Journal of the
Royal Statistical Society
SERIES C
Applied
Statistics

Appl. Statist. (2011)
60, *Part* 1, *pp.* 93–107

Understanding complex legislative and judicial behaviour via hierarchical ideal point estimation

Ying Lu

New York University, USA

and Xiaohui Wang

University of Virginia, Charlottesville, USA

[Received November 2008. Final revision May 2010]

Summary. Ideal point estimation is an important tool to study legislative and judicial voting behaviours. We propose a hierarchical ideal point estimation framework that directly models complex voting behaviours on the basis of the characteristics of the political actors and the votes that they cast. Through simulations and empirical examples we show that this framework holds good promise for resolving many unsettled issues, such as the multi-dimensional aspects of ideology, and the effects of political parties. As a companion to this paper, we offer an easy-to-use R package that implements the methods discussed.

Keywords: Bayesian estimation; Item response theory; Random- and fixed effect models; Roll call data; Vote cast data

1. Introduction

In political science, researchers are often interested in understanding political actors' decision-making behaviour. For example, why do the judges on the Supreme Court vote the way that they do? And why do house members support some bills and reject others? Theorists of legislative and judicial behaviour posit that political actors hold certain policy preferences or ideological values and such preferences underpin their voting behaviour. However, there is usually no explicit data about political actors' preferences. Instead, researchers seek to derive such information from alternative resources such as recorded vote data, speeches of political actors or newspaper editorials.

In this paper, we focus on methods of ideal point estimation that measure political preferences through recorded vote data. A hierarchical statistical framework for ideal point estimation is introduced. Under this framework, researchers can model correlated voting behaviour among groups of individuals and each individual's decisions on related issues. In particular, the hierarchical structure is implemented to allow the elucidation of the characteristics of the decision makers and of the pending bills or cases.

The rest of this paper is organized as follows. In Section 2, we briefly review existing ideal point estimation research and discuss how the complexity in voting behaviour could falsify the commonly adopted assumptions regarding independent voting. Section 3 introduces our model and an illustration of modelling correlated voting behaviour through hierarchical structures.

Address for correspondence: Ying Lu, Department of Humanities and Social Science in the Professions, Kimball Hall, New York University, 246 Greene Street, New York, NY 10003, USA.
E-mail: ying.lu@nyu.edu

 0035–9254/11/60093

Section 4 highlights model estimation. Section 5 presents the results of simulation studies to assess model performance. Section 6 follows with two empirical examples. One example analyses the legislative behaviour of the 109th US House of Representatives and the other analyses the judicial behaviour of the US Supreme Court justices (1919–1996). Finally, Section 7 concludes the paper with a discussion about the potential of this framework.

2. Traditional ideal point estimation and correlated voting behaviour

In political science research, the quantitative measurement of political preference is typically done from ideal points (Epstein and Mershon, 1996; Poole and Rosenthal, 1997; Segal and Spaeth, 1997; Jackman, 2000; Longdregan, 2000; Martin and Quinn, 2002; Clinton *et al.*, 2004; Poole, 2005). The definition of ideal points is based on a theoretical construct of ideological space which represents a liberal–conservative continuum (Poole and Rosenthal, 1997). The main goal of ideal point estimation is to uncover the position of each legislator in the ideological space on the basis of observed vote records. Suppose that there are I political actors making decisions on J different items. The items can be bills discussed in the Congress or cases in the court. The decisions are recorded in a data matrix $\{y_{ij}, i=1, \ldots I, j=1, \ldots, J\}$. When we observe a 'Yea' vote on the jth item by the ith legislator, $y_{ij}=1$; when we observe a 'Nay' vote, $y_{ij}=0$. Following Clinton *et al.* (2004), a unidimensional ideal point estimation model is specified in terms of a latent score t_{ij} with the form

$$t_{ij} = a_j \theta_i - b_j + \varepsilon_{ij}, \tag{1}$$

and

$$y_{ij} = \begin{cases} 1 & \text{if } t_{ij} \geqslant 0, \\ 0 & \text{if } t_{ij} < 0 \end{cases}$$

where θ_i is the ideal point of the ith individual and b_j is the policy position of the jth item that measures how difficult it is for an individual to agree with it. For readers who are familiar with item response theory, b_j is a modified version of the difficulty parameter in item response theory and a_j measures the direction and sensitivity of the jth item in distinguishing individuals' ideal points. Unlike the discrimination parameter in item response theory, the parameter a_j takes values over the entire real line. ε_{ij} is the identically and independently distributed error term.

Under the classical ideal point model (1), there are two assumptions of independence: given item j, every individual votes independently; given ith individual's ideal point, he or she votes independently across all items. The first independence assumption could be violated when voters are influenced by peers, e.g. when party members are influenced by their party to vote towards the party policy line regardless of their ideological values. The second independent assumption is termed *local item independence* in the psychometrics literature. One situation in which it fails is when the unidimensional model is insufficient. For example, an actor can be socially liberal but economically conservative. Another source of local item dependence is related to temporal changes in political preference. In different time periods, legislators could vote differently in response to the changes in political institutions. For example, Lu and McFarland (2007) found that, in the US House of Representatives, there are significant period patterns in voting behaviour of the congressmen under unified democratic, divided and unified Republican government in the last 10 US Congresses. When the independence assumptions are violated, the parameter estimates θ_is, a_j and b_js will be biased and inefficient. There are many references in psychometrics discussing this issue (see Sireci *et al.* (1991), Wainer and Thissen (1996) and Yen (1993)).

In this paper, we generalize equation (1) to allow the characteristics of the political actors and of the cases or bills as well as the context in which the votes are cast to be modelled. Specifically, the generalization takes the form

$$t_{ij} = a_j\theta_{ik} - b_{lj} + \varepsilon_{ij}$$

where we allow the ideal point θ_i to vary across subgroups (denoted by k) of the bills or cases of different contents and the item parameter b_j to vary across subgroups (denoted by l) of individuals of different characteristics. In other words, we introduce covariates into ideal point estimation. This model takes care of the deviation of the assumptions of independence by introducing random-effect terms θ_{ik} and b_{lj}, and allowing them to interact with individuals and cases. A detailed discussion of this model will be presented in Sections 3 and 4.

3. Modelling complex dependent structure

To illustrate our model, we first introduce a set of definitions to denote different types of dependent structures in the recorded vote data. Table 1 illustrates these definitions in a political voting context where there are 10 legislators casting votes on 12 bills. The party affiliations of the legislators are labelled $\mathcal{D}$ for Democratic Party, $\mathcal{R}$ for Republican Party or blank if the legislator does not belong to either party. Moreover, we assume that the contents of the 12 bills can be classified into three different issue areas: economic activities $\mathcal{EA}$, civil liberties issues $\mathcal{CL}$ and political issues $\mathcal{PI}$. A 'Yea' vote is indicated by 1 and a 'Nay' vote is indicated by 0.

We use the term 'allyset' to denote a group of individuals who typically vote together and the term 'voteset' to denote a cluster of items to which each individual's decisions are correlated.

(a) *Allyset*: the hierarchical structure among individuals is defined by allysets. An allyset consists of individuals who tend to influence each other when they vote. For example, individuals who belong to party $\mathcal{D}$ can be considered an allyset, and individuals labelled $\mathcal{R}$ belong to another allyset. Allysets are flexible constructs which can be determined by the characteristics of the political actors. Furthermore, not all individuals need to be included in an allyset; independent individual voters can coexist with allysets.

(b) *Voteset*: the hierarchical structure among items is delineated by votesets. Specifically this term denotes a cluster of items (bills or cases) to which each individual's decisions are correlated; a voteset can be determined by the characteristics of the items such as issue

Table 1. Illustrations of allyset and voteset

I	*Party*	*Types of vote for the following values of J and issues:*											
		1, $\mathcal{EA}$	2, $\mathcal{EA}$	3, $\mathcal{EA}$	4, $\mathcal{EA}$	5, $\mathcal{CL}$	6, $\mathcal{CL}$	7, $\mathcal{CL}$	8, $\mathcal{CL}$	9, $\mathcal{PI}$	10, $\mathcal{PI}$	11, $\mathcal{PI}$	12, $\mathcal{PI}$
1	$\mathcal{D}$	1	1	0	0	1	0	0	0	0	1	1	1
2	$\mathcal{D}$	0	0	1	0	0	0	0	0	1	1	0	0
3	$\mathcal{D}$	0	1	1	0	0	0	0	0	0	1	1	0
4	$\mathcal{D}$	1	0	1	0	0	0	0	1	0	0	0	0
5	$\mathcal{R}$	1	1	0	0	1	1	1	1	1	0	0	0
6	$\mathcal{R}$	0	0	1	1	1	1	1	1	0	1	1	1
7	$\mathcal{R}$	0	1	0	0	1	1	0	0	1	1	0	0
8	$\mathcal{R}$	0	0	0	1	1	0	1	1	0	1	1	1
9		1	0	1	1	1	0	0	1	0	0	0	1
10		1	1	1	0	0	0	0	0	1	0	0	0

areas or time periods. For example, bills 1–4 belong to same voteset $\mathcal{E}A$ because they all concern economic activities, and bills 5–8 and 9–12 belong to two other votesets.

Conditionally on allysets and votesets, we write the hierarchical ideal point estimation model as

$$P\{y_{ij}=1|d(j)=k, p(i)=l\}=\Phi(a_j\theta_{ik}-b_{lj}), \tag{2}$$

with the corresponding latent score,

$$t_{ij}=a_j\theta_{ik}-b_{lj}+\varepsilon_{ij} \tag{3}$$

where ε_{ij} is a standard normal error. Votesets are indexed by $d(j)$; if question j belongs to voteset k, $d(j)=k$. When making decisions, the model assumes that each individual has a unique and voteset-specific ideal point, θ_{ik}. Allysets are indexed by $p(i)$; if individual i belongs to allyset l, $p(i)=l$. The term b_{lj} then measures how difficult it is for members of the lth allyset to agree with the jth item. Later in this paper we also refer to b_{lj} as the lth allyset's policy position for the jth item. When there is only one voteset and one allyset, this model is equivalent to model (1).

We assume that the random effects θ_{ik}, $k=1,\ldots,K$, are randomly distributed with mean 0 and variance–covariance matrix Σ_K,

$$(\theta_{i1},\ldots,\theta_{iK})^{\mathrm{T}}\sim \mathrm{MVN}(0,\Sigma_K), \qquad i=1,\ldots,I, \tag{4}$$

where K is the number of votesets. The mean of θ is set to be 0 for model identification and the diagonal elements of Σ_K are the variance of the votesets, σ_k^2. The off-diagonal terms are the covariance between two votesets, $\sigma_{kk'}$. When there is only one voteset, Σ_K reduces to a scalar σ_θ^2. The random effects b_{lj}, $l=1,\ldots,L$, are assumed to be normally distributed with mean $\mu^{\mathbf{b}}$ and variance Ψ_L,

$$(b_{1j},\ldots,b_{Lj})^{\mathrm{T}}\sim \mathrm{MVN}(\mu^{\mathbf{b}},\Psi_L), \qquad j=1,\ldots,J,$$

where L is the number of allysets. When there is only one allyset, Ψ_L reduces to σ_b^2. We complete the specification of model (3) into a Bayesian hierarchical framework by treating the item parameter a_j as random effects. Specifically we assume that

$$a_j\sim N(\mu_a,\sigma_a^2).$$

Compared with the traditional ideal point model, we can now model various dependences that are introduced by the votesets and allysets. For example, in model (1), the within-subject correlation and within-item correlation of the latent scores are assumed to be constant,

$$\mathrm{corr}(t_{ij},t_{ij'})=\frac{\mu_a^2\sigma_\theta^2}{V(t_{ij})}, \qquad j\neq j',$$

$$\mathrm{corr}(t_{i'j},t_{ij})=\frac{\sigma_b^2}{V(t_{ij})}, \qquad i\neq i',$$

where the variance of t_{ij} is a constant, $V(t_{ij})=1+\sigma_b^2+(\mu_a^2+\sigma_a^2)\sigma_\theta^2$. In the hierarchical ideal point estimation model, the correlations between the latent scores depend on whether items j and j' belong to the same voteset and whether individuals i and i' belong to the same allyset. To illustrate this point, we calculate the correlation assuming only the presence of allysets:

$$\mathrm{corr}(t_{ij},t_{i'j})=\begin{cases}\dfrac{\sigma_l^2}{1+\sigma_l^2+(\mu_a^2+\sigma_a^2)\sigma_\theta^2}, & \text{if } p(i)=p(i')=l,\\[2ex] \dfrac{\sigma_{ll'}}{\sqrt{\{1+\sigma_l^2+(\mu_a^2+\sigma_a^2)\sigma_\theta^2\}}\sqrt{\{1+\sigma_{l'}^2+(\mu_a^2+\sigma_a^2)\sigma_\theta^2\}}}, & \text{if } p(i)=l, p(i')=l', l\neq l'.\end{cases}$$

For two members of the same allyset, the correlation between their responses to the same item j is proportional to the variance of the allyset, σ_l^2. Hence the bigger the variance, the more significant the allyset effect is. However, if the two individuals belong to different allysets, the within-item correlation is proportional to the covariance of the two allysets, $\sigma_{ll'}$. Similarly, we can derive the within-subject correlation in the presence of votesets, $\text{corr}(t_{ij}, t_{ij'})$, to demonstrate that the dependence structure of vote outcomes can be modelled through the covariance matrix of votesets, Σ_K.

$$\text{corr}(t_{ij}, t_{ij'}) = \begin{cases} \dfrac{\mu_a^2 \sigma_k^2}{(\sigma_a^2 + \mu_a^2)\sigma_k^2 + \sigma_b^2 + 1}, & \text{if } d(j) = d(j') = k, \\ \dfrac{\mu_a^2 \sigma_{kk'}}{\sqrt{\{(\sigma_a^2 + \mu_a^2)\sigma_k^2 + \sigma_b^2 + 1\}}\sqrt{\{(\sigma_a^2 + \mu_a^2)\sigma_{k'}^2 + \sigma_b^2 + 1\}}}, & \text{if } d(j) = k, d(j') = k', k \neq k'. \end{cases}$$

4. Hierarchical ideal point estimation

4.1. Identification of the model

Without constraints, model (3) is in general non-identifiable. The quantity $a_j\theta_{ik} - b_{lj}$ is subject to problems of identification. The quantity is invariant to both 'additive aliasing', which is adding any constant to both θ_{ik} and b_{lj}, and to 'multiplicative aliasing', which is multiplying a_j by any non-zero constant d_0 and dividing θ_{ik} by d_0 (Bafumi *et al.*, 2005).

To solve the additive aliasing problem, we set the expected value of θ_{ik} to be 0 in expression (4). To deal with the multiplicative aliasing problem we use an informative prior for θ_{ik} with the prior variance set to be 1. Furthermore, a constraint is put on the rank orders of at least two θ_{ik}-values of each voteset. In a political voting context, this can be easily done by identifying a subset of legislators who are known to be highly conservative or highly liberal in each voteset. The political interpretation of θ_{ik} is not affected by these constraints since they are invariant to the changes in scale.

4.2. Estimation of the models proposed

To minimize the effect of prior specification on the posterior estimation, we assume non-informative Jeffreys prior distributions for the hyperparameters:

$$\Sigma_k \sim |\Sigma_k|^{-(k+1)/2},$$
$$(\mu^{\mathbf{b}}, \Psi_L) \sim |\Psi_L|^{-(L+1)/2},$$
$$(\mu_a, \sigma_a^2) \sim \sigma_a^{-2}.$$

The whole parameter set Ω is

$$\Omega = \{\theta_{ik}, a_j, b_{lj}, i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, K, l = 1, \ldots, L, \Sigma_K, \mu_a, \sigma_a^2, \mu^b, \Psi_L\}.$$

Under the above prior specifications, the posterior estimation of model (3) is fairly straightforward via data augmentation (Tanner and Wong, 1987) and Gibbs sampler. At each step of the Gibbs sampler, the conditional posterior distributions have closed forms. For details see Lu and Wang (2008).

4.3. Assessing model fit

Since the estimation is carried out under a Bayesian framework, we can assess model fit through posterior predictive simulations. The general idea of a posterior predictive check is first to construct some statistics $T(y, \Omega)$ based on the observed values, and then to compare them with

$T(y^{\text{rep}}, \Omega)$ based on replicated (simulated) values from the posterior distribution of Ω. In this paper, we mainly conduct two posterior predictive checks: the p-value of the goodness of fit and model diagnostics based on latent continuous residuals. In general, the posterior predictive p-value of the model is

$$P\{T(y^{\text{rep}}, \Omega) > T(y, \Omega)|y\} = \int P\{T(y^{\text{rep}}, \Omega) > T(y, \Omega)|y, \Omega\} \; f(\Omega|y)\,\mathrm{d}\Omega.$$

When $T(y, \Omega)$ is the deviance statistic, this yields the p-value for testing goodness of fit. In practice, the above integral can be computed by

$$\sum_{m=1}^{M} \mathbf{I}\{T(y^{\text{rep}}, \Omega^m) > T(y, \Omega^m)\}/M,$$

where m represents the total number of posterior draws. In addition, if we are interested in testing specific parameters, such as whether the variances of two allysets are the same, we can also calculate a posterior predictive p-value based on appropriate statistics.

To check whether the distributional assumptions are met, we look at the posterior residuals. The latent continuous residuals ε^s_{ij} (as in equation (3)) are generated conditionally on the sth draw of Ω and the replicated data y^{rep} (Gelman *et al.*, 2000). The normality assumption can be assessed directly on these latent continuous residuals by using tools such as the QQ-plot. The specification of allysets and votesets may be assessed by using the mean absolute predictive error MAPE and deviance information criterion DIC (Spiegelhalter *et al.*, 2002). Specifically, DIC takes the form

$$\text{DIC} = \bar{D} + p_D,$$

where $\bar{D}$, the posterior expectation of the deviance, serves as the Bayesian measure of model adequacy. It can be calculated directly from the Markov chain Monte Carlo chains. p_D is a penalty term that accounts for the complexity of the model and can be interpreted as 'the effective number of parameters'. In general, the smaller the value of DIC, the better the model fit is.

5. Simulation studies

In this section, we present a set of simulation studies to assess the performance of our model in estimating the allyset effect and voteset effect.

There are three different hypothetical scenarios from which simulated data sets are generated. In the first scenario it is assumed that 100 individuals vote on 100 items and there are four votesets, each of which has 25 items, and two allysets, each of which has 50 individuals. The second scenario assumes 200 individuals voting on 200 items. As in the first scenario, there are four votesets of equal numbers of items and two allysets of equal numbers of individuals. Lastly the third scenario involves 400 individuals voting on 400 items. It has the same structure of votesets and allysets.

In each of the scenarios, voteset 1 and voteset 2 are assumed to be correlated with correlation 0.5, and voteset 1 and voteset 3 are correlated with correlation −0.8. The rest of the votesets are assumed to be pairwise independent. Hence, in each simulation, the voteset effects θ_{ik}, $k = 1, \ldots, 4$, are drawn from a multivariate normal distribution as follows:

$$\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \\ \theta_{i3} \\ \theta_{i4} \end{pmatrix} \sim \text{MVN}\left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 & -0.8 & 0 \\ 0.5 & 1 & 0 & 0 \\ -0.8 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right\}.$$

The two allysets are correlated with correlation −0.5 and they have different mean and variance. Specifically, in each simulation, the allyset effects b_{lj}, $l = 1, 2,$ are drawn from a bivariate normal distribution as follows:

$$\begin{pmatrix} b_{1j} \\ b_{2j} \end{pmatrix} \sim \text{BVN}\left\{ \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 & -0.7071 \\ -0.7071 & 1 \end{pmatrix} \right\}.$$

To assess the performance of the Bayesian model estimation proposed, we simulate 100 data sets (owing to the expansive computation of the Bayesian method) for each scenario and fit the Bayesian model for each data set. In Table 2, the estimated expected values of the following parameters are reported: the posterior variance of the voteset effects θ_{ik} and the correlation between every two votesets, the posterior mean and variance of the allyset effects b_{lj} and the correlation between the two allysets, and the posterior mean and variance of item parameter a_j. Table 2 indicates that the estimated entries for the covariance matrix of θ tend to be underestimated when the sample size is small. This is due to the shrinkage effect of the informative prior on θ_{ik}. However, this effect decreases as the sample size increases. In contrast, the estimation of the variance of a_j and b_{lj} is little affected by the non-informative Jeffreys priors.

Table 3 reports the mean-square errors of the individual level voteset-specific ideal point estimates, $\Sigma_{i=1}^{I} (\hat{\theta}_{ik} - \theta_{ik})^2 / I$, the mean-square errors of the item level allyset-specific effects

Table 2. Parameter estimates and standard errors (in parentheses) for three simulated examples†

Parameter	*True value*	*Results for simulation 1, $I=100$, $J=100$*	*Results for simulation 2, $I=200$, $J=200$*	*Results for simulation 3, $I=400$, $J=400$*
Voteset				
$V(\theta_1)$	1	0.876 (0.067)	0.931 (0.040)	0.975 (0.024)
$V(\theta_2)$	1	0.869 (0.072)	0.938 (0.044)	0.969 (0.026)
$V(\theta_3)$	1	0.874 (0.058)	0.939 (0.037)	0.965 (0.028)
$V(\theta_4)$	1	0.875 (0.063)	0.937 (0.038)	0.965 (0.034)
$\rho_\theta(1,2)$	0.5	0.438 (0.086)	0.478 (0.052)	0.486 (0.040)
$\rho_\theta(1,3)$	−0.8	−0.711 (0.042)	−0.758 (0.028)	−0.776 (0.019)
$\rho_\theta(1,4)$	0	−0.000 (0.102)	−0.011 (0.061)	0.000 (0.050)
$\rho_\theta(2,3)$	0	0.006 (0.092)	−0.006 (0.068)	−0.002 (0.049)
$\rho_\theta(2,4)$	0	0.002 (0.102)	−0.005 (0.070)	−0.001 (0.047)
$\rho_\theta(3,4)$	0	−0.013 (0.090)	0.001 (0.069)	0.001 (0.050)
Allyset				
$E(b_1)$	1	1.012 (0.137)	1.004 (0.097)	0.999 (0.075)
$E(b_2)$	−1	−1.002 (0.089)	−1.011 (0.065)	−0.994 (0.043)
$V(b_1)$	2	2.159 (0.399)	2.032 (0.242)	2.058 (0.174)
$V(b_2)$	1	1.058 (0.174)	1.014 (0.114)	1.011 (0.075)
$\rho_b(1,2)$	−0.5	−0.497 (0.081)	−0.501 (0.057)	−0.509 (0.040)
a				
$E(a)$	0	−0.001 (0.136)	0.010 (0.100)	0.000 (0.073)
$V(a)$	2	2.068 (0.437)	2.050 (0.239)	2.013 (0.159)

†The columns report the expected values of the parameter estimates based on averaging the results of the 100 simulations. The standard errors are the means of the standard errors of the corresponding parameter estimates. The estimation is based on model (3).

Table 3. Mean-square errors and standard errors (in parentheses) of the individual and item level parameters of the three simulated examples†

Parameter	*Results for simulation 1, $I=100$, $J=100$*		*Results for simulation 2, $I=200$, $J=200$*		*Results for simulation 3, $I=400$, $J=400$*	
Mse(θ_1)	0.128	0.028	0.066	0.014	0.032	0.005
Mse(θ_2)	0.130	0.037	0.062	0.012	0.032	0.005
Mse(θ_3)	0.127	0.034	0.064	0.013	0.032	0.006
Mse(θ_4)	0.122	0.031	0.063	0.011	0.031	0.004
Mse(b_1)	0.201	0.056	0.112	0.030	0.062	0.012
Mse(b_2)	0.121	0.025	0.068	0.010	0.035	0.005
Mse(a)	0.142	0.031	0.068	0.011	0.034	0.004

†The estimation is based on model (3).

estimates, $\Sigma_{j=1}^{J} (\hat{b}_{lj} - b_{lj})^2/J$, and the mean-square error of a_j, $\Sigma_{j=1}^{J} (\hat{a}_j - a_j)^2/J$. In these calculations, $\hat{\theta}_{ik}$, $\hat{a}_j$ and $\hat{b}_{lj}$ are the posterior means of the corresponding parameters. As the number of questions and the number of individuals double, the mean-square errors of these individual level and item level parameters are also halved. In general, these simulation studies show that the Bayesian estimation gives consistent results.

6. Applications to US judicial and legislative behaviour

In this section, we present two examples of applying the hierarchical ideal point model to analyse legislative and judicial behaviour. In the first example, treating each party as an allyset, we examine to what extent party affiliations affect decisions that were made by the members of the 109th US Congress. The second example explores the issue-specific preferences of the US Supreme Court justices in different issue areas defined by votesets.

To ensure the convergence, all the results that we present are based on multiple chains and past convergence diagnostics (Gelman and Rubin, 1992). Four Markov chains are run in each example; each chain consists of 10000 iterations after a burn-in period of 10000 iterations, and only every 10th draw is kept to reduce the serial correlation of the Markov chains. To assess the distributional assumptions and the goodness of fit of the models, we also conducted posterior predictive checks and the results were satisfactory.

In both data sets, especially in the Supreme Court data, there are many missing values in the data matrices. Since most of the missing values are caused by the finite terms of the judges and legislators rather than based on individual preferences and/or the contents of the items, we treat them as missing at random. In the following two sections, we present the most interesting findings on the basis of our proposed model. For brevity, the details of parameter estimations and posterior predictive check results are relegated to an appendix of the paper, which is available at `http://homepages.nyu.edu/yl46/ipe/Appendices`.

6.1. Party influence in Congress

In the US House of Representatives, legislators may be expected to vote according to their political preference. However, since the US Congress is known to be bipartisan, the question arises whether parties significantly influence their members' votes in addition to their ideological differences. For example, one may expect that members of the minority party in particular might vote in blocks to maximize the party's power. In this data analysis, we examined party influence on voting in the 109th Congress where the Republican Party was the majority party. The vote records were compiled by Lewis and Poole and are available at `http://www.voteview.com`. In this Congress, 1038 non-unanimous votes were cast by 440 representatives, of whom 203 were Democrats and 237 were Republicans. Fig. 1 plots the percentage of 'Yea' vote for Republicans against the same percentage for Democrats. The circles label bills with conservative contents (parameter a_js for these bills are estimated to be greater than 0), and the dots label bills with liberal contents ($a_j < 0$). We can see that, although a sizable number of bills received similar support from members of the two parties, more often, the votes were polarized. Among these bills, 598 were passed, of which about 65% received more votes from the Republicans.

To measure the effect of party influence, we fit a hierarchical ideal point model with two allysets defined by the Democratic and Republican parties. There have been many attempts to assess this effect, but most of them are *ad hoc* (Schickler, 2000; Clinton *et al.*, 2004; Snyder and Groseclose, 2000; Ansolabehere *et al.*, 2001). To keep the discussion succinct, we estimate only the ideal points along the main liberal–conservative continuum without voteset effects,

$$t_{ij} = a_j \theta_i - b_{lj} \tag{5}$$

where $l=1$ for the Democratic Party, and $l=2$ for the Republican Party. To identify the model, we constrain the most liberal member of the house to have a negative θ_i and vice versa for the most conservative member of the house. Consequently, a bill is considered as having liberal content if $a_j < 0$ and vice versa.

Our model reveals that the variance of the allyset Democrat is 6.7 (with posterior standard deviation 0.78) and the variance of allyset Republican is 2.8 (with posterior standard deviation

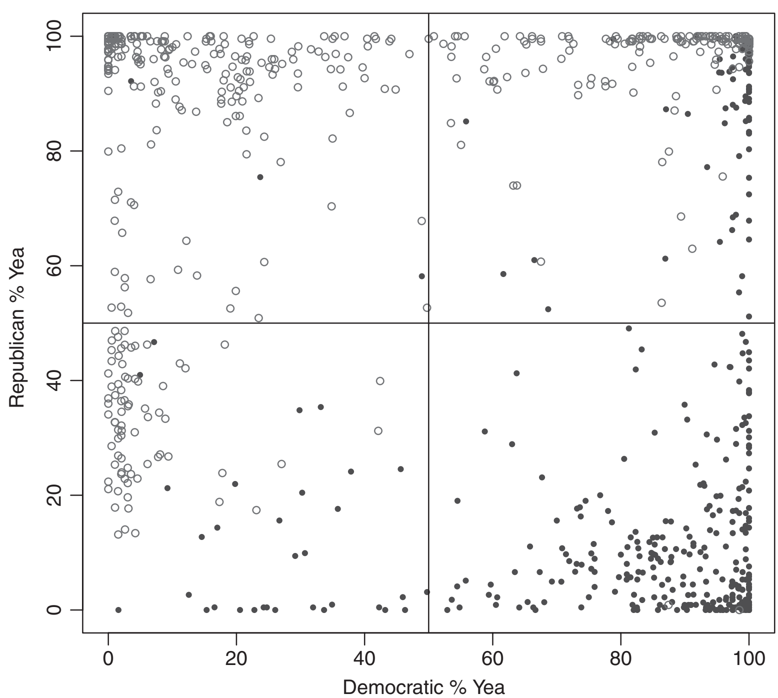

Fig. 1. Roll call outcomes: percentage of 'Yea' votes for Republicans *versus* the same percentage for Democrats in the 109th US Congress

0.32). As discussed in Section 3, this shows that, when voting on each bill, the decisions that are made by Democratic Party members are more correlated than those made by the Republican Party members. This implies that, as the minority party in the 109th Congress, the Democratic Party did have higher influence on their members' votes.

Fig. 2 shows how the votes of the Democratic and Republican Party members are influenced by ideological differences and party membership. The Democratic allyset effects b_{1j} are plotted against the Republican allyset effects b_{2j} in Fig. 2(a). A bill with liberal content ($a_j < 0$) is labelled by a dot and a bill with conservative content ($a_j > 0$) is labelled by a circle. We can see that, for the majority of bills, the policy positions of the two parties are similar. Hence, for those bills, the difference in outcomes of votes mostly reflect the ideological differences of the legislators. In contrast, the policy positions for a substantial number of bills are different between the two parties. For example, the upper left quadrangle of Fig. 2(a) is populated mostly by bills with liberal contents (dots). The policy positions for these bills are different for the two parties. For the Democratic Party, these bills are considered easier to pass ($b_{1j} < 0$) and vice versa for the Republican Party.

A mirror image can be seen in the lower right-hand quadrangle, suggesting that both parties manipulate a significant number of the conservative content roll calls as well. These results are highly suggestive that the political parties can influence their members in the legislative process. Interesting *ad hoc* research can be done to examine the contents of those bills that exhibit greater distance between the policy position of each party, e.g. whether they tend to be bills regarding important 'party' issues, procedural votes or close votes.

In Fig. 2(c), we plot the density of the ideal point estimates of the legislators based on the hierarchical model (5). The full curve represents the density of ideal points for Democratic Party members; the broken curve represents Republican Party members. The underlying political preferences of the legislators are quite polarized with little overlap. In contrast, the density of the ideal points of house members before controlling for the effects of the allysets are presented in Fig. 2(b). We can see that, without identifying the correct dependence structure, the ideal point estimates look artificially polarized with no overlap between members of the two parties.

To compare the allyset model and the null model formally, we present the deviance information criterion DIC and mean absolute predictive errors MAPE of these two models (Table 4). Not surprisingly, the party allyset model performs much better than the null model.

6.2. *Estimating ideal points within different issue areas*

As the individuals with the highest judicial power in the US, the Supreme Court justices receive much attention on their ideological values. Earlier substantive research has suggested that the decisions which justices make could have come from different ideological dimensions such as civil liberties, economic issues and political institutions (Schubert, 1974; Spaeth, 1979). However, more recent research based on the court rulings records under Chief Justice Rehnquist found evidence of a unidimensional court via pattern analysis (Sirovich, 2003). Recently, efforts have been made to develop multi-dimensional ideal point estimation (Jackman, 2001; Rivers, 2003; Poole, 2005). However, such models are all based on the assumption of an orthogonal multi-dimensional space which suffers lack of substantive interpretation of each subdimension and great difficulty in model fitting. In this section, we approach the task of modelling multi-dimensional ideal points via hierarchical model with votesets that are defined on the basis of the contents of the court cases.

The vote records of the US Supreme Court justices are extracted from the Original US Supreme Court judicial database that was compiled by Spaeth (2001). This data set includes

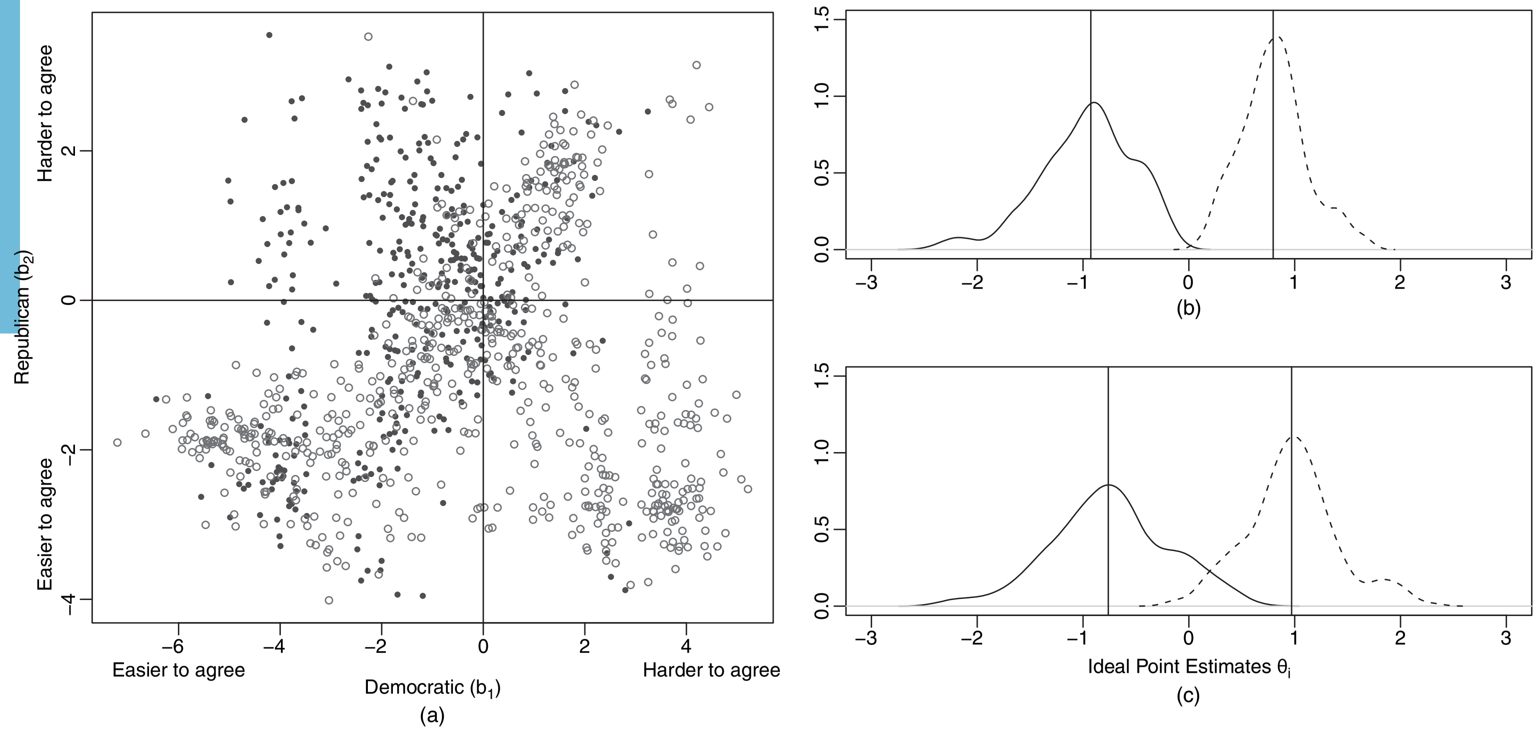

Fig. 2. (a) Allyset effects estimates by contents of the bills, and density of the ideal point estimates θ_i for members of the Democrat (———) and Republican (-------) parties under (b) the null and (c) party models

Table 4. Deviance information criterion and mean absolute predictive errors of the null and party models

Model	*DIC*	$\bar{D}$	p_D	*MAPE*
Null	145248	143565.3	1682.7	0.167
Party	140100.5	137850.3	2250.2	0.159

Table 5. Deviance information criterion and mean absolute predictive errors of five models†

Votesets definitions	*DIC*	$\bar{D}$	p_D	*MAPE*
$\boxed{\text{civil political economic}}$	21348.4	17309.3	4039.1	0.2344
$\boxed{\text{civil political}}$ $\boxed{\text{economic}}$	21123.3	17045.9	4077.4	0.2305
$\boxed{\text{economic civil}}$ $\boxed{\text{political}}$	23696.2	19633.7	4062.5	0.2688
$\boxed{\text{economic political}}$ $\boxed{\text{civil}}$	20988.9	16979.8	4019.1	0.2297
$\boxed{\text{economic}}$ $\boxed{\text{civil}}$ $\boxed{\text{political}}$	22996.8	18884.6	4112.2	0.2307

†A block indicates a voteset.

all court cases and the results of votes from 1953 to 2003. 29 justices and 3069 cases with non-unanimous decisions were debated during this period. In Spaeth's original database, 13 issue areas are defined on the basis of the contents of the cases. We group them into three votesets of major categories: economic activities (including 543 cases related to economic activity, federal taxation, interstate relations and labour unions), civil liberties (including 1180 cases related to civil rights, criminal procedures, due process, the First Amendment and privacy) and federalism (including 325 cases related to attorneys, federalism and judicial power).

To understand whether the justices have different ideological values within different issue areas, we fit five models based on different definitions of the votesets (see Table 5 for definitions; each box refers to a voteset). The models and their performances are presented in Table 5. We can see that a two-voteset model which combines issue areas economic activities and political institutions as one voteset and leaving cases concerning civil liberties as another voteset fits the data best.

Furthermore, our model reveals that there is considerable variation in the justices' ideology toward civil liberties (variance 1.17) compared with the areas of economic and political voteset (variance 0.36). In Fig. 3, we plot the rank orders of the justices in the two votesets. Largely, the ideology of most justices remained consistent in both votesets and the correlation between their rank orders is 0.8. Nevertheless there are exceptions. For example, Judge Clark, who is an avid promoter of the 'New Deal' economic policy, is estimated as having a moderate point of view towards civil liberties issues. In contrast, the model shows that Justices Reed and Minton had very conservative views when deciding cases of civil liberties but were moderate justices when debating cases regarding economic activities and political institutions. The findings about these individual justices are consistent with existing anecdotal research on judicial behaviour.

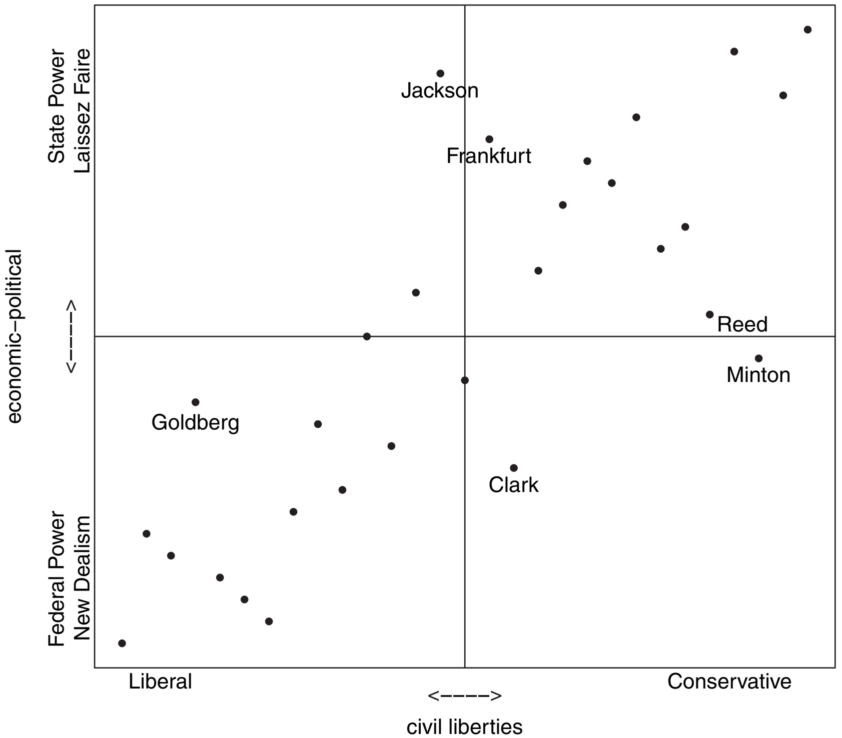

Fig. 3. Ideal point estimates of the Supreme Court justices in different issue areas

7. Discussion and remarks

A unified approach to modelling complex legislative behaviours through votesets and allysets in ideal point estimation has been presented in this paper. In this hierarchical framework, allysets and votesets are defined by the contents of the bills or the cases and the characteristics of the voters as well as the context in which they vote. This enables researchers to incorporate their substantive knowledge about legislative and judicial behaviours into statistical modelling. Furthermore, such definitions of allysets and votesets make them very flexible structures, offering researchers opportunities to test alternative theories of voting.

The model that was proposed in Section 3 represents a general form of modelling hierarchical structures in ideal point estimation and can be easily generalized or reparameterized. For example, when there are independent voters, one can model the true policy position b_j^0 in absence of allysets effects,

$$t_{ij} = a_j\theta_{ik} - (b_j^0 + \varphi_{lj}) + \varepsilon_{ij} \qquad \varphi_{lj} \sim N(0, \sigma_l^2) \tag{6}$$

where φ_{lj} is the allyset-induced effect and is equal to 0 for independent voters, and σ_ls actually quantify the exact party effects. Moreover, one can model time varying party influence through votesets that are defined by time periods and interacting allyset effects and votesets effects (Lu and McFarland, 2007). Similar to the reparameterization (6), if we know that only a subset of items violates the local item independence assumption, we can group them into votesets and model main ideal points of each individual, allowing occasional deviations within votesets,

$$t_{ij} = a_j(\theta_i^0 + \gamma_{ik}) - b_j^0 + \varepsilon_{ij} \qquad \gamma_{ik} \sim N(0, \sigma_k^2) \tag{7}$$

where γ_{ik} is the voteset-induced effect and is equal to 0 for independent items and σ_k actually quantifies the deviation from the main ideological dimension of each voteset. For example, one can test whether Justice Rehnquist influenced the decisions that were made during his tenure as Chief Justice or the conditional government hypothesis of conditional party influence. In reparameterization (7) the model specification is a close variation of the 'testlet effects model' of Bradlow *et al.* (1999) in the field of education testing.

Lastly, we can also consider other forms of correlated voting such as votes affected by interest groups and temporary political coalitions. We denote such structure a 'tactset' of votes. Namely, a tactset is a block (or collections) of correlated decisions that are made by a subgroup of individuals on a selection of bills. They can be incorporated in the hierarchical ideal point estimation model as fixed effects:

$$t_{ij} = a_j(\theta_i + \delta_m) - b_j + \varepsilon_{ij}.$$

Not shown in this paper, simulation studies have been carried out for models with tactsets and consistent results are established. Models with tactsets can be used to study transitory collaborative voting behaviour such as strategic voting, vote trading and agenda setting.

Acknowledgements

Our thanks go to Andrew Martin, Kevin Quinn, Dan Spitzner, Howard Wainer and attendants at the 2006 and 2007 political methodology meetings for their valuable suggestions and comments. We also thank Gary King and the Institute for Quantitative Social Science, Harvard University, for their generous support. This paper is partially supported by National Science Foundation grants DMS-0631652 and DMS-0631639, National Security Agent grant NSA-GG11184 and a developmental grant from the National Institute of Child Health and Human Development funded University of Colorado Population Center (grant R21 HD51146).

References

Ansolabehere, S., Snyder, J. M. and Stewart III, C. (2001) The effects of party and preferences on Congressional Roll-call voting. *Legis. Stud. Q.*, **26**, 533–572.

Bafumi, J., Gelman, A., Park, D. K. and Kaplan, N. (2005) Practical issues in implementing and understanding Bayesian ideal point estimation. *Polit. Anal.*, **14**, 381–396.

Bradlow, E. T., Wainer, H. and Wang, X. (1999) A Bayesian random effects model for testlets. *Psychometrika*, **64**, 153–168.

Clinton, J., Jackman, S. and Rivers, D. (2004) The statistical analysis of roll call data. *Am. Polit. Sci. Rev.*, **98**, 355–370.

Epstein, L. and Mershon, C. (1996) Measuring political preferences. *Am. J. Polit. Sci.*, **60**, 801–818.

Gelman, A., Goegebeur, Y., Tuerlinckx, F. and Van Mechelen, I. (2000) Diagnostic checks for discrete data regression models using posterior predictive simulations. *Appl. Statist.*, **49**, 247–268.

Gelman, A. and Rubin, D. B. (1992) Inference from iterative simulations using multiple sequences (with discussion). *Statist. Sci.*, **7**, 457–511.

Jackman, S. (2000) Estimation and inference are missing data problems: unifying social science statistics via Bayesian simulations. *Polit. Anal.*, **8**, 307–332.

Jackman, S. (2001) Multidimensional analysis of roll call data via Bayesian simulation: identification, estimation, inference and model checking. *Polit. Anal.*, **9**, 227–241.

Longdregan, J. (2000) *Legislative Institutions and Ideology in Chile's Democratic Transition*. New York: Cambridge University Press.

Lu, Y. and McFarland, T. (2007) Reveal Congressional party effects via hierarchical ideal point estimation. *Summer Political Methodology Meet., Penn State University, University Park*.

Lu, Y. and Wang, X. (2008) *Software Manual for IPE: R Package for Item Response Models with Correlated Structures*. New York: New York University. (Available from `http://homepages.nyu.edu/yl46/ipe`.)

Martin, A. D. and Quinn, K. M. (2002) Dynamic ideal point estimation via Markov Chain Monte Carlo for the U.S. Supreme Court. *Polit. Anal.*, **10**, 134–153.

Poole, K. T. (2005) *Spatial Models of Parliamentary Voting*. New York: Cambridge University Press.
Poole, K. T. and Rosenthal, H. (1997) *Congress: a Political-economic History of Roll Call Voting*. New York: Oxford University Press.
Rivers, D. (2003) Identification of multidimensional spatial voting models. *Manuscript*. Department of Political Science, Standford University, Palo Alto.
Schickler, E. (2000) Institutional change in the House of Representatives, 1867-1998: a test of partisan and ideological power balance models. *Am. Polit. Sci. Rev.*, **94**, 269–288.
Schubert, G. (1974) *The Judicial Mind Revisited*. London: Oxford University Press.
Segal, J. A. and Spaeth, H. J. (1997) *The Supreme Court and the Attitudinal Model*. Cambridge: Cambridge University Press.
Sireci, S. G., Wainer, H. and Thissen, D. (1991) On the reliability of testlet-based tests. *J. Educ. Measmnt*, **28**, 237–247.
Sirovich, L. (2003) A pattern analysis of the second Rehnquist U.S. Supreme Court. *Proc. Natn. Acad. Sci. USA*, **100**, 7432–7437.
Snyder, J. and Groseclose, T. (2000) Estimating party influence in congressional roll call voting. *Am. J. Polit. Sci.*, **44**, 193–211.
Spaeth, H. J. (1979) *Supreme Court Policy Making*. San Francisco: Freeman.
Spaeth, H. J. (2001) United States Supreme Court judicial database, 1953-2000. (Available from `http://supremecourtdatabase.org`.)
Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *J. R. Statist. Soc.* B, **64**, 583–639.
Tanner, M. A. and Wong, W. W. (1987) The calculation of posterior distributions by data augmentation. *J. Am. Statist. Ass.*, **82**, 528–540.
Wainer, H., Bradlow, E. T. and Wang, X. (2007) *Testlet Response Theory and Its Applications*. Cambridge: Cambridge University Press.
Wainer, H. and Thissen, D. (1996) How is reliability related to the quality of test score?: what is the effect of local dependence on reliability? *Educ. Measmnt Issue Pract.*, **15**, 22–29.
Yen, W. (1993) Scaling performance assessments: strategies for managing local item dependence. *J. Educ. Measmnt*, **30**, 187–213.