 through p6 are included in the models) indicating what policy area the hearing covered. (As the article explains, I could not estimate models with fixed effects for subcommittees in this analysis because there were too few observations for some subcommittees).

"p0" is commerce; "p1" is finance; "p2" is communications; "p3" is "health/public health," p4 is "transportation," "p5" is "Environment," "p6" is "Energy/Energy and Power."

(8)

"replication for table 2 and figure 1.do" is the Stata do file that uses the datasets from (5) and (6) above to produce the columns in Table 2 and Figure 1.

(9)

"reliability_data.dta" is the dataset with data used to produced Table A1-1 in the online appendix that presents findings from the analysis of the reliability of coding for the primary purpose of a non-legislative hearing. These observations are half of the observations from "da100.dta."

"agency_scrutiny_micah_code" is the code given to the hearing based on the subcommittee chairperson's opening statement by the undergraduate honors student who participated as a coder.

"agency_scrutiny_jason_code" is the code given by the author.

(10)

"reliability_analysis.do" is the file used to produce Table A1-1 in the Online Appendix.

(11)

"Appendix_Figures.do" is the file used to produce the figures in online appendix. The Stata do-file uses the datasets "DoeBudgetOversight.dta" and "CleanAirActOversight.dta" to produce the figures.

(12)

"replication for table 1.do" is the Stata do file providing the syntax to produce Table 1 in the article, using data from "da100.dta" and "da91_115.dta" described above.