Supplementary Materials for "Policy Influence and Private Returns from Lobbying in the Energy Sector"

Karam Kang

June 22, 2015

A. Datasets

A.1. Intermediate Datasets

I collect and construct datasets on legislation, lobbying, and polling from various sources, as described in the data section of the paper. These datasets can be found "Datafortheweb/ IntermediateDatasets," which has three sub-folders, "Policies," "Lobbying," and "Polling." Note that each variable in STATA dta-files is labeled with a short description for reference.

To begin with, *billseclist.dta* in "Policies" folder contains all sections of the bills that were introduced during the 110th Congress and were categorized by the Congressional Research Service as related to energy or environment issues. I followed the policies in the 110th Congress to the subsequent two Congresses, and the results of this exercise can be found in *billseclist_111th.dta* and *billseclist_112th.dta*. The policy position of each lobbying coalition and the content of each policy can be found in *policylist.dta*, and its code book is *codebook.pdf*. All datasets in this folder have been constructed from information freely available online.

In "Lobbying" folder, there are three STATA dta-files. For each lobbying report mandated by the Lobbying Disclosure Act of 1995, a list of bills that are mentioned in the report is extracted and stored in *lobbyingbills_cleaned.dta* for the 110th Congress and in *lobbyingbills_cleaned2.dta* for the 111th and the 112th Congresses. The information on each entity that was engaged in lobbying activities during the 110th Congress can be found in *sdata110th_client.dta*. The main sources for the datasets in this folder are the lobbying reports available on the Senate Office of Public Records website as well as the lobbying data provided by the Center for Responsive Politics.

Polling data can be found in "Polling" folder. The search results for public opinion survey questions on energy and/or environmental topics during the period of 2007–2012 are stored in *poll.csv*. For each polling question in this file, I categorized the content of the question and recorded the respective polling result in *poll_cleaned.csv*.

A.2. Datasets for Estimation and Simulation

Datasets for the Main Analyses The datasets for the estimation and simulation for the main specification of the model can be found in "DataandCode/Data." There are five data files in the folder. First, *data_outcome.txt* contains five dummy variables of equilibrium outcomes (enactment and each lobbying coalitions participation) for each policy. Second,

data_explanatory.txt contains four dummy variables that determine the initial enactment probability (more regulation, less regulation, more spending, and salience) for each policy. Third, *data_explanatory2.txt* contains eight dummy variables on the policy position of and the relevance to each lobbying coalition for each policy. Fourth, *data_polling.txt* contains the polling results for each policy. Lastly, *data_conditoinal.txt* is to facilitate the calculation of conditional moments.

Datasets for Auxiliary Analyses The datasets for bill-level analyses are located in "DataandCode/Auxiliaries/BillEstimation." In 'DataandCode/Auxiliaries/BillAggregation," *data_billid.txt* is used for aggregating the policy-level simulation results to the bill-level. For out-of-sample model fit exercises, the data files of the 111th and the 112th Congresses in "DataandCode/Auxiliaries/OutOfSampleModelFit" can be used. In addition, two sets of data files are used for sensitivity analyses: one regarding the categorization of policies in "DataandCode/Auxiliaries/SensitivityAnalyses/Alternative01" and the other in "DataandCode/Auxiliaries/SensitivityAnalyses/Alternative10" regarding the determination of policy positions.

Random Shocks for Simulation Employing *halton.m* and *shock.m* in "Datafortheweb/RandomShockCode," one can create six text files that contain random shocks for simulation. These files are not provided due to their size, but they can be easily generated by these MATLAB files.

B. Programs

B.1. Parameter Estimation

To estimate the parameters, I use a parallel optimizer, Hybrid Optimization Parallel Search Package (HOPSPACK).[1] This optimizer takes a user-provided program that can evaluate an objective function at a given point. The program used in the estimation, *lvalue* in "DataandCode/Executables," can be generated by running an executable in the same folder, *compile_lvalue*, using six FORTRAN files (*main_lvalue.f90, dataread.f90, esolver.f90, etoolbox.f90, globvar.f90*, and *lhood.f90*) in "DataandCode/FORTRAN." This program calculates the value of the objective function, equation (4.8) in the paper, at any given vector of parameters.

Inputs The inputs in this program include (i) the final data files in "DataandCode/Data", (ii) the random shock files for simulation, and (iii) a file of a vector of parameters.

Output This program generates one output file, which contains the value of the objective function.

[1]This optimizer is developed by Sandia National Laboratories, and is available at https://software.sandia.gov/trac/hopspack/.

B.2. Asymptotic Standard Error Estimation

Employing ten FORTRAN files (*main_stderr.f90, asymvar.f90, csv_file.f90, dataread.f90, esolver.f90, etoolbox.f90, globvar.f90, lhood.f90, mvalue.f90,* and *simulation.f90*) in "DataandCodd/FORTRAN," one can generate a program called *stderr* by running an executable, *compile_stderr*, both of which can be found in "DataandCode/Executables." This program is used to estimate the asymptotic standard errors of given parameters.

Inputs The inputs in this program are the same as those in Section B.1, except that the estimated parameters are used here.

Outputs This program generates two output files. One file stores the estimated parameters and their estimated standard errors. This file also reports some simulated moments. The other file, *var_matrices.csv*, contains an estimator of $A(\theta_0)$ and $B(\theta_0)$, both of which are defined in Appendix C.1 in the paper.

B.3. Confidence Intervals of Simulated Moments

To calculate the 95 percent confidence intervals of the moments in the paper, I use seven FORTRAN files (*main_moments.f90, csv_file.f90, dataread.f90, esolver.f90, etoolbox.f90, globvar.f90,* and *moments.f90*) to generate a program called *smoments* via an executable, *compile_moments*.

Inputs The inputs in this program include those in Section B.2. In addition, *std_mat.txt* is also required for this program to run. This file contains the lower triangular matrix from a Cholesky decomposition of the estimated asymptotic variance-covariance matrix. This file is generated by *cholesky.m* in "DataandCode/FORTRAN," which takes *var_matrices.csv* as an input.

Output The simulated moments and their confidence intervals are stored in an output file.

B.4. Auxiliary Analyses

In addition to the estimation results of the main specification, the paper presents the results of various auxiliary analyses. To perform these additional analyses, one can run the above programs after replacing certain data or FORTRAN files as described below.

Bill-Level Estimation In estimating the model using bill-level data as described in Section 5.4 in the paper, the following files in "DataandCode/Auxiliaries/BillEstimation" are used. One set of the files consists of the data files (*bdata_conditional.txt*, *bdata_explanatory.txt*, *bdata_explanatory2.txt*, *bdata_outcome.txt*, and *bdata_polling.txt*). The corresponding FORTRAN file to read these files, *dataread.f90*, can be found in the folder. Other FORTRAN files (*globvar.f90*, *lhood.f90*, *mvalue.f90*, *simulation.f90*, and *moments.f90*) incorporate the model specification changes suited to the bill-level analysis.

Bill-Level Aggregation To compare the bill-level estimation results with the policy-level ones, I simulate the original model at the policy level and aggregate the results to the bill level. To calculate the moments in Table 15 in the paper, I modify the program for simulated moments and confidence intervals, *smoments*, by replacing *globvar.f90*, *dataread.90* and *moments.f90* with the files in "DataandCode/Auxiliaries/BillAggregation."

Sensitivity Analyses Appendix D of the paper discusses the sensitivity analyses, whose results are presented in Tables 19 and 20. I provide the relevant data and FORTRAN files in "DataandCode/Auxiliaries/SensitivityAnalyses." For Alternative (1), I use a different categorization of policies as described in Appendix D.1. To estimate this specification, the data files provided in "Alterative01" folder are used instead of the final datasets. For Alternative (2), I use a different policy enactment function as described in Appendix D.2. To run this specification, the programs need to be modified by replacing *esearch.f90* with the file provided in "Alternative02" folder. For Alternative (3), a different equilibrium selection rule, as explained in Appendix D.3., is employed when multiple equilibria occur. The corresponding FORTRAN files are provided in "Alternative03" folder. To run the specifications of Alternatives (4) and (5), line 33 of *globvar.f90* needs to be adjusted, where the value of the entry cost is set. Similarly, by modifying line 39 of *main_lvalue.f90* and line 37 of *main_stderr.f90* to set a different value of the weight, λ, in the objective function, one can run the specifications of Alternatives (6) and (7). For Alternatives (8) and (9), different values of total lobbying expenditures can be entered in lines 25–28 of *dataread.f90*. Lastly, I use a different rule for determining the policy positions for each lobbying coalition in Alternative (10), which can be run by replacing *data_explanatory2.txt* with the corresponding file in "Alternative10" folder.

Model Fit for the 111th and 111th Congresses Lastly, Table 24 in the appendix of the paper presents the model fit results for the policies during the 111th an the 112th Congresses. The data files used in these analyses are provided in "DataandCode/Auxiliaries/OutOfSampleModelFit."