Measuring District Level Partisanship with Implications for the Analysis of U.S. Elections: On-line Appendix.

Matthew S. Levendusky,* Jeremy C. Pope,† and Simon D. Jackman‡

1 The Question of Dimensionality

We treat each district's partisanship as an unknown point on a single latent dimension. Our assumption of unidimensionality warrants some brief discussion. District partisanship ranges between two theoretical pure types: a purely Democratic district and a purely Republican district. Each district lies between these two pure types, and hence the latent trait in our model is unidimensional. Were we measuring district *ideology*, say, using survey data, then a multidimensional latent trait might be more appropriate, but this is not the case. The assumption of unidimensionality is consistent with a long tradition in the study of American electoral politics in which two-party competition is the norm, and the preferences of candidates, parties and voters are represented as points on a single dimension. Examples include characterizations of electoral competition and the two-party system (e.g., Downs 1957; Black 1987; Aldrich 1995) and much empirical work on congressional elections (e.g., Ansolabehere, Snyder and Stewart 2001; Canes-Wrone, Cogan and Brady 2002; Jacobson 2004) and is implicit in Converse's (1966) initial formulation of the normal vote. Indeed, in the specific context of congressional elections, the assumption of a unidimensional continuum is typically used without question.

The electoral returns we analyze strongly support the assumption of unidimensionality over the period under study. Table 1 reports the results of an extremely simple first look at the electoral data via principal components analysis (e.g., Joliffe 2002); for each decade, we computed the correlation matrix for the 5 sets of congressional vote shares and 2 or 3 sets of presidential vote shares (all on the log-odds scale), and examined how much variation was accounted for by the first principal component. In every decade, the 1st eigenvalue of the correlation matrix is quite large relative to the number of elections available for analysis (typically greater than 5), and the 2nd eigenvalue is always less than 1.0, strongly suggesting that a single dimension underlies the vote data. Similarly, the amount of variation in the electoral returns data accounted for by the first principal component

*Assistant Professor, Department of Political Science, University of Pennsylvania.

†Assistant Professor, Department of Political Science, Brigham Young University.

‡Professor and Director, Political Science Computational Laboratory, Department of Political Science, Stanford University.

is relatively large, ranging from a minimum of 69% in the 1970s data, to 84% in the 1990s data. We also examined the proportion of variation in a given decade's vote shares that is attributable to cross-district or between-district variation (equivalent to the r^2 from a regression of a decade's vote totals on district fixed effects); in each decade it is clear that bulk of the variation in the votes is cross-district variation, with the elections of the 1990s generating the least amount of within-district variation, and the 1970s elections generating the most. Thus both theory and data support an assumption that levels of district partisanship can be modelled as points on a unidimensional continuum.

	# elections	1st Eigenvalue	2nd Eigenvalue	Proportion Variation Explained By 1st PC	District Fixed Effects
1950s	8	6.53	0.68	0.82	0.68
1960s	7	4.98	0.78	0.71	0.64
1970s	8	5.55	0.91	0.69	0.57
1980s	7	5.47	0.72	0.78	0.70
1990s	8	6.72	0.63	0.84	0.82

Table 1: Common Variance in Vote Data, by decade.

2 The MCMC algorithm

The MCMC algorithm used to estimate our model is initialized with all parameters set to zero, but with latent district partisanship set to the re-scaled average of Presidential vote in the district. These start values are merely used to initialize the algorithm and are of no substantive consequence; further details appear in the next section. The MCMC algorithm quickly moves away from these start values to generate a random tour of the posterior density. After 25,000 burn-in iterations (letting the algorithm step away from the arbitrary initial values) we run another 250,000 iterations of the MCMC algorithm, saving every 250th iteration to get 1,000 samples from the joint posterior distribution. Trace plots and other diagnostics (available upon request) strongly indicate that the MCMC algorithm had converged on the joint posterior density.

3 Conditional Distributions for the Gibbs Sampler

For the purposes of this exposition, our model can be re-written generically as follows:

$$y_{ij} \sim N(\mu_{ij}, \nu_J^2) \quad (1)$$

$$\mu_{ij} = \mathbf{c}'_{ij}\gamma_j + x_i\beta_j \quad (2)$$

$$x_i \sim N(\mathbf{z}'_i\alpha, \sigma^2) \quad (3)$$

where y_{ij} is the log-odds of the Democratic vote share for election j in district i, where $i = 1, \ldots n$ and $j = 1, \ldots, m$; $\mathbf{c}_{ij}$ is a vector of covariates specific to district i in election j (e.g., home-state indicators, incumbency indicators, etc) plus a constant term, and γ_j is a vector of parameters to be estimated, including an election-specific fixed effect (intercept); x_i is the latent preference of district i, and β_j is a discrimination parameter (equivalent to a factor loading); $\mathbf{z}_i$ is a vector of district-specific social-structural and demographic aggregates, and α is a vector of unknown parameters; and ν_j^2 and σ^2 are unknown variance parameters to be estimated. For ease of exposition, assume there is no missing data. Let $\mathbf{\Theta}$ denote the complete set of unknown parameters: i.e., $\mathbf{\Theta} = \{\mathbf{x}, \mathbf{\Gamma}, \beta, \alpha, \nu, \sigma^2\}$, where the vectors and matrices in $\mathbf{\Theta}$ are formed by stacking the corresponding parameters across elections $j = 1, \ldots, m$.

In the Bayesian approach we adopt, we wish to compute the posterior density of the unknown parameters, $p(\mathbf{\Theta}|\mathbf{Y}, \mathbf{Z})$ = $p(\mathbf{x}, \mathbf{\Gamma}, \beta, \alpha, \nu, \sigma^2|\mathbf{Y}, \mathbf{Z}, \mathbf{C})$, where $\mathbf{Y}$, $\mathbf{Z}$ and $\mathbf{C}$ are matrices formed from y_{ij}, $\mathbf{z}_i$, $\mathbf{c}_{ij}$ respectively. Via Bayes' Rule, this posterior density is proportional to the prior density over the parameters multiplied by the likelihood, i.e.,

$$\begin{aligned}
p(\mathbf{x}, \mathbf{\Gamma}, \beta, \alpha, \nu, \sigma^2|\mathbf{Y}, \mathbf{Z}) &\propto \prod_{i=1}^{n}\prod_{j=1}^{m} f(y_{ij}|\mathbf{z}_i, \mathbf{c}_{ij}, \mathbf{\Theta})p(\mathbf{\Theta}) \\
&= \prod_{i=1}^{n}\prod_{j=1}^{m} f(y_{ij}|\mathbf{x}_i, \mathbf{c}_{ij}, \gamma_j, \beta_j, \nu_j^2) f(x_i|\mathbf{z}_i, \alpha, \sigma^2) p(\mathbf{\Theta}_{-\mathbf{x}})
\end{aligned}$$

where $p(\mathbf{\Theta})$ is the prior density of the parameters and $\mathbf{\Theta}_{-\mathbf{x}}$ denotes $\mathbf{\Theta}$ without $\mathbf{x}$. Here we assume that the parameters are *a priori* independent; i.e., $p(\mathbf{\Theta}) = \prod_{i=1}^{n} p(x_i|\mathbf{z}_i, \alpha, \sigma^2) \prod_{j=1}^{m} p(\gamma_j)p(\beta_j)p(\nu_j^2) \cdot p(\alpha)p(\sigma^2)$. As discussed in the text, we use vague normal priors for γ_j, β_j and α parameters, and diffuse inverse-Gamma priors for the variance parameters ν_j^2 and σ^2. Combined with the normal likelihood for both the y_{ij} and x_i, the use of normal and inverse-Gamma priors makes the Bayesian analysis of this model relatively straightforward, at least mathematically: specifically, this is an instance of *conjugate* Bayesian analysis, in that the posterior densities for all parameters will be in the same family as the corresponding prior density.

The only difficulty here is computational, arising from the fact that the posterior density is high-dimensional: n.b., there are as many x_i parameters as there are unique House districts in any given decade. We use computationally-intensive, simulation methods (Markov chain Monte Carlo, or MCMC) to explore this density. MCMC algorithms generate random tours of parameter spaces; the output of each iteration of a MCMC algorithm is a set of values for the unknown parameters, a location in the parameter space. When run long enough, and subject to some mild regularity conditions, MCMC algorithms visit regions in the parameter space with frequency proportional to the posterior probability of a given region.

The Gibbs sampler is the workhorse MCMC algorithm. An iteration of the Gibbs sampler consists of a pass over all parameters in the model, sampling from the conditional distribution of each parameter, where the conditioning is on the other parameters in the model and the data: i.e., at iteration t, sample $\theta_j^{(t)}$ from $p(\theta_j|\theta_1^{(t)}, \ldots \theta_{j-1}^{(t)}, \theta_{j+1}^{(t-1)}, \ldots, \theta_J^{(t-1)}, \mathbf{Y}, \mathbf{Z})$, where j indexes a partitioning of $\mathbf{\Theta}$ into J components. Because of conjugacy, all of these J conditional distributions are either normal or inverse-Gamma distributions, easy to sample from, and with means and variances (or shape and rate parameters, in the case of the Gamma distributions) that are easily computed. For expository purposes, we derive the conditional distribution for the key quantity in our analysis, the latent district partisanship x_i.

To derive conditional distributions for a Gibbs sampler, we exploit the following: (1) almost all statistical models (and, in particular, hierarchical models of the sort we use here) can be represented as conditional independence graphs, or directed acyclic graphs (or DAGs); (2) if θ_j is a node in DAG

$\mathcal{G}$, then its conditional distribution is[1]

$$p(\theta_j|\mathcal{G}_{-\theta_j}) \propto p(\theta_j|\text{parents}[\theta_j]) \times \prod_{\theta_k \in \text{children}[\theta_j]} p(\theta_k|\text{parents}[\theta_k]) \tag{4}$$

where the directed edges in the DAG correspond to parent-to-children relations.

We now use this result to derive the conditional distribution of x_i. Inspection of equation 1, 2 and 3 indicates that y_{ij} are the children nodes of x_i, $j = 1, \ldots, m$, while α and σ^2 are parent nodes of x_i. Let $w_{ij} = y_{ij} - \mathbf{c}'_{ij}\gamma_j$. Then equation 1 can be re-written as $w_{ij} \sim N(x_i\beta_j, \nu_j^2), \ \ j = 1, \ldots, m$, with the w_{ij} children nodes of x_i. Then, via equation 4,

$$p(x_i|\mathcal{G}_{-x_i}) \propto \phi\left(\frac{x_i - \mathbf{z}'\alpha}{\sigma}\right) \prod_{j=1}^{m} \phi\left(\frac{w_{ij} - x_i\beta_j}{\nu_j}\right), \tag{5}$$

where ϕ is the standard normal probability density function. Note that x_i is the regression coefficient in the heteroskedastic regression of the w_{ij} on the β_j, $j = 1, \ldots, m$, with the heteroskedasticity arising because of the different variance terms, the ν_j^2. The MLE of x_i is thus

$$\hat{x}_i = \frac{\sum_j(\beta_j w_{ij}/\nu_j^2)}{\sum_j(\beta_j^2/\nu_j^2)} \tag{6}$$

with variance $\sum_j(\nu_j^2/\beta_j^2)$. A well-known and frequently-used result in Bayesian statistics is that the product of a normal prior over a regression coefficient and the MLE from a normal-linear regression model is itself a normal distribution with (1) mean equal to the precision-weighted average of the means of each distribution and (2) variance equal to the inverse of the sum of the precisions of each distribution, where precision is defined as the inverse of the variance (e.g., Leamer 1978, Theorem 3.9). Applying this result means that the conditional distribution for x_i in equation 5 is a normal distribution with mean

$$\left[\sum_j \left(\frac{\beta_j w_{ij}}{\nu_j^2}\right) + \frac{\mathbf{z}'_i\alpha}{\sigma^2}\right] \cdot \left[\sum_j \left(\frac{\beta_j^2}{\nu_j^2}\right) + \frac{1}{\sigma^2}\right]^{-1}$$

and variance

$$\left[\sum_j \left(\frac{\beta_j^2}{\nu_j^2}\right) + \frac{1}{\sigma^2}\right]^{-1}.$$

Similar derivations yield the conditional distributions for the other parameters in the model.

4 Post-Processing the Gibbs Sampler Output

As written, the model in equations 1 through 3 is not identified. In particular, any re-scaling of x_i yields the same value of the likelihood via offsetting re-scalings of the β_j parameters: i.e., since

[1]See Spiegelhalter and Lauritzen (1990) and Spiegelhalter, Thomas and Best (1996).

$x_i \sim N(\mathbf{z}_i'\alpha, \sigma^2)$, we can re-write $\mu_{ij} = \mathbf{c}_{ij}'\gamma_j + (\mathbf{z}_i'\alpha + \epsilon_i)\beta_j$, where $\text{var}(\epsilon_i) = \sigma^2$. Note that for any $(\alpha, \sigma^2, \beta)$, we can obtain the same value of μ_{ij} with the re-scaled parameters $\alpha^* = g\alpha$, $\sigma^{2*} = g^2\sigma^2$ and $\beta_j^* = \beta_j/g$, for any $g \neq 0$. Likewise, consider a translation of the x_i via an additive change in the α_j to $\alpha_j^* = \alpha_j + \mathbf{d}$, $\mathbf{d} \neq \mathbf{0}$. Then $\mu_{ij}^* = \mathbf{c}_{ij}'\gamma_j^* + \mathbf{z}_i'\alpha\beta_j + \mathbf{z}_i'\mathbf{d}\beta_j + \epsilon_{ij} = \mu_{ij}$, via an offsetting additive shift in the intercept parameter contained in γ_j^* (the election-specific fixed effect); i.e., with $\alpha^* = \alpha + \mathbf{d}$, $\mu_{ij}^* = \mu_{ij}$ if the intercept term $\gamma_{j1}^* = \gamma_{j1} - \mathbf{z}_i'\mathbf{d}\beta_j, j = 1, \ldots, m$.

We actually ignore this lack of identification when implementing the Gibbs sampler for our model. We impose local identification by post-processing the output of the Gibbs sampler. Specifically, we impose the identifying constraint that the x_i have mean zero and standard deviation one across districts, removing the indeterminacy due to scale and translation discussed above.[2] We impose this constraint as follows. At iteration t of the Gibbs sampler: (1) center and scale the $x_i^{(t)}$ to have unit variance: i.e., $\tilde{x}_i^{(t)} = (x_i^{(t)} - a)/b$, where $a = \bar{x}^{(t)}$ and $b = \text{sd}(\mathbf{x}^{(t)})$; (2) let $\tilde{\alpha}^{(t)} = b^{-1}\alpha^{(t)}$; (3) let $\tilde{\sigma}^{2(t)} = b^{-2}\sigma^{2(t)}$; (4) let $\tilde{\beta}_j^{(t)} = b\beta_j^{(t)}, j = 1, \ldots, m$; (5) let $\tilde{\gamma}_{j1}^{(t)} = \gamma_{j1}^{(t)} + \beta_j^{(t)}a, j = 1, \ldots, m$. It is a simple exercise to verify that $\mu_{ij}^{(t)} = \mathbf{c}_{ij}\gamma_j^{(t)} + \beta_j^{(t)}x_i^{(t)} = \tilde{\mu}_{ij}^{(t)} = \mathbf{c}_{ij}\tilde{\gamma}_j^{(t)} + \tilde{\beta}_j^{(t)}\tilde{x}_i^{(t)}$; i.e., we generate the same likelihood contributions as provided by the original model, but from a set of identified parameters. Thus, over many iterations, the transformed values then constitute samples from the joint posterior density defined on the space of *identified* parameters. Figure 1 presents trace plots for selected parameters (and pairs of parameters) in the space of unidentified parameters. In the upper left of Figure 1 is the trace of the mean and standard deviation of latent district preferences in the 1990s data: since latent district partisanship has no "natural" scale, neither of these quantities is identified. Figure 2 presents traces in the space of identified parameters, by imposing the mean zero, variance one restriction. It is apparent that the Gibbs sampler is much better behaved in the space of identified parameters. All inferences, graphs and tables in the body of the paper are summaries of the posterior density of the identified parameters.

5 Validity of the Latent Trait: Member Ideal Points

This section considers some additional validity measures for our latent district partisanship measure. In particular, we display the relationship between latent district partisanship and member ideal points in each decade from 1950-1980 in figure 3, the analogous results for the 1990s are included in the paper.

Figure 3 produces similar graphs from the other four decades we examine. These graphs makes clear the way that roll call voting in the House of Representatives has become increasingly polarized along partisan lines since the 1950s. While this is a compelling and graphically vivid feature of the roll call data, our interest lies in the validating our district partisanship measure. We note that in all decades, there is moderate to very strong relationship between district partisanship and the recovered ideal point of the district's representative. An important exception to this general pattern comes in the 87th House (1961-62), which we use to validate the estimates of district partisanship

[2]This provides local identification (and not global identification) in that it does not rule out a 180 degree rotation of the latent preferences (i.e., setting $g = -1$ in the above discussion). The existence of this "mirror image" set of parameter estimates is of no great practical consequence, since we simply choose the solution that has the latent trait running from Republican to Democratic, and start the Gibbs sampler in the neighborhood of this posterior mode, so that the mirror image mode is never visited by the algorithm. An alternative solution is to also impose the constraint that $\beta_j > 0$.

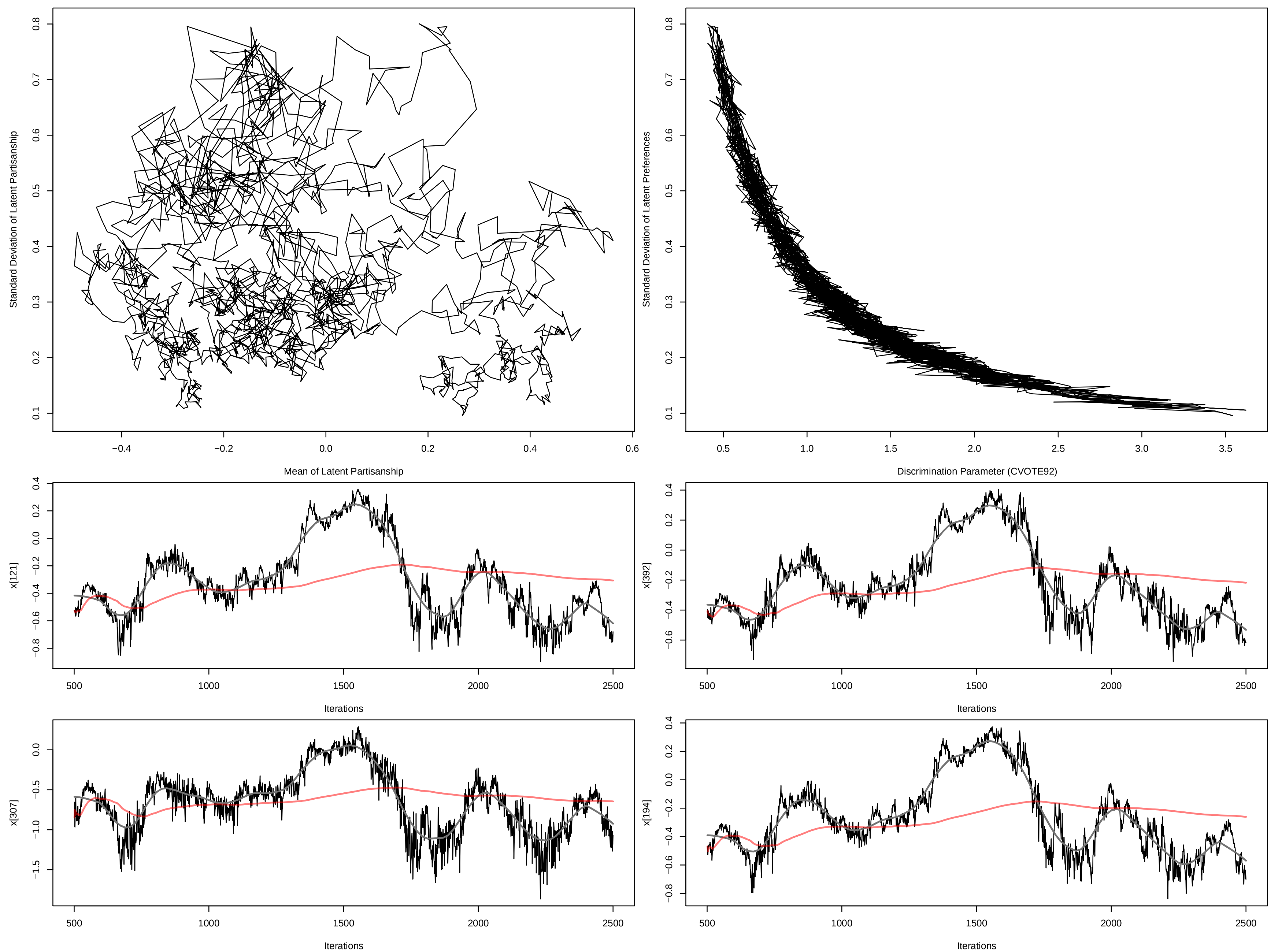

Figure 1: Trace Plots of the Gibbs Sampler in the Space of Unidentified Parameters. For the one dimensional traces, the gray line is a moving average, while the red line is a cumulative average.

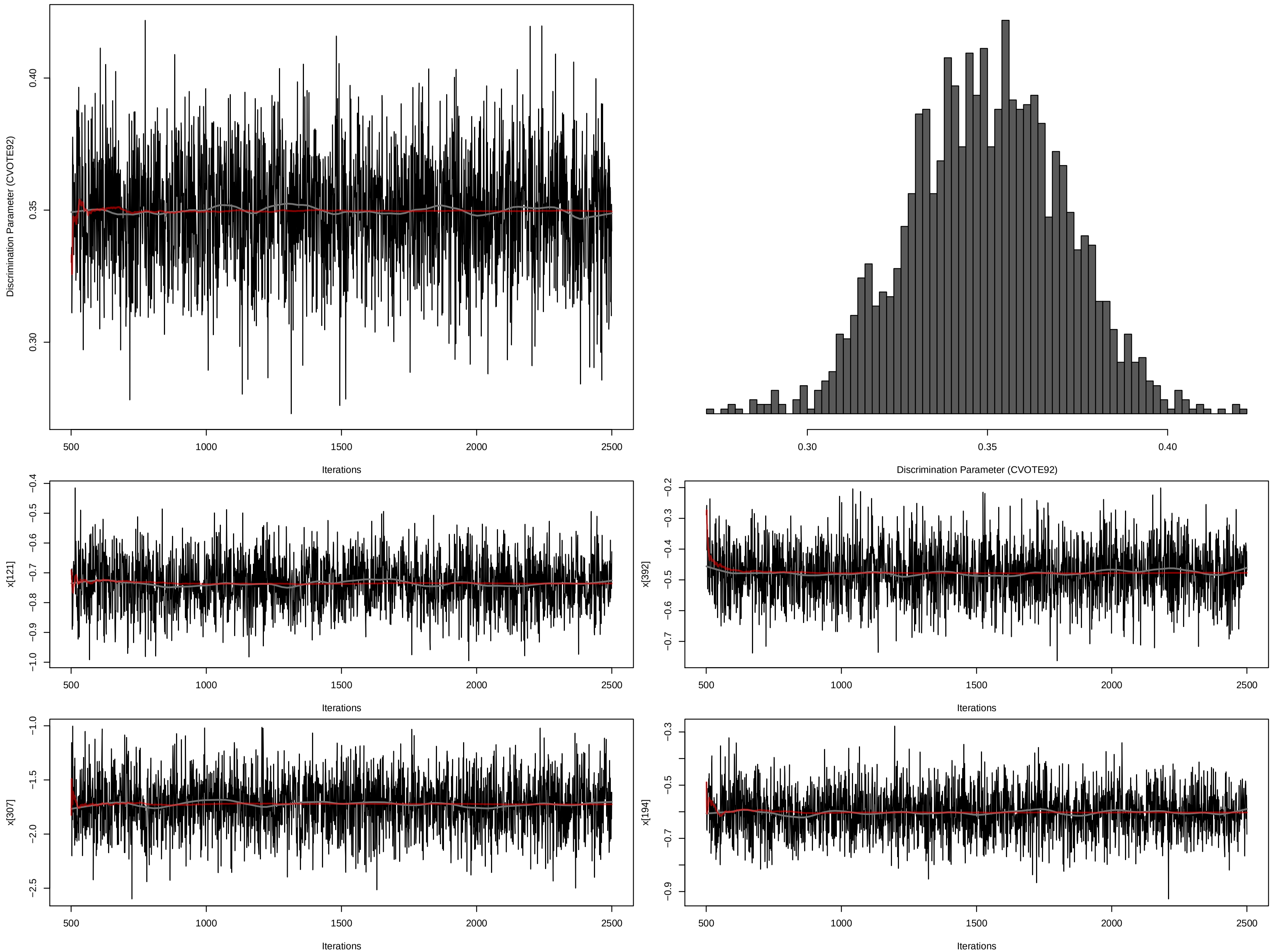

Figure 2: Trace Plots of the Gibbs Sampler in the Space of Identified Parameters and Posterior Summary of γ_{12}.

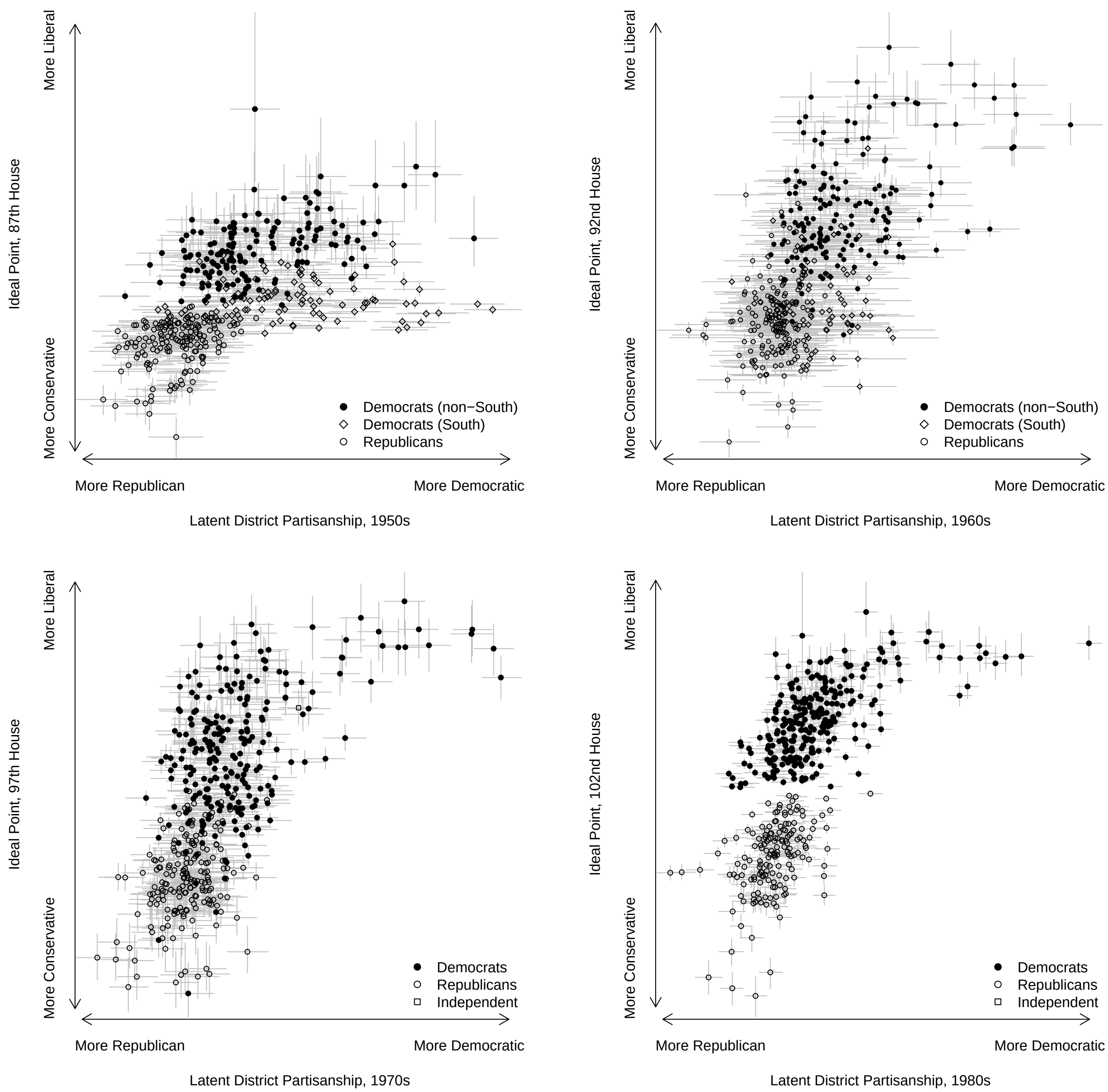

Figure 3: Legislative Preferences (legislators' ideal points) and latent district partisanship, for four decades (1950s, 1960s, 1970s and 1980s) and four Congress immediately following each respective decade: the 87th House (1961-62); the 92nd (1971-72), the 97th (1981-92) and the 102nd (1981-82). Vertical and horizontal lines indicate 95% credible intervals for legislative preferences and district partisanship, respectively.

for the 1950s. In this particular comparison (top left, Figure 3) Democrats representing districts in the South record voting histories that bear no relationship to the levels of partisanship we estimate in their districts, a finding that will come as no surprise to scholars of American politics. Put simply, high levels of Democratic partisanship did not translate into reliably "liberal" voting histories for Southern Democrats; conversely, the ideological and policy

6 Estimates of District Level Partisanship

Figures 4 through 8 show scatterplot of the recovered latent trait and its indicators (presidential and congressional vote shares) for the 1990s. The relationship between the vote shares and the latent trait is as expected. The non-linearities follow from the fact we use log-odds transformations of the vote shares. Outliers are generally more prevalent in the congressional elections scatterplots, resulting from the fact that indicators for incumbency, challenger quality and region (south/non-south) also appear in the measurement model for those elections

Out of Step Members

In the text of the paper, we had a brief discussion of the mebers from the 107th Congress who were "out of step" with their district. Remember that we define "out of step" members as follows. Suppose we break the distribution of district partisanship at its mean value of zero, labelling districts to the left of this point "Republican" districts and districts to the right as "Democratic". If a Republican member represents a "Democratic" district or vice-versa, we call this member out of step. Here, we provide a listing of the out of step members, and give their fates through the 2006 election.

Extension: A Model of District Preferences

In the paper, we built a model of district partisanship using demographic indicators and vote shares. However, for some purposes, this may not be enough: scholars might want to explicitly examine district *preferences* to, say, study representation (Bartels 2002). Our model can accommodate such a request by incorporating additional indicators measuring ideology. In this section, we use survey data aggregated to the district level to provide a measure of ideology.

The type of survey data typically available to most political scientists would be inadequate for this task. As we discussed in the paper, using survey data to study district opinion typically suffers from the "Miller-Stokes" problem of very limited within-district sample sizes. However, during the 2000 election, two polling firms did extensive polling over the course of the campaign, enough to give us reasonable survey sizes per district and coverage over a large number of districts, thereby circumventing the standard problems of small within-district samples and poor cross-district coverage. Knowledge Networks conducted political polling during the course of the campaign, and the Annenberg National Election Study conducted extensive interviews, both in the form of panels around major campaign events (i.e., the New Hampshire and Iowa primaries, Super Tuesday, the conventions, etc.), as well as a rolling cross-section during the course of the campaign. While both surveys asked a variety of political items (and they asked different items to different respondents), they both

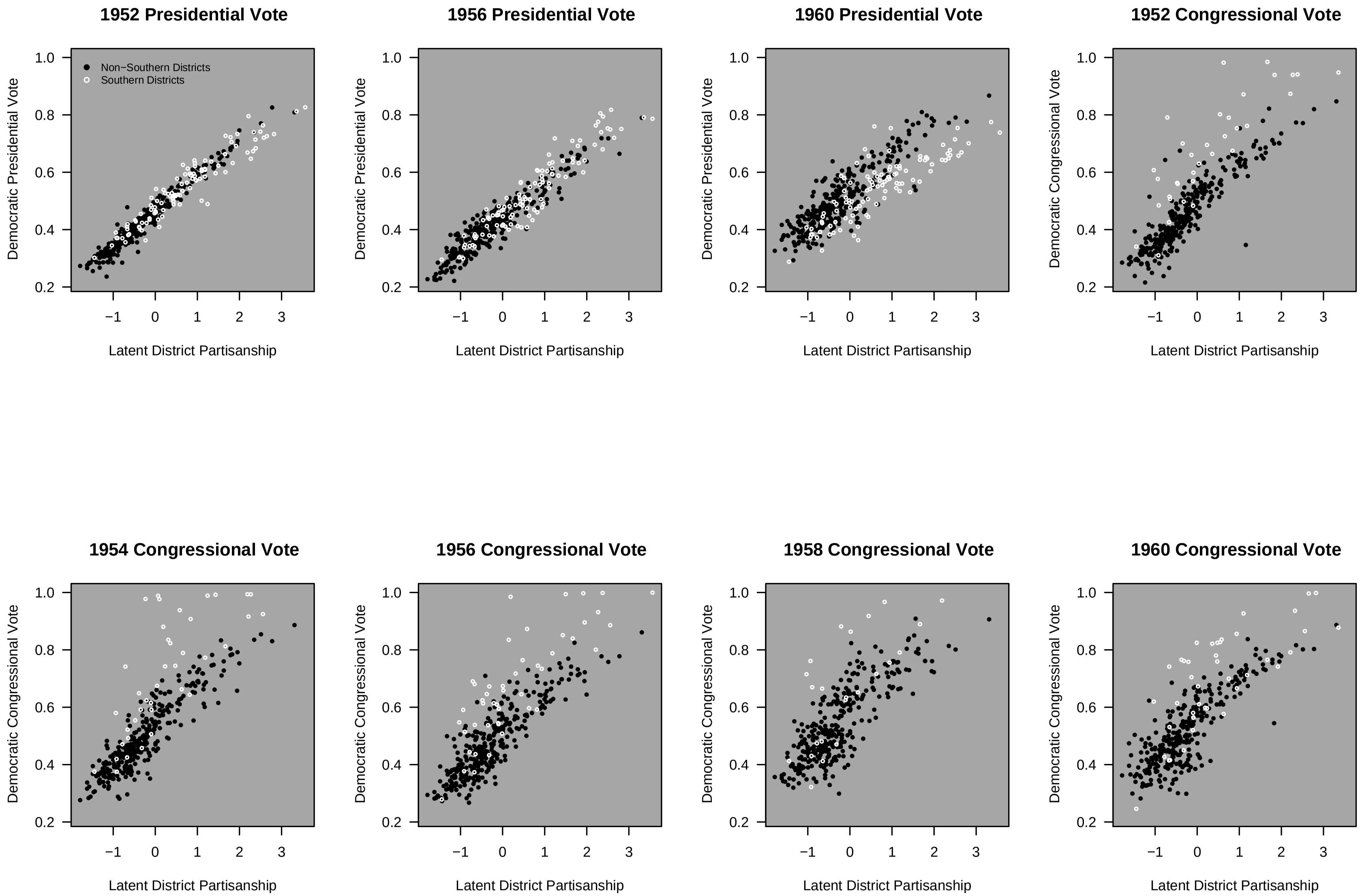

Figure 4: Vote Shares plotted against Latent District Partisanship, 1950s. Presidential election outcomes are modeled as a function of the latent trait plus intercept shifts for home-state effects. Similarly, the model for congressional election outcomes includes intercept shifts for incumbency, region and challenger quality.

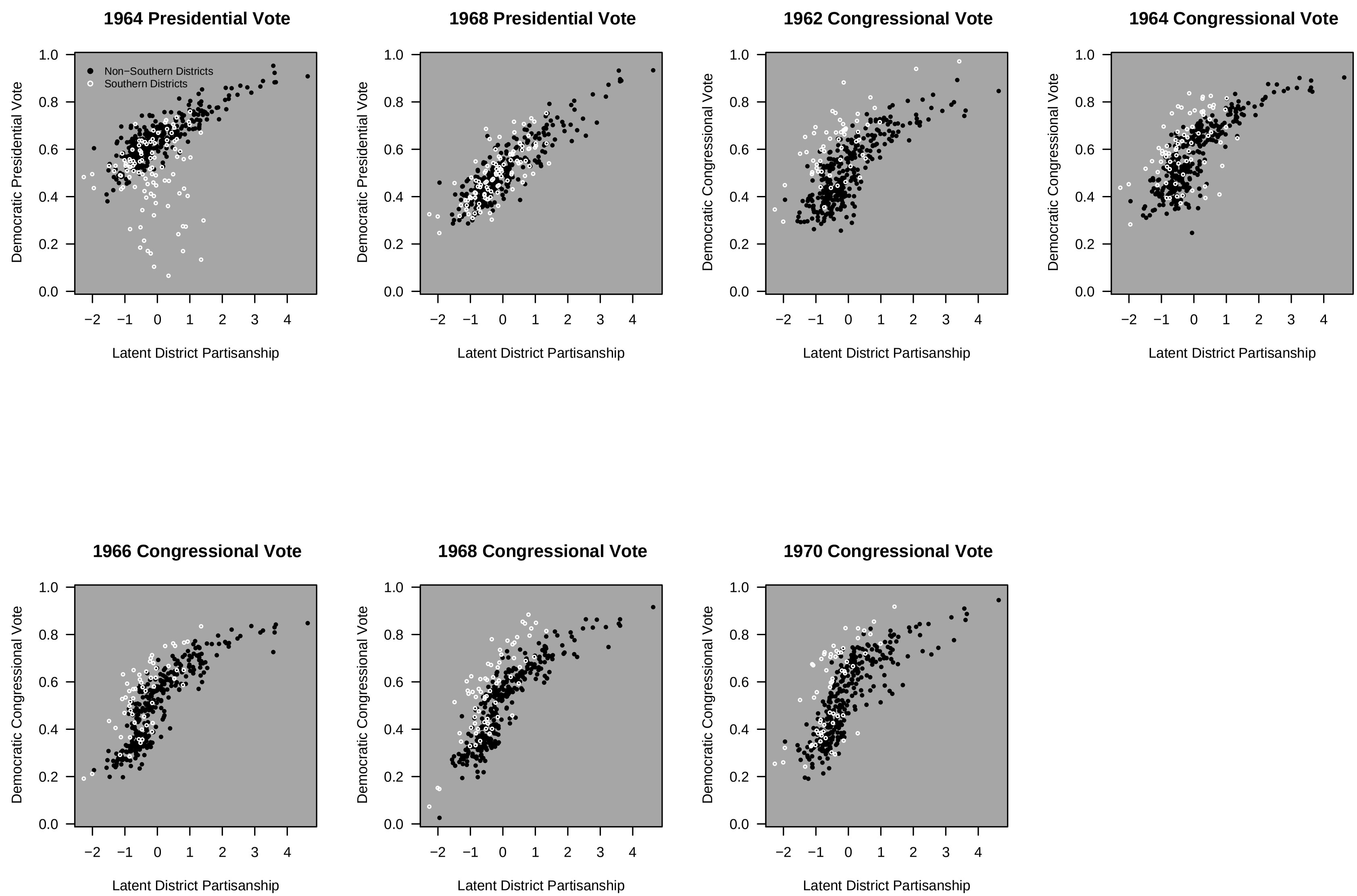

Figure 5: Vote Shares plotted against Latent District Partisanship, 1960s. Presidential election outcomes are modeled as a function of the latent trait plus intercept shifts for home-state effects. Similarly, the model for congressional election outcomes includes intercept shifts for incumbency, region and challenger quality.

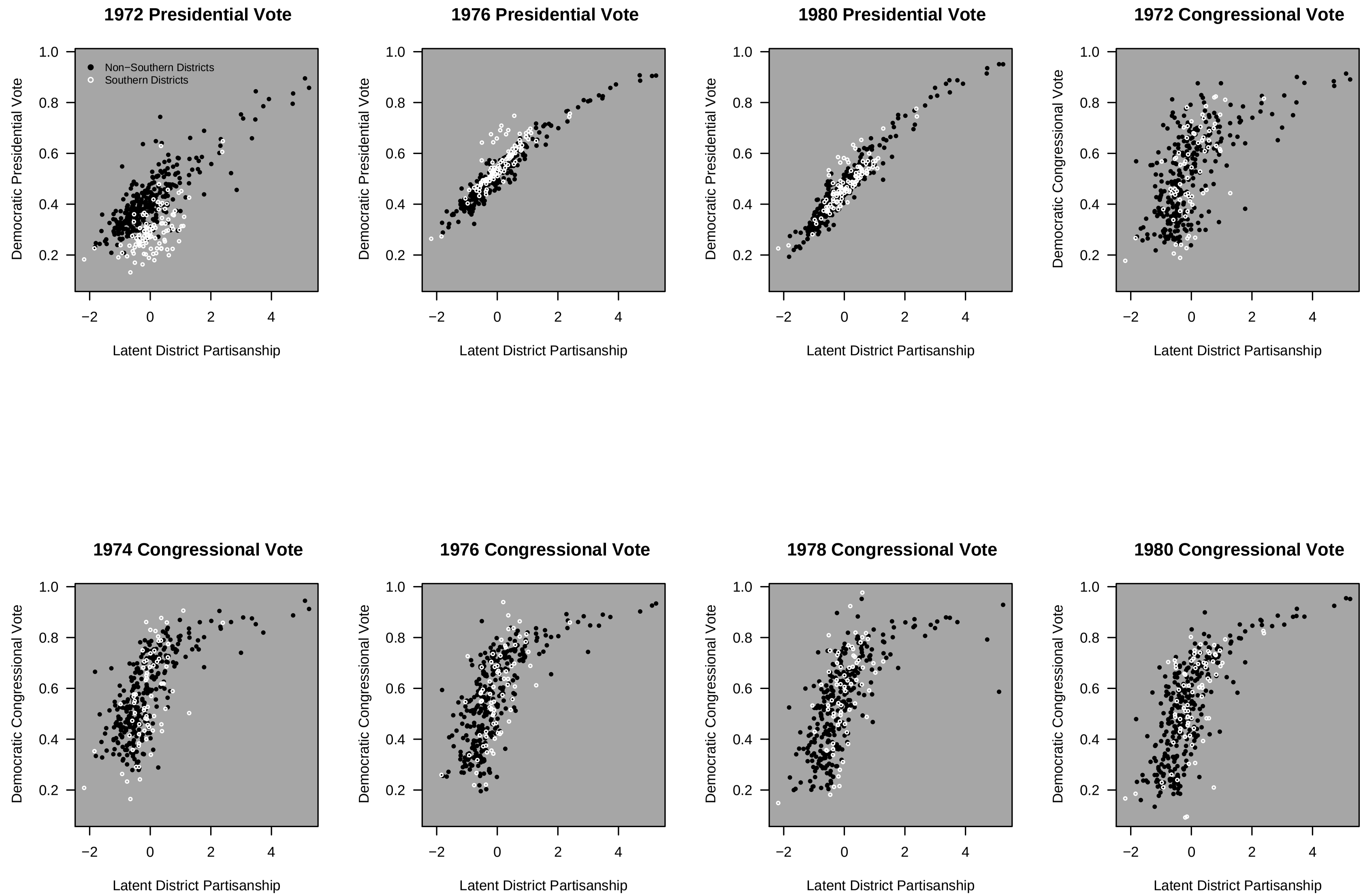

Figure 6: Vote Shares plotted against Latent District Partisanship, 1970s. Presidential election outcomes are modeled as a function of the latent trait plus intercept shifts for home-state effects. Similarly, the model for congressional election outcomes includes intercept shifts for incumbency, region and challenger quality.

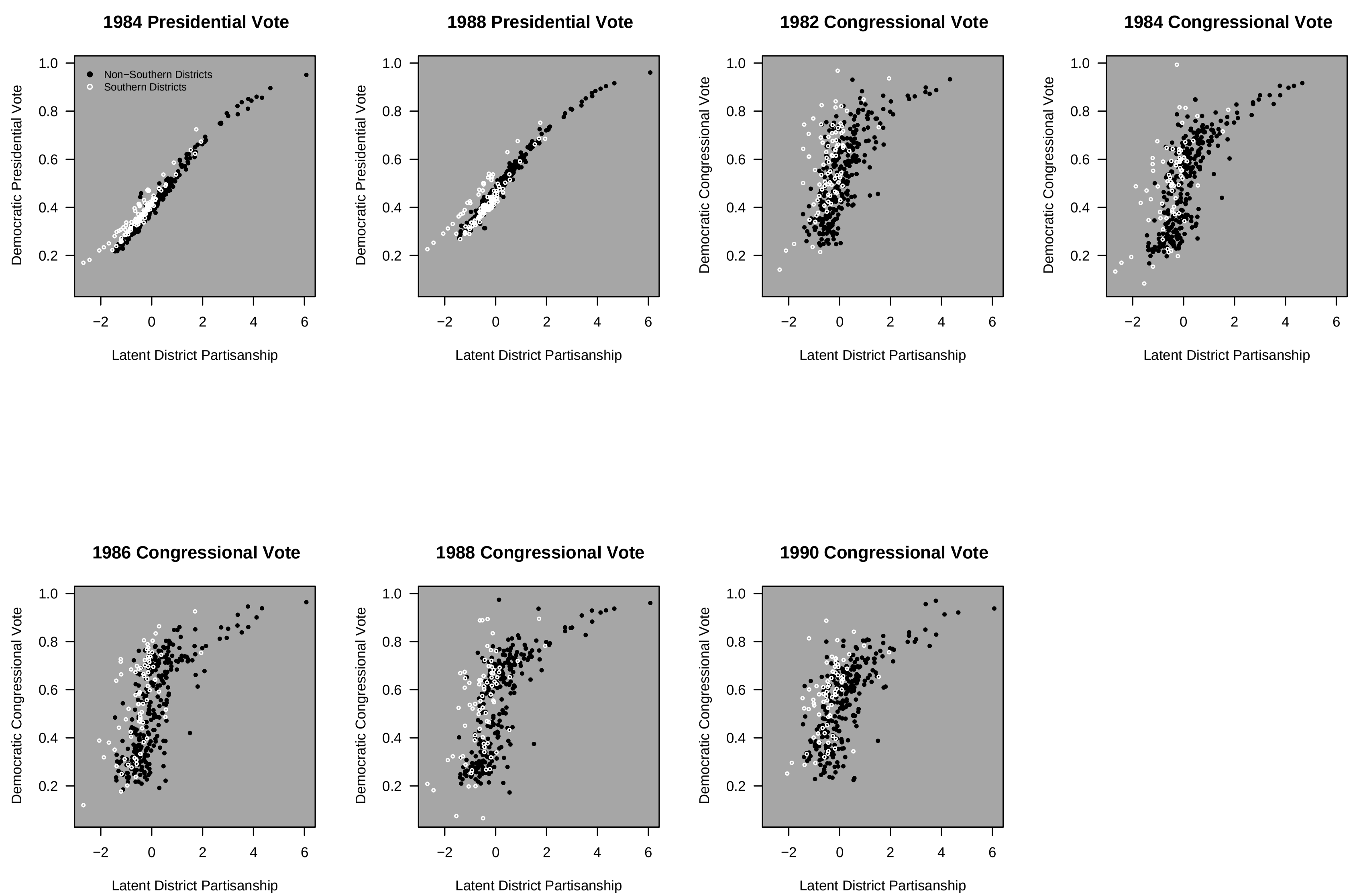

Figure 7: Vote Shares plotted against Latent District Partisanship, 1980s. Presidential election outcomes are modeled as a function of the latent trait plus intercept shifts for home-state effects. Similarly, the model for congressional election outcomes includes intercept shifts for incumbency, region and challenger quality.

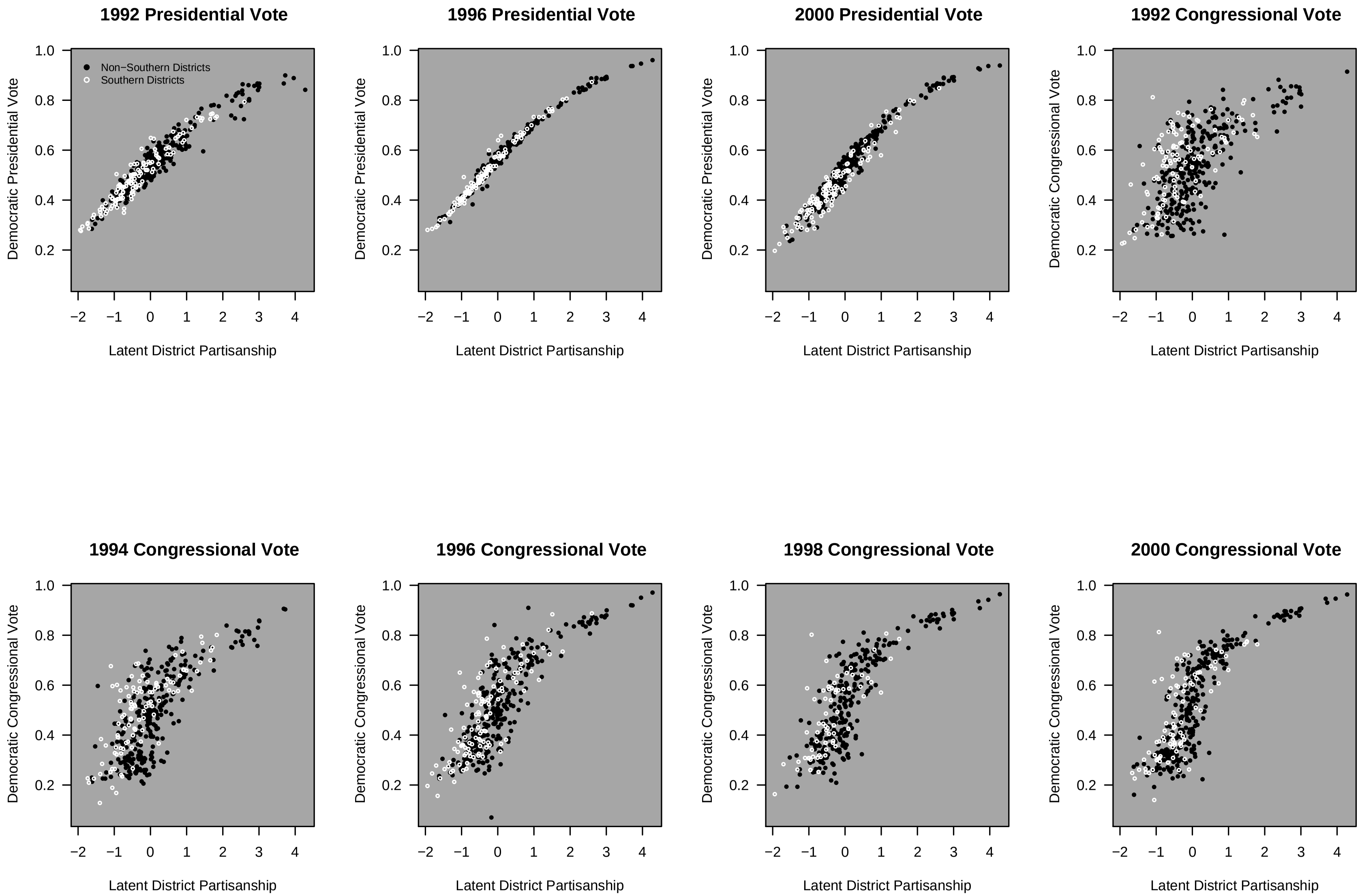

Figure 8: Vote Shares plotted against Latent District Partisanship, 1990s. Presidential election outcomes are modeled as a function of the latent trait plus intercept shifts for home-state effects. Similarly, the model for congressional election outcomes includes intercept shifts for incumbency, region and challenger quality.

Rank	Latent Partisanship	District
Most Democratic:		
1 [1, 3]	3.6 [3.2, 4]	GA:4
2 [1, 4]	3.4 [3, 3.7]	NC:2
3 [1, 4]	3.3 [3, 3.6]	MI:1
4 [3, 10]	2.8 [2.4, 3.2]	GA:8
4 [3, 13]	2.8 [2.4, 3.1]	NY:16
6 [4, 15]	2.7 [2.2, 3]	GA:3
7 [4, 18]	2.6 [2.2, 3]	MS:2
9 [4, 17]	2.6 [2.2, 2.9]	NC:1
8 [5, 17]	2.5 [2.2, 2.9]	NY:23
7 [4, 19]	2.5 [2.1, 2.9]	GA:2
Most Republican:		
1 [1, 12]	-1.8 [-2.2, -1.5]	IL:14
2 [1, 26]	-1.6 [-2, -1.3]	NY:36
2 [1, 25]	-1.6 [-1.9, -1.3]	NY:31
2 [1, 28]	-1.6 [-1.9, -1.3]	NY:37
3 [1, 31]	-1.6 [-1.9, -1.2]	NY:29
3 [1, 35]	-1.5 [-1.9, -1.2]	CA:20
5 [1, 38]	-1.5 [-1.8, -1.1]	NY:1
6 [1, 39]	-1.5 [-1.8, -1.1]	ME:3
7 [2, 42]	-1.5 [-1.8, -1.1]	IL:13
5 [1, 41]	-1.4 [-1.8, -1.1]	TN:1

Table 2: Twenty Most Extreme Districts, 1950s. "Rank" is the mode of the posterior density of the order statistic for latent district partisanship; "district partisanship" is the mean of the posterior density for latent district partisanship. Ninety-five percent confidence intervals (highest posterior density intervals) in brackets.

Rank	Latent Partisanship	District
Most Democratic:		
1 [1, 2]	4.6 [4.1, 5.2]	CA:21 1962-1970
3 [2, 9]	3.6 [3, 4.2]	MI:1 1964-1970
3 [2, 9]	3.6 [3, 4.2]	NY:18 1962-1970
4 [2, 9]	3.6 [3.1, 4.2]	MI:13 1964-1970
4 [2, 10]	3.6 [3, 4.1]	IL:1 1962-1970
2 [1, 18]	3.4 [2.3, 4.6]	GA:2 1962-1962
2 [2, 16]	3.4 [2.3, 4.4]	MI:1 1962-1962
7 [3, 12]	3.3 [2.6, 3.8]	OH:21 1962-1970
8 [3, 13]	3.2 [2.6, 3.7]	IL:7 1962-1970
10 [5, 17]	2.9 [2.3, 3.5]	NY:21 1962-1970
Most Republican:		
1 [1, 5]	-2.3 [-2.8, -1.7]	VA:6 1962-1970
3 [1, 16]	-2 [-2.6, -1.5]	TN:2 1962-1970
2 [1, 21]	-2 [-2.6, -1.3]	TN:1 1962-1970
2 [1, 17]	-2 [-2.5, -1.4]	CA:12 1962-1970
6 [2, 50]	-1.6 [-2.1, -1]	CA:24 1962-1970
8 [1, 44]	-1.5 [-2.1, -1]	CA:35 1962-1970
6 [1, 47]	-1.5 [-2.1, -1]	PA:16 1962-1970
7 [1, 51]	-1.5 [-2, -0.9]	NC:4 1962-1970
12 [2, 52]	-1.5 [-2, -0.9]	CA:32 1962-1970
1 [1, 162]	-1.4 [-2.6, -0.5]	CO:2 1962-1962

Table 3: Twenty Most Extreme Districts, 1960s. “Rank” is the mode of the posterior density of the order statistic for latent district partisanship; “district partisanship” is the mean of the posterior density for latent district partisanship. Ninety-five percent confidence intervals (highest posterior density intervals) in brackets.

Rank	Latent Partisanship	District
Most Democratic:		
1	5.2	MI:1
[1, 3]	[4.9, 5.6]	
2	5.1	IL:1
[1, 3]	[4.8, 5.5]	
3	4.7	MI:13
[2, 4]	[4.4, 5.1]	
4	4.7	NY:21
[2, 4]	[4.3, 5.1]	
5	3.9	NY:19
[5, 7]	[3.6, 4.3]	
6	3.7	OH:21
[5, 9]	[3.4, 4.1]	
7	3.5	NY:12
[6, 10]	[3.1, 3.8]	
8	3.5	MD:7
[5, 10]	[3.1, 3.8]	
9	3.4	IL:2
[6, 11]	[3, 3.7]	
10	3.1	IL:7
[8, 13]	[2.7, 3.4]	
Most Republican:		
1	-2.2	TX:7
[1, 4]	[-2.6, -1.8]	
2	-1.8	TX:3
[1, 10]	[-2.2, -1.5]	
2	-1.8	UT:1
[1, 10]	[-2.2, -1.4]	
3	-1.8	IL:14
[1, 11]	[-2.2, -1.4]	
4	-1.7	NE:3
[2, 15]	[-2, -1.3]	
5	-1.6	IL:12
[2, 18]	[-1.9, -1.2]	
6	-1.6	CA:40
[2, 20]	[-1.9, -1.2]	
8	-1.5	CA:39
[3, 24]	[-1.8, -1.1]	
7	-1.5	ID:2
[3, 30]	[-1.8, -1.1]	
8	-1.4	PA:16
[4, 30]	[-1.8, -1.1]	

Table 4: Twenty Most Extreme Districts, 1970s. “Rank” is the mode of the posterior density of the order statistic for latent district partisanship; “district partisanship” is the mean of the posterior density for latent district partisanship. Ninety-five percent confidence intervals (highest posterior density intervals) in brackets.

Rank	Latent Partisanship	District
Most Democratic:		
1 [1, 1]	6.1 [5.8, 6.3]	IL:1 1982-1990
2 [2, 3]	4.7 [4.4, 4.9]	PA:2 1982-1990
3 [2, 4]	4.3 [4.1, 4.6]	NY:12 1982-1990
4 [3, 5]	4.1 [3.9, 4.4]	MI:1 1982-1990
5 [4, 7]	3.9 [3.7, 4.1]	NY:16 1982-1990
6 [5, 8]	3.8 [3.6, 4]	MI:13 1982-1990
7 [5, 8]	3.8 [3.5, 4]	NY:18 1982-1990
8 [6, 10]	3.5 [3.3, 3.8]	IL:2 1982-1990
10 [8, 10]	3.4 [3.2, 3.7]	NY:11 1982-1990
10 [8, 10]	3.4 [3.1, 3.6]	MD:7 1982-1990
Most Republican:		
1 [1, 3]	-2.7 [-3, -2.4]	TX:7 1984-1990
3 [1, 5]	-2.4 [-2.7, -2.2]	TX:3 1984-1990
1 [1, 12]	-2.4 [-3.4, -1.5]	TX:7 1982-1982
4 [1, 28]	-2.1 [-3.1, -1.1]	TX:3 1982-1982
5 [3, 9]	-2.1 [-2.3, -1.8]	TX:21 1984-1990
7 [4, 13]	-1.9 [-2.1, -1.6]	TX:26 1984-1990
6 [1, 58]	-1.8 [-2.8, -0.9]	TX:21 1982-1982
9 [6, 18]	-1.7 [-2, -1.4]	TX:19 1984-1990
13 [7, 25]	-1.5 [-1.8, -1.3]	LA:1 1984-1990
12 [9, 37]	-1.5 [-1.7, -1.2]	TX:13 1984-1990

Table 5: Twenty Most Extreme Districts, 1980s. “Rank” is the mode of the posterior density of the order statistic for latent district partisanship; “district partisanship” is the mean of the posterior density for latent district partisanship. Ninety-five percent confidence intervals (highest posterior density intervals) in brackets.

Rank	Latent Partisanship	District
Most Democratic:		
1 [1, 1]	4.3 [4.1, 4.4]	NY: 1992-2000
2 [2, 3]	4 [3.8, 4.1]	NY: 1992-2000
3 [3, 4]	3.7 [3.6, 3.9]	NY: 1992-2000
4 [3, 4]	3.7 [3.5, 3.9]	NY: 1992-2000
5 [5, 10]	3 [2.8, 3.2]	MI: 1992-2000
5 [5, 10]	3 [2.8, 3.2]	PA: 1992-2000
7 [5, 11]	3 [2.8, 3.1]	MI: 1992-2000
8 [5, 10]	3 [2.8, 3.1]	IL: 1992-2000
8 [5, 11]	3 [2.8, 3.1]	IL: 1992-2000
10 [6, 13]	2.9 [2.7, 3]	CA: 1992-2000
Most Republican:		
2 [1, 6]	-1.9 [-2.1, -1.8]	TX: 1992-2000
1 [1, 16]	-1.9 [-2.4, -1.5]	TX: 1992-1994
1 [1, 17]	-1.9 [-2.4, -1.4]	TX: 1992-1994
4 [1, 9]	-1.8 [-2, -1.6]	TX: 1996-2000
1 [1, 25]	-1.7 [-2.2, -1.3]	TX: 1992-1994
2 [1, 23]	-1.7 [-2.2, -1.3]	TX: 1992-1994
6 [3, 13]	-1.7 [-1.9, -1.5]	AL: 1992-2000
8 [4, 15]	-1.7 [-1.8, -1.5]	TX: 1996-2000
10 [4, 17]	-1.6 [-1.8, -1.4]	IN: 1992-2000
10 [4, 17]	-1.6 [-1.8, -1.4]	NE: 1992-2000

Table 6: Twenty Most Extreme Districts, 1990s. "Rank" is the mode of the posterior density of the order statistic for latent district partisanship; "district partisanship" is the mean of the posterior density for latent district partisanship. Ninety-five percent confidence intervals (highest posterior density intervals) in brackets.

Member	Fate through the 2006 Election
Ted Strickland (D, OH-6)	Retired in 2006, elected Governor of Ohio
Tim Holden (D, PA-6)	Still in office
Bud Cramer (D, AL-5)	Still in office
Ronnie Shows (D, MS-4)	Defeated in 2002 election
David Price (D, NC-4)	Still in office
Max Sandlin (D, TX-1)	Defeated in 2004 election
Nick Lampson (D, TX-9)	Defeated in 2004 election, although won the 2006 election
Ken Lucas (D, KY-4)	Retired in 2004 (party control flipped)
Gary Condit (D, CA-18)	Defeated in the 2002 primary
Tim Roemer (D, IN-3)	Retired in 2002 (party control flipped)
Dennis Moore (D, KS-3)	Still in office
Earl Pomeroy (D, ND-AL)	Still in office
Vic Snyder (D, AR-2)	Still in office
Gene Taylor (D, MS-5)	Still in office
Jim Turner (D, TX-2)	Retired in 2004
Chet Edwards (D, TX-11)	Still in office
Bart Gordon (D, TN-6)	Still in office
Lois Capps (D, CA-22)	Still in office
Baron Hill (D, IN-9)	Defeated in 2004 election, although won the 2006 election
Collin Peterson (D, MN-7)	Still in office
Sanford Bishop (D, GA-2)	Still in office
Rick Boucher (D, VA-9)	Still in office
Bob Etheridge (D, NC-2)	Still in office
John Spratt (D, SC-5)	Still in office
Ralph Hall (D, TX-4)	Switched to the Republican Party in 2004
Charlie Stenholm (D, TX-17)	Defeated in the 2004 election
Jim Matheson (D, UT-2)	Still in office
Rob Simmons (R, CT-2)	Defeated in 2006 election
Jack Quinn (R, NY-30)	Retired in 2004 (party control flipped)
Jim Leach (R, IA-1)	Defeated in the 2006 election
Clay Shaw (R, FL-22)	Defeated in 2006 election
Connie Morella (R, MD-8)	Defeated in the 2002 election
Steve Horn (R, CA-38)	Retired in 2002 (party control flipped)

Table 7: Members “out of step” with their district based on our measure. For each member, we indicate their fate through the 2006 election.

asked a general liberal-conservative self-placement item to a large number of respondents per district. Clinton (2006) takes the liberal-conservative self-identification measure from each survey and merges these data and aggregates to the level of Congressional district, providing the mean liberal-conservative position, the standard deviation of those responses and the numbers of respondents per congressional district.

The Knowledge Networks item asks respondents

> There has been a lot of talk these days about liberals and conservatives. Would you say that you are: very liberal, liberal, moderate, conservative, very conservative, don't know.

and the responses are coded as $\{-2, -1, 0, 1, 2, \text{NA}\}$. The NAES asks

> Generally speaking, would you describe your political views as: very conservative, conservative, moderate, liberal, or very liberal

and an analogous coding scheme is used here as well. Clinton (2006) merges these data and aggregates to the level of Congressional district, providing (a) the mean liberal-conservative position, (b) the standard deviation of those responses and (c) the numbers of respondents per congressional district. We take the data as reported by Clinton: we have no information as to the representativeness of the data. On average, we have 232 respondents per district (with 50% of the districts having between 178 and 254 respondents per district), and coverage over 432 districts. The smallest number of responses comes from Alaska, with 41 respondents, and the largest number of respondents comes from New Hampshire's 2^{nd} Congressional district, with 2,099 respondents.[3] The liberal conservative scores range from -0.62 (California's 9^{th} Congressional district comprised of Oakland and Berkeley) to 0.53 (Louisiana's 5^{th} Congressional district, located in Northeast Louisiana), with Texas' 19^{th} Congressional district (located in the hills near Midland) the second-most conservative district in the nation.

In order to incorporate the survey data into our model, we proceed as follows. For the 1992-2000 data, we let w_i be the mean liberal-conservative position in the district, and $\sigma^2_{w_i}$ be the sampling variance of w_i, equal to the variance of the w_i divided by the sample size in district i, n_i. The mean survey response is modeled as a function of district partisanship in 2000, i.e., $w_i \sim N(\eta - x_i, \sigma^2_{w_i})$, where η is an intercept parameter to be estimated. Note that the factor loading or discrimination on district partisanship is implicitly set to -1.0. In this way we are letting the information in the survey data contribute heavily to the substantive content of the recovered preference measures, giving our measure more construct validity as a measure of "district ideology" than relying solely on district level vote shares.[4] For that reason, we refer to this quantity of interest as "district preferences" because it more directly taps underlying ideology via the survey data. Otherwise, the model is identical to the model for district partisanship that we presented earlier.

[3] The large number of respondents in New Hampshire is due to the fact that Annenberg conducted a study of the New Hampshire primary, so both districts in New Hampshire have over 2000 respondents in our sample.

[4] This type of restriction is common in confirmatory factor analysis, where setting the loading of indicator j on factor k makes indicator j a "reference item" for factor k, and, for our one dimensional latent variable model, also solves the scale indeterminacy problem (e.g., Bollen 1989, 238-247). With the restriction on the factor loading/discrimination parameter, the only identification issue is invariance to translation (i.e., intercept-shifts). To solve this problem we employ the post-processing strategy outlined earlier in this appendix, although in this case we only need to adjust intercept parameters throughout the model.

Because the survey data is strongly correlated with vote outcomes (see above), our measure of district preferences is extremely similar to our measure of district partisanship. The relationship between the two sets of measures is show in Figure 9. Unsurprisingly, Republican-held seats (red dots) are predominantly more conservative/Republican on both measures, while again there appears to be a pronounced skew in the Democratic tail of the distribution of district partisanship.

As one would expect from figure 9, the two measures correlate extremely highly: the posterior means of the district preferences and district partisanship measures correlate at 0.99, suggesting that according to this model, aggregate district partisanship and preferences are extremely strongly related—even if the micro-level relationship is far weaker. We wish to stress, however, that the strong relationship between our two estimates does not make this additional measure a superfulous exercise. First, the survey data means that the substantive content of the measures differs. Additionally, the fact that the posterior means do not change very much largely stems from the fact that in the 1990s the survey-based ideology measure is very highly correlated with the vote-based measures. Additionally, adding another source of information boosts reliability, reducing ex-post uncertainty about each district's partisanship.

Moreover, the primary substantive point of this exercise was to show how to incorporate additional information into this type of model. As we have shown here, the researcher can easily tailor our model to suit his/her specific substantive needs by adding additional indicators of district partisanship (or even in this case, "district ideology."). The model's flexibility is one of its primary strengths.

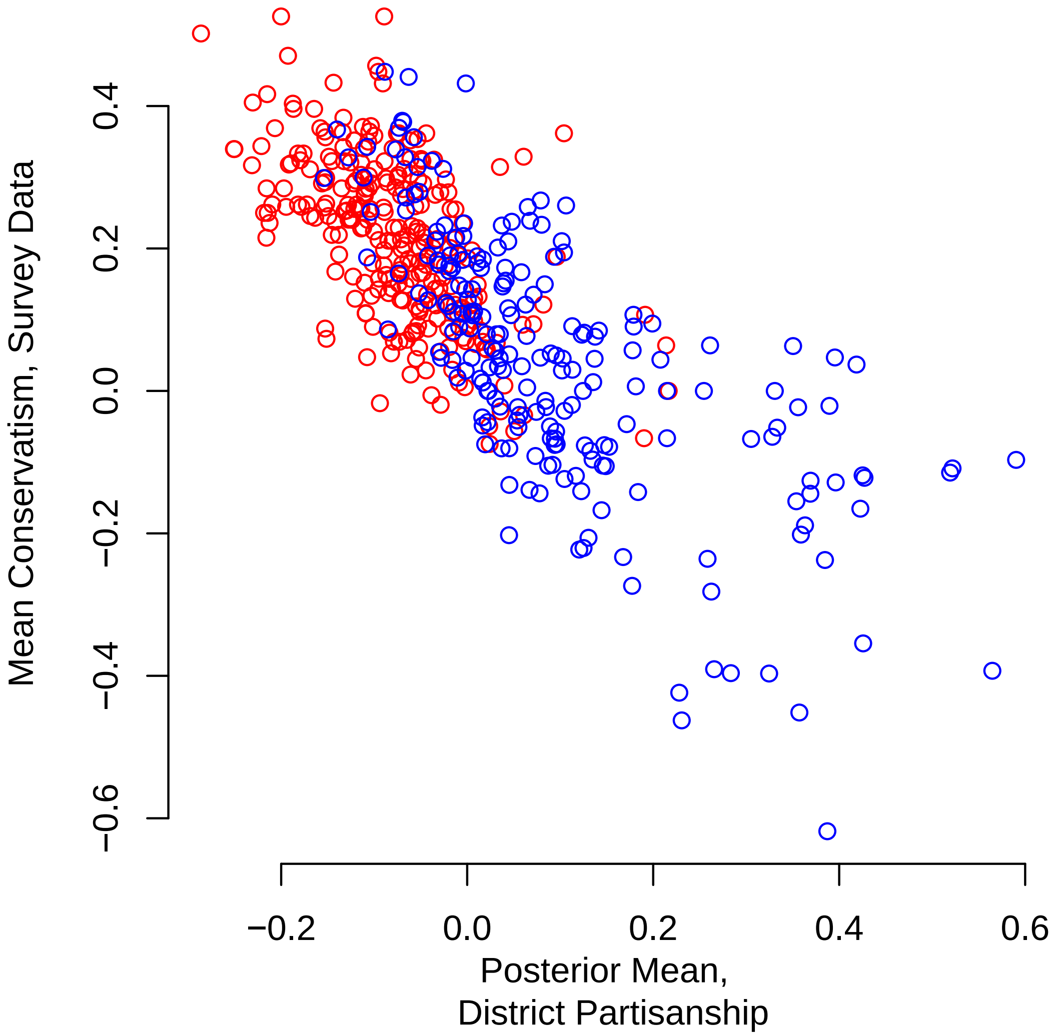

Figure 9: Relationship between Survey Estimates and Estimated District Partisanship (Posterior Means), 2000. Blue circles indicate seats won by Democratic candidates; red circles indicate seats won by Republicans.

References

Aldrich, John. 1995. *Why Parties? The Origin and Transformation of Political Parties in America*. Chicago: University of Chicago Press.

Ansolabehere, Stephen, James M. Snyder Jr. and Charles Stewart III. 2001. "Candidate Positioning in U.S. House Elections." *American Journal of Political Science* 45:17–34.

Bartels, Larry. 2002. "Economic Inequality and Political Representation." Presented at the Annual Meeting of the American Political Science Association, Boston.

Black, Duncan. 1987. *The Theory of Committees and Elections*. Boston: Kluwer Academic Publishers.

Bollen, Kenneth A. 1989. *Structural Equations With Latent Variables*. New York: Wiley.

Canes-Wrone, Brandice, John F. Cogan and David W. Brady. 2002. "Out of Step, Out of Office: Electoral Accountability and House Members' Voting." *American Political Science Review* 96.

Clinton, Joshua. 2006. "Representation in Congress: Constituents and Roll Calls in the 106th House." *Journal of Politics* 68:397–409.

Converse, Philip E. 1966. "The Concept of a Normal Vote." In *Elections and the Political Order*, ed. Angus Campbell, Philip E. Converse, Warren E. Miller and Donald E. Stokes. New York: John Wiley and Sons pp. 9–39.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper Row.

Jacobson, Gary. 2004. *The Politics of Congressional Elections*. Sixth ed. New York: Pearson Longman.

Joliffe, I.T. 2002. *Principal Components Analysis*. Second ed. New York: Springer.

Leamer, Edward. 1978. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York: Wiley.

Spiegelhalter, David J., Andrew Thomas and Nicky G. Best. 1996. "Computation on Bayesian Graphical Models." In *Bayesian Statistics 5*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith. Oxford University Press pp. 407–425.

Spiegelhalter, David J. and S. L. Lauritzen. 1990. "Sequential updating of conditional probabilities on directed graphical structures." *Networks* 20:579–605.