Supplementary Materials for “Dynamic Estimation of Latent Opinion Using a Hierarchical Group-Level IRT Model”

Devin Caughey[*]
Department of Political Science
MIT

Christopher Warshaw[†]
Department of Political Science
MIT

2014/11/27

Contents

[*]Assistant Professor, Department of Political Science, Massachusetts Institute of Technology, caughey@mit.edu

[†]Assistant Professor, Department of Political Science, Massachusetts Institute of Technology, cwarshaw@mit.edu

A. DERIVATION OF GROUP-LEVEL NORMAL OGIVE IRT MODEL

This appendix derives the group-level model in Equation 3 in the main text of the paper. The same result is shown by Mislevy (1983), but our derivation is different.

The model depends on the following assumptions:

1. The responses to question j are independent conditional on θ_{ig}, κ_j, and σ_j.
2. Within each group, the θ_{ig} are normally distributed with group-specific means and common variance: $\theta_{ig} \sim \mathrm{N}(\bar{\theta}_g,\ \sigma_\theta^2)$. Note that the common variance implies homoskedasticity of the group ability distributions.
3. The n_{gj} subjects in group g who answer question j were randomly sampled from that group, independently from the $n_{gj'}$ who answer question $j' \neq j$. (This assumption would be violated if each respondent answered more than one question.)

Equation (2) in the main text implies that respondent i in group g answers item j correctly if and only if

$$(\theta_{ig} - \kappa_j)/\sigma_j + \epsilon_{ij} > 0. \tag{1}$$

Multiplying by σ_j, the inequality in Equation 1 becomes

$$\theta_{ig} - \kappa_j + \epsilon_{ij}\sigma_j > 0. \tag{2}$$

Letting $z_{igj} = \theta_{ig} - \kappa_j + \epsilon_{ij}\sigma_j$, the probability that a randomly sampled member of group g correctly answers question j is

$$\Pr[y_{igj} = 1] = \Pr[z_{igj} > 0]. \tag{3}$$

By Assumption 3, the individual abilities θ_{ig} are distributed $\mathrm{N}(\bar{\theta}_g,\ \sigma_\theta^2)$. Since ϵ_{ij} has a standard normal distribution, the term $\epsilon_{ij}\sigma_j$ is distributed $\mathrm{N}(0,\ \sigma_j^2)$. The sum of two independent normal variables has a normal distribution with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$ (DasGupta, 2011, 326), so:

$$z_{igj} \sim \mathrm{N}(\bar{\theta}_g - \kappa_j,\ \sigma_\theta^2 + \sigma_j^2) \tag{4}$$

Since the CDF of a normal variable $X \sim \mathrm{N}(\mu,\ \sigma^2)$ is $\Phi(\frac{x-\mu}{\sigma})$, the CDF of z_{igj} is:

$$\Pr[z_{igj} \leq x] = \Phi[\frac{x - (\bar{\theta}_g - \kappa_j)}{\sqrt{\sigma_\theta^2 + \sigma_j^2}}] \tag{5}$$

which implies:

$$\begin{aligned}
\Pr[z_{igj} > 0] &= 1 - \Phi[\frac{0 - (\bar{\theta}_g - \kappa_j)}{\sqrt{\sigma_\theta^2 + \sigma_j^2}}] \\
&= 1 - \Phi[-(\bar{\theta}_g - \kappa_j)/\sqrt{\sigma_\theta^2 + \sigma_j^2}] \\
&= \Phi[(\bar{\theta}_g - \kappa_j)/\sqrt{\sigma_\theta^2 + \sigma_j^2}] \\
&= p_{gj} \qquad (6)
\end{aligned}$$

"In other words," writes Mislevy (1983, 278), "if [κ_j] and σ_j are the item threshold and dispersion parameters in the subject-level model, then [κ_j] and $\sqrt{\sigma_\theta^2 + \sigma_j^2}$ are the item threshold and dispersion parameters in the group-level model." The response to each question being a Bernoulli draw with constant probability p_{gj}, the sum of correct answers in group g is distributed $s_{gj} \sim \text{Binomial}(n_{gj}, p_{gj})$, where n_{gj} is the number of valid responses to question j in group g.

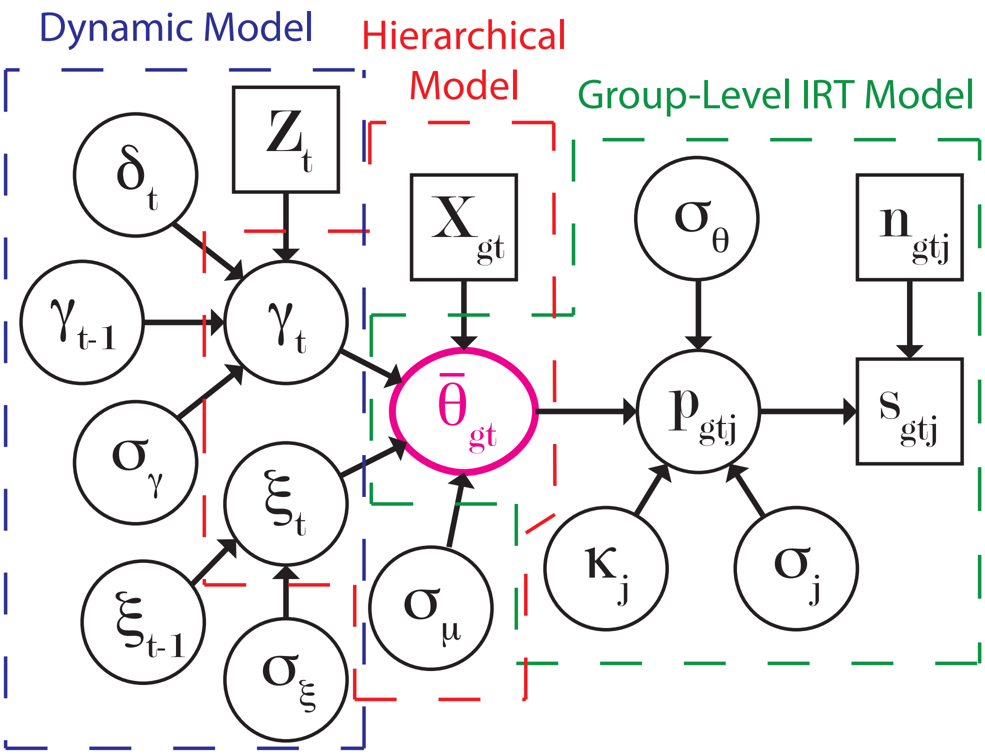

Figure 1: Directed acyclic graph of the dynamic hierarchical group-level IRT model (priors omitted). Squares and circles indicate, respectively, observed and unobserved nodes. Groups are indexed by g, items by j, and time periods by t. The target of inference is $\bar{\theta}_{gt}$: mean latent opinion in each group in each time period.

C. STAN CODE FOR GROUP-LEVEL IRT MODEL

```
data {
    int<lower=1> G; // number of covariate groups
    int<lower=1> Q; // number of items/questions
    int<lower=1> T; // number of years
    int<lower=1> N; // number of observed cells
    int<lower=1> S; // number of geographic units (e.g., states)
    int<lower=1> P; // number of hierarchical parameters, including geographic
    int<lower=1> H; // number of predictors for geographic unit effects
    int<lower=1> H_prior; // number of predictors for geographic unit effects (t=1)
    int<lower=1> D; // number of difficulty parameters per question
    int<lower=0,upper=1> constant_item; // indicator for constant item parameters
    int<lower=0,upper=1> separate_years; // indicator for no over-time smoothing
    int s_vec[N]; // long vector of responses
    int n_vec[N]; // long vector of counts
    int<lower=0> MMM[T, Q, G]; // missingness array
    matrix<lower=0, upper=1>[G, P] XX; // indicator matrix for hierarchical vars.
    row_vector[H] ZZ[T, S]; // data for geographic model
    row_vector[H_prior] ZZ_prior[1, S]; // data for geographic model
}
transformed data {
}
parameters {
    vector[Q] diff_raw[D]; // raw difficulty
    vector<lower=0>[Q] disc_raw; // discrimination
    vector[T] xi; // national mean (common intercept)
    vector[P] gamma[T]; // hierarchical parameters
    vector[T] delta_lag; // weight placed on geo. effects from prev. period
    vector[H] delta_pred[T]; // weight on geographic predictors
    vector[H_prior] delta_pred_prior; // weight on geographic predictors (t=1)
    vector[G] theta_bar[T]; // group mean ability
    vector<lower=0>[T] sd_theta_bar; // sd of group ability means (by period)
    vector<lower=0>[T] sd_theta; // sd of abilities (by period)
    real<lower=0> sd_geo; // prior sd of geographic effects
    real<lower=0> sd_geo_prior; // prior sd of geographic effects (t=1)
    real<lower=0> sd_demo; // sd of demographic effecs
    real<lower=0> sd_innov_delta; // innovation sd of delta_pred and delta_lag
    real<lower=0> sd_innov_logsd; // innovation sd of sd_theta
    real<lower=0> sd_innov_gamma; // innovation sd of gamma, xi, and (opt.) diff
}
transformed parameters {
    vector[Q] diff[D]; // adjusted difficulty
    vector[Q] kappa[D]; // threshold
    vector<lower=0>[Q] disc; // normalized discrimination
    vector<lower=0>[Q] sd_item; // item standard deviation
    vector<lower=0>[Q] var_item; // item variance
    vector<lower=0>[T] var_theta; // variance of abilities
    vector[G] xb_theta_bar[T]; // linear predictor for group means
    vector[G] z[T, Q]; // array of vectors of group deviates
    real prob[T, Q, G]; // array of probabilities
    // Identify model by rescaling item parameters (Fox 2010, pp. 88-89)
```

```
    // scale (product = 1)
    disc <- disc_raw * pow(exp(sum(log(disc_raw))), (-inv(Q)));
    for (q in 1:Q) {
        sd_item[q] <- inv(disc[q]); // item standard deviations
    }
    for (d in 1:D) {
        // location (mean in first year = 0)
        diff[d] <- diff_raw[d] - mean(diff_raw[1]);
        kappa[d] <- diff[d] ./ disc; // item thresholds
    }
    var_item <- sd_item .* sd_item; // item variances
    // Abilities
    var_theta <- sd_theta .* sd_theta; // within-group variances of abilities
    for (t in 1:T) { // loop over years
        xb_theta_bar[t] <- xi[t] + XX * gamma[t]; // Gx1 = GxP * Px1
        for (q in 1:Q) { // loop over questions
            real var_tq; //
            var_tq <- sqrt(var_theta[t] + var_item[q]);
            // Group-level IRT model
            if (constant_item == 0) {
                z[t, q] <- (theta_bar[t] - kappa[t][q]) / var_tq;
            }
            if (constant_item == 1) {
                z[t, q] <- (theta_bar[t] - kappa[1][q]) / var_tq;
            }
            for (g in 1:G) { // loop over groups
                prob[t, q, g] <- Phi_approx(z[t, q, g]); // fast approx. of normal CDF
            } // end group loop
        } // end question loop
    } // end year loop
    // Convert counts and probabilities from array to vector
}
model {
    // TEMPORARY VARIABLES
    real prob_vec[N]; // long vector of probabilities (empty cells omitted)
    int pos;
    pos <- 0;
    // PRIORS
    if (constant_item == 1) {
        diff_raw[1] ~ normal(0, 1); // item difficulty (constant)
    }
    disc_raw ~ lognormal(0, 1); // item discrimination
    sd_geo ~ cauchy(0, 2.5); // sd of geographic effects
    sd_geo_prior ~ cauchy(0, 2.5); // prior sd of geographic effects
    sd_demo ~ cauchy(0, 2.5); // prior sd of demographic parameters
    sd_innov_delta ~ cauchy(0, 2.5); // innovation sd of delta_pred/delta_lag
    sd_innov_gamma ~ cauchy(0, 2.5); // innovation sd. of gamma, xi, and diff
    sd_innov_logsd ~ cauchy(0, 2.5); // innovation sd of theta_sd
    for (t in 1:T) { // loop over years
        if (separate_years == 1) { // Estimate model anew each period
            xi[t] ~ normal(0, 10); // intercept
            for (p in 1:P) { // Loop over individual predictors (gammas)
                if (p <= S) gamma[t][p] ~ normal(ZZ[t][p]*delta_pred[t], sd_geo);
                if (p > S) gamma[t][p] ~ normal(0, sd_demo);
```

```
      }
    }
    if (t == 1) {
      if (constant_item == 0) {
        diff_raw[t] ~ normal(0, 1); // item difficulty
      }
      // Priors for first period
      sd_theta_bar[t] ~ cauchy(0, 2.5);
      sd_theta[t] ~ cauchy(0, 2.5);
      delta_lag[t] ~ normal(0.5, 1);
      delta_pred[t] ~ normal(0, 10);
      delta_pred_prior ~ normal(0, 10);
      if (separate_years == 0) {
        xi[t] ~ normal(0, 10); // intercept
        for (p in 1:P) { // Loop over individual predictors (gammas)
          if (p <= S) {
            gamma[t][p] ~ normal(ZZ_prior[1][p]*delta_pred_prior,
                                 sd_geo_prior);
          }
          if (p > S) gamma[t][p] ~ normal(0, sd_demo);
        }
      }
    }
    if (t > 1) {
      // TRANSITION MODEL
      // Difficulty parameters (if not constant)
      if (constant_item == 0) {
        diff_raw[t] ~ normal(diff_raw[t - 1], sd_innov_gamma);
      }
      // predictors in geographic models (random walk)
      delta_lag[t] ~ normal(delta_lag[t - 1], sd_innov_delta);
      delta_pred[t] ~ normal(delta_pred[t - 1], sd_innov_delta);
      sd_theta_bar[t] ~ lognormal(log(sd_theta_bar[t - 1]), sd_innov_logsd);
      sd_theta[t] ~ lognormal(log(sd_theta[t - 1]), sd_innov_logsd);
      if (separate_years == 0) {
        // Dynamic linear model for hierarchical parameters
        xi[t] ~ normal(xi[t - 1], sd_innov_gamma); // intercept
        for (p in 1:P) { // Loop over individual predictors (gammas)
          if (p <= S) {
            gamma[t][p] ~ normal(delta_lag[t]*gamma[t - 1][p] +
                                 ZZ[t][p]*delta_pred[t], sd_innov_gamma);
          }
          if (p > S) gamma[t][p] ~ normal(gamma[t - 1][p], sd_innov_gamma);
        }
      }
    }
    // RESPONSE MODEL
    // Model for group means
    // (See 'transformed parameters' for definition of xb_theta_bar)
    theta_bar[t] ~ normal(xb_theta_bar[t], sd_theta_bar[t]); // group means
    for (q in 1:Q) { // loop over questions
      for (g in 1:G) { // loop over groups
        if (MMM[t, q, g] == 0) { // Use only if not missing
          pos <- pos + 1;
```

```
                    prob_vec[pos] <- prob[t, q, g];
                }
            } // end group loop
        } // end question loop
    } // end time loop
    // Model for group responses
    s_vec ~ binomial(n_vec, prob_vec); // fully vectorized
}
generated quantities {
    vector<lower=0>[T] sd_total;
    for (t in 1:T) {
        sd_total[t] <- sqrt(variance(theta_bar[t]) + square(sd_theta[t]));
    }
}
```

D. MODEL EXTENSIONS

An advantage of our framework is that our model can (and should) be modified to suit particular analytic purposes. Here, we consider three such possibilities: time-varying item parameters, heterogeneous within-group variances, and a multidimensional latent space. We sketch ways of implementing these extensions to our model and describe applications where they might be useful.

D.1. *Time-Varying Item Parameters*

One possible extension to the model would be to allow the item parameters for each question to evolve over time. The assumption of a constant mapping between the latent θ space and the response probabilities is very useful because it justifies the comparability of estimates over time.[1] For certain items, however, it is clearly implausible, especially with regard to the difficulty parameter α_j.[2]

One possibility is to specify a local-level transition model for $\alpha_{j,t}$:

$$\alpha_{j,t} \sim \text{N}(\alpha_{j,t-1},\ \sigma_\alpha^2). \tag{7}$$

Based on our experimentation with this model, we have found that it helps to identify the model if the difficulty innovation variance σ_α^2 is defined in terms of σ_γ^2, as in $\sigma_\alpha = {}^{\sigma_\gamma}/_{10}$. Substantively, the ratio of σ_α to σ_γ encodes prior beliefs about the magnitude of item-specific change across periods relative to aggregate change in $\bar{\theta}_{gt}$. The downside of allowing $\alpha_{j,t}$ to vary by period is a substantial increase in the number of parameters and as well as in the computational burden of the model. In addition, by altering the mapping between manifest responses and latent opinion, the evolution of item difficulties also complicates the interpretation of opinion estimates from different periods. Whether these additional complexities are worthwhile depends on the application.

D.2. *Heteroskedasticity Across Groups*

As defined in this paper, our model allows $\bar{\theta}_{gt}$ to vary across groups and time but constrains the distributions of $\theta_{i[gt]}$ within groups to be homoskedastic within each period ($\sigma_{\theta,gt} = \sigma_{\theta,t}, \forall g$). This may be misleading if some demographic groups are more hetero-

[1]Other dynamic IRT models, notably Martin and Quinn (2002), achieve comparability over time via a different route. In their Supreme Court application, no cases are repeated and so are not available to bridge across time. Rather, their estimates are comparable across time under the random-walk assumption for the innovation of ideal points and the prior distributions for the item parameters. Other approaches, such as Poole and Rosenthal's (2007) DW-NOMINATE, bridge by constraining ideal-point change to be a polynomial function in time. For an example of cross-period bridging using repeated items, see Asmussen and Jo (2011).

[2]Questions gauging support for gay marriage are an obvious example. While these questions may discriminate well between liberals and conservatives at any point in time, the long-term liberal trend on these questions is not shared by other policy questions. Rather, this issue appears to be governed by idiosyncratic long-term dynamics. Ideally, one would want to account for such issue-specific trends while still allowing the issue to inform cross-sectional differences as well as short-term fluctuations in liberalism.

geneous than others. For example, in many states African Americans may have more homogenous political preferences than other racial groups, particularly if whites and Hispanics are categorized together, as they often are. It is also possible that, over time, demographic groups may become more or less internally diverse. Heterskedasticity of either form may be accommodated by allowing $\sigma_{\theta,t}$ to vary across groups as well as time.[3]

The simplest heteroskedastic specification of the model would simply estimate group-specific values of $\sigma_{\theta,t}$. However, for the same reasons that it makes sense to model $\bar{\theta}_{gt}$ as a function of group covariates, it may also be advantageous to model the $\sigma_{\theta,gt}$. One possible approach is a variance-function regression, which models the variance of the error term as a function of covariates, possibly the same ones as used to model the mean (Park, 1966; see Western and Bloome, 2009 for a Bayesian implementation). One common specification is a log-normal regression. So, for example, the vector of within-group variances of θ_i could be modeled as

$$\boldsymbol{\sigma}_\theta^2 \sim \mathrm{LN}(\mathbf{X}\boldsymbol{\lambda},\ \sigma_{\sigma_\theta}^2), \tag{8}$$

where $\mathbf{X}$ is a matrix of group characteristics (including an intercept), $\boldsymbol{\lambda}$ is a vector of coefficients, and $\sigma_{\sigma_\theta}^2$ is the prior variance of $\boldsymbol{\sigma}_\theta^2$ on the log scale. In addition to potentially providing a better fit to the data, the group variance vector $\boldsymbol{\sigma}_\theta^2$ might be of substantive interest for its own sake.

D.3. *Multidimensionality*

A third natural extension to the model would be to allow for multiple latent dimensions. The question of whether the issue attitudes of the mass public are best modeled with one or multiple dimensions, or possibly none, is an old one and not easily resolved. Between the extremes of unidimensionality (e.g., Jessee, 2009; Tausanovitch and Warshaw, 2013) and little structure at all (e.g., Converse, 1964) lie studies that identify two or three latent dimensions (e.g., Poole, 1998; Peress, 2013). One issue with these multidimensional findings is that secondary dimensions often lack substantive interpretation and do not always correspond to the typical classification of questions into economic, social, and other issue domains (Ellis and Stimson, 2012; cf. Miller and Stokes, 1963; Ansolabehere, Rodden and Snyder, 2008; Treier and Hillygus, 2009).

Adding a second or even a third dimension to the group-level IRT model might shed new light on this long-standing debate. However, it is also likely to exacerbate the computational complexity of the IRT model by greatly increasing the number of parameters (which is approximately proportional to the number of dimensions) as well as the difficulty of identifying the model and mixing through the posterior distributions. As a result, successful estimation of a multidimensional model might require that the model be simplified in other ways (e.g., with groups defined only by state).

[3] On a side note, it would also be possible to allow the variance of the response-level error term e_{ij}, which currently has a standard normal distribution, to vary across groups. This would be equivalent to the approach of Jessee (2010) and Lauderdale (2010), who use such heteroskedasticity to allow for some individuals to behave more "spatially" than others. While the specification would be slightly different, both their approach and the one outlined above would have a similar effect of inflating the denominator of the group-level IRT model (Equation 13 in the main text) with an additional variance component.

E. APPLICATIONS

In this section, we consider two additional applications of our model.

E.1. *Mass Support for the New Deal, 1936–1952*

In the mid-1930s, just as Franklin Roosevelt's New Deal program of liberal reform was reaching its peak, commercial survey firms began fielding the first national opinion polls. The advent of systematic polling was thus well-timed to document the shifts in mass opinion that occurred in the wake of this political watershed. By 1952, when the first American National Election Study was fielded, George Gallup and others had conducted hundreds of commercial opinion polls, querying a total of over one million Americans for their opinions on a multitude of political attitudes and topics (Converse, 1987). Recently, a team led by Adam Berinsky and Eric Schickler has cleaned and standardized the data from these early polls, making them much more accessible to political scientists (Berinsky et al., 2011).

Aside from data issues, a major problem with these early polls is that they were collected with quota-sampling techniques that rendered them unrepresentative of the U.S. population. It is therefore desirable to weight the polls to match known population benchmarks, such as the racial and occupational make-up of each state (Berinsky, 2006). Another difficulty is that a given respondents was rarely asked more than a couple of political questions, and few questions were asked in consistent fashion over many polls. These limitations present a substantial challenge to summarizing the enormous amount of information contained in these polls, either at the individual level (in the form of dimension-reduction techniques) or over time (by, say, tracking consistent question series). These difficulties are what first motivated us to develop the dynamic group-level IRT model described in this paper.

The data for this analysis were derived from quota-sampled Gallup polls fielded between November 1936 and December 1952. These polls contain 453 unique question series asked in identical form across time. Three-quarters of the questions were asked in only a single year; just 18 were asked in more than three different years. Only questions related to such New Deal issues as labor unions, taxation, regulation of the economy, and social welfare were included. A total of 644,370 unique respondents are represented in the data. We coded their responses as either favoring or opposing the New Deal, dichotomizing ordinal responses at an appropriate midpoint.

Respondents were grouped into categories defined by *State* and race-by-region variable (*White South* $\times$ *Black*) with three levels: black, Southern white, and non-Southern white.[4] Within each group, respondents were poststratified to match the joint distribution of *Female* and *Professional* in the population, and the group totals were weighted accrodingly. Including these variables in the model ameliorates the biases introduced by the severe gender, occupational, racial, and regional discrepancies between the poll samples and the population. Mean support for the New Deal in each group was modeled hierarchically as an linear combination of *State* and *White South* $\times$ *Black*. Except in the first year, when the state intercepts were modeled as a function of four-category region, no state-level characteristics

[4]Following Gallup's regional categorization scheme, the South was defined as the eleven states of the former Confederacy plus Kentucky and Oklahoma.

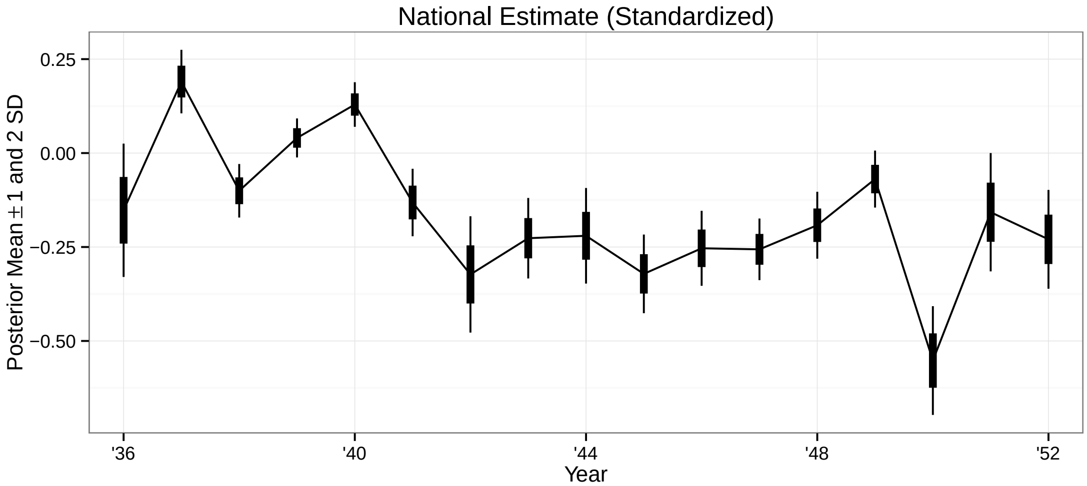

Figure 2: Dynamic group IRT estimates of mean support for the New Deal in the United States, 1936–52. Error bars represent 1 and 2 standard deviations around the mean of the posterior distribution. The estimates have been standardized by the cross-sectional standard deviation of New Deal support in the median year.

were included in the model.

One of the virtues of estimating opinion by group is that the group estimates can be weighted to match whatever the population of interest happens to be. In this case, it is useful to focus not on the U.S. adult population as a whole, but rather on the population minus Southern blacks. We do this for two reasons. First, in this period Southern blacks were almost entirely disfranchised, so they were not part of the potential electorate (Key, 1984[1949]). Second, for this reason, blacks were severely undersampled in Southern states, so our estimates for Southern blacks would be extrapolating heavily (via the multilevel model) from the opinions of Northern blacks. Thus, though black respondents from the South were included in the data used to estimate the model, we poststratify the estimated group means to match the population minus Southern blacks, implicitly given them zero weight in our estimates for Southern states. All estimates below are based on this definition of the U.S. population.

Figure 2 plots estimated mean support for the New Deal in the United States between (the last two months of) 1936 and 1952. The figure displays a large and sharp turn against the New Deal that coincided with U.S. mobilization for the Second World War (1941–42). Since the estimates have been scaled by the standard deviation across individuals in a typical year, the figure implies that that the American public moved almost half a standard deviation to the right between 1940 and 1942. Aside from an anomalous deviation in 1950—which may reflect the small number of questions (seven) in that year—national support for the New Deal was relative stable after 1942.

Now consider the cross-sectional state comparisons presented in 3, which require less-stringent assumptions than do the over-time comparisons. These maps reveal a striking realignment of state opinion between 1937 and 1943, as the South transformed from the region most supportive of the New Deal to the most conservative region. While the South's

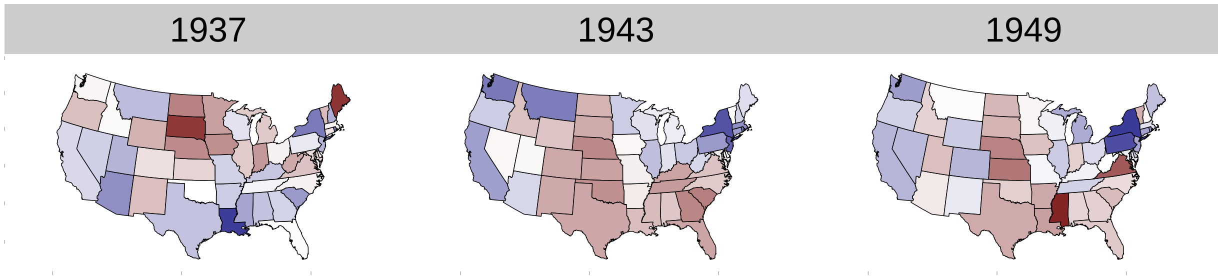

Figure 3: State support for New Deal liberalism in 1937, 1943, and 1949. Estimates have been centered and standardized in each year to accentuate the color contrasts.

turn against liberalism has been noted by scholars (e.g., Ladd and Hadley, 1975), this is the first time its extent and timing has been documented with any precision.

The state-level opinion estimates also permit examination of the relationship between mass liberalism and the voting records of their representatives in Congress. Figure 4 plots the average first-dimension DW-NOMINATE score of state Senate delegations against mean support for New Deal liberalism in the state publics. Two noteworthy patterns emerge from this graph. First, consistent with Figure 3, the Southern white public began the period more liberal than the rest of the nation but quickly become more conservative than average. This sharp regional shift is an exception to the normal pattern of state ideological stability, at least in survey-based measures (for a debate on this point, see Berry et al., 2007 and Erikson, Wright and McIver, 2007).

Second, within each region, mass liberalism and Senate conservatism are quite negatively correlated, especially after 1942. The empirical correspondence between these theoretically related measures provides additional construct validation for our model. The strength of the relationship in the (white) South is somewhat surprising, however, given the absence of partisan competition in the one-party region and the fact that many whites were disfranchised along with blacks (Key, 1984[1949]; Mickey, 2014). But it is consistent with other recent evidence of representatives' responsiveness to the preferences of the potential electorate in the one-party South (Schickler and Caughey, 2011; Caughey, 2012).

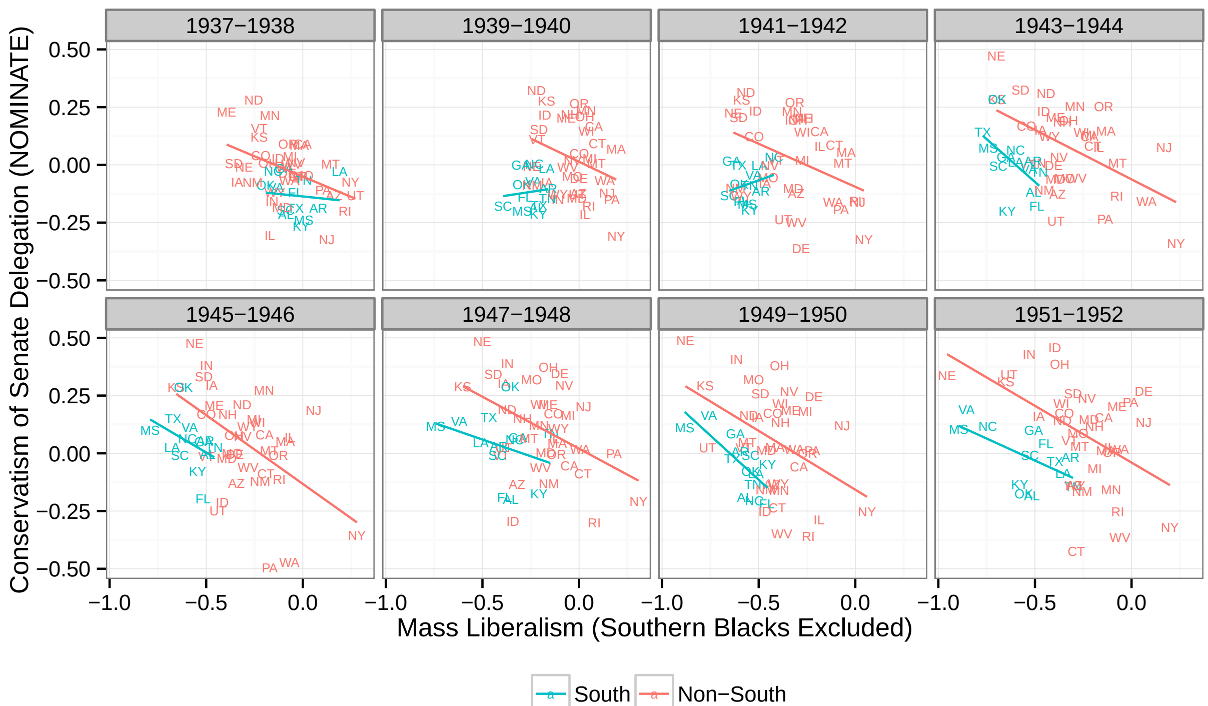

Figure 4: Relationship between mean liberalism of state publics and mean conservatism of state Senate delegations (as measured by first-dimension DW-NOMINATE scores). State opinion estimates exclude Southern blacks, who were disenfranchised at this time. Estimates are pooled within two-year periods corresponding to congressional sessions.

E.2. *State Confidence in the Supreme Court, 1965–2010*

Public opinion on the Supreme Court plays a key role in many theories of judicial politics (for an overview, see Persily, Citrin and Egan, 2008). A major theme in this literature is the effect of public confidence in the Supreme Court on the interaction between the Court and other branches. Because the Court is sensitive to how it is perceived by the public (Baum, 2009), it is more likely to issue unpopular decisions or strike down acts of Congress when it is relatively popular (Caldeira, 1987; Carrubba, 2009; Clark, 2011; Hausseger and Baum, 1999). Congress is also sensitive to how the Court is perceived by the public. Members of Congress are more likely to support legislation that limits the Court's power when public support for the Court is low (Clark, 2009, 2011). In addition, scholars have examined the factors that explain changes in the public's confidence in the Court over time. Mondak and Smithey (1997) find that the Court's support erodes when its decisions diverge from the ideological preferences of the American public.

Previous empirical work on the role of public opinion in judicial politics has been hampered by the difficulty in measuring confidence in the Court either over time or across states. Clark (2009) writes that "public opinion data about the Court are notoriously sparse" (p. 979). Scholars have generally measured support for the Court using aggregated responses to the General Social Survey (GSS) and Harris polls (Caldeira, 1986; Clark, 2009, 2011). But this approach leaves scholars with just a few dozen survey responses in individual states in a given year.[5] Our model builds upon previous approaches by pooling across survey questions and polling firms to estimate latent trust in the Supreme Court at the state-level. Our dynamic model enables us to estimate latent confidence in the Supreme Court even in years with little or no available survey data. This new measure could enable scholars to re-examine whether Senators are more likely to support legislation that limits the Court's power when public support for the Court is low. It also enables scholars to expand our analysis of the interaction between the Court and political officials to new arenas. For instance, scholars could examine whether state-level officials are more likely to challenge the Court when the Court is unpopular in their state.

We use data from 72 polls between 1963 and 2010 with approximately 166,000 total respondents. We use four question series as indicators of confidence in the Court.[6] Some of these questions have multiple ordinal response categories (e.g., "very favorable", "favorable", etc.). To maximize the range of cutpoints with respect to the underlying latent variable, we convert each ordinal variable into a set of dichotomous variables that indicate whether the response was above a given threshold. We model the sum of each of these dichotomous variables, sampling one variable from each respondent so as to avoid having multiple responses from a given individual.

[5] Clark (2011) develops better state-level estimates by using a multi-level regression with poststratification (MRP) model with data from the GSS. But this approach provides no solution to the fact that in some years there is no data at all available from the GSS or Harris surveys. Moreover, it fails to utilize all of the available data from Gallup and other survey firms on judicial approval or confidence.

[6] We use the items: 1) Do you approve or disapprove of the way the Supreme Court is handling its job? 2) In general, what kind of rating would you give the Supreme Court? 3) Would you tell me how much respect and confidence you have in the Supreme Court? 4) Is your overall opinion of the Supreme Court very favorable, mostly favorable, mostly unfavorable, or very unfavorable?

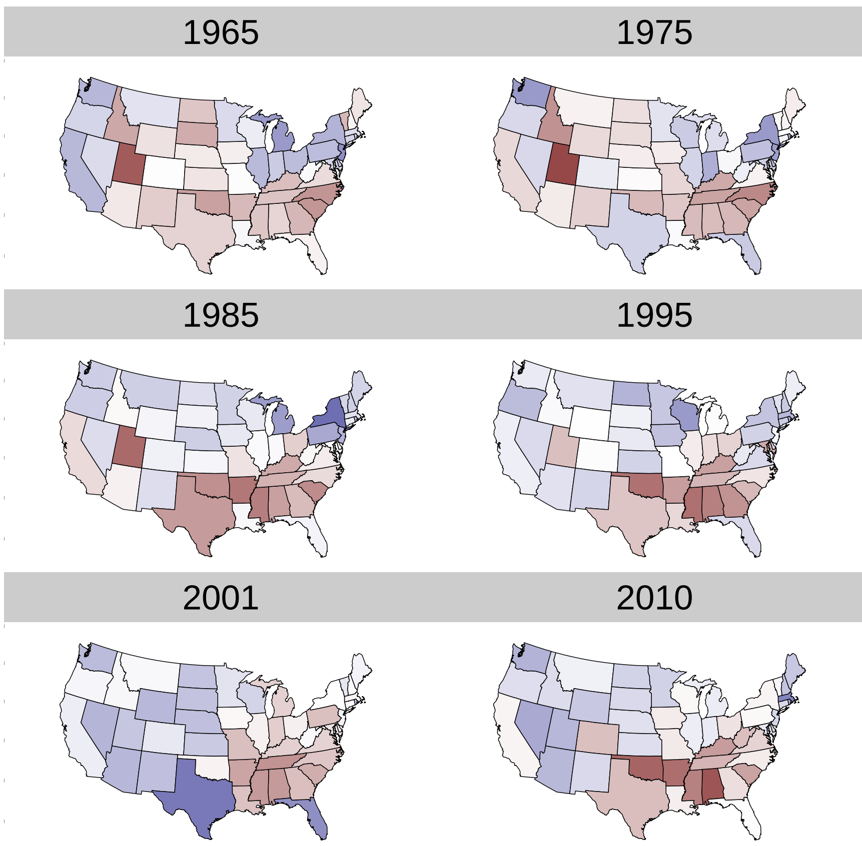

Figure 5: Average state confidence in the Supreme Court, 1965–2010. Blue indicates greater confidence. State estimates have been normalized in each year to highlight cross-sectional differences.

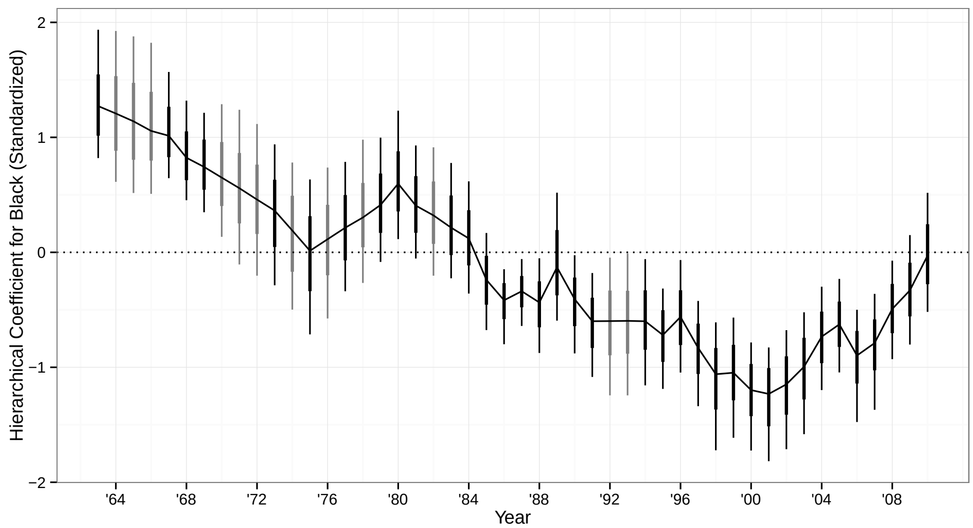

Figure 6: Year-specific estimates of the hierarchical coefficent for the demographic predictor *Black*. The estimates have been standardized by the cross-sectional standard deviation of latent judicial confidence in a typical year. Gray bars indicate years for which no poll data are available.

Figure 5 compares state-level support for the Court across the past five decades. In the early part of the period, there is generally lower support for the Court in the South, which probably reflects Southern whites' dissatisfaction with the Court liberal decisions on school de-segregation and criminal justice. In contrast, there is very strong support for the Court in liberal, northern states during the 1960s and early 1970s. Over time, however, support for the Court drops in northern states and rises in southern states. These changes likely reflect the general shift in the Court's orientation to the ideological right.

A different angle on these same phenomenon is provided by Figure 6, which plots the yearly estimated coefficients for *Black* in the hierarchical model. The estimates have been standardized by the cross-sectional standard deviation of latent judicial confidence in a typical year. In 1963, blacks were predicted to be over a standard deviation more confident in the Supreme Court than non-blacks, conditional on their other demographic and geographic characteristics. Black support dropped as the Court became less closely associated with civil rights and more conservative generally. After bottoming out around 2000, blacks' judicial confidence rebounded, especially after the election of Barack Obama in 2008. These shifts in blacks' relative confidence in the Court highlight the importance of allowing the hierarchical model to evolve over time.

We validate our estimates by using them to predict co-sponsorship of court-curbing bills in Congress. Clark (2011) argues that legislators should be more likely to sponsor court-curbing bills when there is substantial disapproval of the Court in the legislators' constituency. Clark (2011) shows that members of the U.S. House are more likely to sponsor court curbing bills when there is substantial disapproval of the Court in their home state. However, Clark's theoretical logic is actually stronger for the Senate than the House since there should be a closer fit there between state-level estimates of judicial confidence and senators' constituencies.

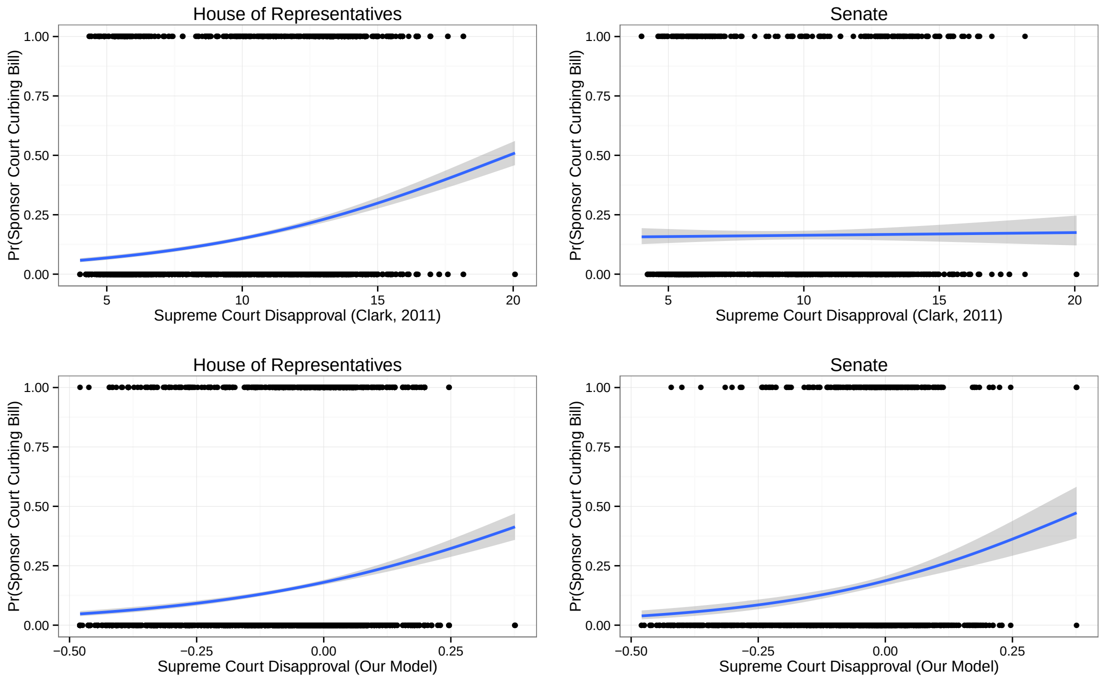

Figure 7: Relationship between probability of sponsoring a court-curbing bill and state public disapproval of the Supreme Court. Top row uses estimates from Clark (2011); bottom row uses our group-level IRT estimates of confidence in the Supreme Court (reverse-coded).

In the top-row of figure 7, we replicate Clark's results for members of the U.S. House using his MRP-based measure of judicial confidence and co-sponsorships of court curbing bills by representatives. As the top-right panel shows, however, Clark's state-level estimates of Court disapproval are uncorrelated with senatorial support for court-curbing. The bottom row of Figure 7 conducts the same analysis using our measure of state-level confidence in the Court (reverse-coded). Our estimates not only predict House court-curbing as well as Clark's, but they also predict it in the Senate. These results both reinforce the validity of our measure of confidence in the Supreme Court and demonstrate our model's empirical usefulness for studying constructs other than policy preferences.

REFERENCES

Ansolabehere, Stephen, Jonathan Rodden and James M. Snyder, Jr. 2008. "The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting." *American Political Science Review* 102(2):215–232.

Asmussen, Nicole and Jinhee Jo. 2011. "Anchors Away: A New Approach for Estimating Ideal Points Comparable Across Time and Chambers." Unpublished manuscript. Available for download at http://my.vanderbilt.edu/nicoleasmussen/files/2011/08/Anchors-Away-updated-March-28-2011.pdf.

Baum, Lawrence. 2009. *Judges and Their Audiences: A Perspective on Judicial Behavior*. Princeton, NJ: Princeton University Press.

Berinsky, Adam J. 2006. "American Public Opinion in the 1930s and 1940s: The Analysis of Quota-Controlled Sample Survey Data." *Public Opinion Quarterly* 70(4):499–529.

Berinsky, Adam J., Eleanor Neff Powell, Eric Schickler and Ian Brett Yohai. 2011. "Revisiting Public Opinion in the 1930s and 1940s." *PS: Political Science & Politics* 44(3):515–520.

Berry, William D., Evan J. Ringquist, Richard C. Fording and Russell L. Hanson. 2007. "The Measurement and Stability of State Citizen Ideology." *State Politics & Policy Quarterly* 7(2):111–132.

Caldeira, Gregory. 1987. "Public Opinion and the U.S. Supreme Court: FDR's Court-Packing Plan." *American Political Science Review* 81(4):1139–1153.

Caldeira, Gregory A. 1986. "Neither the Purse Nor the Sword: Dynamics of Public Confidence in the Supreme Court." *American Political Science Review* 80(4):1209–26.

Carrubba, Clifford James. 2009. "A Model of the Endogenous Development of Judicial Institutions in Federal and International Systems." *Journal of Politics* 71(1):55–69.

Caughey, Devin. 2012. "Congress, Public Opinion, and Representation in the One-Party South, 1930s–1960s." PhD dissertation, UC Berkeley.

Clark, Thomas S. 2011. *The Limits of Judicial Independence*. New York: Cambridge University Press.

Clark, Tom S. 2009. "The Separation of Powers, Court Curbing, and Judicial Legitimacy." *American Journal of Political Science* 53(4):971–89.

Converse, Jean M. 1987. *Survey Research in the United States: Roots and Emergence*. Berkeley: University of California Press.

Converse, Philip E. 1964. The Nature of Belief Systems in Mass Publics. In *Ideology and Discontent*, ed. David E. Apter. London: Free Press pp. 206–261.

DasGupta, Anirban. 2011. *Fundamentals of Probability: A First Course*. Springer (PDF ebook).

Ellis, Christopher and James A. Stimson. 2012. *Ideology in America*. New York: Cambridge UP.

Erikson, Robert S., Gerald C. Wright and John P. McIver. 2007. "Measuring the Public's Ideological Preferences in the 50 states: Survey Responses versus Roll Call Data." *State Politics & Policy Quarterly* 7(2):141–151.

Hausseger, Lori and Lawrence Baum. 1999. "Inviting Congressional Action: A Study of Supreme Court Motivations in Statutory Interpretation." *American Journal of Political Science* 43(1):162–85.

Jessee, Stephen A. 2009. "Spatial Voting in the 2004 Presidential Election." *American Political Science Review* 103(1):59–81.

Jessee, Stephen A. 2010. "Partisan Bias, Political Information and Spatial Voting in the 2008 Presidential Election." *Journal of Politics* 72(2):327.

Key, Jr., V. O. 1984[1949]. *Southern Politics in State and Nation*. Knoxville: University of Tennessee Press.

Ladd, Everett Carll and Charles D. Hadley. 1975. *Transformations of the American party system: Political Coalitions from the New Deal to the 1970s*. New York: Norton.

Lauderdale, Benjamin E. 2010. "Unpredictable Voters in Ideal Point Estimation." *Political Analysis* 18:151–171.

Martin, Andrew D. and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999." *Political Analysis* 10(2):134–153.

Mickey, Robert W. 2014. *Paths Out of Dixie: The Democratization of Authoritarian Enclaves in America's Deep South*. Princeton, NJ: Princeton UP.

Miller, Warren E. and Donald E. Stokes. 1963. "Constituency Influence in Congress." *American Political Science Review* 57(1):45–56.

Mislevy, Robert J. 1983. "Item Response Models for Grouped Data." *Journal of Educational Statistics* 8(4):271–288.

Mondak, Jeffrey J. and Shannon Ishiyama Smithey. 1997. "The Dynamics of Public Support for the Supreme Court." *Journal of Politics* 49(4):1114–42.

Park, R. E. 1966. "Estimation with Heteroscedastic Error Terms." *Econometrica* 34(4):888.

Peress, Michael. 2013. "Candidate Positioning and Responsiveness to Constituent Opinion in the U.S. House of Representatives." *Public Choice* 156(1-2):77–94.

Persily, Nathaniel, Jack Citrin and Patrick J. Egan, eds. 2008. *Public Opinion and Constitutional Controversy*. New York: Oxford University Press.

Poole, Keith T. 1998. "Recovering a Basic Space From a Set of Issue Scales." *American Journal of Political Science* 42(3).

Poole, Keith T. and Howard Rosenthal. 2007. *Ideology & Congress.* New Brunswick, NJ: Transaction Publishers.

Schickler, Eric and Devin Caughey. 2011. "Public Opinion, Organized Labor, and the Limits of New Deal Liberalism, 1936–1945." *Studies in American Political Development* 25(October):1–28.

Tausanovitch, Chris and Christopher Warshaw. 2013. "Measuring Constituent Policy Preferences in Congress, State Legislatures and Cities." *Journal of Politics* 75(2):330–342.

Treier, Shawn and D. Sunshine Hillygus. 2009. "The Nature of Political Ideology in the Contemporary Electorate." *Public Opinion Quarterly* 73(4):679–703.

Western, Bruce and Deirdre Bloome. 2009. "Variance Function Regressions for Studying Inequality." *Sociological Methodology* 39(1):293–326.