APPENDIX

for

Comparable Preference Estimates across Time and Institutions for the Court, Congress and Presidency

American Journal of Political Science Vol. 51, No. 3, July 2007, pp. 433-448.

Michael A. Bailey
Department of Government and Public Policy Institute
Georgetown University,
baileyma@georgetown.edu

Contents

1 Statistical Details

The latent variable specification in Equation 3 is derived from a random utility framework. Let i = 1,...,N index individuals and v = 1,...,V index votes. The utility of actor i of voting for the conservative alternative is

$$u_i(\lambda_t^C) \quad = \quad -(\theta_{it} - \lambda_v^C)^2 + \eta_{iv}^C \tag{1}$$

where λ_v^C is the spatial location of the conservative alternative, θ_{it} is the ideal point of the actor at the time of proposal t and η_{iv}^C is a random shock. The utility of voting for the liberal alternative with spatial location of λ_v^L is analogous.

Let $\tilde{y}^*_{itv}$ be the utility difference between the conservative and liberal alternatives. It is

$$\begin{aligned} \tilde{y}^*_{itv} &= -(\theta_{it} - \lambda_v^C)^2 + \eta_{iv}^C + (\theta_{it} - \lambda_v^L)^2 - \eta_{iv}^L \\ &= 2\theta_{it}(\lambda_v^C - \lambda_v^L) + \lambda_v^{L2} - \lambda_v^{C2} + \eta_{iv}^C - \eta_{iv}^L \\ &= (\lambda_v^C - \lambda_v^L)(2\theta_{it} - (\lambda_v^L + \lambda_v^C)) + \eta_{iv}^C - \eta_{iv}^L \end{aligned} \tag{2}$$

Let $\hat{\epsilon}_{iv} = \eta_{iv}^C - \eta_{iv}^L$ be a mean-zero random variable with variance $= \sigma_v^2\sigma_i^2$. The variance has been decomposed into an element associated with the vote (σ_v^2) and an element associated with the individual (σ_i^2). This is useful in order to allow the variance of the shock to vary across votes (and, by extension, across institutions) and also produces substantively interesting estimates of individual level variation. Dividing Equation 2 through by σ_v yields

$$y^*_{itv} \quad = \quad \alpha_v(\theta_{it} - \kappa_v) + \epsilon_{iv} \tag{3}$$

where $y^*_{itv} = \frac{\tilde{y}^*_{itv}}{\sigma_v}$, $\kappa_v = \frac{\lambda_v^L + \lambda_v^C}{2}$ is the vote "cutpoint," $\alpha_v = \frac{2(\lambda_v^C - \lambda_v^L)}{\sigma_v}$ is the vote "discrimination parameter" and $\epsilon_{iv} = \frac{\eta_{iv}^C - \eta_{iv}^L}{\sigma_v}$ is a mean-zero random variable with variance σ_i^2.

Observed votes (as opposed to unobserved latent values above) are denoted by y_{itv}. I address rotational identification (e.g. liberals can have high values or low values) by coding votes in the conservative direction as $y_{itv} = 1$. I identify the location and scale of ideal points by assuming they have mean 0 and variance 1; this is equivalent to fixing two individuals at arbitrary points (see, e.g., Bafumi, Gelman, Park and Kaplan 2005).

The estimation process uses a Gibbs sampler algorithm. This algorithm allows us to draw samples from the posterior distribution of the parameters (Gelman et al 1995, 326; see also Johnson and Albert 1999, 194-197). To start the process, I set provisional ideal points to be the percent of conservative votes by the individual minus 0.5; this anchors conservative ideologies as high values and liberal ideologies as low values. After a "burn in" period, the following iterative procedure will produce random samples from the underlying posterior distribution.

1. Equation 3 implies that y^*_{itv} (where i indicates individual, t indicates term and v indicates vote) will be distributed according to one of the two truncated distributions (see e.g. Jackman 2000, 311)

$$y^*_{itv}|y_{itv} = 1 \sim N(\alpha_v(\theta_{it} - \kappa_v), \sigma_i^2)I(y^*_{itv} > 0) \tag{4}$$
$$y^*_{itv}|y_{itv} = 0 \sim N(\alpha_v(\theta_{it} - \kappa_v), \sigma_i^2)I(y^*_{itv} \leq 0) \tag{5}$$

   where I is an indicator function that serves to truncate distributions above or below zero.

2. Generate γ_i (a 5 x 1 vector) on an individual-by-individual basis. Substituting Equation 4 from the paper into Equation 3 and rearranging yields

$$\frac{y^*_{itv}}{\alpha_v} + \kappa_v = X'_{it}\gamma_i + \frac{\epsilon_{itv}}{\alpha_v}. \tag{6}$$

   Because of the heteroscedasticity implied by the above, use GLS results to calculate the distribution of γ to be

$$\gamma_i \sim N((X_i'\Sigma_i^{-1}X_i)^{-1}X_i'\Sigma_i^{-1}\tilde{y}, \sigma_i^2(X_i'\Sigma_i^{-1}X_i)^{-1}) \tag{7}$$

   where X_i is a $V_i \times 5$ matrix of covariates for individual i (and V_i the number of observations for individual i), $\tilde{y} = \frac{y^*_{itv}}{\alpha_v} + \kappa_v$ and Σ_i is a $V_i \times V_i$ covariance matrix with $\frac{1}{\alpha_v^2}$ down the diagonal. I impose a N(0,Ω) prior on γ to identify the preferences of individuals who vote conservatively or liberally all the time. Without this prior, their estimated ideal points would become unbounded. The implementation of the prior follows Gelman et al (1995, 260). As discussed in the paper, I restrict the higher order elements of γ to be 0 for individuals who served relatively short periods of time. For individuals with fixed preferences Equation 7 plus the prior simplifies to $\gamma_{0i} \sim N(\frac{\sum \alpha_v^2\tilde{y}}{1+\sum \alpha_v^2}, \frac{1}{1+\sum \alpha_v^2})$.

3. Simulate σ_i^2 for each individual from an inverse χ^2 distribution. A problem with variance parameters is that they can become unbounded (creating a situation in which the variance is very high and the preference parameters are essentially meaningless). To prevent this, I include an Inv-$\chi^2(n_{0i}, 0.36)$ prior which is equivalent to having observed n_{0i} observations with variance equal 0.36 (Gelman, Carlin, Stern and Rubin (1995, 261). I let n_{0i} equal 10 percent of the total observations for individual i. A variance of 0.36 implies that an individual with an ideal point of 1 (toward the conservative edge of the spectrum) would have about a 5 percent chance of voting liberal on a vote for which an individual with an ideal point of 0 would have a 50 percent chance of voting liberal.

The posterior distribution is σ_i^2 Inv $\chi^2(n_{0i}+n_i, \frac{n_{0i}0.36+n_i s^2}{n_{0i}+n_i})$. To draw from this distribution, draw Z_i from the $\chi^2_{n_{0i}+n_i}$ and let $\sigma_i^2 = (n_{0i}+n_i)s^2/Z$ (where $s^2 = \frac{\sum(y^*_{itv}-\alpha_v\theta_{it}+\alpha_v\kappa_v)^2}{n_i}$) (see Gelman, Carlin, Stern and Rubin 1995, 480). The code for generating the χ^2 random variable is from Marsaglia and Tsang (2000).

4. Generate $\alpha, \alpha\kappa$ on a vote-by-vote basis. If we let $\beta_v = [\alpha_v, \alpha_v\kappa_v]'$ and $\Theta_v = [\theta_{it}, -1]$ (indicating the preference parameter of individual i for vote v which occurred during term t)we can re-write Equation 3 as

$$y^*_{itv} = \Theta_v\beta_v + \epsilon_{iv}. \tag{8}$$

By standard GLS results,

$$\beta_v \sim N((\Theta_v'\Sigma_v^{-1}\Theta_v)^{-1}\Theta_v'\Sigma_v^{-1}y_v^*), (\Theta_v'\Sigma_v^{-1}\Theta_v)^{-1})$$

where Σ_v is a $V_v \times V_v$ covariance matrix with the individual level variance parameters (σ_i^2) of the individuals who voted on vote v on the diagonal and zero elsewhere (with V_v being the number of votes cast on vote v), and y_v^* is a vector of all votes for vote v.

The discrimination parameter is, in part, a measure of vote-specific variance and, as a variance parameter is subject to becoming unbounded as discussed above (see also Baker 1992, 97-98; Mislevy and Bock 1990, 8). I therefore incorporate normal priors on α following Gelman, Carlin, Stern and Rubin (1995, 254, 260). For more guidance on priors in these models, see Johnson and Albert (1999, 192).

The knowledge we have about the relations between vote cutpoints is incorporated in the following manner. Knowing that a case had a cutpoint lower than another (in the manner discussed in the body of the paper) implies that the cutpoint parameter distribution is truncated at the cutpoint of that other case. Given that cutpoints are jointly distributed with discrimination parameters, I sample from this truncated distribution by drawing from the truncated joint distribution of both vote parameters via rejection sampling.

A model is unidentified "if the same likelihood function is obtained for more than one choice of the model parameters" (Gelman et al 1995, 422). For fixed-preference one-dimensional models, it is sufficient to fix polarity (meaning, for example, conservative preferences are high values and liberal preferences are low values) and two observations (which I do by setting the mean $\theta = 0$ and variance of $\theta = 1$) (see discussions in Clinton, Jackman and Rivers (2004, 356) and Bafumi, Gelman, Park and Kaplan (2005)).

I deal with unbounded discrimination and preference parameters by using a prior and setting maximum values. Missing values are not imputed as a computational convenience that does not affect estimation. In order to facilitate convergence to the true conditional densities I deleted the first $80,000$ iterations (often referred to as the "burn in" period) and took every 40th sample produced thereafter until I had $1,000$ MCMC samples.

2 Data

2.1 Guide to Data Distribution File

The first tab of "XTI_Bailey_DataDistribution_March2007.xls" contains ideal points for members of Congress, senators, presidents and Supreme Court justices from 1951 to 2002. The ID number for representatives, senators and presidents are the ICPSR numbers reported in Poole and Rosenthal data distributions, with .1 appended to senators ICPSR numbers in order to distinguish service in the senate from service in the house for individuals who served in both. There are no votes for a representative from Pennsylvania 14 (ID = 3604) in the 1951 votes in the sample.

Identification numbers for justices all have a prefix of "20" and have individual identification numbers for each justice after the decimal. There are no votes in this sample for justices 20.06 (Rutledge) and 20.07 (Murphy).

The second tab of "XTI_Bailey_DataDistribution_March2007.xls" contains medians for each institution across time. In calculating congressional medians by year, one needs a policy for dealing with instances in which individuals serve only part of the year. I used the member from the seat who voted on the most roll call votes within a given year.

2.2 Sources for bridge observations

The data on amicus filings come from Gibson (1997) for the period 1953 through 1987 and from Lexis-Nexis Academic Universe and the Solicitor General's website thereafter. Only amicus filings on merit are included. For more details on the selection of amicus filing and Senate roll calls, see Bailey and Chang (2001).

Comments in the Senate and House were taken from the *Congressional Record*. For 1989 to present, I used the Thomas.gov database to search for entries with "Supreme Court." For years before that my research assistants and I investigated every entry under "Supreme Court" in the annual indices. I also investigated other sources, such as Eskridge (1991), *Congressional Almanac* and sources cited therein. We also investigated congressional actions that were ruled on in Supreme Court cases. As described below, one has to be

One must be careful when using roll call votes to ascertain members' of Congress positions on Supreme Court cases. First, provisions that address court cases are often embedded in broader legislation. This makes it impossible to know if the vote indicates an opinion on the court case or some other matter. Second, votes are seldom directly on a court ruling. For example, in the 1957 *Mallory* case the court threw out a conviction of a rapist due to delay before his arraignment. Sen. John Marshall Butler (R, MD) sought to overrule the court with legislation providing that evidence could not be ruled inadmissible due to delay. This legislation was amended to say evidence could not be deemed inadmissible because of "reasonable delay." Butler voted against the amended legislation, deeming it too weak, even as most of the supporters of the legislation viewed it as a vote against the court decision (Murphy 1962, 195, 207).

Another example is *Denver Area Educational Telecommunications Consortium v Federal Communications Commission* (518 U.S. 727) (1996) which struck some elements and upheld other elements of the Cable Television Consumer Protection and Competition Act of 1992. This act was passed over the veto of President Bush with nearly universal support of Democrats and substantial support of Republicans (although 85 of the 114 votes against it in the House on October 5, 1992 came from Republicans). The court ruled only on one small part of the bill, a part that put various restraints on cable operators in the interest of controlling "indecent" programming. Using a vote on the overall bill as an indicator of congressional positions on the issue addressed by the Supreme Court would not be reasonable. However, it turns out that the Court explicitly addressed Sections 10 (a) and (b) of the law (upholding the first and striking the second) and that these were added in an amendment by Sen. Helms (R, NC) that passed 95-0. I use the vote on the amendment, but not a vote on passage. Section 10(c) of the law was also explicitly addressed by the court. There was no roll call vote on this, but the legislative history reveals that Sen. Fowler (D, GA) and Sen. Wirth (D, CO) sponsored this language, meaning that the position of these two on this section is clear.

2.3 Selection of Supreme Court cases

I use the Spaeth database and limit cases to those VALUE ≤ 6 (criminal procedure, civil rights, First Amendment, due process, privacy and attorneys). I use citations as the unit of analysis (ANALU =0 in Spaeth's data set) and add split-vote decisions (ANALU = 4) when we have bridging observations. *Bakke* is a prominent example of a case with a split votes and many members of Congress taking positions on one or the other (or both) of the main holdings. I do not include memorandum cases and decrees (DEC TYPE = 3 or 4).

Important cases are those for which at least one of the following is true: discussed directly in the *Harvard Law Review*'s annual court review, included as a landmark case in the Legal Information Institute's database of cases (see supct.law.cornell.edu/supct/cases/name.htm), coded as a salient case in Epstein and Segal (2000), included in the CQ's key cases list, a president or senator or future justice took a position on the case, the case has clear cutpoint relation to another case.

There are a small number of instances in which I do not use Spaeth's coding of the liberal/conservative directionality of a decision. *Buckley* is one such instance, as Spaeth codes decisions to restrict campaign expenditures as conservative (in the sense, I presume, that such a decision limits free speech), when it is clear by the coalition on the court and in Congress that expenditure limits were a liberal reform targeting wealthy contributors.

2.4 Data totals

The total number of legislators, roll calls and Supreme Court cases is indicated in Table A1 (note that for one of the presidents (President Truman) there is extremely limited information).

Table A1: Numbers of individuals and votes

	Total
Legislators	2730
Roll call votes	1342
Presidents	11
Justices	31
Supreme Court cases	1408

Because most individuals do not vote on most roll calls and court cases (because they were not in the institution at the time), this raises the question of whether the bridge observations provide enough data to pin down the relative preferences across the institutions. To address this question, I simulated ideal points and vote parameters for the actual numbers of individuals and votes in the data set. Based on these parameters, I then simulated votes such that I have a simulated observation for every case in which I have an actual observation in the data set, including bridge observations. This simulated data has the same data structure as the actual data, only here the true parameter values are known. I then estimated the model for this simulated world five times. The results indicate that one can identify inter-institutional preferences quite accurately with this data structure. For example, in the five simulations using quadratic preferences, the estimated γ_1 parameter correlates with true γ_1 at 0.94 once and 0.95 four times. Averaging the estimated γ_1 parameters across the five simulations yields an estimate that correlates with the true values at 0.97.

3 DW-Nominate Scores and Court-related studies

In the body of the paper, I discuss how first dimension Common Space scores have been, to some degree, purged of the divisions over race that are highly relevant when thinking about the Supreme Court in the fifties, sixties and seventies. A similar issue arises with DW-Nominate scores. To take one of many examples, civil rights icon Sen. Hubert Humphrey (D-MN) was more conservative on the DW-NOMINATE dimension one than Sen. William Fulbright (D-AR) who signed the Southern Manifesto and opposed major civil rights legislation. In addition, DW-NOMINATE first dimension scores for southern segregationist Senators (such as Smathers (D-FL), Hill (D-AL), Sparkman (D-AL), Ellender (D-LA), Talmadge (D-GA) and Stennis (D-MS)) were more liberal on the first dimension than northern Republicans who were, in fact, liberal on race such as Hatfield (R-OR), Percy (R-IL), Weicker (R-CT) and Packwood (R-OR).

4 Comparison of XTI Measure with ELSW Measure

In the paper I compare results based on XTI and ELSW approaches to preference estimation using two non-nested tests. The Schwarz Criterion approximates the log of the Bayes Factor (Kass and Raftery 1995, 778, 789). Bayes Factors can be used to assess relative value of two hypotheses (Greene 4e, 411). Bayes factors are "the posterior odds of one hypothesis when the prior probabilities of the two hypotheses are equal" (Kass and Raftery 1995, 773). That is, when we have no a priori reason to believe one model is better than another, the Bayes Factor for the two models is the ratio of the probability of the data given model 1 over the probability of the data given model 2. If model 1 does a much better job explaining the data, the Bayes Factor should be very high as the probability of the data under it is much higher than the probability of the data under model 2. This ratio looks similar to a likelihood ratio that is familiar in the likelihood ratio test.[1]

[1] The difference is that the Bayes Factor is not simply to likelihood of the data given the value of the estimated parameters, but it is the probability of the data over all possible values of the parameter space. This introduces integration over the parameter space and prior assumptions about the parameter space and thereby makes for a quite complicated calculation.

The advantage of the Schwarz Criterion is that it requires only the values of the likelihood at the maximized parameter values, eliminating the need for priors on the parameters and complex efforts to approximate the multidimensional integral in the Bayes Factor. The Schwarz Criterion $= \log \Pr(D|\hat{\theta}_1, H_1) - \log \Pr(D|\hat{\theta}_2, H_2) - \frac{1}{2}(d_1 - d_2)log(n)$, where D is the observed data, θ_k are the MLE estimates of the parameters under hypothesis k, H_k is hypothesis k, d_k is the dimensionality of the parameter space under H_k and n is the sample size. When testing two measures against each other, the dimensionality of the parameter space is the same, meaning that the Schwarz Criterion reduces to a comparison of the likelihoods at the maximized values.

Table A2: Non-nested Tests of XTI vs. ELSW Measures

	All Votes (incl voice votes) N = 2247	Recorded Votes N = 1669	Contested Votes N =1184
Schwarz Criterion	57.8	61.75	62.85
Vuong Test	0.0016	0.0001	0.000075

See text for interpretation of values

The first row of Table A2 presents the Schwarz-based tests of the two measures. Three sample selection rules are used: in the first, I use all votes, including voice votes (for which ELSW code every Senator voting for the nominee). Because we cannot be sure who was present for voice votes, I look only at recorded votes in the second specification. Finally, as it is possible that unanimous recorded votes (sometimes referred to as "hurrah" votes) are not representative of contested votes, I also look at all non-unanimous recorded votes (thereby eliminating votes on Blackmun, Stevens, Scalia, O'Connor and Kennedy). The results are based on models that include ELSW variables on the qualifications of the nominee, whether the Senator was in party of president and whether the president's party controlled the Senate. In all specifications, the two models are compared based on identical samples. Kass and Raftery (1995, 777) provide a useful rule of thumb that whenever the Schwarz Criterion is greater than 10 there is strong evidence in favor of model 1 versus model 2; by this standard, there is very strong evidence that the XTI ideology scores do a better job explaining Senate voting.

Another test for non-nested models is the Vuong test (Vuong 1989). The null hypothesis here is that the two theories are equivalent. As Clarke (2001, 734) explains

> if this null hypothesis is true, the average value of the log-likelihood ratio should be zero. If H_f is true, the average value of the log-liklohood ratio should be significantly greater than zero. If the reverse is true, the average value of the log-likelihood ratio should be significantly less than zero. In other words, the Vuong test statistic is simply the average log-likelihood ratio suitably normalized.

The second row of Table A2 reports the p-value from Vuong tests across the specifications defined above.[2] The entries in each cell are the probabilities under the null of observing a likelihood ratio in favor of the XTI measures under the null hypothesis that there is no difference between the measures. Here again, the evidence strongly favors the XTI measure.

[2]The Vuong test statistic is calculated using the "pscl" package in R available from pscl.stanford.edu/content.html.

References for Appendix

Bafumi, Joseph, Andrew Gelman, David Park and Noah Kaplan. 2005. Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation. *Political Analysis* 13: 171-187.

Baker, Frank. 1992. *Item Response Theory.* New York: Marcel Dekker.

Epstein, Lee and Jeffrey Segal. 2000. Measuring Issue Salience. *American Journal of Political Science*, 44: 66-83.

Eskridge, William. 1991. Overriding Supreme Court Statutory Interpretation Decisions. *Yale Law School Journal* 101: 331- 365.

Gelman, Andrew, John Carlin, Hal Stern and Donald Rubin. 1995. *Bayesian Data Analysis.* Boca Raton: Chapman Hall.

Johnson, Valen and James Albert. 1999. *Ordinal Data Modeling.* New York: Springer.

Kass, Robert E and Adrian E. Raftery. 1995. Bayes Factors. *Journal of the American Statistical Association* 90: 773- 795.

Marsaglia, G. and W.W. Tsang. 2000. A Simple Method for Generating Gamma Variables. *ACM Transactions on Mathematical Software* 26: 363-372.

Mislevy, Robert and R. Darrell Bock. 1990. BILOG 3: Item Analysis and Test Scoring with Binary Logistic Models [computer program]. Mooresville IN: Scientific Software, Inc.

Murphy, Walter. 1962. *Congress and the Court.* Chicago: University of Chicago Press.

Vuong, Quang. 1989. Likelihood Ratio Tests of Model Selection and Non-Nested Hypotheses. *Econometrica* 57: 307-333.