Journal of the
Royal Statistical Society
SERIES C
Applied
Statistics

Appl. Statist. (2012)
61, *Part* 5, *pp.*

A sparse factor analytic probit model for congressional voting patterns

P. Richard Hahn

University of Chicago, USA

and Carlos M. Carvalho and James G. Scott

University of Texas at Austin, USA

[Received December 2010. Final revision December 2011]

Summary. The paper adapts sparse factor models for exploring covariation in multivariate binary data, with an application to measuring latent factors in US Congressional roll-call voting patterns. This straightforward modification provides two advantages over traditional factor analysis of binary data. First, a sparsity prior can be used to assess the evidence that a given factor loading may be exactly 0, realizing a principled unification of exploratory and confirmatory factory analysis. Second, incorporating sparsity into existing factor analytic probit models effects a favourable bias–variance trade-off in estimating the covariance matrix of the multivariate Gaussian latent variables. Posterior summaries from this model applied to the roll-call data provide novel metrics of partisanship of a given Senate.

Keywords: Covariance estimation; Factor models; Multivariate probit models; Voting patterns

1. Introduction

In this paper, we extend the Bayesian multivariate probit model (Chib and Greenberg, 1998) to encompass a sparse factor analytic approach for inference about the underlying correlation structure of binary data. We apply the proposed method to study partisan patterns in 60 years of non-lopsided roll-call votes from the US Senate, refining the similar analyses of Jackman (2001) and Clinton *et al.* (2004a). Our results show that the role of partisanship has risen distinctly in the last few decades, after hitting a low around 1970.

Our goal is not to propose a model that explains why senators cast their votes; such a model would need to consider not only party membership but also geography, incumbency, committee membership and much more besides. Rather, we undertake an exploratory analysis of the Senate roll-call data. Our statistical approach differs from previous methods in that it allows practitioners to assesses whether individual elements of the factor loadings matrix—intuitively, the mapping between latent traits and observed votes—are identically equal to 0. In substantive terms, an inferred 0 in the (j, g) entry of the factor loadings matrix corresponds to the judgement that the gth latent trait plays no discernible role in predicting whether senators are likely to vote for or against bill j. In our approach, moreover, such judgements need not be encoded beforehand. Instead, they emerge naturally in the final analysis, assuming that they are supported by the data.

Address for correspondence: James G. Scott, Division of Statistics and Scientific Computing and McCombs School of Business, University of Texas at Austin, Austin, TX 78712, USA.
E-mail: james.scott@mccombs.utexas.edu

 0035–9254/12/61000

The individual components of our model are the multivariate probit model, Gaussian factor models and point mass sparsity priors, each of which have been introduced in previous literature. In extending this literature by incorporating sparsity priors within a probit model, our paper has two motivating goals.

(a) *One-pass factor analysis*: factor analysis has traditionally been used in two distinct modes, an exploratory phase, which is used to generate further hypotheses about the forces at play in the data, and a confirmatory phase where elements of the factor loadings are set to 0 to reflect presumed conditional independences. Our Bayesian framework with sparsity priors collapses these methods into a one-pass approach using ideas from Bayesian model averaging (Hoeting *et al.*, 1999).
(b) *Regularization*: imposing a factor structure on a covariance matrix stabilizes estimation, which is critical when the number of variables p is large relative to the sample size n (Rajaratnam *et al.*, 2008). Such regularization is even more crucial when the estimated covariance matrix corresponds to an unobservable latent variable as it does in a multivariate probit model. Accordingly, our sparse factor model increases the degree of regularization by shrinking the elements of the factor loadings matrix towards 0. A simulation study demonstrates that sparse factor models effect a highly favourable bias–variance trade-off. This favourable trade-off persists even when there is no particular reason to suspect an underlying factor structure.

1.1. Background: latent factor probit models

The latent factor probit model that is described in this section is well known in some research communities, going by various names—ideal point estimation and item response modelling being two of the most common. Use of such models is currently widespread in many fields: political science (Jackman, 2001, 2009; Clinton *et al.*, 2004a; Quinn, 2004), statistics (Song and Lee, 2001, 2005), biostatistics (Bock and Gibbons, 1996) and marketing (Elrod and Keane, 1995). Two fine general references are Johnson and Albert (1999) and Bartholomew (1987).

Our presentation focuses on a normal latent variable representation and the associated covariance estimation problem because this perspective is well suited to the computational methods that we employ. On this view, the model may be thought of as a Gaussian factor model embedded inside a multivariate probit model. But our approach can equally well be construed as an individual level model; as Clinton *et al.* (2004a) observed, our probit model corresponds to the assumption that individual senators have a quadratic utility function with normal errors and occupy specific locations (or 'ideal points') in a multi-dimensional Euclidean space.

1.1.1. Multivariate probit model

Let $\mathbf{Y} = (\mathbf{y}_1 \ldots \mathbf{y}_n)^{\mathrm{T}}$, where each $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,p})^{\mathrm{T}}$ represents p binary observations on a single subject. The multivariate probit model (Chib and Greenberg, 1998; Ashford and Sowden, 1970) induces a probability distribution on y_i via an unobserved continuous quantity (utility) z_i, which is given a multivariate Gaussian distribution:

$$\begin{aligned} z_i &\sim N(\alpha, \boldsymbol{\Sigma}), \\ y_{ij}|z_{ij} &\equiv \begin{cases} 0 & \text{if } z_{ij} \leqslant 0, \\ 1 & \text{if } z_{ij} > 0. \end{cases} \end{aligned} \tag{1}$$

In this way each of the 2^p possible binary vectors is associated with an orthant in $\mathbb{R}^p$ and assigned probability according to the corresponding multivariate Gaussian cumulative distribution function. Marginally,

$$\begin{aligned} y_{ij} &\sim \text{Ber}(\rho_j), \\ \rho_j &= \Pr(z_{ij} > 0). \end{aligned} \tag{2}$$

Note that $\boldsymbol{\Sigma}$ is identified only up to its correlation structure, because scaling the latent utilities preserves the distribution of y_i as can be deduced from expression (1). Without loss of generality, we denote the mean of z_i by α, with the understanding that this may be a conditional predictor involving covariates (i.e. $\alpha_i \equiv \alpha(x_i)$).

The multivariate probit model reduces the problem of estimating 2^p probabilities to the problem of estimating the $p(p-1)/2$ pairwise correlations which compose $\boldsymbol{\Sigma}$. The price of this reduction is the normality assumption on the latent utilities, which implies (among other things) a linear dependence structure. For many applications these assumptions are unobjectionable, and indeed the multivariate probit model is widely used (see, for example, the examples that were discussed in Lesaffre and Molenberghs (1991)).

1.1.2. Gaussian factor models

Whereas inference in a multivariate probit model is reduced to estimation of a correlation matrix, this task presents challenges of its own. Standard estimators are liable to be unstable when p is large compared with n and can provide a distorted picture of the eigenstructure of $\boldsymbol{\Sigma}$ (Sun and Berger, 2006). These difficulties are magnified in the multivariate probit model, where the covariance matrix corresponds to an unobserved random variable.

A popular approach to address this instability is by imposing a factor structure, letting

$$\text{cov}(z_i) = \mathbf{B}\mathbf{B}^{\mathrm{T}} + \boldsymbol{\Psi}, \tag{3}$$

where $\boldsymbol{\Psi}$ is $p \times p$ diagonal with non-negative elements and $\text{rank}(\mathbf{B}) = k < p$. We may rewrite this model by augmenting the representation to include latent factor scores f_i. Conditional on $\mathbf{B}$ and f_i, the elements of z_i are independent:

$$\begin{aligned} z_i = \alpha + \mathbf{B} f_i + \varepsilon_i, \qquad & \varepsilon_i \sim N(0, \boldsymbol{\Psi}), \\ f_i \sim N(0, \mathbf{I}_k). & \end{aligned} \tag{4}$$

For $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^{\mathrm{T}} + \boldsymbol{\Psi}$ to have a unique solution in $\mathbf{B}$, $\mathbf{B}$ must be constrained in some manner. In particular, two sorts of unidentifiability must be addressed: sign indeterminacy and rotational indeterminacy. Traditional solutions to this problem include forcing $\mathbf{B}$ to have mutually orthogonal columns or $\mathbf{B}^{\mathrm{T}}\boldsymbol{\Psi}\mathbf{B}$ to be diagonal. Another approach, which was popularized in the Bayesian community by Geweke and Zhou (1996) and adopted here, is to constrain $\mathbf{B}$ to be zero for upper triangular entries $\{b_{js} = 0 : s > j\}$ for $s \leqslant k$ (addressing rotational indeterminacy) and positive along the diagonal $\{b_{jj} > 0\}$ for all j (addressing sign indeterminacy).

It is additionally possible to permit correlated factors so that $f_i \sim N(0, \mathbf{V})$. In this case the covariance is expressible as $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{V}\mathbf{B}^{\mathrm{T}} + \boldsymbol{\Psi}$ and identification of $\mathbf{B}$ and $\mathbf{V}$ can be achieved by setting certain $k(k-1)/2$ additional elements of $\mathbf{B}$ to 0 (Bartholomew, 1987). Correlated factors are often preferable from the standpoint of interpretability. We revisit this topic in Section 2, where correlated factors combine with sparsity priors to provide valuable *a priori* biases towards easily interpretable factor rotations.

Factor models have been a topic of research for more than 100 years. A seminal reference is Spearman (1904), whereas Press (1982) and Bartholomew (1987) are key contemporary references. Bayesian factor models for continuous data have been developed by many researchers,

including Geweke and Zhou (1996), Aguilar and West (2000) and West (2003). A comprehensive bibliography may be found in Lopes (2003).

1.1.3. Latent factor probit models

Putting these pieces together—by assuming that the covariance parameter in a multivariate probit model admits a factor decomposition as in equation (3)—yields a latent factor probit model. This is equivalent to the individual level model

$$\Pr(y_{ij} = 1 | \alpha_j, b_j, f_i) = \Phi\{(\alpha_j + b_j^{\mathrm{T}} f_i)/\sqrt{\psi_j}\}, \tag{5}$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function and b_j denotes the jth row of $\mathbf{B}$.

In this context, the factor decomposition allows a simple identification strategy for the correlation structure, which is to let $\mathbf{\Psi} = \mathbf{I}$. Priors on element b_{js} of $\mathbf{B}$ then induce a prior distribution on the correlation coefficient ρ_{js} for $j \neq s$ (Rossi *et al.*, 2006); the magnitudes of these elements describe the amount of the variation attributable to the factor structure as opposed to idiosyncratic noise. Thus, the scale of the prior distribution governs prior expectations about the strength of the factor structure in terms of describing observed patterns of covariation.

1.2. Model for roll-call votes

Political scientists have long sought to understand the historical forces leading to the entrenched partisan rancour of modern American politics. Untangling the relative contributions of various polarizing factors is the subject of a vigorous scholarly debate, one that is too vast to recapitulate here. A recent book length treatment and a long list of references may be found in McCarty *et al.* (2006). An important first step, which is the present focus, is simply to measure (rather than to explain) ideological polarization.

For example, if we know that the majority whip votes against a particular bill, then we may believe that most other members of the majority will vote against it, also. The goal then becomes to quantify the strength of this association by estimating the amount of variation in observed voting records that can be accounted for by a so-called 'partisanship factor'. For this, we analyse publicly available US Congressional roll-call data, restricting our attention to votes in the US Senate between 1949 and 2009.

In our analysis we let $y_i = (y_{i,1}, \ldots, y_{i,p})^{\mathrm{T}}$ denote the vector of yea (1) or nay (0) votes of senator i. Our main data set contains the 30 closest votes in each 2-year Senate term ($p = 30$). The close votes are typically the most interesting and also allow us to sidestep the many near-unanimous votes which tend to be wholly unrelated to major policy decisions (Jackman, 2001). Missing data in the form of 'no' votes are easily handled in our framework, as will be described later.

Often party affiliation is not included specifically in a factor analysis of roll-call votes; rather it emerges implicitly in that senators from different parties tend to cluster in factor space (Jackman, 2001; Clinton *et al.*, 2004a). Our approach is to include this information explicitly by forcing the first factor f_{1i} to be left or right truncated depending on senator i's party affiliation. This ensures that the first factor is 'pure' partisanship. Note that this approach is equivalent to a factor model with $p + 1$ observed dimensions, the first of which is a pseudovote recording each senator's party affiliation, where the residual variance on that element is set to 0 ($\psi_1 = 0$); then $z_{1i} = f_{1i}$ up to a scale factor. To distinguish this privileged role, we denote the partisanship factor for senator i by γ_i and the remaining latent factors by

f_i. Similarly, we relabel the factor loadings column that is associated with γ as $\boldsymbol{\lambda}$, so that

$$E(z_i|f_i,\gamma_i,\alpha,\lambda,\mathbf{B}) = \alpha + \lambda\gamma_i + \mathbf{B}f_i,$$

and we define the dummy vote R_i recording whether senator i is a Republican as $R_i = \mathbb{1}(\gamma_i > 0)$.

Under this parameterization, large positive entries in λ can be interpreted as a constellation of Republican-supported issues whereas large negative loadings on λ correspond to Democrat-supported issues. Likewise, the inferred γs for each senator can be interpreted as individual measures of partisanship: a large positive γ_i indicates a tendency for senator i to vote for Democrat-supported issues with high probability. Meanwhile **B** encodes commonalities in voting behaviour that are independent of party membership. These patterns can be used to generate hypotheses about why senators vote the way that they do, irrespective of party affiliation.

2. Sparse factor probit model

A sparse model is one in which certain of the parameters are permitted to be exactly 0. The canonical example is a linear model in which subsets of the regression coefficients may be estimated to be 0. The sparsity framework spans the areas of regularized prediction, hypothesis testing and model selection, depending on whether it is viewed as a means, an end or both.

Sparsity can be achieved in a variety of ways, such as direct testing or thresholding. Here we take an implicit testing approach via sparsity priors which affix a point mass probability at zero (George and McCulloch, 1993; Mitchell and Beauchamp, 1988). A detailed discussion of this and similar Bayesian approaches to model selection in linear regression can be found in Ishwaran and Rao (2005). Our development closely follows West (2003) and Carvalho *et al.* (2008), who developed sparse factor models for continuous data in the context of gene expression studies. These models assume that each latent factor will be associated with only a small number of observed variables, yielding a more parsimonious covariance structure. Specifically, the prior on the loadings matrix **B** takes the form

$$(b_{js}|\nu_s, q_s) \sim q_s N(0,\nu_s) + (1-q_s)\delta_0(b_{js}), \tag{6}$$

$$\nu_s \sim \mathrm{IG}(c_s/2, c_s d_s/2), \tag{7}$$

$$q_s \sim \mathrm{beta}(1,1) \tag{8}$$

where there is a different variance component v_s and prior inclusion probability q_s associated with each column ($s = 1,\ldots,k$) of the loadings matrix. Here IG refers to an inverse gamma distribution. The second term in the mixture, $\delta_0(b_{js})$, denotes a point mass measure at $b_{js} = 0$. The hyperparameters c_s and d_s modulate prior expectations about the scale of each column.

Although q_s is a prior probability that is shared by elements in the sth column of **B**, this common prior parameter does not result in shared posterior inclusion probabilities across elements in a given column (excepting the uninteresting extremal cases of $q_s = 0$ or $q_s = 1$, which preclude posterior learning of sparsity patterns altogether). Even with a shared prior inclusion probability, different elements of the sth column of **B** will emerge with distinct posterior probabilities. By treating the prior inclusion probabilities as model parameters to be estimated from the data, this model induces a strong multiplicity correction, automatically adjusting for the

multiple-testing problem that is implicit in trying to learn the non-zero entries in each column of $\mathbf{B}$ (Scott and Berger, 2006, 2010). Analogous priors are placed on the elements of λ (which is simply a renamed column of $\mathbf{B}$).

Note also that, in the context of sparsity, allowing correlated factors confers an additional advantage beyond possible ease of interpretation, which is that a greater range of covariance structures admit sparse representations in $\mathbf{B}$ if $\mathrm{cov}(f_i) = \mathbf{V}$ is free to differ from the identity. To see this, observe that under the lower triangular identification scheme we may move freely between independent factors and dependent factors via the transformation

$$\tilde{\mathbf{B}} = \mathbf{BL} \tag{9}$$

where $\mathbf{V} = \mathbf{LL}^{\mathrm{T}}$ denotes the (lower) Cholesky decomposition of the factor covariance. Both $\tilde{\mathbf{B}}$ and $\mathbf{B}$ are lower triangular, because the group of lower triangular matrices is closed under multiplication, and the two models are equivalent because

$$\mathbf{BVB}^{\mathrm{T}} = \tilde{\mathbf{B}}\tilde{\mathbf{B}}^{\mathrm{T}}. \tag{10}$$

But, in general, $\tilde{\mathbf{B}}$ will be much less sparse, i.e., if $\tilde{\mathbf{B}}$ is sparse, setting $\mathbf{V} = \mathbf{I}$ gives a sparse $\mathbf{B}$, whereas a sparse $\mathbf{B}$ does not similarly yield a sparse $\tilde{\mathbf{B}}$. The columns of $\tilde{\mathbf{B}}$ are linear combinations of those of $\mathbf{B}$ and sums of sparse vectors need not be sparse.

Thus, sparsity priors on $\mathbf{B}$, combined with a prior on $\mathbf{V}$ centred at the identity matrix, serve to capture the inherent trade-off between parsimonious factor loadings and independent factors, automatically shrinking estimates towards more interpretable configurations.

As suggested in the previous section, we chose to let the partisanship factor be independent of all other factors, while allowing those additional factors to be correlated:

$$\begin{aligned} \begin{pmatrix} \gamma_i \\ f_i \end{pmatrix} &\sim N(0, \tilde{\mathbf{V}}), \\ \tilde{\mathbf{V}} &= \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{V} \end{pmatrix}, \end{aligned} \tag{11}$$

where $\mathbf{V}$ is an arbitrary correlation matrix. Though this approach may result in a potentially non-sparse partisanship loading λ, it has the interpretive advantage that a zero element of λ means that partisanship is wholly non-predictive of the corresponding roll-call vote. Meanwhile, allowing the remaining partisanship-independent factors to be correlated fosters sparser loadings, as described above.

We may think of the resulting inferences as coming from a mixture over different candidate 'confirmatory' models, where each model is weighted by its ability to describe the observed data (Hoeting *et al.*, 1999). Crucially, by fixing certain elements of $\mathbf{B}$ to 0 for identification purposes, we guarantee that all the models being mixed over observe these same constraints. This means that, although different patterns of sparsity may emerge during posterior sampling, the interpretation of the factors within each of these models (as defined by the identification constraints) remains fixed.

2.1. Synopsis: the sparse factor model

For ease of reference we write down the full model here, including prior distributions and hyperparameter specifications. The model can be thought of as having three parts: the data $\mathbf{Y} = (y_1, y_2, \ldots, y_n)$ given the latent utilities $\mathbf{Z} = (z_1, z_2, \ldots, z_n)$, the latent factors $\mathbf{F} = (f_1, f_2, \ldots, f_n)$ and $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n)$ given the model parameters and hyperparameters $\{\alpha = (\alpha_1, \ldots, \alpha_p)^{\mathrm{T}},$

$\mathbf{B}, \mathbf{V}, \boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)^{\mathrm{T}}$, $\mathbf{q} = (q_1, \ldots, q_k)$, $\nu = (\nu_1, \ldots, \nu_k)\}$, and finally the prior over the model parameters. Altogether, for $j = 1, \ldots, p$, $i = 1, \ldots, n$ and $s = 1, \ldots, k$, we have

$$
\begin{aligned}
R_i &= \mathbb{1}(\gamma_i > 0), \\
y_i | z_i &= \mathbb{1}(z_i > 0), \\
z_i &\sim N(\alpha + \lambda\gamma_i + \mathbf{B} f_i, \mathbf{I}_p), \\
f_i &\sim N(0, \mathbf{V}), \\
\gamma_i &\sim N(a, 1), \\
\alpha_j &\sim N(0, \nu_\alpha), \\
a &\sim N(0, \nu_a), \\
\mathbf{V} &\sim \mathrm{IW}(\mathbf{V}_0, h), \\
(b_{js} | \nu_s, q_s) &\sim q_s N(0, \nu_s) + (1 - q_s)\delta_0(b_{js}), \\
\nu_s &\sim \mathrm{IG}(c_s, c_s d_s/2), \\
q_s &\sim \mathrm{beta}(1, 1), \\
(\lambda_j | \nu_\lambda, q_\lambda) &\sim q_\lambda N(0, \nu_\lambda) + (1 - q_\lambda)\delta_0(\lambda_j), \\
\nu_\lambda &\sim \mathrm{IG}(c_s, c_s d_s/2), \\
q_\lambda &\sim \mathrm{beta}(1, 1).
\end{aligned}
$$

Here $\mathrm{IW}(\mathbf{V}_0, h)$ denotes an inverse Wishart distribution with mode matrix $\mathbf{V}_0$ and degrees of freedom h. Note that $\Phi(a)$ can be interpreted as the latent fraction of Republicans among hypothetical senators.

Recall that this model has structural 0s in $\mathbf{B}$ for identification purposes. For example, in a three-factor model ($k = 3$), the identification structure could be achieved by setting elements b_{41}, b_{52} and b_{61} to 0, allowing $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{V}\mathbf{B}^{\mathrm{T}} + \mathbf{I}$ to be solved uniquely for $\mathbf{B}$ and $\mathbf{V}$, given $\boldsymbol{\Sigma}$. Having 1s on the diagonal permits $\mathbf{V}$ to be a covariance matrix, rather than being restricted to a correlation matrix. One should note that fixing this structure beforehand is convenient notationally but renders the ordering of the votes a non-trivial modelling decision.

Our analysis of the roll-call data (which is summarized in Section 3) was performed with $\nu_\alpha = \nu_a = 10$, $h = 6$, $\mathbf{V}_0 = \mathbf{I}_k$ and $c_s = 2$ and $d_s = 1$ for all s. We used a three-factor model (the partisanship factor γ, plus $k = 2$ additional factors). Larger models are possible, but in practice we did not encounter a case where a fourth factor exhibited significant loadings.

2.2. Posterior sampling

We employ a Gibbs sampler to draw correlated samples from the joint posterior distribution of all parameters (Gelfand and Smith, 1990; Geman and Geman, 1984). In what follows we describe how to sample from each of the full conditional distributions.

Step 1: draw the latent observation matrix $\mathbf{Z} = (z_{ij})$ by drawing each $(z_{ij}|-) \sim N(\alpha_j + \lambda_j\gamma_i + b_j f_i, 1)$ truncated above at 0 if $y_{ij} = 0$ and below at 0 if $y_{ij} = 1$.
Step 2: sample γ_i as $(\gamma_i|-) \sim N[(1 + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\lambda})^{-1}\{\nu_a^{-1}a + \boldsymbol{\lambda}^{\mathrm{T}}(z_i - \alpha - \mathbf{B}f_i)\}, (1 + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\lambda})^{-1}]$, truncated above at 0 if senator i is Republican and below at 0 if senator i is Democrat. Handling Independents is as simple as treating the sign of their latent partisanship as missing data and simply bypassing the truncation.
Step 3: sample a as $(a|-) \sim N\{(\nu_a^{-1} + n)^{-1}\Sigma_i\gamma_i, (\nu_a^{-1} + n)^{-1}\}$.
Step 4: sample the elements of $\boldsymbol{\alpha}$ as $(\alpha_j|-) \sim N(m_j, M)$ with

$$\left.\begin{aligned} \tilde{z}_i &= \boldsymbol{\Psi}^{1/2}(z_i - \lambda\gamma_i - \mathbf{B}f_i), \\ M &= (\nu_\alpha^{-1} + n)^{-1}, \\ m_j &= M\sum_i \tilde{z}_{ji}. \end{aligned}\right\} \tag{12}$$

Step 5: sample the vectors of factor scores independently as $(f_i|-) \sim N(m, \mathbf{M})$, with

$$\begin{aligned} \mathbf{M} &= \mathbf{V} - \mathbf{V}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}(\mathbf{B}\mathbf{V}\mathbf{B}^{\mathrm{T}} + \mathbf{I}_p)^{-1}\mathbf{B}\mathbf{V}, \\ &= (\mathbf{V}^{-1} + \mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}, \\ m &= \mathbf{M}\mathbf{B}^{\mathrm{T}}(z_i - \alpha - \lambda\gamma_i). \end{aligned}$$

Step 6: sample $(\mathbf{V}|-) \sim \mathrm{IW}(\mathbf{V}_0 + \mathbf{F}\mathbf{F}^{\mathrm{T}}, n + h)$.
Step 7: to sample the unconstrained loadings of column s, define $\tilde{z}_i = z_i - \alpha - \lambda\gamma_i - \mathbf{B}_{-s}f_{-s,i}$ where $\mathbf{B}_{-s}$ is $\mathbf{B}$ with column s removed and $f_{-s,i}$ is the vector of factor scores for individual i with factor score s removed. Sample

$$(b_{js}|-) \sim (1 - \hat{q}_{js})\delta_0 + \hat{q}_{js}N(\hat{b}_{js}, \hat{\nu}_{js}),$$

where

$$\begin{aligned} \hat{\nu}_{js} &= \left(\sum_{i=1}^{n} f_{s,i}^2 + \nu_s^{-1}\right)^{-1}, \\ \hat{b}_{js} &= \hat{\nu}_{js}\left(\sum_{i=1}^{n} f_{s,i}\tilde{z}_{ij}\right), \\ \hat{r}_{js} &= \frac{N(0|0, \nu_j)}{N(0|\hat{b}_{js}, \hat{\nu}_{js})}, \\ \hat{q}_{js} &= \frac{\hat{r}_{js}}{(1 - q_s)/q_s + \hat{r}_{js}}. \end{aligned}$$

Step 8: let m_s be the number of (unconstrained) elements in $\mathbf{B}_s$ currently set to non-zero. Draw

$$\nu_s \sim \mathrm{IG}\{(c_s + m_s)/2, (c_s d_s + \mathbf{B}_s^{\mathrm{T}}\mathbf{B}_s)/2\}.$$

Step 9: draw $(q_s|-) \sim \mathrm{Be}(1 + m_s, 1 + \tilde{m}_s - m_s)$, where m_s is as in the previous step and $\tilde{m}_s$ is the maximum possible number of non-zero elements for column s (given the fixed identification constraints).
Step 10: the elements of $\boldsymbol{\lambda}$ are drawn analogously to those of the B_s: define $\tilde{z}_i = z_i - \alpha - \mathbf{B}f_i$ and samples

$$(\lambda_j|-) \sim (1 - \hat{q}_{\lambda,j})\delta_0 + \hat{q}_{\lambda,j}N(\hat{\lambda}_j, \hat{\nu}_\lambda),$$

where

$$\begin{aligned} \hat{\nu}_{\lambda,j} &= \left(\sum_{i=1}^{n} \gamma_i^2 + \nu_\lambda^{-1}\right)^{-1}, \\ \hat{\lambda}_j &= \hat{\nu}_{\lambda,j}\left(\sum_{i=1}^{n} \gamma_i\tilde{z}_{ij}\right), \\ \hat{r}_{\lambda,j} &= \frac{N(0|0, \nu_\lambda)}{N(0|\hat{\lambda}_j, \hat{\nu}_\lambda)}, \\ \hat{q}_{\lambda,j} &= \frac{\hat{r}_{\lambda,j}}{(1 - q_\lambda)/q_\lambda + \hat{r}_{\lambda,j}}. \end{aligned}$$

Step 11: likewise, let m_λ be the number of elements in λ that are currently set to non-zero, and draw

$$(\nu_\lambda|-) \sim \mathrm{IG}\{(c_\lambda + m_\lambda)/2, (c_\lambda d_\lambda + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\lambda})/2\}.$$

Step 12: again, similarly, draw $(q_\lambda|-) \sim \mathrm{Be}(1 + m_\lambda, 1 + p - m_\lambda)$.

In this sampler, as highlighted in Song and Lee (2005) in the context of an analogous EM algorithm, it is not necessary to draw from a high dimensional truncated multivariate normal distribution; all the dependence between the elements of z_i is encoded in $\mathbf{B}$ so the truncations arising from the observed data y_{ij} can be handled independently, leading to a series of easier univariate truncations.

As with independent senators, missing data (uncast votes, in this case) may be accommodated by simply drawing the corresponding latent utilities without truncation in the first step of the sampler, under the assumption of non-informative missingness. Informative missingness may also be incorporated by truncating with some predetermined probability, possibly as a function of additional covariates.

2.3. Simulation study

This section shows via simulation that sparse factor models result in a highly favourable bias–variance trade-off. We compare three models of the covariance structure: the Wishart model, a $k = 6$ factor model and a $k = 6$ sparse factor model. We examine the performance of each of these models under four distinct regimes:

(a) data drawn with underlying covariance matrix which has a factor structure with three factors;
(b) data drawn with underlying covariance matrix which has a factor structure with 10 factors;
(c) data drawn with underlying covariance matrix with no factor structure (equivalently, with the number of factors equal to the number of dimensions);
(d) data drawn with underlying covariance matrix given by the identity matrix.

Specifically, for a given covariance matrix $\boldsymbol{\Sigma}$ and mean vector $\boldsymbol{\alpha}$, we let

$$\mathbf{C} = \mathbf{D}^{-1/2}\boldsymbol{\Sigma}\mathbf{D}^{-1/2}, \tag{13}$$

$$\mathbf{D} = \mathrm{diag}(\boldsymbol{\Sigma}), \tag{14}$$

$$z_i \sim N(\alpha, \mathbf{C}), \tag{15}$$

$$y_i = \mathbb{1}(z_i > 0). \tag{16}$$

In all regimes $\boldsymbol{\alpha}$ was drawn as $N(0, 0.2\mathbf{I})$.

For all simulations the number of observations was fixed at $n = 50$. An estimated correlation matrix $\tilde{\mathbf{R}}$ was obtained for $p = 20$ and $p = 100$ and the mean Frobenius and Stein losses were computed over 100 replications. Frobenius and Stein losses are given respectively as

$$L_{\mathrm{F}}(\tilde{\mathbf{C}}, \mathbf{C}) = \mathrm{tr}\{(\tilde{\mathbf{C}} - \mathbf{C})^2\}, \tag{17}$$

$$L_{\mathrm{S}}(\tilde{\mathbf{C}}, \mathbf{C}) = \mathrm{tr}(\tilde{\mathbf{C}}\mathbf{C}^{-1}) - \ln\{\det(\tilde{\mathbf{C}}\mathbf{C}^{-1})\} - p. \tag{18}$$

The regimes that are examined here include cases where $p > n$ and also $n > p$, cases where the assumed factor structure has both too few and too many included factors, and both the factor models (sparse and non-sparse) and the Wishart model are centred at the identity matrix (since

Table 1. Mean Stein and Frobenius losses suffered in reconstructing the true correlation matrix **R** in various configurations

Loss function	*True model*	*Losses for the following fitted models:*		
		Wishart	*6 factor*	*Sparse 6 factor*
Stein	$p = 20, k = 3$	74.7	13.9	9.9
	$p = 20, k = 10$	91.0	24.0	29.7
	$p = 20, k = 20$	53.8	12.1	18.0
	$p = 20$, identity	3.6	2.9	0.4
Frobenius	$p = 20, k = 3$	89.6	14.6	12.9
	$p = 20, k = 10$	40.3	12.3	14.0
	$p = 20, k = 20$	37.6	14.6	13.0
	$p = 20$, identity	8.1	6.7	0.89
Stein	$p = 100, k = 3$	503.1	136.7	43.4
	$p = 100, k = 10$	1323.2	357.4	394.2
	$p = 100, k = 100$	827.8	454.2	667.3
	$p = 100$, identity	28.3	26.2	1.0
Frobenius	$p = 100, k = 3$	2573.8	430.5	234.0
	$p = 100, k = 10$	1143.1	403.8	410.0
	$p = 100, k = 100$	305.7	275.7	160.9
	$p = 100$, identity	94.6	136.3	2.1

$E(\mathbf{B}) = 0$). As such, this battery provides a good snapshot of the performance of the three models across a range of plausible real data scenarios. Results are reported in Table 1.

The differences between the various models when $n > p$ are modest, but the factor model is seen to dominate the Wishart model. Meanwhile, the difference between sparse *versus* non-sparse factor models can be attributed to which of these models is closest *a priori* to the generating model—so the sparse model performs better for the identity and for the $k = 3$ true models, whereas the non-sparse model does better for the $k = 10$ and $k = 20$ generated data. However, with $p = 100$ and $n = 50$ the slight penalty that the sparse model incurs for underestimating the number of factors is shown to be relatively minor. In this setting, the benefit over the Wishart distribution becomes more stark. Naturally, whichever model favours the truth (*a priori*) still comes out on top. For instance, the sparse model on the identity matrix gives outstanding performance.

That said, the $p = 100$, $k = 10$ (second row), results are most interesting: since six factors are insufficient to reconstruct $\mathbf{\Sigma}$ perfectly in this case, it is striking that the factor model still outperforms the Wishart model. Furthermore, incorporating the sparsity component does not suffer much at all in this case, whereas we can see that, when the true number of factors is less than 6, adding the sparsity offers a substantial benefit (first row). In short, it would appear that the bias that is induced by the factor structure assumption is more than compensated by the reduced variance when $p > n$.

3. Analysis of US Senate roll-call votes, 1949–2009

3.1. Overview of approach

Our analysis of US Senate roll-call votes focuses on three different, complementary methods for characterizing partisan voting behaviour. First, we can rank individual senators by partisanship, along the lines of Clinton *et al.* (2004b). This is done by examining the posterior distribution of each senator's partisanship factor score, $p(\gamma_i|\mathbf{Y})$. We observe that there are typically large

differences between senators in both the mean of this distribution as well as the associated uncertainty (see, for example, Fig. 1).

Second, we can ascertain which roll-calls—if any—were plausibly free of party influence. We do this by examining the posterior probabilities that each roll-call loads on each of the $k+1$ factors.

Third, we can study historical changes in the importance of partisanship for predicting roll-call votes for the Senate as a whole (rather than for individual senators, as above). This can be operationalized in two ways for each Senate:

(a) as the percentage of variation in roll-call votes attributable to the partisanship factor and
(b) as the fraction of non-zero elements in the partisanship factor loadings vector $\boldsymbol{\lambda}$.

In principle, this permits two different notions of increasingly partisan behaviour to be detected: individual roll-calls becoming more severely partisan, *versus* more roll-calls overall becoming at least somewhat partisan.

3.2. Partisanship over time

The three characterizations of partisanship that were sketched above may be estimated for any Congress in aggregate by considering the posterior distribution of quantities involving all votes. Specifically, for each of the last 30 Congresses, spanning the 60-year period from 1949 to 2009, we consider the posterior distributions of the following quantities:

(a) the overall proportion of variation attributable to partisanship, defined as

$$p^{-1}\sum_j \frac{\lambda_j^2}{\lambda_j^2 + b_j \mathbf{V} b_j^{\mathrm{T}} + 1}; \tag{19}$$

(b) the overall probability of a non-zero partisanship loading, given by the parameter q_λ;
(c) the average dimensionality across the bills, defined as

$$p^{-1}\sum_{j,g} \mathbb{1}(b_{jg} \neq 0). \tag{20}$$

These summaries provide a snapshot of time varying partisanship. A key feature of the inferred partisanship trajectory that is shown in Fig. 2(a) is that Senate voting patterns of the 1960s and 1970s suggest forces at play beyond those captured by party affiliation; this much has long been recognized. Our analysis shows, however, that, as partisanship has fluctuated noticeably, the dimensionality of the voting landscape has remained relatively stable. Justifying this finding from a theoretical political science perspective would be an interesting line for future inquiry.

An important issue in high dimensional Bayesian analysis is prior sensitivity. Throughout, our focus has been on specifying a default, or conventional, set of hyperparameters in an attempt to minimize the influence of prior inputs. These choices follow closely along the lines of Carvalho *et al.* (2008). In many cases such default choices are natural—e.g. a uniform distribution for the prior inclusion probabilities q_s. It is important to observe that improper priors cannot be used for the variances ν_s for each column of the factor loadings matrix, since these variances enter an implicit Bayes factor computation in the sparse model. In a very real sense, no 'objective' choice for this prior is possible. In this case our use of an IG(1,1) prior recognizes the need for propriety, while simultaneously attempting to exhibit as little influence as possible over the final answers.

Space—and the sheer combinatorial explosion of hyperparameters—both preclude a full discussion of prior robustness, but we observe two important generalities. First, the prior for q_s

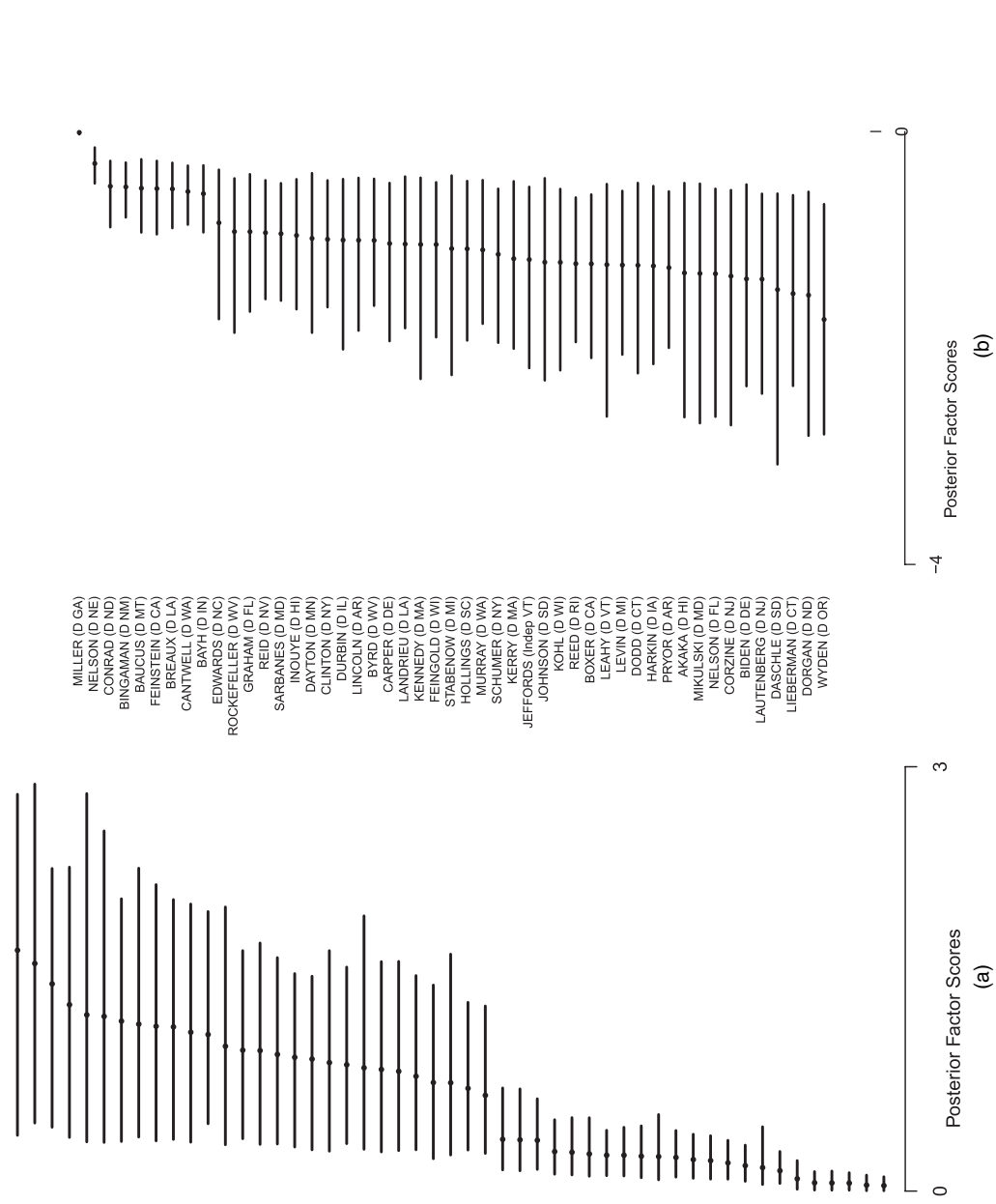

Fig. 1. Posterior mean partisanship scores for each senator of the 108th Congress, along with associated 95% credible intervals: (a) Republican partisanship scores; (b) Democrat partisanship scores

(the prior inclusion probabilities) makes the greatest amount of difference in posterior inference, which is consistent with the findings of Scott and Berger (2006) in the parallel context of multiple-hypothesis testing. In our experiments, we observed that changing the hyperparameters that govern q_s could predictably change the corresponding posterior inclusion probabilities, but very rarely change other measures of partisanship such as percentage variation explained. Given the existence of an obvious default (uniform) prior for q_s, this would seem untroubling. Second, results for later Congresses are less sensitive to the prior than results for earlier Con-

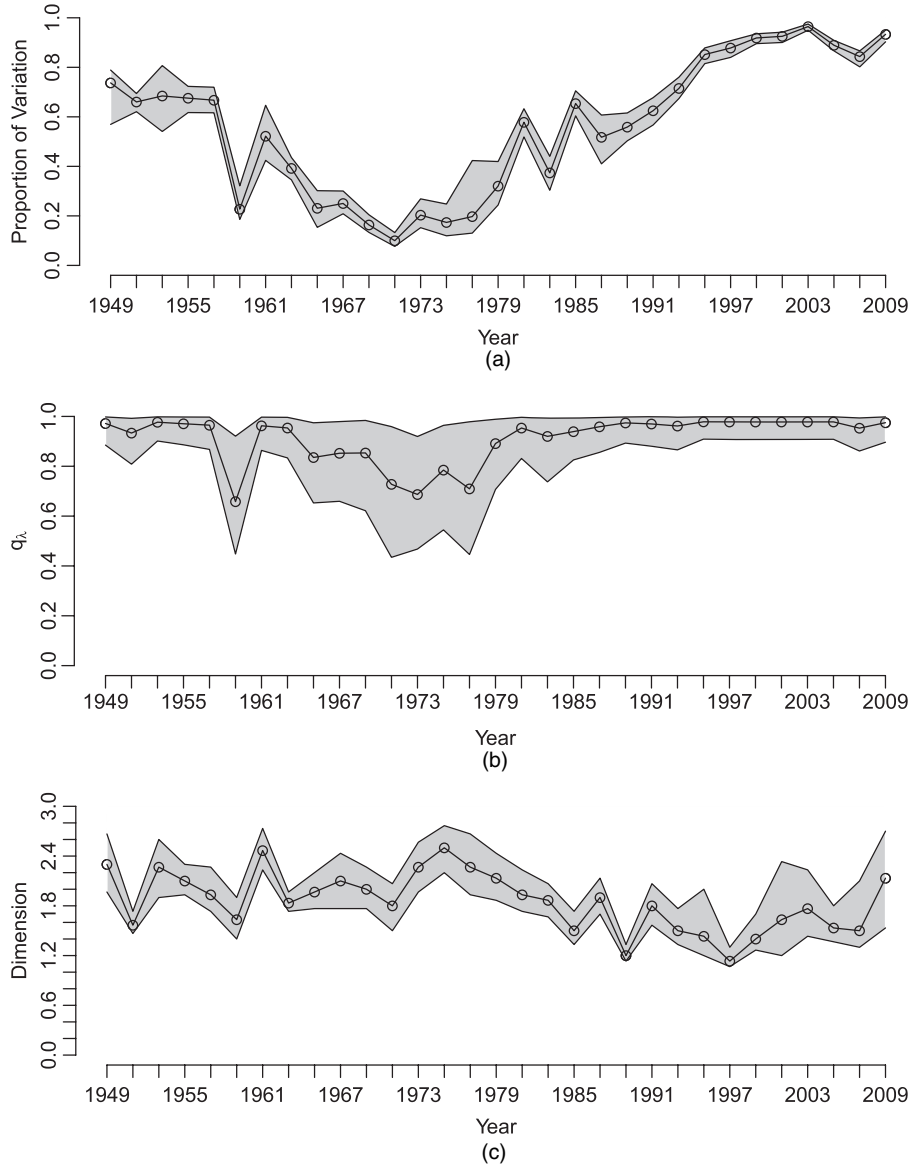

Fig. 2. Three measures of partisan behaviour in Senate roll-call votes, 1949–2009: (a) variation attributed to partisanship by year; (b) probability of non-zero partisanship loading; (c) average dimension of a bill

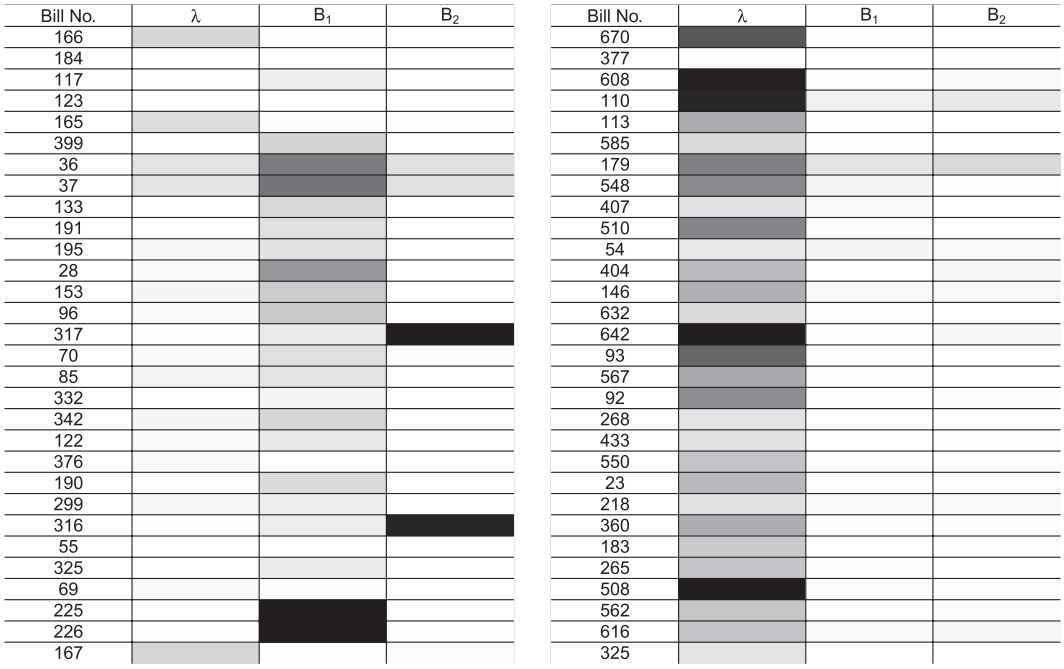

Fig. 3. Greyscale depictions of the relative magnitudes of posterior mean factor loadings: (a) the 86th Senate has two prominent factors unrelated to partisanship, whereas (b) the 111th Senate shows that party affiliation accounts for the majority of variation in the observed votes

gresses. This is fully consistent with the notion that the statistical signature of partisanship has become more pronounced.

3.3. Comparing the 86th and 111th Congresses

Within a given congress, individual senators and individual bills can be evaluated and ranked in terms of their inferred partisanship. In terms of ranking senators, we recapitulate the basic findings of Clinton *et al.* (2004a) for the 108th Congress. Our findings differ from theirs mainly in the larger posterior intervals, probably owing simply to the fewer number of votes that we considered, but also perhaps to the increased flexibility of our sparse model. We also include the most recent 111th Congress for comparison as an on-line supplementary file.

In terms of analysing the bills themselves, we compare inferences from the most recent 111th Senate and the 86th Senate from 1960–1962 (Fig. 3). As witnessed in the time series above, we expect to see a less fervent partisanship signature in the 1960s votes, and indeed we do. By looking more closely at the inferred factor loadings we can venture interpretations of the non-partisan columns on the basis of the emergent sparsity patterns. In particular we find that the first non-partisan factor loads heavily on two labour union bills and two bills concerning appropriations for teachers' salaries. These findings are reported in Table 2.

4. Discussion

The sparse factor probit model represents a principled approach to bridging the gap between exploratory and confirmatory factor analyses by using ideas from Bayesian model averaging.

Table 2. Posterior summaries for the 30 closest Senate votes in the 86th and 111th Congresses†

Results for 86th Senate				*Results for 111th Senate*			
Bill	*PV*	*PIP*	*MPD*	*Bill*	*PV*	*PIP*	*MPD*
166	0.95	1.00	1.37	670	1.00	1.00	1.58
184	0.02	0.32	0.39	377	0.12	0.86	1.42
117	0.01	0.23	1.33	608	1.00	1.00	2.13
123	0.02	0.32	0.60	110	0.98	1.00	2.67
165	0.92	1.00	1.46	113	0.98	1.00	1.94
399	0.01	0.26	1.32	585	0.95	1.00	1.93
36	0.34	1.00	2.92	179	0.95	1.00	2.73
37	0.34	1.00	2.93	548	0.98	1.00	2.43
133	0.01	0.25	1.31	407	0.91	1.00	2.16
191	0.01	0.24	1.34	510	0.98	1.00	2.35
195	0.36	1.00	2.13	54	0.82	1.00	2.30
28	0.07	0.87	1.96	404	0.96	1.00	2.41
153	0.27	1.00	2.08	146	0.97	1.00	2.40
96	0.10	0.83	1.93	632	0.95	1.00	1.99
317	0.01	0.52	1.93	642	0.99	1.00	2.45
70	0.26	0.86	2.11	93	0.99	1.00	2.11
85	0.49	1.00	2.07	567	0.99	1.00	2.05
332	0.05	0.44	1.42	92	0.99	1.00	2.11
342	0.31	1.00	1.26	268	0.92	1.00	1.95
122	0.23	0.97	2.03	433	0.93	1.00	2.00
376	0.38	1.00	1.26	550	0.97	1.00	2.00
190	0.06	0.58	1.67	23	0.98	1.00	2.13
299	0.46	1.00	2.03	218	0.84	1.00	2.40
316	0.01	0.49	1.87	360	0.98	1.00	2.07
55	0.00	0.15	0.25	183	0.97	1.00	1.91
325	0.01	0.21	1.30	265	0.97	1.00	2.14
69	0.38	0.98	1.51	508	0.99	1.00	2.16
225	0.01	0.39	1.51	562	0.97	1.00	2.02
226	0.02	0.39	1.57	616	0.96	1.00	2.27
167	0.94	1.00	1.44	325	0.93	1.00	1.77

†PV, proportion of variation attributed to partisanship; PIP, posterior inclusion probability; MPD, mean posterior dimension.

The result is a one-pass approach to uncovering covariance patterns with ready substantive interpretations, as our application to US Senate roll-call votes demonstrates.

Our simulations also demonstrate the beneficial regularizing properties of both the factor structure and the sparsity prior. Together, these allow the multivariate probit model to be effective even when the dimension p is quite large. Many other approaches to covariance estimation in this setting, such as banding or l^1-regularization, do not offer the interpretational benefits of our method; nor do they easily accommodate additional modelling structure—e.g. time series or spatial models.

As noted in Jackman (2001), extending Bayesian models of voting is relatively straightforward and our sparse model naturally inherits this property. The use of covariates in the linear predictor $\alpha(x_i)$ could easily be incorporated to sharpen the investigation of hypotheses that are suggested by an initial analysis. A most interesting extension of the method—in light of our approach to incorporating partisanship—would be to add an auto-correlation component, be it spatial or temporal, to the factor scores. This would account for senators serving in consec-

utive Congresses, or senators in nearby states, and would intrinsically handle the interesting (but relatively infrequent) case of senators switching party affiliation (Clinton *et al.*, 2004a). (In our framework the switched truncation in consecutive Congresses for such a senator, in combination with temporal auto-correlation, would pull those factor score estimates towards neutral 0.) This is just one example of how larger models could be constructed that would allow flexible borrowing of information across spatial and temporal dimensions, all within a factor analytic framework.

Taken together, these reasons suggest that the sparse factor probit model can be a key exploratory tool in the increasingly common situation of high dimensional, correlated categorical data.

Acknowledgements

The authors thank the Joint Editor, Associate Editor and two referees for their many helpful suggestions in improving the paper.

References

Aguilar, O. and West, M. (2000) Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *J. Bus. Econ. Statist.*, **18**, 338–357.
Ashford, J. and Sowden, R. (1970) Multivariate probit analysis. *Biometrics*, **26**, 535–546.
Bartholomew, D. (1987) *Latent Variable Models and Factor Analysis*. London: Griffin.
Bock, R. D. and Gibbons, R. D. (1996) High-dimensional multivariate probit analysis. *Biometrics*, **52**, 1183–1194.
Carvalho, C. M., Lucas, J., Wang, Q., Nevins, J. and West, M. (2008) High-dimensional sparse factor modelling: applications in gene expression genomics. *J. Am. Statist. Ass.*, **103**, 1438–1456.
Chib, S. and Greenberg, E. (1998) Analysis of multivariate probit models. *Biometrika*, **85**, 347–361.
Clinton, J., Jackman, S. and Rivers, D. (2004a) The statistical analysis of roll call data. *Am. Polit. Sci. Rev.*, **98**, 355–370.
Clinton, J., Jackman, S. and Rivers, D. (2004b) "The most liberal senator"?: analyzing and interpreting congressional roll calls. *Polit. Sci. Polit.*, **37**, 805–811.
Elrod, T. and Keane, M. P. (1995) A factor-analytic probit model for representing the market structure in panel data. *J. Marktng Res.*, **32**, 1–16.
Gelfand, A. E. and Smith, A. F. M. (1990) Sampling based approaches to calculating marginal densities. *J. Am. Statist. Ass.*, **85**, 398–409.
Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattn Anal. Mach. Intell.*, **6**, 721–741.
George, E. I. and McCulloch, R. E. (1993) Variable selection via Gibbs sampling. *J. Am. Statist. Ass.*, **88**, 881–889.
Geweke, J. and Zhou, G. (1996) Measuring the pricing error of the arbitrage pricing theory. *Rev. Finan. Stud.*, **9**, 557–587.
Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinksy, C. T. (1999) Bayesian model averaging: a tutorial. *Statist. Sci.*, **14**, 382–417.
Ishwaran, H. and Rao, J. S. (2005) Spike and slab variable selection: frequentist and Bayesian strategies. *Ann. Statist.*, **33**, 730–773.
Jackman, S. (2001) Multidimensional analysis of roll call data via Bayesian simulation: identification, estimation, inference, and model checking. *Polit. Anal.*, **9**, no. 3.
Jackman, S. (2009) *Bayesian Analysis for the Social Sciences*. New York: Wiley.
Johnson, V. and Albert, J. (1999) *Ordinal Data Modeling*. New York: Springer.
Lesaffre, E. and Molenberghs, G. (1991) Multivariate probit analysis: a neglected procedure in medical statistics. *Statist. Med.*, **10**, 1391–1403.
Lopes, H. (2003) Factor models: an annotated bibliography. *Bull. Int. Soc. Baysn Anal.*, Aug., 1–4.
McCarty, N., Poole, K. T. and Rosenthal, H. (2006) *Polarized America: the Dance of Ideology and Unequal Riches*. Cambridge: MIT Press.
Mitchell, T. and Beauchamp, J. (1988) Bayesian variable selection in linear regression (with discussion). *J. Am. Statist. Ass.*, **83**, 1023–1036.
Press, S. (1982) *Applied Multivariate Analysis: using Bayesian and Frequentist Methods of Inference*, 2nd edn. New York: Krieger.

Quinn, K. M. (2004) Bayesian factor analysis for mixed ordinal and continuous responses. *Polit. Anal.*, **12**, 338–353.
Rajaratnam, B., Massam, H. and Carvalho, C. M. (2008) Flexible covariance estimation in graphical Gaussian models. *Ann. Statist.*, **36**, 2818–2849.
Rossi, P. E., Allenby, G. M. and McCulloch, R. (2006) *Bayesian Statistics and Marketing*. New York: Wiley.
Scott, J. G. and Berger, J. O. (2006) An exploration of aspects of Bayesian multiple testing. *J. Statist. Planng Inf.*, **136**, 2144–2162.
Scott, J. G. and Berger, J. O. (2010) Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *Ann. Statist.*, **38**, 2587–2619.
Song, X. and Lee, S. (2001) Bayesian estimation and test for factor analysis model with continuous and polytomous data in several populations. *Br. J. Math. Statist. Psychol.*, **54**, 237–263.
Song, X.-Y. and Lee, S.-Y. (2005) A multivariate probit latent variable model for analyzing dichotomous responses. *Statist. Sin.*, **15**, 645–664.
Spearman, C. (1904) General intelligence, objectively determined and measured. *Am. J. Psychol.*, **15**, 201–293.
Sun, D. and Berger, J. O. (2006) Objective priors for the multivariate normal model. In *Bayesian Statistics 8* (eds J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West). Oxford: Oxford University Press.
West, M. (2003) Bayesian factor regression models in the "large p, small n" paradigm. In *Bayesian Statistics 7* (eds J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West), pp. 723–732. Oxford: Oxford University Press.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Supporting Information for "A sparse factor-analytic probit model for Congressional voting patterns"'.

Please note: Wiley–Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the author for correspondence for the article.