Supplementary Material for

Flexible Estimation of Policy Preferences for Witnesses in Committee Hearings

Kevin M. Esterling
Professor
School of Public Policy and Department of Political Science
UC–Riverside
900 University Ave.
Riverside, CA 92506
Email: kevin.esterling@ucr.edu
https://profiles.ucr.edu/kevin.esterling

Ju Yeon (Julia) Park
Assistant Professor
Department of Political Science
The Ohio State University
http://www.juyeonpark.com

February 1, 2024

Contents

$A.1$	Simulation to Demonstrate Model Identification	3
$A.2$	Coding of the Hearings Transcripts	9
$A.3$	Are Health Care Roll Calls Different?	12
$A.4$	Implementation of the Statistical Model and Estimation	13
$A.5$	Results	16
$A.6$	Text Analysis Details	32
	SM References	32

A Supplementary Material

This supplementary material (SM) provides technical detail that space limitations do not allow in the main text.

A.1 Simulation to Demonstrate Model Identification

Here we offer a simulation to demonstrate identification of our flexible model that allows agent preferences to shift and rotate relative to legislator preferences (e.g., measured in `DW-NOMINATE`), and to evaluate the bias and efficiency properties of the flexible model when compared to the constrained-regression model that does not permit shifts or rotation. The simulation examines identification, bias and efficiency for the model structural parameters (that is, the model coefficients) as well as for the estimates of agent preferences, under both approaches.

The simulation demonstrates that our proposed model exactly and uniquely recovers the target benchmarks for every model parameter, that is, all of the equation structural parameters as well as the parameters imputing the missing agent preferences. Convergence on benchmark parameters in the computational model is sufficient to prove theoretical identification of the model. If a model is not theoretically identified, the model will not have unique solutions for the parameters and as a result the computational model will not converge to a single vector of estimates for the parameters. Convergence is not necessary to prove identification, however, since a theoretically-identified model might not be empirically identified.

As we describe in the text, the flexible model is composed of two submodels, an outcome component that is needed to estimate witness preferences and to test the preference-distance hypotheses, and an optional bridging component that can recover the geometric relationship between the witness and legislator preference dimensions. We reproduce the system here as equations 3a, 3b and 3c.

$$
\begin{aligned}
O_{ij}^m \sim & \text{Poisson}(\widetilde{\lambda_{ij}^m}) && \text{(3a)}\\
& ln\widetilde{\lambda_{ij}^m} = \lambda_{ij}^m = \beta_0^m + \beta_1^m d(L_j,\zeta_i) + \beta_2^m R_i + \beta_3^m d(L_j,\zeta_i) R_i + \eta_{1_{ij}} + \eta_{2_j} + \eta_{3_i} \\
L_j \sim & \text{Normal}(\mu_j^L, \sigma^L) && \text{(3b)}\\
& \mu_j^L = \alpha_0 + \alpha_1 \psi_j \\
\zeta_i \sim & \text{Normal}(\mu_i^\zeta, \sigma^\zeta) && \text{(3c)}\\
& \mu_i^\zeta = (\alpha_0 + \alpha_2) + (\alpha_1 + \alpha_3)\psi_i.
\end{aligned}
$$

The notation is described in the text. For simplicity in the simulation, we assume only one outcome equation ($m = 1$), and omit all random effects and coefficients from the simulation's statistical model, which are not relevant for considerations of identification.[22] In the bridging equation 3b we set the legislator-specific transformation parameters $\alpha_0 = 0$ and $\alpha_1 = 1$, which is the parameter that governs the relationship between the common space ideology scale ψ and preferences L among legislators. To generate agent preferences, the simulation considers three cases: parallel preference dimensions for agent and legislator ($\theta = 0$), an oblique rotation ($\theta = \frac{\pi}{4}$) and an orthogonal rotation ($\theta = \frac{\pi}{2}$). The agent-specific transformation parameter α_3 in equation 3c governs rotation in the relationship between ψ and agent preferences ζ. To implement the three cases, we vary the amount of rotation by setting $\alpha_3 = 0$ for the parallel case, $\alpha_3 = -0.5$ for the oblique case, and $\alpha_3 = -1$ for the orthogonal case. In all cases we set the shift parameter $\alpha_2 = 0.5$.

There are $N_l = 500$ legislators that are used for bridging and $N_a = 500$ agents. We draw the $N_l + N_a$ observations of the ideology common space ψ from a standard normal distribution. We set the standard deviation σ_l of ϵ_j^L to 0.25 and draw the N_l observations of L from equation 3b assuming a normal distribution for each ϵ_j^L. Likewise we set the standard deviation σ_a of ϵ_j^ζ to 0.25 and draw the N_a observations of ζ from equation 3c assuming a normal distribution for each ϵ_j^ζ. We also assume that ϵ_j^L and ϵ_j^ζ are independent.

[22]And, in any case, simulation models that we run that include random effects and nested data converge equally well as the ones we report here.

For the outcome equation 3b, we set $\beta_1 = -1$; k indexes dyads with 10 committee members per agent (that is, each agent is nested within 10 dyads). We set $\beta_0 = 2$. We draw observations for O_{ij} by drawing each committee member's preferences L_j from a standard normal distribution, computing the distance $d(L_{l_j}, \zeta_i)$ using the committee member's L_j and the "true" (unobserved) agent preference ζ_i and the parameters in equation 3a. After creating the outcomes, we set each of the "observed" agent preferences ζ^{obs} to missing and retain the "true" agent preferences ζ_i as the benchmarks for posterior predictive checks (Gelman et al., 1996).

A.1.1 Estimation

We estimate the model in `MultiBUGS` using Bayesian MCMC methods (Goudie et al., 2020). We run the model until the posterior distribution of the structural estimates are stationary, and then sample from the posterior distribution to create marginal distributions of each parameter of interest. These models converge almost instantly and show excellent mixing. For each model, we then sampled 10k iterations saving each 10 for 3 chains (3000 replicates from the posterior).

A.1.2 Simulation Results

We estimate two versions of the model for each rotation (parallel, oblique and orthogonal), one version for the model based on the constrained-regression approach that estimates the bridge and outcome equations separately, and one for the flexible model.

The simulation shows that the flexible model recovers unique, unbiased estimates for all of the structural parameters in the model but the constrained-regression approach does not. The flexible model remains unbiased under each degree of rotation, but the degree of bias increases for the constrained-regression model as the rotation increases. Both models lose efficiency and estimate witness preferences with increasing error as the rotation increases, but the measurement error is always worse for the constrained regression case.

Structural Parameters. First consider the results for the structural parameter estimates for the three rotation cases. We begin by noting that in all three cases, in the flexible model the bridging equation intercept and rotation parameters for agent preferences ζ are recovered exactly: The estimates for the bridging model intercept $\widehat{\alpha_2}$ are precisely centered at 0.5 for each of the three cases (SE 0.02); the estimates for the rotation parameters $\widehat{\alpha_3}$ are precisely zero for the parallel case, -0.5 for the oblique case, and 1.0 for the orthogonal case (all with SE 0.01). Of course the α_2 and α_3 parameters are not identified for the constrained-regression model so there is nothing to report for that case.

Next consider the structural parameters in the outcome model, shown in figure 5. These parameters are identified for both the constrained and flexible models. The benchmark target value for each parameter is shown by a horizontal line. The estimates are shown in a color that matches the color of the corresponding target line. To get a sense of the variability in the parameter estimates, we estimate a modification of the model that allows 10 independent estimates of each parameter within each run.

We jitter the placement of the parameter estimates on the horizontal axis in figure 5 so that the estimates can be seen distinctly. The variation in the vertical direction indicates the amount of uncertainty across replicates of the estimates. Note that the flexible model retrieves unbiased point estimates for each parameter and with extremely low uncertainty. In contrast, the degree of bias in the constrained regression estimates increases with the degree of rotation.

Posterior Predictive Checks. Following the advice of Gelman and his coauthors (1996), we make use of posterior predictive checks to verify the model estimates for $\widehat{\zeta^{obs}}$ track the "true" agent preferences ζ.

Figure 6 shows the expected values of $\widehat{L^{obs}_{a_i}}$ (averaged across the full set of simulations)

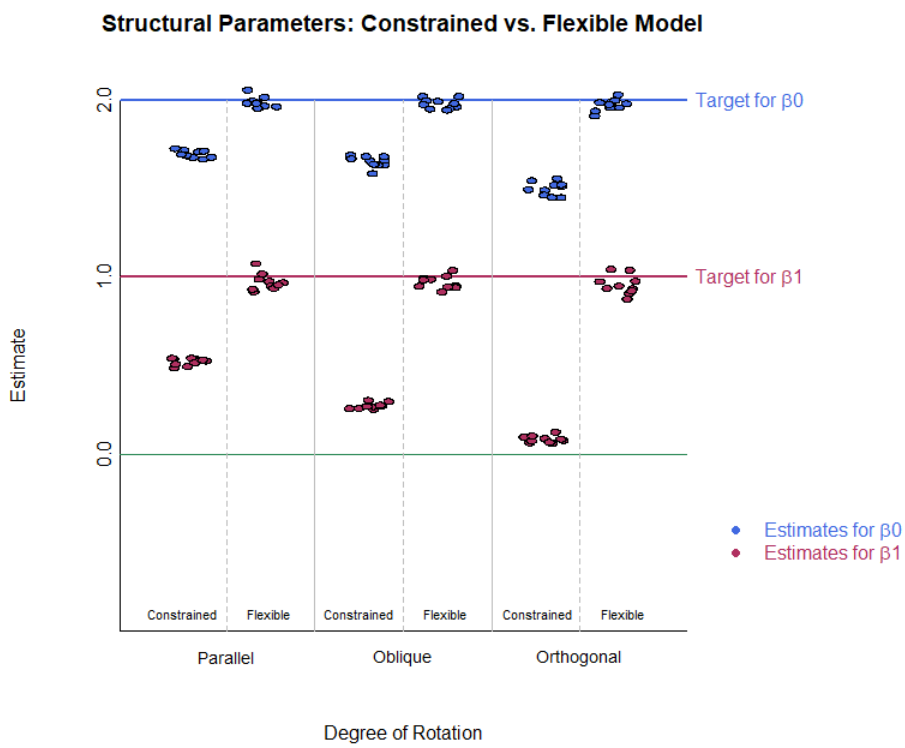


Figure 5: Simulation results for the structural parameter estimates with 10 unbiased committee members, for three example rotations. Notice that the flexible model results remain unbiased and reasonably efficient across rotations, while the constrained-regression estimates increase in bias and inefficiency with the degree of rotation.

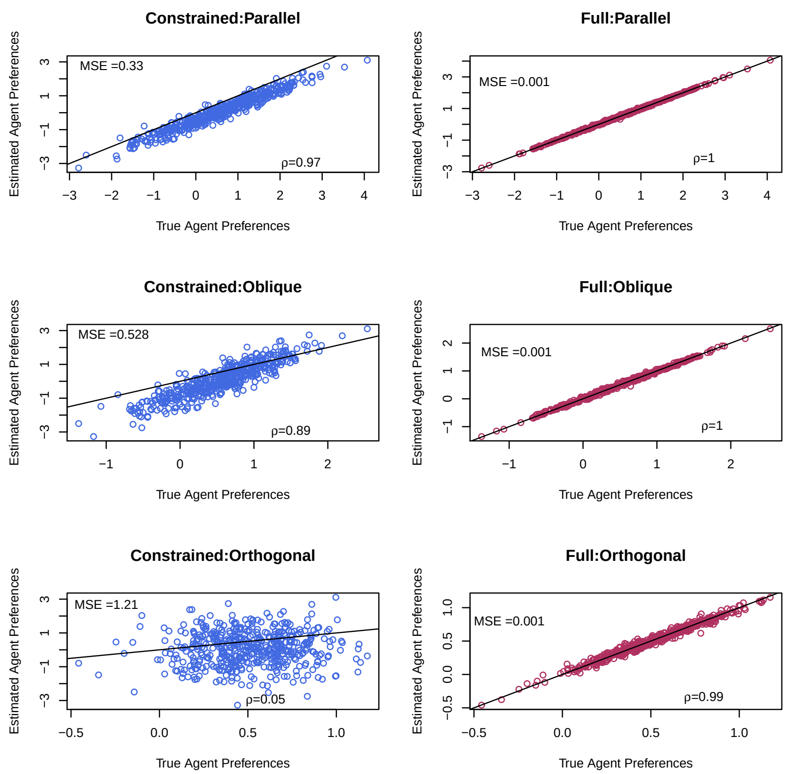


Figure 6: Posterior predictive check 10 unbiased committee members: scatter plots of point estimates for agent preferences versus true, for each model and each degree of rotation. The correlation coefficient ρ indicates the flexible model predicts the outcome data across the rotations, with a slight decrease in efficiency as the rotation increases, while the constrained-regression model fails to predict the outcome data as the degree of rotation increases. The mean square error (MSE) indicates the flexible model is more efficient in each case.

plotted against the corresponding true agent preference, along with descriptive statistics for the correlation and mean squared error. The flexible model recovers efficient and unbiased estimates in all three rotation cases. The constrained regression model shows both inefficiency and bias across all three cases, seen in the MSE and correlation statistics; for example, the MSE of the parallel (best case) constrained regression is nearly 20 times that of the flexible model. The scatter plot indicates that in the orthogonal case, the estimates of the constrained regression model have no correspondence with the true data; instead, the estimates are like a random draw, while the preference estimates in the flexible model are unbiased and measured with relatively little error.

A.2 Coding of the Hearings Transcripts

In the application, we develop a set of hypotheses regarding the different types of statements that legislators make within committee hearings. To collect the outcome data regarding the types of sentences stated in committee hearings, a research assistant and one of the authors hand coded the sentences (questions and statements) that legislators direct to specific witnesses at the sampled committee hearings, as recorded in committee hearing prints, making use of the hearings transcript coding rules in Esterling (2007). The coding rules contain 20 separate codes to mutually exclusively and exhaustively classify the types of statements and questions committee members ask of witnesses at the hearing. In addition to a set of "miscellaneous" statements, we follow the procedures in Esterling (2007) to organize these codes under the labels "falsifiable," "opinion," and "anecdotal" sentences.

Falsifiable sentences contain analytical policy information and are stated in a form that could be made into an operational research statement. Unlike opinion sentences, they are asserted as positive effects of a program or factual descriptions of the real world. Unlike anecdotal sentences, they are asserted as general and systematic, rather than local or personal. Examples of falsifiable sentences include:

- Verifiable factual statement. "Cataracts is one of the most significant causes for decreased vision."

- Description of how a program, policy, or organization operates at a general level. "From 1966 and until the mid-1990s, claims billing errors by hospitals across the country were handled through normal external audit process[es]."

- Causal implication or argument about the effect of a current policy or program. "The SMI Trust Fund, which in balance on an annual basis, shows a rate of growth of costs which is clearly unsustainable."

- The hypothetical future effects of a proposal. "If Medicaid payments to managed care plans... are set below market rates to achieve savings, the participation of mainstream plans could be compromised."

- Description of past actions in the policy process and intents of political actors. "Congress intended for payment reform to neither increase nor decrease overall Medicare payments to physicians."

Opinion sentences are normative or non-falsifiable statements, or statements that are explicitly qualified as the author's own belief or opinion. In all cases, these statements are not asserted as "true" or empirically demonstrable. Examples include:

- A policy position. "Medicare beneficiaries should be provided with a range of health plan choices, and those choices should be accompanied by incentives to select the more cost effective alternatives."

- A policy recommendation. "Delinking public health care programs from public cash assistance programs is good public policy."

- A normative argument (fairness, ideology). "No reason has been shown why the pharmaceutical industry should be singled out from others that freely negotiate the prices of their products with the DVA and the other departments and agencies of the Federal Government."

- The speaker's belief, feeling, or desire. "The proposed 16 percent reduction in the conversion factor results from a misinterpretation by HCFA of the mandate for budget neutrality contained in OBRA-89, as well as from inappropriate and demeaning assumptions about anticipated physician behavior in response to payment reform."

- A rhetorical question or political advice. "The Congress may want to create a process to adjust future conversion factors based on actual billing experience."

Anecdotal sentences reference only the speaker's immediate experience, or the immediate experience of the witness's organization with a policy or program, or only make reference to local conditions such as conditions within a specific congressional district. Examples include:

- A person's or organization's particular experience in a program. "In December our accountant received a list of more than 10,000 alleged billing errors during those five years."

- Likely effects from program or an alternative generalized from personal experience. "Dr. Russell Snow, and eye, ear, nose, and throat doctor from Caldwell, Idaho, says his colleagues are so frightened by federal enforcement provisions that many more are [going to drop Medicare patients]."

- Information about a congressional district or locality. "In my State, the hospitals that are okay are the ones that are doing cardiac."

- Statements about length of personal experience with a policy area

- Quote from well-known figure, adage, what "other people" are saying

Reliability tests for coding. A research assistant independently recoded a testbed random sample of sentences stated in Medicare committee hearings ($N = 578$) for an inter-coder reliability test, and one of the authors re-coded a second random sample ($N = 711$) one year after completing the first round of coding to conduct an intra-coder reliability test. The Cohen's Kappa reliability statistic for the intercoder reliability test is 0.57 with a 71 percent agreement rate (32 percent expected, $p < 0.0001$), and the intra-coder reliability is 0.79 with an 85 percent agreement rate (30.5 percent expected, $p < 0.0001$). While there are no established thresholds for reliability, a kappa statistic in the range of 0.75 to 0.80 is widely considered excellent agreement beyond chance, and 0.40 to 0.75 fair to good agreement beyond chance (Nuendorf, 2002, 143). All member sentences in the hearing transcripts sampled for this project were double-coded by both the research assistant and the principal investigator, with the latter resolving disagreements.

A.3 Are Health Care Roll Calls Different?

The topics of the 29 sampled hearings include headline grabbing issues such as prescription drug benefits, comprehensive health financing reform, and the solvency of the Medicare trust fund, and also less visible issues such as prospective payment systems for health providers, competition and managed care, billing fraud, medical savings accounts, risk adjustment, coverage information for beneficiaries, prevention and disease management, telemedicine, long term care, billing relations between the VA and Medicare, and demonstrations involving military retirees and the Federal Employee Health Benefits Program. The results of the main paper demonstrate that the preference space for health care witnesses across these topics is along a quality-cost dimension, a dimension that rotates orthogonally to the legislator roll-call preference space recovered in first dimension `DW-NOMINATE` scores.

The `DW-NOMINATE` procedure uses a cross section of roll calls, and it is reasonable to wonder if legislators' roll-call preferences estimated from the subset of health care roll calls also would rotate away from `DW-NOMINATE` scores. To evaluate this, we extracted the subset of health-specific roll calls taken in the House in the 106^{th}, 107^{th} and 108^{th} congresses (corresponding to the time period of the present study) by identifying all bills listed on Lewis et al. (2018) with "Social Welfare" in the Clausen Code, and then among those we searched the CRS policy topic codes for health-related topics. To enable convergence we dropped all votes where there was more than 90 percent agreement in the chamber, along with a handful of other votes that empirically prevented convergence.[23] This yielded 62 health-specific votes across the three congresses. We developed a model that is similar in structure to the equation set 1b but substituting the roll calls for the survey items to estimate the common space ψ.

In this model, the correlation between the first dimension `DW-NOMINATE` scores and our estimated scores from the health-only roll calls is 0.87 among the committee members in our sample, which indicates that a scale recovered from health-care votes is nearly an exact prediction of their scores based on all votes. So the assumption that preferences estimated from the full set of roll calls, such as those in `DW-NOMINATE`, are appropriate for estimating legislative preferences in this study.

A.4 Implementation of the Statistical Model and Estimation

The flexible statistical model is composed of two submodels that are estimated jointly, and linked by the common parameter ζ_i. Submodel A is a measurement model that places former members and witnesses into a personal common ideology space ψ and bridges these scores into the two-dimensional legislative preference space L and ζ. Submodel B contains the outcome equations of substantive interest. Here we set out the full flexible model, and then implement the constrained-regression approach as a set of restrictions on the flexible model. We estimate the flexible model in a Bayesian framework with likelihood:

[23] Roll call votes are not validated as items useful for scaling, and it is not unusual to find empirical convergence problems in such data.

Likelihood for Submodel A (Bridging Equations):

$$\left.\begin{array}{rcl} Markets_u & \sim & \text{OrderedLogit}(-1 \times \psi_u) \\ Companies_u & \sim & \text{OrderedLogit}(\gamma_{11}\psi_u) \\ HelpPoor_u & \sim & \text{OrderedLogit}(\gamma_{12}\psi_u) \\ Access_u & \sim & \text{OrderedLogit}(\gamma_{13}\psi_u) \\ Incomes_u & \sim & \text{OrderedLogit}(\gamma_{14}\psi_u) \\ L_u & \sim & \text{Normal}(\mu_{0u}, \sigma_L) \text{ if } u \text{ is a legislator} \\ \zeta_u & \sim & \text{Normal}(\mu_{0u}, \sigma_L) \text{ if } u \text{ is an agent} \\ \mu_{0u} & = & \alpha_0 + \alpha_1\psi_u \\ & & +\alpha_2 Agent_u + \alpha_3\psi_u Agent_u \end{array}\right\} 1 \leq u \leq N_1$$

Likelihood for Submodel B (Outcome Equation):

$$\left.\begin{array}{rcl} Distance_k & = & (L_{l_j} - \zeta_{i_k})^2 \\ Falsifiable_k & \sim & \text{Poisson}(\lambda_{1k}) \\ ln\lambda_{1k} & = & \beta_{10} + \beta_{11}Distance_k + \beta_{12}ResOrg_{i_k} + \\ & & \beta_{13}(ResOrg_{i_k} \times Distance_k) + \eta_{1_k} + \eta_{2_{j_k}} + \eta_{3_{i_k}} \\ Opinion_k & \sim & \text{Poisson}(\lambda_{2k}) \\ ln\lambda_{2k} & = & \beta_{20} + \beta_{21}Distance_k + \beta_{22}ResOrg_{i_k} + \\ & & \beta_{23}(ResOrg_{i_k} \times Distance_k) + \gamma_1\eta_{1_k} + \eta_{2_{j_k}} + \eta_{3_{i_k}} \\ Anecdote_k & \sim & \text{Poisson}(\lambda_{3k}) \\ ln\lambda_{3k} & = & \beta_{30} + \beta_{31}Distance_k + \beta_{32}ResOrg_{i_k} + \\ & & \beta_{33}(ResOrg_{i_k} \times Distance_k) + \gamma_2\eta_{1_k} + \eta_{2_{j_k}} + \eta_{3_{i_k}}. \end{array}\right\} 1 \leq k \leq N_2$$

where u indexes N_1 former members and witnesses in the combined matrix, i indexes witnesses and j indexes legislators in the dyadic matrix, k indexes N_2 legislator-witness dyads that occur across all of the committee hearings in the sample, i_k is the k^{th} dyad's witness, j_k is the k^{th} dyad's legislator, ζ_{i_k} is imputed for each witness in the k^{th} dyad.[24]

Submodel A is the measurement model that places the witnesses who appeared at the congressional committee hearings into the legislative preference space. The equation set in submodel A estimates ψ, via the estimated difficulty parameters (factor coefficients)

[24]Specifically, legislators are indexed as congress-committee-legislator, so Henry Waxman has separate indexes as a member of the Subcommittee on Health in the 106 and 107 Congresses, and yet another as a member of the Subcommittee on Government, Information and Technology in the 106th Congress. The legislator random effects $\eta_{2_{j_k}}$ are estimated over these indexes.

γ_{11} to γ_{14} and the ordered probit threshold parameters.[25] Submodel A then identifies the transformation from ψ to the legislative roll call space L (`DW-NOMINATE`) with the linear transformation given by α_0 and α_1 (and optional covariates $\mathbf{X}$). The posterior agent preferences ζ from the flexible model are modeled as a witness-level hierarchical equation that estimates the α_2 and α_3 structural parameters, the *Agent* indicator (1 if witness, 0 if former member of Congress), and the interaction between *Agent* and ψ.

The outcome equations are contained in submodel B. Each within-dyad question type count is modeled as Poisson-distributed, conditional on the within-dyad distance in legislative preference space and the random effects, allowing separate β parameters for dyads that have a witness from a *Research organization* (= 1) such as a university or think tank, and for dyads where the witness is not from a research organization (= 0). Distances can be measured in any way, but here we use quadratic distance.

The m outcome equations share a common legislator-specific random effect η_2 that captures the committee member's propensity to ask questions and make statements of all types to witnesses; a witness-specific random effect η_3 that captures the witness's propensity to attract questions and comments from legislators; and a dyad-specific random effect, η_1, that captures omitted variables that govern the dyadic interaction. These three random effects capture any omitted legislator-, witness-, or dyad-specific covariates. In particular, the random effects account for any structural features of the committee process, such as if the majority party has the ability to schedule witnesses that are more favorable to majority committee members. The dyad-level random effect also accounts for over dispersion that comes from added variance in the count data (Skrondal and Rabe-Hesketh, 2007).

To complete the Bayesian model, we set the priors for the $\lambda_{\cdot}$ parameters as distributed

[25]For identification we constrain the difficulty parameter for the *Markets* survey item to be negative one; the negative sign is because this item has an opposite ideological direction from the other items.

Uniform(0, 10) in order to ensure the known direction labeling in the factor model. The σ_L prior is Uniform(0,100). ψ and each η have a standard normal prior. All other priors are unrestricted, normally distributed with mean zero and standard deviation 1000. Other than setting the correct direction labels for the items in the IRT model, the model in no way depends on the priors for identification.

The flexible model estimates both A and B submodels simultaneously, so the bridging submodel and the outcomes submodel jointly inform the posterior distribution over each ζ_i^1. The constrained-regression approach estimates the two models separately and so the outcome equations do not update the preferences of agents, yielding agent preference estimates ζ^0 that are derived under the assumption that legislator and agent preferences are constrained to a single dimension.

Estimation. We estimate the model in `OpenBUGS` using Bayesian MCMC methods (Spiegelhalter et al., 1996). We run the model until the posterior distribution of the structural estimates are stationary, and then sample 3,000 replicates from the posterior distribution to create marginal distributions of each parameter of interest. For each model, we use a 100k burn in period, and then sampled 300k iterations saving each 300 for 3 chains.

A.5 Results

In this section we describe the full sets of model results that are summarized in the paper. We first describe the measurement model that recovers the ideology common space ψ, and then the constrained-regression results, and then the flexible structural equations model results.

Measurement Model Here we describe the results of the model that measures personal ideological preferences ψ, the common space ideology measure that appears in the bridging equations. Table 3 shows the results of the measurement model using only the data from the former members' sample (i.e., for now not using the responses from witnesses). The measurement model is able to test the convergent validity of the indicators we use to estimate ψ, which is indicated by large and statistically significant discrimination parameters (factor coefficients) $\boldsymbol{\lambda}$; these are reported in the first two columns of results in table 3 and clearly indicate convergent validity. The measurement model results shown in table 4 show the discrimination parameters from the pooled (witness plus former member) sample. Comparing the discrimination parameter results across these two tables, one can see that the parameters are equivalent in the pooled and unpooled case. This equivalence shows that the responses to these items is invariant across the two groups and so serve as good measures of a common space (Jessee, 2016).

The next two sets of columns in table 3 regress the personal ideology scale ψ on former members' roll-call vote first dimension `DW-NOMINATE` scores (Lewis et al., 2018) in order to construct the bridge that we use in the constrained-regression approach. The middle columns regress former members' `DW-NOMINATE` scores on their ideology (ψ) and a set of covariates which test whether there are organizational or employment-based characteristics that would affect the bridging transformation among members. Here we include covariates for party identification, whether the former member was a senator, the member's tenure in office, and whether the former member returned the survey from the DC area (likely indicating a lobbyist versus a true retiree). Note that only party identification is significantly different from zero. The fixed effect of party is large, 0.47, and the posterior is far to the right of zero.

The final set of columns in table 3 re-estimates the bridging model among the former members but excluding covariates. Including the party covariate reduces the mapping coefficient α_1 point estimate from 0.26 to 0.11. That party is a statistically significant predictor of legislators' roll-call scores is no surprise. Party plays a central role in organizing congressional politics. The literature on Congress leaves it as an open question whether it is theoretically sensible to include this covariate in the bridging model (Kre-

Table 3: Bridge Equation Posterior Results (former members only)

	ψ Model Only		with Covariates		without Covariates	
	Mean	SE	Mean	SE	Mean	SE
Bridge Equation						
Ideology (ψ)			0.10	0.03	0.26	0.03
DC Area			-0.01	0.04		
Chamber			-0.02	0.07		
Party			0.47	0.05		
Tenure			-0.00	0.02		
Constant			-0.21	0.03	-0.05	0.04
Sigma			0.15	0.01	0.20	0.02
Ideology Factor Loadings						
Markets	1		1		1	
Companies	1.64	0.41	1.60	0.40	1.62	0.38
Help Poor	2.85	0.88	2.65	0.75	2.53	0.65
Access	2.33	0.56	2.51	0.57	2.50	0.53
Incomes	1.84	0.46	1.88	0.44	1.74	0.38

Posterior structural parameter estimates for the bridge equation (only equation 1b), modeling `DW-NOMINATE` roll-call preference scores for the former member (FMC) subsample ($N = 87$). Note that party is the only covariate other than ideology that predicts roll-call preferences. Comparing factor loadings with those in table 4 shows the items load identically on the ideology factor across member and witness subsamples.

hbiel, 2000). However, including covariates in the bridging equation or omitting them does not affect the estimates of the main model parameters and so we choose to omit them from our analyses.

Relationship between Legislators' Ideology Scores and `DW-NOMINATE` We use the former members' responses to the survey items and the IRT portion of the statistical model to estimate their personal left-right ideological ideal point ψ. Figure 7 illustrates the relationships between the personal ideology space ψ, party identification, and the roll call preference space `DW-NOMINATE` among the set of former members in the sample (in a figure similar to Shor and McCarty, 2011, 535). In this figure, Republicans are indicated with filled circles, Democrats with empty circles, the pooled relationship between ψ and `DW-NOMINATE` is indicated by the solid black line, and the relationships adjusted for party are indicated by dashed lines. In the left hand panel, we include only an intercept shift for party, while the right hand panel we also include a term interacting party with ψ.

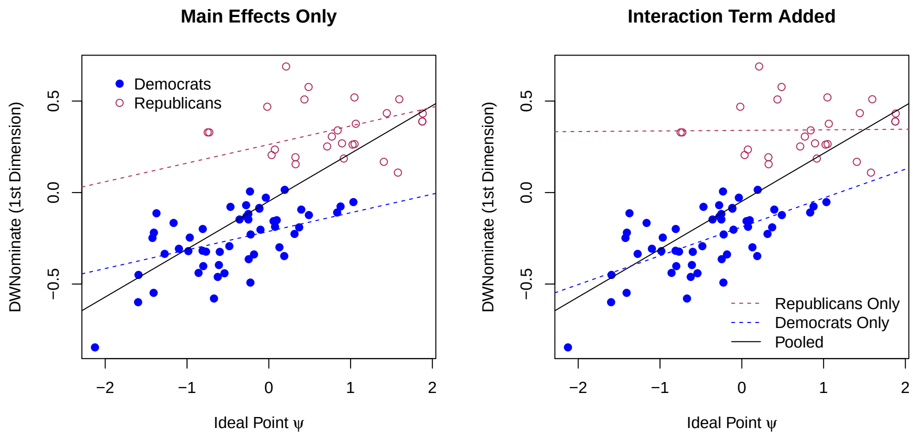


Figure 7: The relationship between personal ideology ψ and roll call preferences L among former members of Congress. $N = 77$. The solid line is the unconditional regression while the dashed lines account for differences by party.

If one ignores the party classification there is a clear linear relationship between ψ and `DW-NOMINATE`. If one considers party, the members' common space preferences ψ remain unimodal but the roll call preferences separate into a bimodal distribution. Further, if one allows the full interaction of the right hand panel, it appears that among Republican members, personal ideology has no relationship with roll call voting, while ideology matters for Democrats. One could conclude that ideology does not matter for Republicans, although there is no suggestion in the Congress literature that would support this asymmetric relationship, and the full interaction is likely an over fit to the sample.

Imputing Party Preferences of Witnesses To use party as a covariate in the bridging equation, we need the party identification of both legislator and agent. Unfortunately, we do not have party identification for the witnesses in the sample; in addition, the vast majority of groups in the sample do not make campaign contributions and so contribution data cannot help either. To impute this missing data, we use a data-driven approach to classify witnesses by party where we regress estimates for witness's personal ideology ψ on a classification for the type of organizations that the witness works for.[26] In this descriptive regression, we examine which employer types have witnesses that are statistically different on ψ compared to those who work at the baseline category of for-profit trade associations, organizations that are firmly in the Republican constituency. In this analysis, we find that professional associations, corporations, law firms, partisan think tanks, and industry coalitions are statistically similar to for-profit trade associations, while labor unions, universities, non-profit NGOs, state and federal agencies, not-for-profit trade associations, nonpartisan think tanks, and hospitals are statistically more liberal. Witnesses in the first set we label as in the Republican constituency, and the latter set in the Democratic constituency.

Results for the Constrained-Regression Models In this section we report the full set of results for the sentence count outcome equations from the constrained-regression approach to bridging, which constrains agent preferences to the legislative roll call preference space under an assumption that these always coincide within a single dimension.

To estimate the model in the first columns of table 4 we first extract the expected values of the witness preferences $\widehat{L^0_{a_i}}$ based on the regression results reported in the second and third columns of table 3 and witnesses' estimated ideology score $\widehat{\psi_{a_i}}$, relying on the linear assumptions of the regression, and then we use these estimates to construct the distance measures between $\widehat{\zeta^0_i}$ and L_j within each dyad. We then jointly regress each question count on distance, the research organization indicator, the interaction between distance and research organization, and a constant, using an Poisson likelihood for each count type conditional on random effects. These three outcome models are estimated jointly and each equation includes random effects for witness, legislator and dyad (not shown).

[26]This regression is on the full sample of 165 witnesses who returned surveys as a part of the larger project, not just the ones in the sample of hearings.

Table 4: Model Results

	Constrained Mean	Constrained SD	Flexible Mean	Flexible SD
Bridge Equation				
Ideology (ψ)	0.25	0.03	0.22	0.03
Agent			-0.12	0.06
Ideology X Agent			-0.19	0.04
Constant	-0.07	0.03	-0.07	0.03
Sigma	0.20	0.02	0.23	0.02
Ideology Factor Loadings				
Markets	1		1	
Companies	1.58	0.30	1.59	0.31
Help Poor	2.09	0.44	2.10	0.46
Access	2.01	0.38	1.98	0.39
Incomes	1.55	0.29	1.56	0.32
Outcome Equations				
Falsifiable Statement Count				
Distance	0.12	0.18	-2.58	0.89
ResearchOrg	0.26	0.35	0.77	0.38
Interaction	0.01	0.29	-2.82	1.34
Constant	-1.75	0.25	-1.07	0.29
Opinion Statement Count				
Distance	-0.27	0.27	-3.00	1.20
ResearchOrg	-0.48	0.43	0.47	0.50
Interaction	0.126	0.50	-5.58	2.47
Constant	-3.24	0.33	-2.52	0.39
Anecdotal Statement Count				
Distance	0.08	0.30	-0.33	1.27
ResearchOrg	-0.95	0.54	1.50	0.71
Interaction	-0.74	0.74	-25.6	6.78
Constant	-4.78	0.47	-4.51	0.52
Dyad-Specific Random Effect				
Falsifiable Equation	1		1	
Opinion Equation	1.71	0.12	1.66	0.12
Anecdotal Equation	2.49	0.21	2.49	0.22
WAIC				
Falsifiable Model	1016.59		954.78	
Opinion Model	575.08		542.75	
Anecdotal Model	489.71		421.13	

Structural parameter estimates for the constrained-regression model that imputes agent preferences based on the unidimensional assumption, comparing results with two distance functions and with and without covariates. $N_{dyads} = 669, N_{witnesses} = 67, N_{members} = 87$.

Results for the Flexible Model We report for the flexible model in the second set of columns in table 4, which employs the rotated posterior preference measures for witnesses in the outcome equation distance functions. In contrast to the constrained-regression approach, the flexible model estimates the bridge equation model and the outcome models jointly and hence uses the full information in both models for identification and to improve estimation. We report the results of the model that includes a soft constraint $\alpha_1 + \alpha_3 \geq 0$, which prevents the rotation of the witness preferences from achieving an obtuse rotation, which is theoretically implausible and would likely reflect an over fit to the sample, but the results without this additional constraint are virtually identical.

The flexible model is able to estimate agent-specific bridging parameters (attached to the *Agent* and $Ideology \times Agent$ rows) that govern the degree of shift and rotation of the witness preference space away from the roll-call preference space. The estimate indicates an approximate orthogonal rotation. More importantly, note that the hypotheses based on cheap talk information theory regarding preference distance, research organization, and their interaction are largely confirmed. That the results are significant in this table but not in the constrained-regression approach strongly suggests that the constrained-regression approach generates preference scores for witnesses that are measured with error.

Agent Preference Estimates and Posterior Model Checks In this section we compare the estimates of agent preferences (and their precision) between the constrained-regression approach to bridging and the flexible simultaneous equations model that estimates agent preferences based on observed behavior of members in committees. The constrained model assumes that agents' and legislators' preference spaces coincide, while the flexible structural equations model allows the agent preference space to be distinct. We examine the effect of including and relaxing this constraint on agent preference estimates. The specifications of each model are those described above, with a linear distance function and no covariates in the bridging equation.

Figure 8 shows a histogram of the change in the standard deviation of each witness's preference estimate, where the standard deviation of the preferences from the constrained-regression model is subtracted from the standard deviation of the posteriors from the flexible simultaneous equations model, so negative changes indicate that the unconstrained preference is estimated more precisely. While the degree of change in precision varies, the histogram indicates that the flexible, unconstrained preferences are estimated typically with more precision than the constrained.

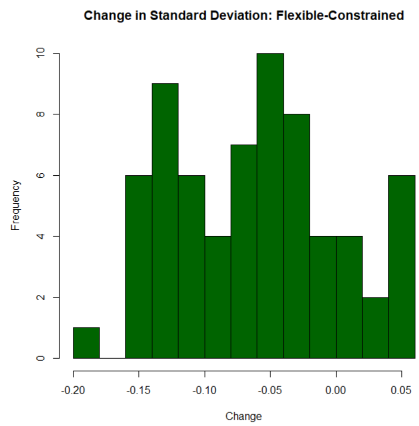
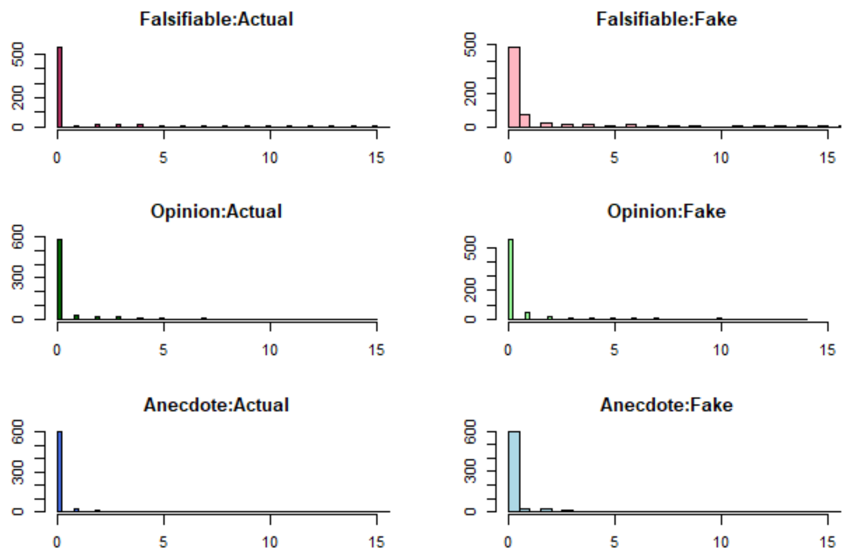


Figure 8: Agents' preferences tend to be estimated with greater precision in the flexible simultaneous equations model than in the constrained-regression model.

Next, we report the results of posterior predictive checks for the flexible simultaneous equations model, to verify that the model can reproduce the observed distributions of the outcome data. We draw 3,000 replicates of the three outcome variables – the counts of falsifiable, opinion and anecdotal statements and questions that members direct to witnesses – derived from the posterior estimates of each model. This creates $3,000 \times 3$ "fake data" replication data sets.

As we describe in the paper, the vast majority of witnesses receive zero questions so the median for each outcome is zero; by comparison, the 9,000 replicated data sets also all have median zero. Because it is such an easy target, the median is not a good test for this reproducibility of the original data. We report the means and standard deviations



FigAR10

Figure 9: Posterior predictive check for the flexible simultaneous equations model: actual and fake outcome data, showing a good replication of the data.

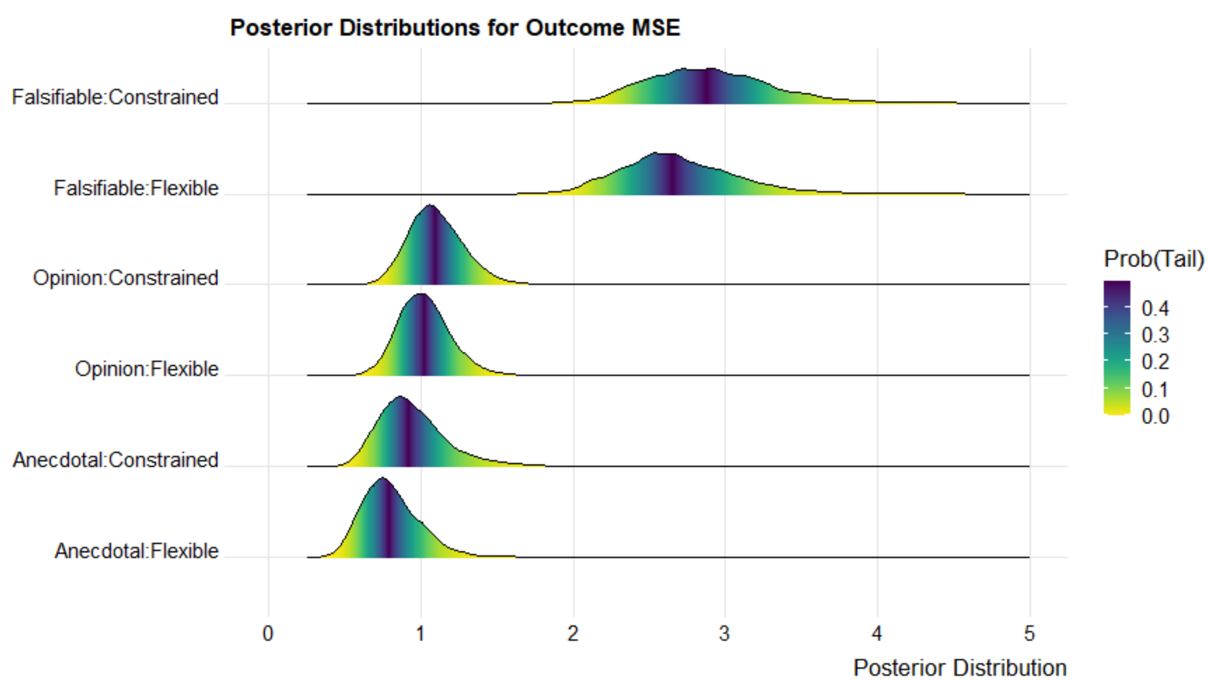
Table 5: Flexible Model Posterior Predictive Check Tests

Outcome	Mean	SD
Falisfiable:TrueData	1.15	3.15
Falisfiable:FakeData	1.15	3.12
Opinion:TrueData	0.46	1.58
Opinion:FakeData	0.46	1.58
Anecdotal:TrueData	0.38	1.72
Anecdotal:FakeData	0.38	1.75

for the true and fake data in table 5, and these also very closely match. In this table, the fake data results are the averages of the means and standard deviations across all 3,000 replicates. Finally, figure 9 shows the histograms of the true data (in dark colors) and for comparison we show the histogram of a single replicate fake data set (in light colors) for each of the outcomes. (The other replicates are very similar.) Overall, the flexible model does a good job replicating the original data based on the posterior parameter estimates, matching not only the means and standard deviations but also the pronounced skewness of the outcome data.

Goodness of Fit Test Our primary objective is to estimate witness preferences and to conduct the hypothesis test for the distance-test parameter, rather than the typical objective of a regression analysis to best predict or explain the outcomes. At the same time, the outcome counts in the analysis serve as the only ground truth for the models we use in the application, and hence a goodness of fit test helps to assess the quality of each model in replicating the data.

To conduct this assessment, we first take 3,000 samples from the model posterior distribution, use each to create predicted count outcomes for the model for each of our three outcomes, and then calculate the mean squared error (MSE) for each replication compared to the observed outcome. We plot the posterior distribution of the MSE for each outcome, first for the constrained model then for the flexible model. We show these posteriors in figure 10.



Fig

Figure 10: Mean squared error posterior distributions for each outcome, for each model. The error is the difference between the predicted outcome and the actual outcome.

Note that on average the flexible model has lower MSE, but that there is not much difference between the two models. The reason for this similarity is that each model, both the flexible and the constrained, are saturated with random effects that explain most of the variance.

Simulation Analysis to Assess Measurement Error The simulations we report above are designed to evaluate the identification of the statistical model. Here we report the results of a simulation analysis that is designed to mimic the properties of the observed dataset for our application to Medicare hearings. The Medicare hearing count outcomes each have relatively low means, and hence the statistical models have relatively low power. At the same time, our statistical model estimates the witness preference parameter ζ by embedding the parameter in three outcomes (the counts of falsifiable, opinion and anecdotal questions). Hence the ζ parameter reflects the information contained in the three outcome equations taken together, which increases the statistical power.

In order to assess the underlying degree of measurement error in our application results, we create a dataset that replicates the features of our application data, reflecting each of the three count means along with added cluster-level error at the dyad, witness and legislator level. We then estimate the identical flexible model we use for the application using these data. For details, see the replication package.

The results show that the model we estimate in the application has relatively low measurement error. In the simulation, we find a correlation between the estimated $\widehat{\zeta_i^1}$ parameters and the "true" benchmark witness preferences is 0.84. The results are shown in figure 11.

Considering confounds for the witness preference space. Given that the posterior preferences for witnesses rotate away from a well-defined and familiar roll-call preference space, it is a reasonable question to ask if the agent posterior parameters $\widehat{\zeta^1}$ measure a preference space, as opposed to a confounding space such as a degree of technical expertise or topic or some other dimension. For example, it is possible that legislators at one end

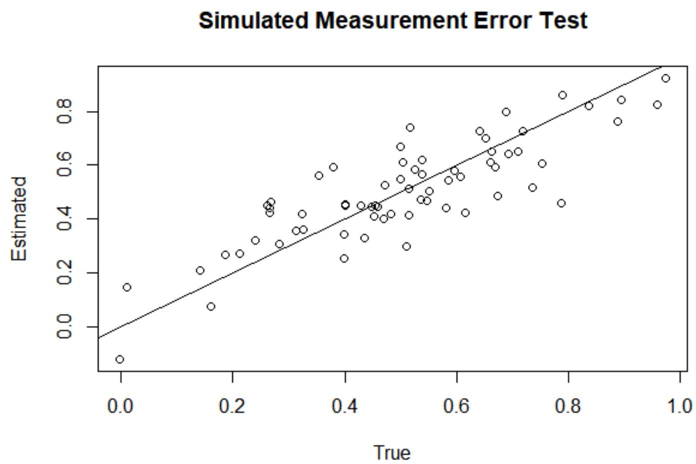


Figure 11: Comparing true versus estimated data in a simulation that mimics the properties of our Medicare Hearing data, to give a sense of the measurement error in the model. Correlation is 0.86.

of the ideological spectrum value expertise more than the other and the agent space is simply an expertise space. However, we examine the correlation between a wide variety of measures of the witness' technical expertise (such as the presence of data tables and figures and the number of cites to academic research in the testimony) and the posterior witness preferences and there is none.

Perhaps another confound is topic, where perhaps liberals ask more questions to witnesses from one topic area and conservatives from another. To test this we organized the hearings by topic and find no differences in the distributions of witness preference locations across topics. This is illustrated in figure 12, where the large center dot is the average of witness preferences for the topic and the smaller dots are the individual estimates, for topics where there are six or more witnesses. Note that the averages closely cluster around the center and show little variability, while the individuals vary significantly around the average for all topics.

Replication using Bonica CF Scores As we explain above, identification of the flexible model requires a measure for a personal ideological space ψ that is a common

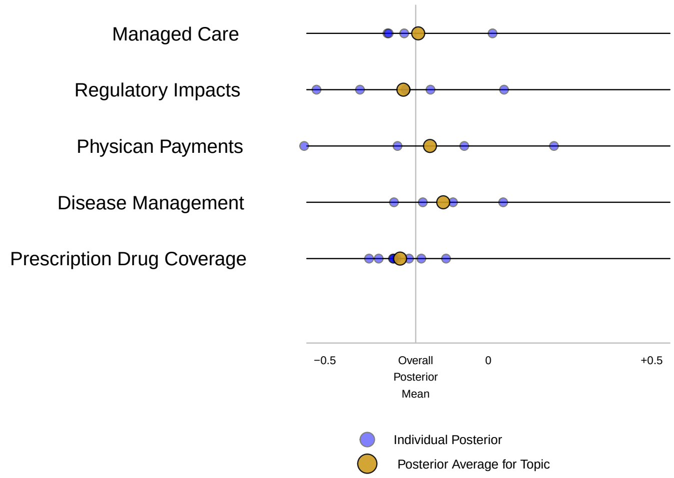


Figure 12: Distribution of agent preferences across selected topics using estimates of $L^1_{a_i}$ from the simultaneous equations model, showing preferences are unrelated to topics.

preference space shared by both agents and legislators. We use this common space to map between personal ideological preferences and individuals' policy preferences and is necessary for the model's identification. Our application to Medicare hearings was designed to measure this common ideological space using standard survey items and an IRT model, which is the most preferred method as the survey items have been validated to measure personal ideology. Sending out surveys to legislators and witnesses is time consuming and costly, however. In this section we demonstrate that one can achieve nearly identical results using the existing measures of ideological common space developed by Bonica (2013), called CF scores, which enables researchers to use off-the-shelf data to estimate flexible policy preferences for witnesses at any set of hearings on any topic.

We accessed the Bonica CF scores at https://data.stanford.edu/dime. For the former members of Congress we used the recipient CF scores for the member's final year in office, finding exact matches using the member's ICPSR number. For the witnesses, we first extracted all individuals in the contributor database that had the matching last name, and either the matching first name or initial or initials. We then manually examined the matches for each individual to match using other attributes such as the witness's employer or location. This manual matching could yield one of three results for each witness. If the search yielded a unique exact match on all attributes we record the witness CF contributor score. If the search yielded no matches or only matches that lack identifying attributes (such as if there were many "smith, tom" entries but none with a matching employer) then we do not record a CF score for that witness. Finally, if the search yielded several exact matches (such as if there were separate entries for "esterling, kevin" and "esterling, kevinm" both employed at UCR) then we took the weighted average of the corresponding CF scores, where the weights were determined by the number of contributions supporting each score, so if the first entry had 3 contributions and the second had 133 contributions, the corresponding weights would be (3/136) and 133/136). We do not provide the worksheets for the matches in the replication material since these contain the Bonica identification numbers, but we can make the worksheets available on request with IRB approval.

These matching procedures produced CF scores for all but one of the former members of Congress, and 33 out of 67 witnesses. If an analyst wished to rely exclusively on CF scores for the common space, then the estimation sample would need to be limited to witnesses who have matching CF scores. However, the present application was designed around a survey-based approach to measuring the common space and so the survey responses determined the estimation sample. To keep the remaining 34 witnesses and one member who did not have a matching CF score, we use the survey common space measure to impute the missing witness CF scores. The survey items and the CF scores load uniquely on a single factor and hence this imputation is simply using one ideology common space (the survey measure) to impute scores on another ideology common space (the CF scores), using a standard bridging model, so the imputation is straightforward.

Figure 13 shows the results in which we replicate the model underlying figure 3 of the main text, but relying only on the point estimate Boncia CF scores as the common space

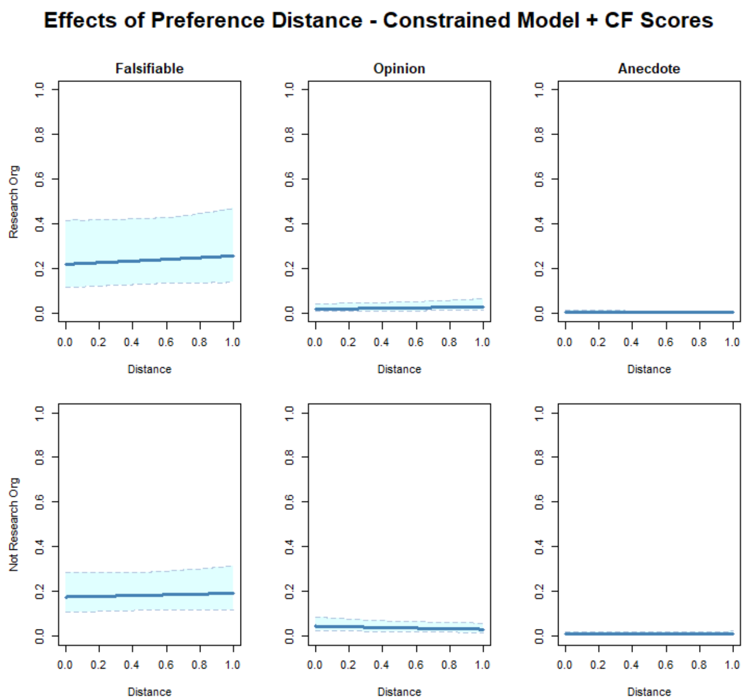


Figure 13: Replication of Figure 3 results from the constrained-regression model but using Bonica CF scores for the common space measure. The results are essentially identical to those using survey items to measure the common space.

in place of the IRT-estimated common space scores – that is, we use the CF score point estimates instead of the IRT model. The results are essentially identical to those we find using survey items.

Figure 14 shows the results in which we replicate the model underlying figure 4 of the main text, using the flexible model, but again relying only on the Bonica CF scores as the common space. The results are very nearly identical, with the exception that the Bonica replication results have slightly higher standard errors and WAIC scores, which reflects the methodological issue that the Bonica CF score point estimates have measurement error while the scores in our preferred model are IRT scores that are latent scales. Thus the preferred results here are based on our survey rather than CF scores, but we are able to show here that one gets essentially the same results either way. We note that the similarity of results across the two different common space measures is no surprise since the common space measure is only used for identification in the model and does not

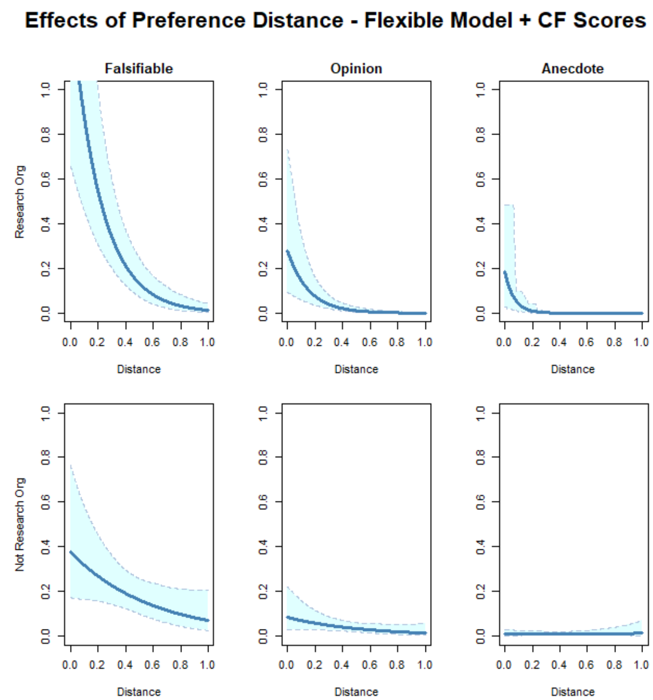


Figure 14: Replication of Figure 4 results from the flexible model but using Bonica CF scores for the common space measure. The results are exactly similar except the standard errors and the model WAIC scores are a bit larger because of the measurement error in the Bonica CF score point estimates relative to the survey-based IRT model.

directly play a substantive role. Thus, an analyst could rely on Bonica CF scores for the common space measure and not administer a survey in order to implement our model.

Finally, we re-estimated the constrained model using the exact setup as in figure 1 of the main text, but instead using CF scores for both the committee members (those who participated in the hearings) and the witnesses. The results under this replication are essentially identical to those found in figure 1 so we omit the results here. This null result too is no surprise since substituting CF scores for `DW-NOMINATE` scores does not change the model in any way other than changing the scale of the distance measure, which has no econometric impact, and since CF scores rely on the same unidimensional structure as `DW-NOMINATE` scores.

A.6 Text Analysis Details

The latent, agent-specific preference scale we recover in the application is orthogonal to the preference scale of legislators that is recovered using roll-call votes. In the paper we describe the text analysis procedure where we analyze the testimony of witnesses who score either high or low on the agent preference scale to describe the content of the preference dimension.

To conduct the text analysis, we grouped the hearings by topic, and identified the witnesses who had the highest and lowest posterior preference scores ζ^1 for the following set of topics: managed care, regulatory impacts, physician payments, disease management, and prescription drug coverage, and we pooled the testimony for those with high values together and those with low values together. We stratified by topic in order to ensure comparability across the two sets of documents and since word frequencies are governed by the content of documents. We made use of the text analysis tools recommended in Silge and Robinson (2017).[27] We stemmed and removed common English language stop words, and then we deleted health-care specific stop words that were common to both sets of documents and hence did not distinguish between documents that score high and low on the witness preference dimension. These words are "medicare," "health," "program," "managed," "patient," "physician," and "plan."

References

Bonica, A. (2013). Ideology and Interests in the Political Marketplace. *American Journal of Political Science 57*(2), 294–311.

Esterling, K. M. (2007). Buying Expertise: Campaign Contributions and Attention

[27]Available at https://www.tidytextmining.com/.

to Policy Analysis in Congressional Committees. *American Political Science Review 101*(Feb.), 93–109.

Gelman, A., X.-L. Meng, and H. Stern (1996). Posterior Predictive Assessment of Model Fitness Via Realized Discrepancies. *Statistica Sinica 6*, 733–807.

Goudie, R., R. Turner, D. De Angelis, and A. Thomas (2020). MultiBUGS: A Parallel Implementation of the BUGS Modeling Framework for Faster Bayesian Inference. *Journal of Statistical Software 95*, 1–20.

Jessee, S. (2016). (How) Can We Estimate the Ideology of Citizens and Political Elites on the Same Scale. *American Journal of Political Science 60*(4), 1108–1124.

Krehbiel, K. (2000). Party Discipline and Measures of Partisanship. *American Journal of Political Science 44*(2), 212–227.

Lewis, J. B., K. T. Poole, H. Rosenthal, A. Boche, A. Rudkin, and L. Sonnet (2018). Voteview: Congressional Roll-Call Votes Database.

Nuendorf, K. A. (2002). *The Content Analysis Guidebook*. Thousand Oaks, Ca.: Sage Publications.

Shor, B. and N. McCarty (2011). The Ideological Mapping of American Legislatures. *American Political Science Review 105*(3), 530–551.

Silge, J. and D. Robinson (2017). *Text Mining with R: A Tidy Approach*. Sebastopol, CA: O'Reilly.

Skrondal, A. and S. Rabe-Hesketh (2007, 10). Redundant overdispersion parameters in multilevel models for categorical responses. *Journal of Educational and Behavioral Statistics 32*(4), 419–430.

Spiegelhalter, D., A. Thomas, N. Best, and W. Gilks (1996). BUGS 0.5: Bayesian Inference Using Gibbs Sampling Manual (version ii). Technical report, MRC Biostatistics Unit.