

Congress and Its Experts

Dataset Codebook

PI: Jesse M. Crosson*

co-PIs: Alexander Furnas[†] Timothy LaPira[‡]

February 21, 2018

Background Information and Data Collection

Citation

Please cite these data as: Crosson, JM, Furnas, A & LaPira, T, *Congress and Its Experts*, electronic dataset, University of Michigan.

Project Background

These data were collected as part of a major data collection project entitled “Congress and Its Experts,” a project supported through research assistance provided primarily by the University of Michigan Department of Political Science and the University of Michigan Undergraduate Research Opportunities Project (UROP). Research assistance was also provided by undergraduates at James Madison University, supervised by Prof. Tim LaPira. Funding

*Ph.D. Candidate, Department of Political Science, University of Michigan

[†]Ph.D. Candidate, Department of Political Science, University of Michigan

[‡]Associate Professor, Department of Political Science, James Madison University

for the project has been provided by the Hatcher Graduate Library at the University of Michigan, and by the Michigan UROP program.

“Congress and Its Experts” seeks to collect panels of salary and responsibility data for every individual who worked in U.S. House of Representatives personal offices, from 1994-2013. These data are useful for a variety of purposes, but the primary usefulness of these data lie in their ability to track member investments in specific legislator behaviors over time. Given scarcity of time and resources, members of Congress face tradeoffs in which types of behaviors they prioritize (see Mayhew (1974) and Fenno (2002)). These trade-offs may be captured directly by measuring members’ investments of their budgetary allowances (referred to as Members’ Representational Allowances, or MRAs; see Brudnick (2014) and Brudnick (2017) for additional information), but data availability challenges have precluded scholars from making full use of such data for studying member priorities. By systematically recording salary and responsibility data, “Congress and Its Experts” addresses common data challenges and provides a reliable across-time measure of member investments.

Data Description and Coding Process

Raw Data Sources

The project draws its data from two sources: the U.S. House Statements of Disbursement (the official accounting documents of the House) and Legistorm (a company that sells reformatted and enhanced disbursement data, among other personnel data in Washington, D.C.). Legistorm has data from its founding (2000) to present day, and these data were provided to the PI as part of a negotiated data purchase through the Hatcher Graduate Library Data Grant Program at the University of Michigan. The purchase includes data from 2000-2013. In addition to this commercial data, research assistants for the project have transcribed and coded historical, print records from the U.S. House, for the years 1994-1999.¹ Both the Legistorm and historical House disbursements are staffer-level datasets that document the names of individuals who worked for Congress in each quarter of each year, which representative’s office(s) they worked for, what their title was, and how much they were paid.

¹Data for years preceding 1994 are available in print, but given the time-consuming nature of transcribing and coding such data, data collection efforts have reached only back to 1994.

Given that members of Congress face few restrictions on how to spend their budget allotment, members vary in which kinds of investments they make. Indeed, while some members spend over 90 percent of their allotted funds on personnel compensation, others choose to spend their money on opening additional offices in the district or sending a large amount of franked mail. Even among members who spend similar amounts on staff overall, offices differ considerably in how those staff dollars are spent, with some offices focusing spending in the district and others in Washington. This dataset makes use of this salary data to capture not only the overall investment of members in staff, but to capture their priorities as they invest money differentially across various staffer responsibilities.

Responsibility Coding Procedure

Using the aforementioned data, a hybrid human- and machine-based coding algorithm was applied to the data to assign primary responsibilities to each staffer in the dataset—a total of over 200,000 coding decisions,² and counting. According to this protocol, certain job titles receive automatic coding decisions, which are assigned via a simple algorithm in Python. However, for more ambiguous job titles, research assistants investigated the staffer’s responsibilities for the specified year and quarter in greater detail. This additional investigation involved searching for staffers in quarterly volumes of the *Congressional Yellowbooks*,³ where factors such as the staffer’s office location (Washington versus the district), policy portfolio (if one exists), and (occasionally) more descriptive job titles are listed. This information was incorporated systematically into the assistants’ coding decisions, as delineated in the coding protocol. The coding protocol is included at the end of this document, and questions regarding this protocol may be directed to Jesse Crosson at jessemc@umich.edu.

While some studies have opted to fully automate similar coding decisions, such automation is highly likely to encourage both measurement error and systematic bias. First, many common job titles contain little to no useful information regarding the responsibilities of the staffer—despite the fact that staffers of said title often possess highly disparate responsibil-

²Note that this does not represent the total number of unique staffers who worked in Congress over this period, as data are reported quarterly (and staffers occasionally work for more than one member at a time).

³See <https://www.leadershipdirectories.com/Products/LeadershipinPrint/Government/CongressionalYellowBook> for more information.

ities from office to office (Petersen, 2011). For instance, despite the presence of a Chief of Staff in nearly every modern member office, the title “Chief of Staff” in itself does not convey a great deal of substantive information. Indeed, in some offices, the Chief of Staff doubles as a member’s legislative director or senior legislative counsel, while in others the Chief of Staff focuses his/her energy on political endeavors, such as serving as the member’s primary scheduler and gatekeeper. Were an algorithm to code all chiefs of staff identically, it would artificially overestimate members’ legislative investments in some offices or understate such investments in others. By carefully investigating the policy responsibilities of Chiefs of Staff (and similarly ambiguous job titles), the coding procedure in this study avoids large-scale measurement error due to job title ambiguity. Second, beyond the overall measurement error encouraged by job title ambiguity, systematic differences in job title usage over time could lead fully automated procedures to measure responsibilities in biased fashion. For example, while Chief of Staff is typically the title given to member’s most trusted senior adviser in today’s Congress, members of Congress have previously listed other titles for staffers with identical responsibilities. In particular, the title “Administrative Assistant” was used commonly in the 1990s in place of “Chief of Staff,” even though the title itself would seem to denote that the staffer in question focused more on clerical work than on high-level political or legislative strategy. Here again, a fully automated algorithm is likely miss this nuance, leading to systematic biases between time periods. Conversely, careful human coding can capture these changes and more accurately report staffers’ responsibilities. Therefore, this study opts for a hybrid approach that harnesses the efficiency gains of automated coding without forfeiting the nuance provided by human coding.

Description of Resulting Dataset

According to the coding protocol, staff are classified into five groups: **legislative staff**, **political management staff**, **communications staff**, **office management staff**, and **constituency service staff**. **Legislative Staff** are staff whose primary responsibilities are to advise the member of Congress on matters pertaining to the legislative process. Responsibilities may include drafting new bills, reading existing ones, offering voting or cosponsorship advice, or providing expertise on the legislative process. **Political Management Staff** are

staff whose primary responsibilities are to manage the member's relationships with other elites in Washington, such as leaders of political parties and issue caucuses, lobbyists, and major donors. **Communications Staff** focus their energy on interacting with the media on the member's behalf, scheduling television appearances, drafting speeches, and submitting op-eds to newspapers. **Office Management Staff** deal primarily in clerical responsibilities, such as providing IT support to other staffers or handling the member's office accounting. Finally, **Constituency Service Staff** deal primarily in relations with the member's constituents. Responsibilities include handling bureaucratic casework for elderly citizens or immigrants, answering constituent phone calls or mail, and alerting the member of (or organizing) pertinent local activities within the district. While overlap between these categories undoubtedly exists in some cases, they are designed to capture Mayhew's primary member activities for reelection (advertising, credit-claiming, and position-taking; Mayhew (1974)) and goals (reelection, influence in the chamber, and good public policy; Fenno (1973)).⁴

These categories allow the resulting dataset to capture member-level investments in a wide variety of ways, as is summarized in the variables below. Such summaries include overall investment in each responsibility classification (and staff overall), percentages of an office's money dedicated to each responsibility, average quarterly payments to staff within each responsibility classification, and counts of staff in each classification. These measurements capture not only a member's investment in legislative work, communications, etc., but they also provide an illustration of the member's hiring philosophy. That is, while two members may both allocate \$60,000 a quarter to legislative work, one member may do so by paying three staffers \$20,000 each, while another pays six staffers \$10,000 each. Although both staffers have invested the same amount of money in legislative work, these differences in spending patterns could indicate different kinds of legislative goals. Such differences could have implications for studies of legislative effectiveness and other related fields.

Beyond the member-level spending variables aggregated from staff salary data, the dataset also leverages some of the additional variables provided by Legistorm. Legistorm gathers a variety of staffer-level variables for inclusion on their staffer web profiles, many of which are

⁴Note that, on occasion, a staffer may have more than one title associated with her name. In those cases, after consultation with the *Congressional Yellowbooks*, this staffer's salary is split evenly between the two (or more) responsibilities associated with the job titles.

geared at improving the networking prowess of their subscribers. These include membership in college fraternities, receipt of awards, hometown, and even interesting facts about staffers. These variables differ considerably in terms of missingness within the dataset, and some of them are likely not particularly useful for social scientific research. Others, however, may be of interest to social scientists and exhibit sufficient data coverage to merit inclusion in this member-level dataset. These include staffer sex and level of education. After some cleaning and categorization, these variables are presented here as averages for each office, among staffers for whom data was available.

Data Coverage

The dataset current covers the years 1994 - 2008. Coding on years 2009-2013 is still in progress, with each year ranging from 40 to 70 percent completeness. A data update including these more recent years will be provided in the coming months, with an imputed version of the data for these years provided in the interim. Within the 1994 - 2008 range, however, there is some missingness, due poor data availability. First, the year 2000 is not included, because Legistorm only collected partial data, since the company had just been founded. Second, years 2005-2006 are not included, because of a wide-ranging set of clerical errors made by the House of Representatives. For those years, the House Statements of Disbursement listed either no titles for staffers or listed all staffers as “staff assistants.” Consequently, the coding protocol could not be applied without extreme measurement error, precluding the inclusion of these years in the final dataset. No other years exhibit this problem.

All staff characteristic variables (namely, demographic variables) are available only for the portion of the data coded from Legistorm records: 2001 forward. Spending and count variables, on the other hand, are available for the full time series. Additionally, because this dataset was merged with the Center for Effective Lawmaking dataset, all non-staff variables (delineated below) are also available for the full time series.

As a final note, all yearly spending totals are provided as estimates, because accounting procedures in the U.S. House render it difficult to offer a definitive spending total for staffers and offices (due in part to how offices deal with the period from Jan. 1-2, during which offices are gearing up for the next congressional sessions). However, because all data were

collected in the same fashion (i.e., data from all years was split into quarters in the same way), the totals are comparable across years, even if they may differ from real-world totals by a few hundred dollars.

Variables

Identification Variables

- *thomas_number* — Numeric counter for order in THOMAS data from 93rd - 110th Congresses; drawn from TheLawmakers.org.
- *thomas_name* — Legislator name, as given in THOMAS; drawn from TheLawmakers.org.
- *st_name* — postal code of member's state; drawn from TheLawmakers.org
- *congress* — Congress number
- *icpsr* — ICPSR legislator code
- *year* — year of payment

Spending on Staff

- *est_total_spending* — estimated yearly total for member spending on personnel by year
- *est_total_legis_spending* — estimated yearly total for member spending on legislative staff by year
- *prop_legis_spending* — proportion of total personnel spending accounted for by legislative spending for each MC in each year
- *est_total_pol_spending* — estimated yearly total for member spending on political management staff by year
- *prop_pol_spending* — proportion of total personnel spending accounted for by political management spending for each MC in each year
- *est_total_comm_spending* — estimated yearly total for member spending on communications staff by year

- *prop_comm_spending* — proportion of total personnel spending accounted for by communications spending for each MC in each year
- *est_total_off_spending* — estimated yearly total for member spending on office management staff by year
- *prop_off_spending* — proportion of total personnel spending accounted for by office management spending for each MC in each year
- *est_total_constit_spending* — estimated yearly total for member spending on constituency service staff by year
- *prop_constit_spending* — proportion of total personnel spending accounted for by constituency service spending for each MC in each year
- *legis_avg_payment* — average quarterly payment per legislative staffer, by each MC in each year
- *pol_avg_payment* — average quarterly payment per political management staffer, by each MC in each year
- *comm_avg_payment* — average quarterly payment per communications staffer, by each MC in each year
- *off_avg_payment* — average quarterly payment per office management staffer, by each MC in each year
- *constit_avg_payment* — average quarterly payment per constituency service staffer, by each MC in each year
- *avg_payment* — average payment per staffer in each MC office, each year

Staff Count Variables

- *est_num_legis* — estimated number of payments made to legislative staffers in each MC office in each year. This is not a variable that captures the total number of unique

legislative staffers employed by the member; rather, it is the total number of payments made to legislative staff, meaning that there could be up to four payments for the same person.

- *est_num_pol* — estimated number of payments made to political management staffers in each MC office in each year (see *est_num_legis*).
- *est_num_comm* — estimated number of payments made to communications staffers in each MC office in each year (see *est_num_legis*).
- *est_num_off* — estimated number of payments made to office managements staffers in each MC office in each year (see *est_num_legis*).
- *est_num_constit* — estimated number of payments made to constituency service staffers in each MC office in each year (see *est_num_legis*).
- *est_total_payments* — estimated total number of staff payments made by each MC in each year. (see *est_num_legis*)
- *est_avg_office_size* — average size of member’s office over the course of the specified year. This variable accounts for the “repeated” staffer entries capture in the “*est_num*” variables.
- *est_legis_staff_size* — estimated number of legislative staff working for a member in an average quarter within the specified year. This variable accounts for the “repeated” staffer entries capture in the “*est_num*” variables.
- *est_pol_staff_size* — estimated number of political management staff working for a member in an average quarter within the specified year. This variable accounts for the “repeated” staffer entries capture in the “*est_num*” variables.
- *est_comm_staff_size* — estimated number of communications staff working for a member in an average quarter within the specified year. This variable accounts for the “repeated” staffer entries capture in the “*est_num*” variables.

- *est_off_staff_size* — estimated number of office management staff working for a member in an average quarter within the specified year. This variable accounts for the “repeated” staffer entries capture in the “*est_num*” variables.
- *est_constit_staff_size* — estimated number of constituency service staff working for a member in an average quarter within the specified year. This variable accounts for the “repeated” staffer entries capture in the “*est_num*” variables.
- *pct_legis_staff* — percentage of staff payments made to legislative staffers, by MC office and year
- *pct_pol_staff* — percentage of staff payments made to political management staffers, by MC office and year
- *pct_comm_staff* — percentage of staff payments made to communications staffers, by MC office and year
- *pct_off_staff* — percentage of staff payments made to office management staffers, by MC office and year
- *pct_constit_staff* — percentage of staff payments made to constituency service staffers, by MC office and year

Staffer Characteristics

- *total_staff_experience* — sum total of all years in the dataset within which a members’ staffers appeared.
- *total_legis_staff_experience* — sum total of all years in the dataset within which a members’ legislative staffers appeared
- *pct_female* — percentage of total staff that is female in each MC office, each year. Available only for years drawn from Legistorm data (2001 forward).

- *avg_education* — average educational attainment of a member’s staff, by year. Available only for years drawn from Legistorm data (2001 forward). Staffer education is coded in the following manner: 0 for some college or below, 1 for a bachelor’s degree, 2 for some graduate school, and 3 for graduate degree. Note that Legistorm did not collect this variable for all staff.

Turnover

- *perc_holdover* — percentage of member i’s staff employed at time t that was also employed by member i at t-1.
- *num_holdover* — count of member i’s staff employed at time t that was also employed by member i at t-1.
- *perc_policy_holdover* — percentage of member i’s legislative staff employed at time t that was also employed by member i at t-1
- *num_policy_holdover* — count of member i’s legislative staff employed at time t that was also employed by member i at t-1.

Center for Effective Lawmaking Variables

Note that descriptions for all variables below are provided by the Center for Effective Lawmaking at TheLawmakers.org (Volden and Wiseman, 2014).

- *cd*
- *dem*
- *majority*
- *elected*
- *female*

- *afam*
- *latino*
- *votepct*
- *votepct_sq*
- *dwnom1*
- *meddist*
- *majdist*
- *deleg_size*
- *speaker*
- *chair*
- *subchr*
- *power*
- *budget*
- *seniority*
- *sensq*
- *state_leg*
- *state_leg_prof*
- *maj_leader*
- *min_leader*
- *south*
- *south_dem*

- *ss_bills*
- *ss_aic*
- *ss_abc*
- *ss_pass*
- *ss_law*
- *s_bills*
- *s_aic*
- *s_abc*
- *s_pass*
- *s_law*
- *c_bills*
- *c_aic*
- *c_abc*
- *c_pass*
- *c_law*
- *all_bills*
- *all_aic*
- *all_abc*
- *all_pass*
- *all_law*
- *les*
- *leslag*

References

- Ida A Brudnick. Congressional salaries and allowances: In brief. *United States Congressional Research Service*, www.senate.gov/CRSReports/crs-publish.cfm, 2014.
- Ida A Brudnick. Members' representational allowance: History and usage. *United States Congressional Research Service*, <https://fas.org/sgp/crs/misc/R40962.pdf>, 2017.
- Richard F Fenno. *Congressmen in committees*. Little, Brown, 1973.
- Richard F Fenno. *Home style: House members in their districts (Longman Classics series)*. Longman Publishing Group Harlow, 2002.
- David R Mayhew. *Congress: The electoral connection*. Yale University Press, 1974.
- R Eric Petersen. Congressional staff: Duties and functions of selected positions. *Current Politics and Economics of the United States, Canada and Mexico*, 13(3):401, 2011.
- Craig Volden and Alan E Wiseman. *Legislative Effectiveness in the United States Congress: The Lawmakers*. Cambridge University Press, 2014.