Bayesian Inference for Political Science Panel Data*

Andrew D. Martin
Washington University
admartin@artsci.wustl.edu

Kyle L. Saunders
Northern Illinois University
ksaun@niu.edu

August 25, 2002

[WORK IN PROGRESS – COMMENTS ARE APPRECIATED]

Abstract

To answer substantive questions regarding individual change, it is necessary to collect information from each subject multiple times. For decades, political scientists have collected a great deal of panel data, relating to mass behavior, comparative politics, and international relations. Unfortunately, the most commonly used method of analyzing panel data – linear models with individual fixed effects – oftentimes masks important quantities that can be estimated using alternative strategies. In this paper, we review the literature on the general linear panel model, and discuss Bayesian estimation strategies using Markov chain Monte Carlo methods. The model is extremely flexible, allowing for multiple fixed and random effects, and can be estimated using standard Gibbs sampling. To illustrate the utility of the approach, we model party identification from the 1992-1996 American National Election Study panel. In addition, we provide easy-to-use software to estimate the models as part of the `MCMCpack` package for the R language.

*Paper prepared for presentation at the 2002 Annual Meeting of the American Political Science Association. This research is supported by the National Science Foundation Law and Social Sciences and Methodology, Measurement, and Statistics Sections, Grant SES-0135855 to Washington University. R code to estimate the model is part of the `MCMCpack` package. A preliminary version of which can be downloaded from `http://adm.wustl.edu/tmp/`.

1 Introduction

There are many puzzles of politics and political behavior which cannot be answered with cross-sectional data analysis. To wit, for decades political scientists have devoted vast resources to collect panel data. In the field of political behavior, panel datasets exist in the United States, Canada, Russia, South Africa, and various West European democracies. Additionally, in fields such as international relations and international political economy, scholars often analyze data collected on a set of countries repeatedly measured over time.

While the importance of panel data is well-understood, in political science far less attention has been paid to statistical models of panel data. Indeed, the dominant empirical approach is to employ subject-specific fixed effects to account for between subject heterogeneity (Baltagi, 2001). This is the dominant econometric approach (Hsiao, 1986), and is the one most discussed in political science applications (see Wawro, 2002; Green and Yoon, 2002). The approach is conceptually simple; for each subject, include a dummy variable to allow it to have its own mean. In the case of time-series cross-sectional data (Beck and Katz, 1995), these fixed effects can be estimated because the number of subjects is typically small. In cross-sectionally dominated panel datasets (Stimson, 1985), it is impossible to estimate these effects because the number of subjects is large while the number of time periods is small. This is easily fixed by differencing the system (e.g., taking a two-strobe panel and modeling Δy_i instead of both $y_{1,i}$ and $y_{2,i}$), thus accounting for between subject differences. But this approach begs some questions: Why are there differences between subjects? And, more importantly, what explains those differences? If there are differences in the mean level for each subject, might there also be differences in the marginal effects across subjects?

In this paper, we utilize a general linear panel model (or mixed model) that allows for multiple subject-specific fixed and random effects (Laird and Ware, 1982; Chib and Carlin, 1999; Pinheiro and Bates, 2000). The model is easily extended to allow for random coefficients across subjects, with the inclusion of subject-specific covariates to model random coefficients across subjects (Western, 1998; Raudenbush and Bryk, 2002; Goldstein, 2002). This hierarchical model can be estimated using the same algorithm as the general linear panel model after some simple matrix algebra. We adopt a Bayesian inferential approach, which affords a number of advantages, including the ability to estimate any quantity of interest, the ability to analyze small datasets, and to comfortably interpret random effects.[1] We use Markov chain Monte Carlo methods to estimate the model, and provide easy-to-use software for the R language as part of the `MCMCpack` package. See Appendix A for software details.

The paper proceeds as follows. In the following section, we discuss the general linear panel model, detailing both our estimation strategy and the extension of the model to include random coefficients. Section 3 contains a discussion of our application – party identification in the 1992-1996 American National Election Study panel. We present results from two models in Section 4. The first model is a baseline model with a single random effect, and a second model is a hierarchical model that explores the interactive effects of race, gender, and income on party identification. In Section 5 we return to the model, and discuss how it can be extended in a number of ways, including modeling time series processes in the errors, dealing with dichotomous response variables, and handling panel attrition. The final section concludes.

2 The General Linear Panel Model

We begin our analysis with a general linear panel model (Laird and Ware, 1982; Chib and Carlin, 1999; Pinheiro and Bates, 2000). This model, presented below, is extremely flexible, and allows for all types of between-subject heterogeneity through the use of multiple, correlated random effects. Suppose we observe data from a total of n subjects, and for each subject we observe k responses.[2] Let p denote the number of covariates in the fixed effects part of the models. That is, the part

[1] In the frequentist tradition, where parameters are fixed and unknown (and, for that matter, unknowable) quantities, interpreting random effects requires a great deal of mental gymnastics.

[2] It is easy within this context to deal with unbalanced panels; i.e., k_i differs across subjects. In most political science cases, panels will be balanced, and the software requires the use of balanced panels.

of the model where the marginal effect of a covariate is assumed to be constant across subjects. Further, let q denote the number of covariates in the random effects part of the model; i.e., the part of the model where there exists heterogeneity across subjects, both in the error structure and in the mean level.

2.1 The Model

The general linear panel model takes the following form:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{W}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i \tag{1}$$

The response vector for each subject $\mathbf{y}_i$ is of order $(k \times 1)$. The observed covariates for the fixed effects are included in the matrix $\mathbf{X}_i$, which is a $(k \times p)$ matrix. Typically this matrix contains a constant (although, when extending the model to allow for random coefficients, there is no constant in this matrix). The observed covariates for the random effects are included in the matrix $\mathbf{W}_i$, which is of dimensionality $(k \times q)$. In the simplest case, $\mathbf{W}_i = \mathbf{1}$ which is a single random effect that does not covary with any explanatory variables. The $\mathbf{X}_i$ and $\mathbf{W}_i$ should share no covariates, although each can contain a constant (Both intercepts are identified because the random effects are assumed to have mean zero). As discussed below in Section 2.3, one can form the $\mathbf{X}_i$ and $\mathbf{W}_i$ matrices to allow for random coefficients; i.e., a $\boldsymbol{\beta}_i$ for each individual, that can then be modeled hierarchically.

We assume homoscedastic, independent errors:

$$\boldsymbol{\varepsilon}_i \sim \mathcal{N}_k(\mathbf{0}, \sigma^2\mathbf{I}_k)$$

And further assume that the random effects are distributed:

$$\mathbf{b}_i \sim \mathcal{N}_q(\mathbf{0}, \mathbf{D})$$

Notice that this setup allows for heteroscedasticity across individuals (so-called panel heteroscedasticity), since the total variance for each subject is a sum of the overall variance $\sigma^2\mathbf{I}_k$ and the between-subject variance $\mathbf{D}$. However, this model does not allow for serially correlated errors. We discuss extensions of our model that allow for autocorrelated errors in Section 5.

We are interested in summarizing the posterior distribution $f(\boldsymbol{\beta}, \sigma^2, \mathbf{D}|\mathbf{y})$, where $\mathbf{y} = \{\mathbf{y}_i\}_{i=1}^n$ denotes the observed data. To do so, it is necessary to posit prior beliefs about the parameters of interest. We assume *a priori* that each parameter is independent, which allows us to factor $f(\boldsymbol{\beta}, \sigma^2, \mathbf{D}) = f(\boldsymbol{\beta})f(\sigma^2)f(\mathbf{D})$. For each of these, we employ standard conjugate priors. For the $(p \times 1)$ vector of fixed effects parameters $\boldsymbol{\beta}$ we assume a Gaussian prior:

$$\boldsymbol{\beta} \sim \mathcal{N}_p(\mathbf{b}_0, \mathbf{B}_0^{-1})$$

For the scalar conditional error precision σ^{-2} we assume a Gamma prior:

$$\sigma^{-2} \sim \mathcal{G}(\nu_0/2, \delta_0/2)$$

Finally, for the order $(q \times q)$ precision matrix $\mathbf{D}^{-1}$ for the random effects, we assume a Wishart prior:

$$\mathbf{D}^{-1} \sim \mathcal{W}(\eta_0, \mathbf{R}_0)$$

The researcher can posit her prior beliefs about the parameter values by assigning numbers to each of the hyperparameters, each of which is subscripted by zero to denote that it is for a prior distribution.

2.2 Estimation via Markov chain Monte Carlo

The target of inference is the posterior distribution:

$$f(\boldsymbol{\beta}, \sigma^2, \mathbf{D}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \mathbf{D}) \times f(\boldsymbol{\beta})f(\sigma^2)f(\mathbf{D})$$

With the priors assumed above, this posterior distribution does not take a closed form. We thus must turn to Markov chain Monte Carlo (MCMC) methods to simulate from it (Tanner and Wong, 1987; Gelfand and Smith, 1990). See Jackman (2000) or Gill (2002) for an introduction to Bayesian inference and MCMC geared toward political scientists. With a large enough number of draws from the posterior, we can fully characterize the distribution, and report quantities of interest, such as posterior means and standard deviations, Bayesian credible intervals, and so forth.

A great deal of basic research has dealt with MCMC methods for the general linear panel model (see Chib and Carlin, 1999, for a review). One can apply Gibbs sampling to each block parameters, augmenting the sampler with draws for $\{\mathbf{b}_i\}$. This method, due to Gelfand and Smith (1990), is known to mix extremely slowly, thus requiring a huge number of draws from the posterior distribution with large thinning intervals to perform reliable inference. This is the sampler that would result from using BUGS for the analysis of this model (Spiegelhalter et al., 1997). Chib and Carlin (1999) demonstrate that the BUGS results are unreliable. To wit, Chib and Carlin (1999) propose three additional algorithms with varying blocking structures to estimate the model. We adopt the Algorithm 2 (p. 19), which requires standard Gibbs sampling. Chib and Carlin (1999) show that this algorithm mixes well, and thus provides good estimates of both variance parameters. We have implemented this algorithm in an R (Ihaka and Gentleman, 1996) package called `MCMCpack`.

The sampler we implement takes the following form (Chib and Carlin, 1999):

1. Simulate $\boldsymbol{\beta}$ and $\{\mathbf{b}_i\}$ from $f(\boldsymbol{\beta}, \{\mathbf{b}_i\} \mid \mathbf{y}, \sigma^2, \mathbf{D})$ by sampling:

$$\begin{array}{lcl} (i) & \boldsymbol{\beta} & \mid \mathbf{y}, \sigma^2, \mathbf{D} \\ (ii) & \{\mathbf{b}_i\} & \mid \mathbf{y}, \boldsymbol{\beta}, \sigma^2, \mathbf{D}. \end{array}$$

2. Simulate $\mathbf{D}^{-1}$ from $f(\mathbf{D} \mid \mathbf{y}, \boldsymbol{\beta}, \{\mathbf{b}_i\}\, \sigma^2)$.

3. Simulate σ^{-2} from $f(\sigma^2 \mid \mathbf{y}, \boldsymbol{\beta}, \{\mathbf{b}_i\}\, \mathbf{D})$.

What remains is to write down the full conditional distributions for each of these quantities.

The first conditional distribution $f(\boldsymbol{\beta}, \{\mathbf{b}_i\} \mid \mathbf{y}, \sigma^2, \mathbf{D})$ in step one is factored into the product of the unconditional distribution of $\boldsymbol{\beta}$ and the conditional distribution of the random effects $\{\mathbf{b}_i\}$. The first update is multivariate Normal (Lindley and Smith, 1972):

$$\boldsymbol{\beta} \mid \mathbf{y}, \sigma^2, \mathbf{D} \sim \mathcal{N}_p(\tilde{\mathbf{b}}, \tilde{\mathbf{B}})$$

To define these quantities, let $\mathbf{V}_i = \sigma^2 \mathbf{I}_k + \mathbf{W}_i \mathbf{D} \mathbf{W}_i'$, which is the total variance for observation i. With this quantity, we can write the variance covariance matrix for the draw as:

$$\tilde{\mathbf{B}} = \left[\mathbf{B}_0 + \sigma^{-2} \sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i \right]^{-1}$$

The mean of the draw takes the familiar form:

$$\tilde{\mathbf{b}} = \tilde{\mathbf{B}} \left[\mathbf{B}_0 \mathbf{b}_0 + \sigma^{-2} \sum_{i=1}^{n} \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{y}_i \right]$$

The conditional distribution of the random effects $\{\mathbf{b}_i\}$ is also multivariate Normal. We simulate the random effect for each subject separately, which implies that for each Gibbs iteration n random effects are drawn from the following distribution. Due to the large number of these random effects, our software does not store these quantities. The conditional distribution for each $\mathbf{b}_i$ is:

$$\mathbf{b}_i | \mathbf{y}, \boldsymbol{\beta}, \sigma^2, \mathbf{D} \sim \mathcal{N}_q(\tilde{\mathbf{b}}_i, \tilde{\mathbf{B}}_i)$$

The forms are similar to those above:

$$\tilde{\mathbf{B}}_i = \left[\mathbf{D}^{-1} + \sigma^{-2} \mathbf{W}_i' \mathbf{W}_i \right]^{-1} \quad \text{and} \quad \tilde{\mathbf{b}}_i = \sigma^{-2} \tilde{\mathbf{B}}_i \mathbf{W}_i' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})$$

There is no prior precision multiplied by the prior mean in the latter quantity because the mean of the random effects is forced to be zero.

The precision matrix for the random effects $\mathbf{D}^{-1}$ can be simulated from the Wishart distribution:

$$\mathbf{D}^{-1} | \mathbf{y}, \boldsymbol{\beta}, \{\mathbf{b}_i\} \, \sigma^2 \sim \mathcal{W}(\tilde{\eta}, \tilde{\mathbf{R}})$$

Where the shape parameter $\tilde{\eta} = \eta_0 + n$ and the scale matrix is:

$$\tilde{\mathbf{R}} = \left[\mathbf{R}_0^{-1} + \sum_{i=1}^{n} \mathbf{b}_i \mathbf{b}_i' \right]^{-1}.$$

The final update is for the conditional error precision σ^{-2}. With the conjugate prior discussed above, the full conditional distribution of this parameter is Gamma distributed:

$$\sigma^{-1} | \mathbf{y}, \boldsymbol{\beta}, \{\mathbf{b}_i\} \, \mathbf{D} \sim \mathcal{G}(\tilde{\nu}, \tilde{\delta})$$

The scale parameter is:

$$\tilde{\nu} = \frac{\nu_0 + nk}{2}.$$

The shape parameter is:

$$\tilde{\delta} = \frac{\delta_0 + \sum_{i=1}^{n} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{W}_i \mathbf{b}_i)'(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{W}_i \mathbf{b}_i)}{2}.$$

We iteratively sample from these distributions a large number of times, after preliminary burnin iterations, to summarize the posterior density. See Appendix A for a discussion of the particular software employed, as well as for implementation issues.

2.3 Random Coefficients Extension

If it is the case that subjects are heterogeneous (which is why one would want to study them in the first place!), and that heterogeneity extends beyond differences in mean levels explained solely by the coefficients in the $\mathbf{X}_i$ matrix, it would be desirable to *explain* that heterogeneity using a hierarchical, or multi-level, model (Goldstein, 2002; Raudenbush and Bryk, 2002). Such a model would allow for coefficients to vary across subjects, and then use subject-specific covariates to model the random coefficients. See Western (1998) for a political science application of hierarchical models. It turns out that we can use the *same estimation algorithm* for the general linear panel model to estimate a two-level hierarchical model after performing some matrix algebra.

Suppose we are interested in estimating the following random coefficients model:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_1 + \mathbf{Z}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i \qquad \boldsymbol{\varepsilon}_i \sim \mathcal{N}_k(\mathbf{0}, \sigma^2 \mathbf{I}_k) \tag{2}$$

Here there are two blocks of covariates. The covariates included in the $(k \times p)$ $\mathbf{X}_i$ matrix are assumed to have fixed marginal effects across all individuals. These are the same as the fixed effects matrix in Equation 1. However, there is not a constant in the $\mathbf{X}_i$ matrix. For the other block of covariates, included in the $(k \times q)$ $\mathbf{Z}_i$ matrix (which does include a constant), we assume that the marginal effects vary across individuals.

To explain this between-subject heterogeneity, we model the random coefficients $\boldsymbol{\beta}_i$ with a multivariate regression model:

$$\boldsymbol{\beta}_i = \mathbf{A}_i\gamma + \mathbf{b}_i \qquad \mathbf{b}_i \sim \mathcal{N}_q(\mathbf{0}, \mathbf{D}) \tag{3}$$

The $(q \times l)$ matrix $\mathbf{A}_i$ contains individual-specific covariates to model the varying marginal effects across individuals. We allow the errors at the second level of the hierarchy to be correlated. Note the difference between this specification and the one in Equation 1 – that the random effects have a zero mean. This is why for the hierarchical model we cannot include a constant in both blocks of covariates.

It turns out that this hierarchical model can be re-written as a general linear panel model after performing some matrix algebra. Substituting Equation 3 into Equation 2, and rearranging, yields the following:

$$\mathbf{y}_i = [\mathbf{X}_i \ \mathbf{Z}_i\mathbf{A}_i] \begin{bmatrix} \boldsymbol{\beta}_1 \\ \gamma \end{bmatrix} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i \qquad \boldsymbol{\varepsilon}_i \sim \mathcal{N}_k(\mathbf{0}, \sigma^2\mathbf{I}_k)$$

With the standard distribution for the random effects:

$$\mathbf{b}_i \sim \mathcal{N}_q(\mathbf{0}, \mathbf{D})$$

This model is precisely the same as the one written in Equation 1. All that is required to estimate a random coefficients model is to create the proper data matrices. The fixed effects matrix is now a $[k \times (p+l)]$ matrix, formed from the usual fixed effects covariates, and the random effects covariates multiplied by the data from the second level of the hierarchy. The random effects covariates are the same as before. The utility of this two-level hierarchical model is solely limited by the creativity of the researcher and available data.

3 Application: Party Identification in the ANES Panel, 1992-1996

To illustrate the general linear panel model, we turn to data from the American National Election Study 1992-1996 Panel (Shapiro et al., 1999). Our dependent variable is the party identification of the respondent, measured in 1992, 1994, and 1996.[3] Since the publication of *The American Voter*, political scientists have generally divided the factors that influence voting decisions and election outcomes into two types: short-term forces and long-term forces (Campbell et al., 1960). Short-term forces include the issues, candidates, and conditions peculiar to a given election, while the most important long-term force is the distribution of party identification within the electorate. Campbell et al. (1960) find that party identification is far more stable than attitudes toward issues and candidates. As a result, party identification exerts a strong influence on individual voting decisions both directly and indirectly, through its influence on attitudes toward the candidates and issues. It is for this reason that understanding party identification is so important. If we can understand party identification, then we can understand elections (and thus governance).

The stability of party identification is entrenched in the American politics canon. Most recent research confirms that party identification is more stable than other political attitudes (Converse and Markus, 1979; Fiorina, 1981; Abramson and Ostrom, 1991) and exerts a much stronger influence on these attitudes than they exert on party identification during the course of a single election campaign (Green and Palmquist, 1990, 1994). This implies that the distribution of party identification remains a key influence on the outcomes of elections in the United States. However, the question of what explains partisan attitudes and what falls where in the chain of causality of these attitudes remains under debate. To truly understand the stability (or instability) of party identification in the electorate, one cannot rely on multiple cross-sections of data. Rather, it is necessary to turn to a panel design, and look at repeated measures on the same individuals over time.

[3]We treat party identification as if it were interval-level. It is not, and is probably best modeled as if it were an ordinal variable. Nonetheless, mass behavior scholars have almost universally treated party identification as a continuous variable (e.g. Campbell et al., 1960; Green and Palmquist, 1994; Abramowitz and Saunders, 1998).

3.1 Party Identification, Race, and Gender

What explains party identification in the electorate? The literature suggests that there are two types of factors: enduring long-term factors (which are typically demographic), and short-term factors, which relate to issues, candidates, and campaigns. In the models we estimate below, we include the standard litany of demographic factors found to be associated with (and argued by Campbell et al. to be *causally* related to) party identification: region, age, education, income, union membership, and religious affiliation. The two enduring factors that we pay particular attention to are gender and race.

Since the early 1990s, scholars have documented a "gender gap" in the American electorate (Greenberg, 2000). The explanations are that women's shared interest in activist government and supporting Democratic candidates has been reflected in their partisan attitudes and the Democratic Party's ability to relate to "women's concerns." This is actually a new phenomenon; prior to the early 1970s, the Republican party was most closely identified with women's issues (Wolbrecht, 2000). Still others have noted that the reason for the gender gap is most notably that women are staying with the Democratic Party and men, mostly white and middle-to-upper class, are leaving for the Republicans (Kaufmann and Petrocik, 1999; Norrander, 1999; Greene and Elder, 2001). This literature suggests that all things being equal, women should identify more strongly with the Democratic party with men. Moreover, we would expect the influence of income to vary between men and women.

Another important factor is the trend of African Americans identifying in great proportions with the Democratic party (Tate, 1993; Dawson, 1994). This too is a relatively recent phenomenon; before the 1960s, the Republican party had nearly universal support from black America (Carmines and Stimson, 1987). Today, African Americans are the most strongly identified group with the Democratic party, and overwhelming supported Clinton in the 1996 election. Race remains central to political behavior, but in ways that are independent from and interactive with socio-economic status and values. Dawson (1994) argues that class does not explain black political behavior, rather African Americans rely on racial group identification or a "black utility heuristic" to understand politics (see also Tate, 1993). While blacks identify at greater than 65% with the Democratic party (and vote at greater than 90%), it is worthwhile to more closely scrutinize the data to gauge the importance of race when controlling for other relevant factors. And, due to the interaction with socio-economic status, we expect income effects to also be mediated by race.

3.2 Ideology and Party Identification

A considerable body of research has demonstrated that party identification is also influenced by policy preferences (see, e.g. Page and Jones, 1979; Luskin et al., 1989; Franklin, 1992). These findings imply that changes in the parties' policy stands or the salience of these policy stands could, over the course of several election cycles, alter the distribution of party loyalties in the electorate as individuals respond to these changes by bringing their party loyalties into line with their policy preferences.

Using data from the 1976-1994 American National Election Studies and the 1992-94 ANES Panel, Abramowitz and Saunders (1998) demonstrate that the outcomes of the 1992 and 1996 elections reflected just this – a long-term shift in the bases of support and relative strength of the two parties. This shift in the party loyalties of the electorate was based on the increased ideological polarization of the two parties during the Reagan and post-Reagan eras. Clearer differences between the parties ideological positions made it easier for citizens to choose a party identification based on their policy preferences. The result, then, was a secular realignment of party loyalties along ideological lines (Abramowitz and Saunders, 1998).

One of the pieces of evidence offered by Abramowitz and Saunders are results that demonstrate the changing comparative influences of parental partisanship and ideology on partisan affiliation. They show the increasing importance of ideology and the decreasing importance of parental socialization on an individual's partisanship over time, controlling for many of these demographic trends.

To wit, in our study, we re-examine the hypothesis that the influence of parental partisanship has declined in comparison to the effect of an individual choosing their partisanship based on their ideological and issue predilections.

3.3 The Data and Model Specification

While aggregate change does not necessarily imply change in the individual level, it is important to first document that party identification exhibits some aggregate level dynamics. Figure 1 demonstrates that partisan affiliation indeed does have a dynamic component, and that the American electorate has steadily, but slowly, moved toward the Republican party. Over the time period from 1980 to 2000, the proportion of both Democratic and Republican identifiers in the electorate has fluctuated quite a bit. Democratic identification has fluctuated from 55% to 47% over the time period, with one of the lows occurring in 1994, concurrent with the high in Republican identification at 42% in 1994. Republican identification has fluctuated from 32% to 42% over the two decades in this figure.

The analyses reported in the remainder of this paper utilize panel survey data collected in the ANES in 1992, 1994, and 1996, giving us three strobes of data for each respondent (Shapiro et al., 1999). We chose to look at these data, not only because this is a time period of dramatic change in American electoral politics with the large and relatively unexpected win of the Republicans in the House of Representatives in 1994, but it also allows us to capture the short-term changes in partisanship and to explain those changes within and across individuals. The dependent variable $\mathbf{y}_i$ is the vector of party identification responses for each respondent in 1992, 1994, and 1996 respectively. We listwise delete missing values, which drops our sample size to $n = 533$. As discussed in Section 5, panel attrition can be handled with an extension of the model.[4] We expect party identification to be related to a handful of long-term factors, including region, age, gender, education, family income, union membership, religious affiliation, and parental party identification.

[4] Dealing with missing covariates is a more difficult problem, which we do not broach here.

To capture short-term factors, and to re-test the Abramowitz and Saunders (1998) findings, we also include a measure of ideology. See Appendix B for a detailed description of all coding decisions.

4 Results

In this section, we report findings from two panel models of party identification. The first model serves as our baseline. For this model, we have $n = 533$ respondents with $k = 3$ strobes. We include fourteen fixed effects covariates (and a constant) in the $\mathbf{X}_i$ matrix, which we assume to have constant marginal effects across all individuals in the study. This implies that $p = 16$. To allow for between-subject heterogeneity, we include a single random effect ($q = 1$): $\mathbf{W}_i = \mathbf{1}$. To estimate the model, we also must characterize our prior beliefs about the parameters of interest. For the $\boldsymbol{\beta}$ vector, we use uninformative priors, setting $\mathbf{b}_0 = \mathbf{0}$ and $\mathbf{B}_0^{-1} = 4 \cdot \mathbf{I}_p$. *A priori* this places most of the prior mass between -8 and 8 for all of the elements of $\boldsymbol{\beta}$.

Assigning priors for the variance parameters is a bit more difficult. We used the unconditional variance of the dependent variable (4.37) as a reference when choosing priors. Most of the prior mass for both the σ^2 parameter and the $\mathbf{D}$ parameter should be to the left of that value. For σ^{-2}, we set $\nu_0 = 6$ and $\delta_0 = 10$. For $\mathbf{D}^{-1}$, we set $\eta_0 = 6$ and $\mathbf{R}_0 = 6$. We have plotted both of these prior distributions in Figure 2. While the priors for the model are assigned to the precisions, we have plotted the priors for the variances (using an inverse-Gamma and inverse-Wishart distribution respectively). It is important to note that we have estimated the model using many different prior parameterizations for the variance terms, and the substantive results do not change. We ran the Gibbs sampler for 25000 iterations after 1000 burnin iterations. We thin every 25 iterations when summarizing the posterior density.

4.1 The Baseline Model

The results from the baseline party identification model are reported in Table 1. Some interesting findings emerge. When controlling for all other factors, including income, higher levels of education make one more likely to become a Republican. Not surprisingly, the presence of a union member in the household makes one more likely to become a Democrat. The income variable is marginally significant; with 96% posterior probability, higher incomes led respondents to more likely identify with the Republicans. Gender and race both exert strong effects; women and blacks are more strongly identify with the Democratic party. Ideology also exerts a strong effect; those ideologically more conservative closely identify with the Republican party. Parental party identification is also strongly positive. Since both are measured on a comparable scale, one can compare the magnitude of these coefficients to assess the Abramowitz and Saunders (1998) argument. Ideology exerts a stronger effect than parental party identification, *ceteris paribus*, which is consistent with the Abramowitz and Saunders (1998) finding.

In addition to learning about the marginal effects, we have also learned about the variance parameters. In Figure 3, we plot the sampler traceplots and the posterior kernel densities of both parameters. Upon visual inspection, it is clear that the sampler is mixing well for both parameters. Chib and Carlin (1999) documents that other estimation strategies, such as the parameter-by-parameter approach used by BUGS, mixes extremely slowly. In the kernel density plots, we have overlaid the prior density over the posterior density. It is clear that the data is *very* informative in this regard. We estimate the conditional error variance to be 0.86. This tells us that on average, after accounting for unobserved heterogeneity, the average residual falls within ± 1.85 of the fitted values. This is a reasonable fit for a measure on a seven point scale, although one could surely do better. In the second kernel density plot, it is clear that the data also speaks very loudly about the variance of the random effects. We estimate the variance of the random effects to be 1.89, which suggests that there is a great deal of unobserved heterogeneity worthy of further exploration.

4.2 A Hierarchical Model

To illustrate how the model can be adapted to fit a two-level hierarchical model, we next fit a model where we allow the effect of income to differ across individuals. We specifically expect gender and race to mediate the effect of income across individuals. We expect high-income whites to self-identify with the Republican party, while high-income blacks to self-identify with the Democratic party (Dawson, 1994). And, we would expect higher incomes to push men more toward Republicanism than women (Greenberg, 2000). In the baseline model, we assume that income has a fixed, constant effect across individuals. We now relax that assumption.

We have thus specified a hierarchical model, as written in Equations 2 and 3. We use the same covariates in the fixed effects part of the model as in the baseline model, except for income. Note that the $\mathbf{X}_i$ matrix does not contain a constant. Because we expect its effect to vary across individuals, the random effects covariates in the $\mathbf{Z}_i$ matrix contain a constant and our measure of income for each individual. We suspect the differences across individuals in these marginal effects can be explained by the gender and race of the respondent. Thus, the matrix of second-level covariates is:

$$\mathbf{A}_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & Gender_i & Race_i \end{bmatrix}$$

This allows gender and race to explain the random coefficients on the income and variable at the first level of the hierarchy. With the $\mathbf{X}_i$, $\mathbf{Z}_i$, and $\mathbf{A}_i$ matrices in hand, it is straightforward to form the data matrices for the software. Note that the $\boldsymbol{\gamma}_1$ coefficient is now the constant term, $\boldsymbol{\gamma}_2$ captures the direct effect of income, and $\boldsymbol{\gamma}_3$ and $\boldsymbol{\gamma}_4$ capture the interactive effects of income and gender and race respectively. But for the correlation across the random effects, this model is the same as a model with interaction terms. However, the multilevel approach allows for much greater flexibility, both in the mean and the variance specification, than using simple interactive terms.

Again we have $n = 533$ and $k = 3$ strobes. After transforming the data, the number of fixed effects covariates $p = 13 + 4 = 17$, and there are $q = 2$ random effects covariates. We use priors commensurate with those from the baseline model: $\mathbf{b}_0 = \mathbf{0}$, $\mathbf{B}_0^{-1} = 4\ \mathbf{I}_p$, $\nu_0 = 6$, and $\delta_0 = 10$. Choosing an appropriate scale matrix for the Wishart prior was a bit more difficult. Since the income variable is measured on a five point scale and party identification is measured on a seven point scale, we expected the variance of the marginal effects to be small. We thus set $\eta_0 = 6$ and:

$$\mathbf{R}_0 = \begin{bmatrix} 5.0 & 0.0 \\ 0.0 & 0.1 \end{bmatrix}$$

Our experience suggests that if the prior on the precision matrix $\mathbf{D}$ contains little support near the value suggested by the data data, the algorithm breaks down and does not mix well. For example, setting $\mathbf{R}_0 = 5 \cdot \mathbf{I}_q$ causes the chain to wander about the parameter space and not converge. We ran the sampler for 100000 Gibbs iterations after 5000 burnin iterations. We thin every 25 iterations when summarizing the posterior density.

Table 2 contains the posterior density summary for the hierarchical model. Many of the significant direct effects are the same as those from the baseline model: education, union membership, gender, ideology, and parental party identification. Race is now insignificant, with 87.0% posterior probability of exerting a negative direct effect. The more interesting findings are in the $\boldsymbol{\gamma}$ hyperparameters. Income exerts a marginally significant effect in this model, with 94.5% posterior probability of being positive. The interaction with gender is insignificant, which means that the positive marginal effect for income does not vary among men and women. For both men and women, after controlling for race, having high incomes is associated with identifying more strongly with the Republican party. The interactive effect with race is more interesting. The $\boldsymbol{\gamma}_4$ hyperparameter is significant, with magnitude far greater than the direct effect. Indeed, the posterior mean for the quantity $(\boldsymbol{\gamma}_2 + \boldsymbol{\gamma}_4)$ – the total effect of income for blacks – is -0.125. It is also marginally statistically significant, with 92.5% posterior probability of being negative. This suggests that for whites, income contributes to stronger support for the Republican party, but for blacks, income contributes to stronger support for the Democratic party. This is contrary to conventional wisdom about the black middle class. Most white voters have been called pocketbook voters, where economic issues far outweigh social issues. Many Republicans have hoped that the monolith of black Democratic partisanship and voting behavior would be cracked with the continued rise if the white middle class. Our results indicate otherwise for blacks, that income may further consolidate black racial identification overriding economic factors among blacks. The effect of income in the baseline model is thus attenuated due to this racial distinction.

The variance parameters are also interesting. First, the conditional error variance σ^2 is smaller than that from the previous model, suggesting an overall better model fit. The variance of the random effects is large; indeed, the variance of the unmodeled heterogeneity across subjects $D_{1,1}$ is actually larger in this model than in the previous model. Further interactive effects need to be explored. The constant random effect and the income random effect are negatively related; if an individual is "strange" with regard to income, she is likely to be more "normal" with regard to the remaining covariates, and *vice versa*. The $D_{2,2}$ parameter is the estimated variance of the marginal effects of income across respondents; the rather large estimate is consistent with the sign flip with respect to race discussed above.

5 Model Extensions

NOTE: This section is a work in progress – and has yet to be written. We have also not yet implemented these extensions in the software. We plan to fit these models in future versions of the paper.

The general linear panel model can be extended in a number of ways to handle other types of data, and to estimate other quantities of interest. In this section, we discuss three potentially useful extensions of the method.

5.1 Modeling Time Series Error Processes

We are interested in relaxing the model to allow:

$$\boldsymbol{\varepsilon}_i \sim \mathcal{N}_k(\mathbf{0}, \boldsymbol{\Omega})$$

We assume a conjugate Wishart prior on the precision:

$$\boldsymbol{\Omega}^{-1} \sim \mathcal{W}(\nu_1, \mathbf{R}_1)$$

Note that $\mathbf{\Omega}$ has $\frac{k(k+1)}{2}$ free elements. If k is small (less than about four), then with a reasonable sample size one can just estimate $\mathbf{\Omega}$ by replacing Step 3 of the sampler with:

$$\mathbf{\Omega}^{-1}|\mathbf{y}, \boldsymbol{\beta}, \{\mathbf{b}_i\}\, \mathbf{D} \sim \mathcal{W}(\nu_1 + nk, \tilde{\mathbf{R}}_1)$$

Where:

$$\tilde{\mathbf{R}}_1 = \left[\mathbf{R}_1^{-1} + \sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{W}_i\mathbf{b}_i)(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{W}_i\mathbf{b}_i)'\right]^{-1}$$

If k is large, then it is necessary to put some restrictions on $\mathbf{\Omega}$ to make the model estimable, such as assuming the errors follow an $ARMA(p, q)$ process. See Chib (1993) and Chib and Greenberg (1995) for details.

5.2 Dichotomous Dependent Variables

Suppose that $\mathbf{y}_i$ takes only zeros or ones, and we would thus want to fit a panel probit model. Once can use a data augmentation approach to easily estimate such a model. CITES: Albert and Chib (1993), Chib and Carlin (1999), and Albert and Chib (1996).

Assume that the observed data $\mathbf{y}_i$ is explained by a latent variable $\mathbf{y}_i^*$, such that if the latent variable is positive, the manifest variable takes a value one, and if it is negative, then the manifest variable takes a value zero. Conditional on the latent variables $\mathbf{y}_i^*$, the panel probit model uses precisely the same algorithm as before, except that $\sigma^2 = 1$ to identify the scale of the latent variable.

One only has to add a step to the algorithm to simulate the latent utilities. Albert and Chib (1993) show that these can be drawn element-by-element from a truncated Normal distribution:

$$y_{i,k}^* \sim \mathcal{TN}(\mathbf{x}_{i,k}'\boldsymbol{\beta} + \mathbf{w}_{i,k}'\mathbf{b}_i, 1)$$

Where the limits of truncation are $(-\infty, 0)$ if the observed variable is zero, and $(0, \infty)$ if the observed variable is one.

5.3 Panel Attrition

As is typically the case in political science panels, suppose that there is a great deal of panel attrition. Further assume that the drop-off is not systematically related to the phenomenon of interest. Given these assumptions, one can easily fix panel attrition by adding a data augmentation step to the estimation algorithm (Tanner and Wong, 1987).

Suppose the data $\mathbf{y} = \{\mathbf{y}^{(mis)}, \mathbf{y}^{(obs)}\}$, where $\mathbf{y}^{(mis)}$ denotes the missing data, and $\mathbf{y}^{(obs)}$ denotes the observed data. The data augmentation step would entail simulating $\mathbf{y}^{(mis)}$ from the estimated parameters of the model in Equation 1.

6 Conclusion

The purpose of this paper is to introduce an alternative modeling strategy for panel data to political scientists. In applied statistics, biostatistics, and psychometrics, linear mixed models have been used for decades to analyze repeated measures data. Most political science, on the other hand, have used fixed effects approaches developed in economics. As argued above, fixed effects approaches mask a great deal of individual-level heterogeneity. We argue that a better approach is, at a minimum, to employ random effects, and ideally to *model* the random effects using suitable subject-specific covariates with hierarchical models. This allows the researcher to not only identify the heterogeneity, but to go one step further and actually *explain* the heterogeneity. We depart from the standard statistical literature by adopting a Bayesian approach, which is particularly suitable for the oftentimes small panel datasets that have been collected in political science. To facilitate the use of these methods by other scholars, we have provided easy-to-use software. Ultimately, the utility of this models, just as any other model, is determined by the substantive research questions and the creativity of others.

A Appendix. `MCMCpack` Software

We have written software to estimate the general linear panel model in Equation 1 using R (Ihaka and Gentleman, 1996). The software is part of the `MCMCpack` package, a preliminary version of which is available from: `http://adm.wustl.edu/tmp/`. On a Linux or MacOS X workstation, simply download the package, log in a superuser, and type the following at the command line:

```
$ R INSTALL MCMCpack_0.1.2.tar.gz
```

This installs the package. To use the function in a R session or in an R program, issue a `library(MCMCpack)` at the R prompt, and all of the functions will become available. At this point, `MCMCpack` contains the general linear panel model, a linear regression model, and some additional density functions and random number generators useful when specifying priors.

The function to estimate a general linear panel model is called `MCMCpanel()`, and is used in the following fashion:

```
MCMCpanel(obs, Y, X, W, burnin = 1000, gibbs = 25000, thin = 25,
    b0 = 0, B0 = .25, nu0=6/2, delta0=10/2, eta0=6, R0=5)
```

Each of the parameters in the function is documented in R. To view the documentation, type `?MCMCpanel` at the R command prompt. The notation used in the documentation corresponds to the notation used in the paper. The `obs` parameter is an $(nk \times 1)$ vector that contains an identification number for each subject. The first subject should be labeled one, the second two, and so forth. For example, when $k = 3$ and $n = 6$:

```
obs <- c(1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6, 6)
```

The `Y`, `X`, and `W` matrices are $\mathbf{y}_i$, $\mathbf{X}_i$, and $\mathbf{W}_i$ respectively, stacked one observation on top of another, corresponding to the `obs` vector. `Y` is thus $(nk \times 1)$, `X` is $(nk \times p)$, and `W` is $(nk \times q)$. In future versions of the software, the model definition statements will correspond to those in the `nlme()` library (Pinheiro and Bates, 2000). The function returns an `mcmc` object that can be directly loaded into the `coda` library for posterior density summary (Best et al., 1997). The functions in `MCMCpack` use R for data management, error checking, and reporting. All samplers are coded in `C++` using the Scythe Statistical Library (Martin and Quinn, 2001). The speed gains over coding the samplers in R are profound. For the general linear panel model, the `C++` version of the code completed the simulation over forty times faster (clock time) than code written in solely in R. In the near future `MCMCpack` will have many more functions, including linear regression with Student-t errors, probit, logit, and item response models, and will be publicly available on `CRAN`.

B Appendix. Measures

Party Identification. The dependent variable in this analysis is the standard 7-point NES party identification scale, ranging from 0 (strong Democrat) to 6 (strong Republican). [Discrete 0-6]

Parental Partisanship. We created a measure of parental partisanship that combines the recalled party identification of the respondents mother and father at the time that he or she was growing up. This measure ranged from 0 (both parents Democrats) to 6 (both parents Republicans) with a middle category of 3 (both parents independents or one parent of each party). This type of recall measure may tend to exaggerate agreement between respondents and their parents (Jennings and Niemi, 1974). Thus our results may somewhat underestimate the true extent of intergenerational change in party identification. This question was only asked in the first strobe of the survey; for strobes two and three, the responses for each respondent were coded the same as in the first strobe. [Discrete 0-6]

Ideology. For our analysis of the 1992-1996 panel survey, we constructed an index of ideology using fourteen individual questions that were asked of panel respondents in all three strobes of the panel. All fourteen items either already were measured in 7-point scales, or were recoded into 7-point scales, with the most liberal response coded as 1 and the most conservative response coded as 7. Respondents with no opinion on an item were placed in the middle category (Abramowitz and Saunders, 1998).

The measures in this index included: liberal conservative self-placement, government vs. personal responsibility for jobs and living standards, government help for disadvantaged minority groups, government vs. private responsibility for health insurance, attitudes toward defense spending, support for affirmative action programs in hiring decisions, a question on abortion policy, and seven questions dealing with the level of government spending on domestic programs (environmental programs, social security, welfare, AIDS research, public schools, food stamps, and child care). By combining these fourteen items, we were able to construct a general measure of ideology with a high degree of reliability (Cronbach's $\alpha > .7$ for all three strobes). [Continuous 1-7]

Religious Affiliation and Commitment. All the coding of religion variables closely followed the guidelines laid out in the Appendix of Layman (2001). Respondents were then placed into a religious tradition based upon their indicated denomination and dummy variables for overall protestant, evangelical, and protestants, based on the denomination of religion that the respondent identified with in the questions. A religious commitment variable was computed using an additive scale based on frequency of church attendance, frequency of prayer, and subjective importance of religion. [Continuous 1-15]

Other Control Variables. For these variables, missing data was put into the median category.

- Region. Coded as South=1, non-South=0. [Discrete 1-0]
- Age. [Continuous 18-99]
- Education. Coded into seven levels. [Discrete 1-7]
- Union. Presence of a union member in the household. [Discrete 0-1]
- Income. Coded into quintiles. [Discrete 0-5]
- Gender. Coded as women = 1, men = 0. [Discrete 0-1]
- Race. Coded as blacks = 1, everyone else = 0. [Discrete 0-1]
- Urban/Rural. Coded as urban = 0, suburban = 1, rural = 2. [Discrete 0-1]

References

Abramowitz, Alan I., and Kyle L. Saunders. 1998. "Ideological Realignment in the U.S. Electorate." *Journal of Politics* 60(3):642–661.

Abramson, Paul R., and Charles W. Ostrom. 1991. "Macropartisanship: An Empirical Reassessment." *American Political Science Review* 85(1):181–192.

Albert, James, and Siddhartha Chib. 1996. "Bayesian Modeling of Binary Repeated Measures Data with Application to Crossover Trials." In *Bayesian Biostatistics* (Donald A. Berry, and Dalene K. Stangl, editors), New York: Marcel Dekker.

Albert, James H., and Siddhartha Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88(June):669–679.

Baltagi, Badi H. 2001. *Econometric Analysis of Panel Data*. New York: Wiley, second edition.

Beck, Neal, and Jonathan N. Katz. 1995. "What To Do (And Not To Do) with Time-Series Cross-Section Data." *American Political Science Review* 89(September):634–647.

Best, Nicky, Mary Kathryn Cowles, and Karen Vines. 1997. *Convergence Diagnostics and Output Analysis Software for Gibbs Sampling Output, Version 0.4*. Cambridge: MRC Biostatistics Unit.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter*. New York: John Wiley & Sons.

Carmines, Edward G, and James A. Stimson. 1987. *Issue Evolution: Race and the Transformation of American Politics*. Princeton, NJ: Princeton University Press.

Chib, Siddhartha. 1993. "Bayes Regression with Autoregressive Errors: A Gibbs Sampling Approach." *Journal of Econometrics* 58(August):275–294.

Chib, Siddhartha, and Bradley P. Carlin. 1999. "On MCMC Sampling in Hierarchical Longitudinal Models." *Statistics and Computing* 9(January):17–26.

Chib, Siddhartha, and Edward Greenberg. 1995. "Hierarchical Analysis of SUR Models With Extensions to Correlated Serial Errors and Time-varying Parameter Models." *Journal of Econometrics* 68(August):339–360.

Converse, Philip E., and Gregory B. Markus. 1979. "Plus Ca Change: The New CPS Election Study Panel." *American Political Science Review* 73(1):32–49.

Dawson, Michael C. 1994. *Behind the Mule: Race and Class in African-American Politics*. Princeton, NJ: Princeton University Press.

Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven, CT: Yale University Press.

Franklin, Charles H. 1992. "Measurement and the Dynamics of Party Identification." *Political Behavior* 14(3):297–310.

Gelfand, Alan E., and Adrian F. M. Smith. 1990. "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association* 85(December):398–409.

Gill, Jeff. 2002. *Bayesian Methods for the Social and Behavioral Sciences*. London: Chapman & Hall.

Goldstein, Harvey. 2002. *Multilevel Statistical Models*. London: Edward Arnold, third edition.

Green, Donald P., and Bradley Palmquist. 1990. "Of Artifacts and Partisan Instability." *American Journal of Political Science* 34(3):872–901.

Green, Donald P., and Bradley Palmquist. 1994. "How Stable is Party Identification?" *Political Behavior* 16(4):437–466.

Green, Donald P., and David H. Yoon. 2002. "Reconciling Individual and Aggregate Evidence Concerning Partisan Stability: Applying Time-Series Models to Panel Survey Data." *Political Analysis* 10(Winter):1–24.

Greenberg, Anna. 2000. "Why Men Leave: Gender and Party Politics in the 1990s." Paper presented at the 2000 Annual Meeting of the American Political Science Association.

Greene, Steven, and Laurel Elder. 2001. "Gender and the Psychology of Partisanship." *Women and Politics* 22:63–84.

Hsiao, Cheng. 1986. *Analysis of Panel Data*. New York: Cambridge University Press.

Ihaka, Ross, and Robert Gentleman. 1996. "R: A Language for Data Analysis and Graphics." *Journal of Computational and Graphical Statistics* 5(3):299–314.

Jackman, Simon. 2000. "Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo." *American Journal of Political Science* 44(April):369–398.

Jennings, M. Kent, and Richard G. Niemi. 1974. *The Political Character of Adolescence*. Princeton, NJ: Princeton University Press.

Kaufmann, Karen M., and John R. Petrocik. 1999. "The Changing Politics of American Men: Understanding the Sources of the Gender Gap." *American Journal of Political Science* 43(July):864–887.

Laird, N. M., and J. H. Ware. 1982. "Random-effects Models for Longitudinal Data." *Biometrics* 38:363–374.

Layman, Geoffrey. 2001. *The Great Divide: Religious Cultural Conflict in American Party Politics*. New York: Columbia University Press.

Lindley, D. V., and A. F. M. Smith. 1972. "Bayes Estimates for the Linear Model." *Journal of the Royal Statistical Society B* 34(1):1–41.

Luskin, Robert C., John P. McIver, and Edward G. Carmines. 1989. "Issues and Transmission of Partisanship." *American Journal of Political Science* 33(2):440–458.

Martin, Andrew D., and Kevin M. Quinn. 2001. "Scythe Statistical Library, Release 0.1." `http://scythe.wustl.edu/`.

Norrander, Barbara. 1999. "Evolution of the Gender Gap." *Public Opinion Quaterly* 63(Winter):566–576.

Page, Benjamin I., and Calvin C. Jones. 1979. "Reciprocal Effects of Policy Preferences, Party Loyalties and the Vote." *American Political Science Review* 73(4):1071–1089.

Pinheiro, José C., and Douglas M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS*. New York: Springer.

Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage, second edition.

Shapiro, Virginia, Steven J. Rosenstone, Donald R. Kinder, Warren E. Miller, and the National Election Studies. 1999. *American National Election Studies, 1992-1997: Combined File [Computer file]*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research, second edition.

Spiegelhalter, David J., Andrew Thomas, Nicky Best, and Wally R. Gilks. 1997. *BUGS 0.6: Bayesian Inference Using Gibbs Sampling*. Cambridge: MRC Biostatistics Unit.

Stimson, James A. 1985. "Regression in Space and Time: A Statistical Essay." *American Journal of Political Science* 29(November):914–947.

Tanner, M. A., and W. Wong. 1987. "The Calculation of Posterior Distributions by Data Augmentation." *Journal of the American Statistical Association* 82(June):528–550.

Tate, Katherine. 1993. *From Protest to Politics: The New Black Voters in American Elections*. Cambridge, MA: Havard University Press.

Wawro, Gregory. 2002. "Estimating Dynamic Panal Data Models in Political Science." *Political Analysis* 10(Winter):25–48.

Western, Bruce. 1998. "Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modelling Approach." *American Journal of Political Science* 42(October):1233–1259.

Wolbrecht, Christina. 2000. *The Politics of Women's Rights: Parties, Positions, and Change*. Princeton, NJ: Princeton University Press.

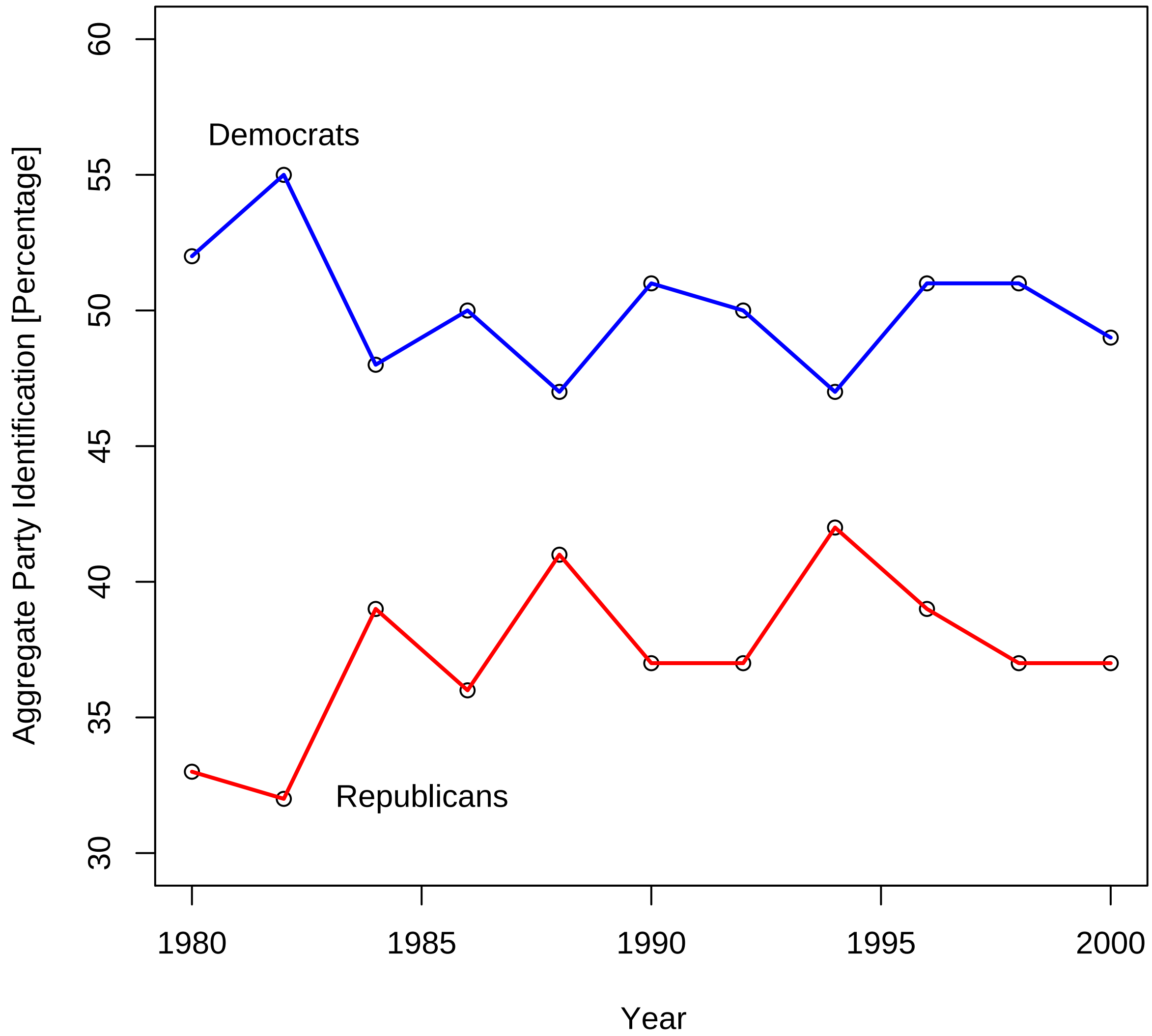

Figure 1: Aggregate levels of party identification from the National Election Studies, 1980-2000.

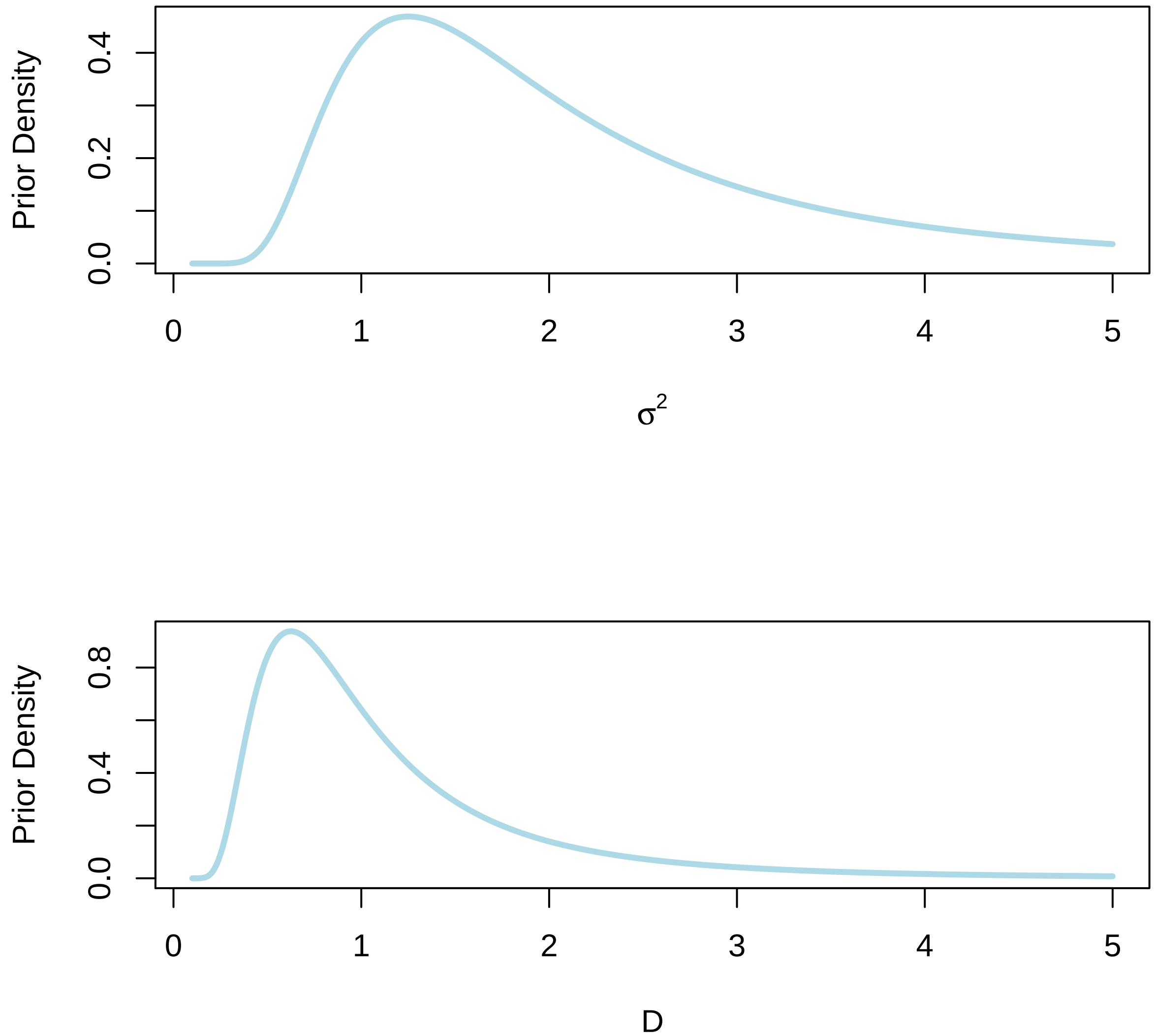

Figure 2: Prior densities for the variance parameters for the baseline model.

	Parameter	Mean	StdDev	2.5%	97.5%
β_1	Constant	0.260940	0.273380	−0.286822	0.802672
β_2	Region	−0.025339	0.113376	−0.243601	0.194003
β_3	Age	−0.001047	0.003336	−0.007418	0.005596
β_4	Education	0.092320	0.033409	0.026426	0.157859
β_5	Union	−0.277949	0.100535	−0.471160	−0.089237
β_6	Income	0.052520	0.031352	−0.008848	0.112955
β_7	Gender	−0.486510	0.110584	−0.703794	−0.263138
β_8	Race	−0.633509	0.171170	−0.962941	−0.298406
β_9	Urban Rural	0.087976	0.078858	−0.059092	0.239450
β_{10}	Ideology	0.377325	0.043021	0.294828	0.460622
β_{11}	Religious Commitment	−0.002593	0.009961	−0.022980	0.016556
β_{12}	Protestant	0.364625	0.195596	−0.010515	0.760642
β_{13}	Evangelical	−0.083025	0.187097	−0.462108	0.271029
β_{14}	Mainline	0.224619	0.200569	−0.167885	0.615816
β_{15}	Parental Party ID	0.310905	0.025708	0.260158	0.361871
D		1.887819	0.143603	1.635254	2.187282
σ^2		0.859414	0.039593	0.788004	0.939718

Table 1: Posterior density summary for the baseline party identification model from the ANES Panel, 1992-1996. $n = 533, k = 3$. The sampler was run for 25000 iterations (thinned every 25) after 1000 burn-in iterations. Standard converge diagnostics suggest that the chain has reached steady state.

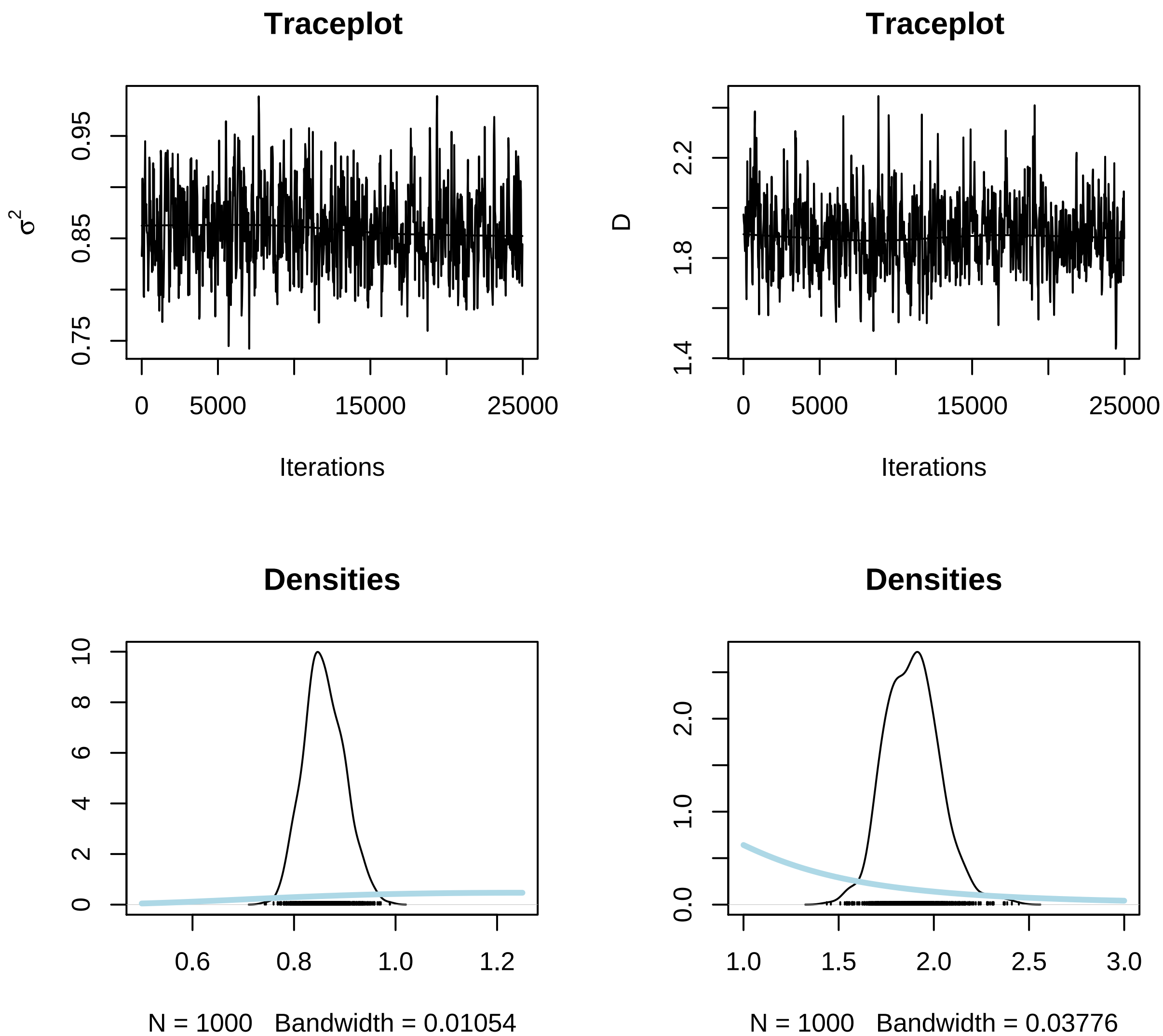

Figure 3: Traceplots, prior densities, and posterior densities for the variance parameters for the baseline model.

	Parameter	Mean	StdDev	2.5%	97.5%
β_1	Region	-0.034238	0.112287	-0.253800	0.188390
β_2	Age	-0.001110	0.003307	-0.007665	0.005467
β_3	Education	0.102010	0.033094	0.038594	0.168271
β_4	Union	-0.235353	0.096941	-0.422598	-0.047707
β_5	Gender	-0.517039	0.193584	-0.892862	-0.130019
β_6	Race	-0.229612	0.207650	-0.638732	0.180548
β_7	Urban Rural	0.073560	0.081598	-0.083251	0.233009
β_8	Ideology	0.343289	0.041703	0.261349	0.425765
β_9	Religious Commitment	-0.006200	0.009309	-0.024564	0.012157
β_{10}	Protestant	0.403121	0.197288	0.016118	0.792332
β_{11}	Evangelical	-0.149510	0.187562	-0.512047	0.211617
β_{12}	Mainline	0.144219	0.203509	-0.250888	0.545213
β_{13}	Parental Party ID	0.317661	0.025931	0.267753	0.370222
γ_1	Constant	0.317929	0.268576	-0.203718	0.848304
γ_2	Income	0.077174	0.048680	-0.018959	0.173158
γ_3	Income $\times$ Gender	0.010603	0.061644	-0.113055	0.131289
γ_4	Income $\times$ Race	-0.202641	0.082098	-0.359634	-0.044830
$D_{1,1}$		2.510101	0.405466	1.778962	3.380774
$D_{1,2}$		-0.398460	0.103462	-0.616741	-0.212742
$D_{2,2}$		0.209307	0.030368	0.156301	0.274683
σ^2		0.775626	0.037152	0.705837	0.853598

Table 2: Posterior density summary for the hierarchical party identification model from the ANES Panel, 1992-1996. $n = 533, k = 3$. The sampler was run for 100000 iterations (thinned every 25) after 5000 burn-in iterations. Standard converge diagnostics suggest that the chain has reached steady state.