Biometrika (2009), **96**, 4, *pp.* 821–834
© 2009 Biometrika Trust
Printed in Great Britain

doi: 10.1093/biomet/asp049
Advance Access publication 9 October 2009

Bayesian analysis of matrix normal graphical models

By HAO WANG and MIKE WEST

Department of Statistical Science, Duke University, Durham, North Carolina 27708, U.S.A.
hao@stat.duke.edu mike@stat.duke.edu

Summary

We present Bayesian analyses of matrix-variate normal data with conditional independencies induced by graphical model structuring of the characterizing covariance matrix parameters. This framework of matrix normal graphical models includes prior specifications, posterior computation using Markov chain Monte Carlo methods, evaluation of graphical model uncertainty and model structure search. Extensions to matrix-variate time series embed matrix normal graphs in dynamic models. Examples highlight questions of graphical model uncertainty, search and comparison in matrix data contexts. These models may be applied in a number of areas of multivariate analysis, time series and also spatial modelling.

Some key words: Gaussian graphical model; Graphical model search; Hyper-inverse Wishart distribution; Marginal likelihood; Matrix normal model; Matrix-variate dynamic graphical model; Parameter expansion.

1. Introduction

We introduce and analyze matrix normal graphical models; that is, matrix normal distributions (Dawid, 1981; Gupta & Nagar, 2000) in which each of the two characterizing covariance matrices reflects conditional independencies consistent with an underlying graphical model (Whittaker, 1990; Lauritzen, 1996). We present fully Bayesian analysis of the matrix normal model as the special case of full graphs, and develop computational methods for marginal likelihood computation on a specified graphical model. This enables graphical model search and comparison for posterior inferences about conditional independence structures. The random sampling framework is extended to matrix-variate time series models that inherit the graphical model structure to represent conditional independencies in matrix time series. We focus on decomposable graphs although the general approach will also apply to nondecomposable models.

Matrix-variate normal distributions have been studied in analysis of two-factor linear models for cross-classified multivariate data (Finn, 1974; Galecki, 1994; Naik & Rao, 2001), in spatio-temporal models (Mardia & Goodall, 1993; Huizenga et al., 2002) and other areas. Some computational and inferential developments, including iterative calculation of maximum likelihood estimates (Dutilleul, 1999; Mitchell et al., 2006) and empirical Bayesian methods for Procrustes analysis with matrix models (Theobald & Wuttke, 2006) have been published. Our work appears to be the first to develop fully Bayesian analysis of the basic matrix normal model alone, though that is only a necessary first step to the broader framework of matrix graphical models.

In time series, graphical modelling of the covariance matrix of multivariate data appears in Carvalho & West (2007a, 2007b). Here we generalize that earlier work to time series of matrix data, providing fully Bayesian inference and graphical model search related to both row and column intra-dependencies in the cross-sectional structure of a matrix-valued time series.

2. Matrix variate normals, graphs and notation

The $q \times p$ random matrix Y is matrix normal, $Y \sim N(M, U, V)$, with mean M $(q \times p)$, column and row covariance matrices $U = (u_{ij})$ $(q \times q)$ and $V = (v_{ij})$ $(p \times p)$, respectively, when

$$p(Y) \equiv p(Y \mid U, V) = k(U, V) \exp[-\mathrm{tr}\{(\mathrm{Y}-\mathrm{M})'\mathrm{U}^{-1}(\mathrm{Y}-\mathrm{M})\mathrm{V}^{-1}/2\}], \tag{1}$$

where $k(U, V) = (2\pi)^{-qp/2}|U|^{-p/2}|V|^{-q/2}$. The rows $y_{i\star}$ $(i = 1, \ldots, p)$, and columns $y_{\star j}$ $(j = 1, \ldots, q)$, have margins $y_{i\star} \sim N(m_{i\star}, u_{ii}V)$ and $y_{\star j} \sim N(m_{\star j}, v_{jj}U)$ with precision matrices $\Lambda = V^{-1} = (\lambda_{ij})$ and $\Omega = U^{-1} = (\omega_{ij})$, respectively. The normal conditional distributions have mean vectors and covariance matrices

$$E(y_{i\star} \mid y_{-i\star}) = m_{i\star} - \omega_{ii}^{-1} \textstyle\sum_{s \in (1,\ldots,q \setminus i)} \omega_{is}(y_{s\star} - m_{s\star}), \quad \mathrm{cov}(\mathrm{y}_{i\star} \mid \mathrm{y}_{-i\star}) = \omega_{ii}^{-1} V,$$

$$E(y_{\star j} \mid y_{-\star j}) = m_{\star j} - \lambda_{jj}^{-1} \textstyle\sum_{t \in (1,\ldots,p \setminus j)} \lambda_{tj}(y_{\star t} - m_{\star t}), \quad \mathrm{cov}(\mathrm{y}_{\star j} \mid \mathrm{y}_{-\star j}) = \lambda_{jj}^{-1} U,$$

for rows $i = 1, \ldots, q$ and columns $j = 1, \ldots, p$. Zeros in Λ and Ω define conditional independencies. If $(i, j) \neq (s, t)$ then y_{ij} and y_{st} may, conditional upon $y_{-(ij,st)}$ be dependent through either rows or columns; conditional independence is equivalent to: at least one zero among λ_{tj} and ω_{is} when $s \neq i$, $j \neq t$; $\omega_{is} = 0$ when $s \neq i$, $j = t$; $\lambda_{jt} = 0$ when $s = i$, $j \neq t$. With no loss of generality, in this section we set $M = 0$.

Undirected graphical models can be applied to each of Λ and Ω to represent strict conditional independencies. A graph G_V on nodes $\{1, \ldots, p\}$ has edges between pairs of column indices (j, t) for which $\lambda_{jt} \neq 0$; Λ has off-diagonal zeros corresponding to within-row conditional independencies. Similarly, a graph G_U on nodes $\{1, \ldots, q\}$ lacks edges between row indices (i, s) for which $\omega_{is} = 0$. We focus here on decomposable graphs G_U and G_V. The theory of graphical models can be now overlaid to define conditional factorizations of the matrix normal density over graphs. Over G_V, for example, we have

$$p(Y \mid U, V, G_V, G_U) = \prod_{P_V \in \mathcal{P}_V} p(Y_{\star P_V} \mid U, V_{P_V}) \Big/ \prod_{S_V \in \mathcal{S}_V} p(Y_{\star S_V} \mid U, V_{S_V}), \tag{2}$$

where $\mathcal{P}_V$ is the set of complete prime components, or cliques, of G_V and $\mathcal{S}_V$ is the set of separators. For each subgraph $g \in \{\mathcal{P}_V, \mathcal{S}_V\}$, $Y_{\star g}$ is the $q \times |g|$ matrix with variables from the $|g|$ columns of Y defined by the subgraph, and V_g the corresponding submatrix of V. Each term in equation (2) is matrix normal, $Y_{\star g} \sim N(0, U, V_g)$ with $\Lambda_g = V_g^{-1}$ having no off-diagonal zeros. We can similarly factorize the joint density over G_U.

Now, U and V are not uniquely identified since, for any $c > 0$, $p(Y \mid U, V) = p(Y \mid cU, V/c)$. There are a number of approaches to imposing identification constraints such as $\mathrm{tr}(V) = p$ (Theobald & Wuttke, 2006), and possible strategies that use unconstrained parameters; we discuss the latter in § 8. Our use of hyper-Markov priors over each of U and V with underlying graphical models, discussed below, makes it desirable to adopt an explicit constraint and we enforce $v_{11} = 1$ from here on.

3. Matrix graphical modelling

Hyper-inverse Wishart priors are conjugate for covariance matrices in multivariate normal graphical models (Dawid & Lauritzen, 1993). Hyper-inverse Wishart distributions are compatible and consistent across graphs, which is critical when admitting uncertainty about graph structures (Giudici & Green, 1999; Jones et al., 2005). On decomposable graphs, the implied priors on sub-covariance matrices on all components and separators are inverse Wishart. Use of independent hyper-inverse Wishart priors for U, V in the current context is a natural choice, and maintains compatibility and consistency across graphs G_U, G_V. To incorporate the identification constraint $v_{11} = 1$, we use a parameter expansion approach. Parameter expansion involves expanding the parameter space by adding new nuisance parameters, and has been used purely algorithmically to accelerate Markov chain Monte Carlo samplers (Liu et al., 1998; Liu & Wu, 1999), but can also be used to induce new priors (Gelman, 2004, 2006) as is germane here.

We assume the prior $p(U, V) = p(U)p(V)$ where, using the hyper-inverse Wishart notation of Giudici & Green (1999) and Jones et al. (2005), the margins are defined by

$$U \sim \text{HIW}_{G_U}(b, B), \quad V = V^*/v^*_{11}, \quad V^* \sim \text{HIW}_{G_V}(d, D). \tag{3}$$

The density function for U is, following Dawid & Lauritzen (1993),

$$p(U) = \prod_{P_U \in \mathcal{P}_U} p(U_{P_U} \mid b, B_{P_U}) \Big/ \prod_{S_U \in \mathcal{S}_U} p(U_{S_U} \mid b, B_{S_U}),$$

where each component is an inverse Wishart density; $p(V^*)$ has a similar form.

The parameter expansion concept relates to v^*_{11} as an added parameter that converts column scales in V to those relative to the scale of the first column. As we move across graphs G_V, the priors $p(V \mid G_V)$ have the same induced priors over subgraph correlation structures but are no longer in complete agreement for $V = V^*/v^*_{11}$ due to the different parameterizations and interpretations. This is natural and appropriate. Suppose G_V and G'_V are two graphs with a common clique C. Each element in $\text{diag}(V_C)$ represents the relative scale of variance of that column to the variance of the first column so that, if G_V and G'_V imply different conditional dependencies between the first column and columns linked to C, then the induced priors over V_C should indeed be different.

The prior $p(V)$ is obtained by transformation from V^*. On any graph G_V, V is determined only by those free elements appearing in the submatrices corresponding to the cliques of the graph, and the nonfree elements of V are deterministic functions of the free elements (Carvalho et al., 2007). Let ν be the number of free elements; then the transformation from V^* to (V, v^*_{11}) has Jacobian $(v^*_{11})^{\nu-1}$ leading to

$$p(V, v^*_{11}) = \text{HIW}_{G_V}(v^*_{11} V \mid d, D)(v^*_{11})^{\nu-1}.$$

Coupled with the prior $p(U)$ on G_U, this defines a class of conditionally conjugate priors in the expanded parameter space.

4. Posterior and marginal likelihood computation

4·1. *Gibbs sampling on given graphs*

Assume an initial random sampling context with $q \times p$ data matrices Y_i $(i = 1, \ldots, n)$, drawn independently from equation (1), and write Y for the full set of data. It is easy to see that, on specified graphs (G_U, G_V), the posterior $p(U, V, v^*_{11} \mid Y)$ has conditional distributions:

$$(U \mid V, v^*_{11}, Y) \sim \text{HIW}_{G_U}\left(b + np, B + \sum_{i=1}^{n} Y_i V^{-1} Y_i'\right),$$

$$(V \mid U, v^*_{11}, Y) \sim \text{HIW}_{G_V}\left(d + nq, D/v^*_{11} + \sum_{i=1}^{n} Y_i' U^{-1} Y_i\right) I(v_{11} = 1),$$

$$(v^*_{11} \mid U, V, Y) \sim \text{IG}\{a/2 - \nu, \text{tr}(DV^{-1})/2\},$$

where $a = \sum_{P_V} |P_V|(2|P_V| + d) - \sum_{S_V} |S_V|(2|S_V| + d)$. These distributions form the basis of Gibbs sampling for the target posterior $p(U, V, v^*_{11} \mid Y)$. This involves iterative resampling from the hyper-inverse Wishart, inverse gamma and new conditional hyper-inverse Wishart distributions. Simulation of the former is based on Carvalho et al. (2007), while sampling the latter can be done as follows. From Lemma 2.18 of Lauritzen (1996), we can always find a perfect ordering of the nodes in G_V so that node 1 is in the first clique, say C, and then initialize the hyper-inverse Wishart sampler of Carvalho et al. (2007) to begin with a simulation of the implied conditional inverse Wishart distribution for the covariance matrix on that first clique. Sampling V_C from an inverse Wishart distribution conditional on the first diagonal element set to unity is straightforward.

4·2. *Marginal likelihood*

Exploration of uncertainty about graphical model structures involves consideration of the marginal likelihood function over graphs. For any pair (G_U, G_V), this is

$$p(Y) \equiv p(Y \mid G_U, G_V) = \int p(Y \mid U, V)p(U)p(V)\mathrm{d}U\,\mathrm{d}V.$$

The priors in the integrand depend on the graphs although for clarity we drop that from the notation. In multivariate models, marginal likelihoods can be evaluated in closed form on decomposable graphs (Giudici, 1996; Giudici & Green, 1999; Jones et al., 2005; Carvalho & West, 2007a, 2007b). In our matrix models, the integral cannot be evaluated but we can generate useful approximations via use of the candidate's formula (Besag, 1989; Chib, 1995). Write $\Theta = \{U, V, v^*_{11}\}$ for all parameters, and suppose that we can evaluate $p(\theta \mid Y)$ for some subset of parameters $\theta \in \Theta$; the candidate's formula gives the marginal likelihood via the identity $p(Y) = p(Y \mid \theta)/p(\theta \mid Y)$. Applying this requires that we estimate components of the numerator or denominator. Choosing θ to maximally exploit analytic integration is key, and different choices that integrate over different subsets of parameters will lead to different, parallel approximations of $p(Y)$ that can be compared. We use two approximations based on marginalization over desirably disjoint parameter subsets, namely, (A): $p(Y) = p(Y, v^*_{11}, U)/p(v^*_{11}, U \mid Y)$ at any chosen value of $\theta = \{v^*_{11}, U\}$, and (B): $p(Y) = p(Y, V)/p(V \mid Y)$ at any value of $\theta = V$. We estimate the components of these equations that have no closed form, then insert chosen values U, V, v^*_{11}, such as approximate posterior means, to provide two estimates of $p(Y)$.

For (A), first rewrite as

$$p(Y) = \frac{p(Y, v^*_{11}, U)p(V \mid v^*_{11}, U, Y)}{p(v^*_{11}, U \mid Y)p(V \mid v^*_{11}, U, Y)} = \frac{p(Y \mid V, v^*_{11}, U)p(U)p(V \mid v^*_{11})p(v^*_{11})}{p(v^*_{11}, U \mid Y)p(V \mid v^*_{11}, U, Y)}.$$

The numerator terms are each easily computed at any $\{V, v^*_{11}, U\}$. The second denominator term $p(V \mid v^*_{11}, U, Y)$ has an easily evaluated closed form, as in the Gibbs sampling step. The first denominator term may be approximated by

$$\begin{aligned} p(v^*_{11}, U \mid Y) &= \int p(v^*_{11} \mid Y, V)p(U \mid Y, V, v^*_{11})p(V \mid Y)\mathrm{d}V \\ &\approx \frac{1}{M}\sum_{j=1}^{M} p(v^*_{11} \mid Y, V_j)p(U \mid Y, V_j, v^*_{11}), \end{aligned}$$

where the sum is over posterior draws V_j; this is easy to compute as it is a sum of the product of inverse gamma and hyper-inverse Wishart densities.

For (B), the numerator can be analytically evaluated as

$$p(V, Y) = \int p(Y, U, V, v_{11}^*) \mathrm{d}U\, \mathrm{d}v_{11}^*$$
$$= \frac{q_V (2\pi)^{-nqp/2} H(b, B, G_U) H(d, D, G_V)}{H(b + np, B + \sum_{i=1}^{n} Y_i V^{-1} Y_i', G_U) H\{c, \mathrm{tr}(DV^{-1}), 1\}},$$

where

$$q_V = \prod_{P_V \in \mathcal{P}_V} |V_{P_V}|^{-(nq+d+2|P_V|)/2} \Big/ \prod_{S_V \in \mathcal{S}_V} |V_{S_V}|^{-(nq+d+2|S_V|)/2},$$

the $H(\cdot, \cdot, G_\cdot)$ terms are normalizing constants of the corresponding hyper-inverse Wishart distributions (Giudici & Green, 1999; Jones et al., 2005) and

$$c = \sum_{P_V \in \mathcal{P}_V} |P_V|(2|P_V| + d) - \sum_{S_V \in \mathcal{S}_V} |S_V|(2|S_V| + d) - 2\nu.$$

The density function in the denominator is approximated as

$$p(V \mid Y) = \int p(V \mid v_{11}^*, U, Y) p(v_{11}^*, U \mid Y) \mathrm{d}v_{11}^*\, \mathrm{d}U \approx \frac{1}{M} \sum_{j=1}^{M} P(V \mid Y, U_j, v_{11,j}^*),$$

where the sum over posterior draws $(U_j, v_{11,j}^*)$ can be easily performed, with terms given by conditional hyper-inverse Wishart density evaluations.

4·3. *Graphical model uncertainty and search*

Now admit uncertainty about graphs (G_U, G_V) using sparsity-encouraging priors in which edge inclusion indicators are independent Bernoulli variates (Dobra et al., 2004; Jones et al., 2005). We now extend Markov chain Monte Carlo simulation for multivariate graphical models (Giudici & Green, 1999; Jones et al., 2005) to learning on (G_U, G_V) in the above matrix model analysis. Our analysis generates multiple graphs with values of approximate posterior probabilities, using the Markov chain simulation for model search. This relies on the computation of the unnormalized posterior over graphs, $p(G_U, G_V \mid Y) \propto p(Y \mid G_U, G_V) p(G_U, G_V)$ involving the marginal likelihood value for any specified model (G_U, G_V) at each search step. For the latter, we average the approximate marginal likelihood values from methods (A) and (B). The work in Jones et al. (2005) includes evaluation of the performance of various stochastic search methods in single multivariate graphical models; for modest dimensions, they recommend simple local-move Metropolis–Hastings steps. Here, given a current pair (G_U, G_V), we can apply local moves in G_U space based on the conditional posterior $p(G_U \mid Y, G_V)$, and vice-versa. A candidate G_U' is sampled from a proposal distribution $q(G_U'; G_U)$ and accepted with probability

$$\alpha = \min\{1,\ p(G_U' \mid Y, G_V) q(G_U; G_U') / p(G_U \mid Y, G_V) q(G_U'; G_U)\};$$

our examples use the simple random add/delete edge move proposal of Jones et al. (2005). We then couple this with a similar step using $p(G_V \mid Y, G_U)$ at each iteration. This requires a Markov chain analysis on each graph pair visited in order to evaluate marginal likelihood, so implying a substantial computational burden.

5. EXAMPLE: A SIMULATED RANDOM SAMPLE

A sample of size $n = 48$ was drawn from the $(q = 8) \times (p = 7)$ dimensional $N(0, U, V)$ distribution, where, using $\cdot$ to denote zeros to highlight structure, the precision matrices are

$$\Lambda = \begin{pmatrix} 1{\cdot}85 & -0{\cdot}09 & -0{\cdot}65 & \cdot & -0{\cdot}24 & 0{\cdot}45 & \cdot \\ -0{\cdot}09 & 0{\cdot}21 & 0{\cdot}08 & \cdot & \cdot & 0{\cdot}14 & -0{\cdot}13 \\ -0{\cdot}65 & 0{\cdot}08 & 0{\cdot}58 & 0{\cdot}10 & \cdot & -0{\cdot}30 & \cdot \\ \cdot & \cdot & 0{\cdot}10 & 0{\cdot}48 & \cdot & -0{\cdot}10 & \cdot \\ -0{\cdot}24 & \cdot & \cdot & \cdot & 0{\cdot}70 & -0{\cdot}17 & \cdot \\ 0{\cdot}45 & 0{\cdot}14 & -0{\cdot}30 & -0{\cdot}10 & -0{\cdot}17 & 0{\cdot}61 & -0{\cdot}36 \\ \cdot & -0{\cdot}13 & \cdot & \cdot & \cdot & -0{\cdot}36 & 3{\cdot}72 \end{pmatrix},$$

$$\Omega = \begin{pmatrix} 0{\cdot}99 & \cdot & \cdot & -0{\cdot}33 & \cdot & 0{\cdot}05 & \cdot & \cdot \\ \cdot & 3{\cdot}65 & 0{\cdot}33 & \cdot & -0{\cdot}39 & -0{\cdot}41 & \cdot & -0{\cdot}03 \\ \cdot & 0{\cdot}33 & 2{\cdot}23 & \cdot & \cdot & -0{\cdot}38 & \cdot & \cdot \\ -0{\cdot}33 & \cdot & \cdot & 1{\cdot}65 & \cdot & \cdot & \cdot & \cdot \\ \cdot & -0{\cdot}39 & \cdot & \cdot & 2{\cdot}91 & -0{\cdot}30 & \cdot & \cdot \\ 0{\cdot}05 & -0{\cdot}41 & -0{\cdot}38 & \cdot & -0{\cdot}30 & 4{\cdot}71 & -0{\cdot}13 & -0{\cdot}40 \\ \cdot & \cdot & \cdot & \cdot & \cdot & -0{\cdot}13 & 1{\cdot}07 & -0{\cdot}26 \\ \cdot & -0{\cdot}03 & \cdot & \cdot & \cdot & -0{\cdot}40 & -0{\cdot}26 & 1{\cdot}45 \end{pmatrix}.$$

First consider analysis on the true graphs under priors with $b = d = 3$ and $B = 5I_8$, $D = 5I_7$ and simulation sample size 8000 after an initial, discarded burn-in of 2000 iterations. Convergence is rapid and apparently fast-mixing in this as in other simulated examples. The corresponding posterior means of the precision matrices are

$$\hat{\Lambda} = \begin{pmatrix} 1{\cdot}86 & -0{\cdot}11 & -0{\cdot}68 & \cdot & -0{\cdot}28 & 0{\cdot}44 & \cdot \\ -0{\cdot}11 & 0{\cdot}28 & 0{\cdot}14 & \cdot & \cdot & 0{\cdot}16 & -0{\cdot}21 \\ -0{\cdot}68 & 0{\cdot}14 & 0{\cdot}68 & 0{\cdot}16 & \cdot & -0{\cdot}33 & \cdot \\ \cdot & \cdot & 0{\cdot}16 & 0{\cdot}59 & \cdot & -0{\cdot}15 & \cdot \\ -0{\cdot}28 & \cdot & \cdot & \cdot & 0{\cdot}75 & -0{\cdot}14 & \cdot \\ 0{\cdot}44 & 0{\cdot}16 & -0{\cdot}33 & -0{\cdot}15 & -0{\cdot}14 & 0{\cdot}71 & -0{\cdot}45 \\ \cdot & -0{\cdot}21 & \cdot & \cdot & \cdot & -0{\cdot}45 & 4{\cdot}14 \end{pmatrix},$$

$$\hat{\Omega} = \begin{pmatrix} 0{\cdot}90 & \cdot & \cdot & -0{\cdot}27 & \cdot & -0{\cdot}02 & \cdot & \cdot \\ \cdot & 3{\cdot}23 & 0{\cdot}50 & \cdot & -0{\cdot}35 & -0{\cdot}22 & \cdot & -0{\cdot}12 \\ \cdot & 0{\cdot}50 & 2{\cdot}14 & \cdot & \cdot & -0{\cdot}37 & \cdot & \cdot \\ -0{\cdot}27 & \cdot & \cdot & 1{\cdot}46 & \cdot & \cdot & \cdot & \cdot \\ \cdot & -0{\cdot}35 & \cdot & \cdot & 2{\cdot}88 & -0{\cdot}41 & \cdot & \cdot \\ -0{\cdot}02 & -0{\cdot}22 & -0{\cdot}37 & \cdot & -0{\cdot}41 & 4{\cdot}20 & -0{\cdot}29 & -0{\cdot}08 \\ \cdot & \cdot & \cdot & \cdot & \cdot & -0{\cdot}29 & 0{\cdot}91 & -0{\cdot}26 \\ \cdot & -0{\cdot}12 & \cdot & \cdot & \cdot & -0{\cdot}08 & -0{\cdot}26 & 1{\cdot}58 \end{pmatrix}.$$

Figure 1 gives an implementation check on the concordance of the two marginal likelihood estimates. These are very close and differ negligibly on the log probability scale even at small Monte Carlo sample sizes.

Consider graphical model uncertainty with prior edge inclusion probabilities $2/(q-1)$ for G_U and $2/(p-1)$ for G_V. Repeat explorations suggest stability of the marginal likelihood estimation using smaller Monte Carlo sample sizes, and we use 2000 draws within each step of the model search. The add/delete Metropolis-within-Gibbs was run for 20 000 iterates starting from empty

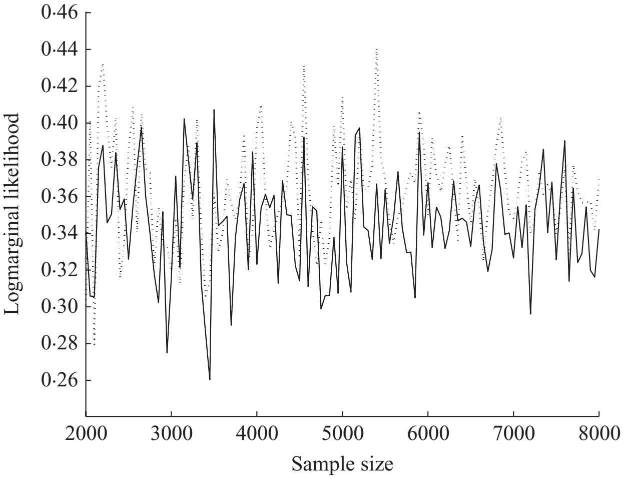

Fig. 1. Logmarginal likelihood values in the example of § 5. The two estimates A (full line) and B (dashed line) of § 4·2 were successively re-evaluated and plotted here at differing simulation sample sizes. The vertical scale has been adjusted by addition of 3583 for clarity.

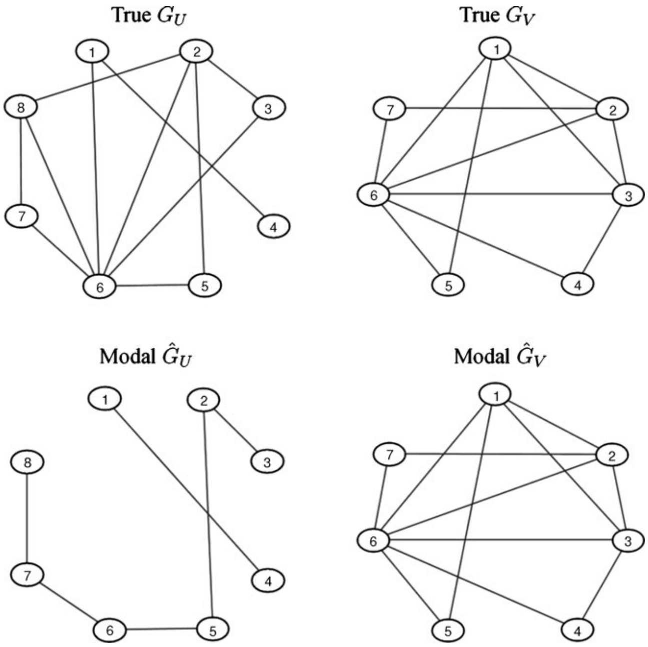

Fig. 2. True graphs in the simulated data example together with graphs of highest posterior probability identified from the analysis.

graphs. Results are essentially replicated starting at the full graphs. The most probable graphs visited, $(\hat{G}_U, \hat{G}_V)$, are shown in Fig. 2; these are local modes and also have greater posterior probability than the true graphs also displayed, and this model was first visited after 2614 Markov chain steps. The edges in $(\hat{G}_U, \hat{G}_V)$ generally have higher posterior edge inclusion probability than those not included; the lowest probability included edge has probability 0·52, while the highest probability excluded edge has probability 0·59. Thus, graphs discovered by highest posterior probability and by aggregating high probability edges are not dramatically different. The modal $\hat{G}_U$ is sparser than the true G_U, reflecting the difficulties in identifying very weak signals; for example, the modal graph lacks an edge corresponding to the true $\Omega_{1,6} = 0{\cdot}05$, and the posterior probability of that edge is naturally low. One measure of inferred sparsity is the posterior mean of the proportion of edges in each graph; these are about 28%, 59·6% for G_U, G_V, respectively. Additional posterior summaries and exploration of the posterior samples suggest clean convergence of the simulation analysis and the Metropolis–Hastings steps over graphs had good empirical acceptance rates of about 26%, 9% for G_U, G_V, respectively.

6. Dynamic matrix-variate graphical models for time series

Using the theory and methods for matrix normal models developed above, we are now able to extend our ideas to matrix time series involving two covariance matrices and associated graphical models. In the notation below, the work of Carvalho & West (2007a, 2007b) is the special case of vector data with $q = 1$, U fixed, and inference on (V, G_V) only.

A $q \times p$ matrix-variate times series Y_t follows the dynamic linear model

$$\begin{aligned} Y_t &= (I_q \otimes F_t')\Theta_t + \nu_t, \quad \nu_t \sim N(0, U, V), \\ \Theta_t &= (I_q \otimes G_t)\Theta_{t-1} + \Upsilon_t, \quad \Upsilon_t \sim N(0, U \otimes W_t, V) \end{aligned}$$

for $t = 1, 2, \ldots,$ where (a) $Y_t = (Y_{t,ij})$, the $q \times p$ matrix observation at time t; (b) $\Theta_t = (\Theta_{t,ij})$, the $qs \times p$ state matrix comprised of $q \times p$ state vectors $\Theta_{t,ij}$ each of dimension $s \times 1$; (c) $\Upsilon_t = (\omega_{t,ij})$, the $qs \times p$ matrix of state evolution innovations comprised of $q \times p$ innovation vectors $\omega_{t,ij}$ each of dimension $s \times 1$; (d) $\nu_t = (\nu_{t,ij})$, the $q \times p$ matrix of observational errors; (e) W_t is the $s \times s$ innovation covariance matrix at time t; (f) for all t, the s-vector F_t and $s \times s$ state evolution matrix G_t are known. Also, Υ_t follows a matrix-variate normal distribution with mean 0, left covariance matrix $U \otimes W_t$ and right covariance matrix V. In terms of scalar elements, we have $q \times p$ univariate models with individual s-vector state parameters, namely,

$$\begin{aligned} Y_{t,ij} &= F_t'\Theta_{t,ij} + \nu_{t,ij}, \quad \nu_{t,ij} \sim N(0, u_{ii}v_{jj}), \\ \Theta_{t,ij} &= G_t\Theta_{t-1,ij} + \omega_{t,ij}, \quad \omega_{t,ij} \sim N(0, u_{ii}v_{jj}W_t), \end{aligned} \tag{4}$$

for each i, j and t. Each of the scalar series shares the same F_t and G_t elements, and the reference to the model as one of exchangeable time series reflects these symmetries. In the example below, $F_t = F$ and $G_t = G$, as in many practical models, but the model class includes dynamic regressions when F_t involves predictor variables. This form of model is a standard specification (Quintana & West, 1987; West & Harrison, 1997) in which the correlation structures induced by U and V affect both the observation and evolution errors; for example, if u_{ij} is large and positive, vector series $Y_{t,i\star}$ and $Y_{t,j\star}$ will show concordant behaviour in movement of their state vectors and in observational variation about their levels. Specification of the entire sequence of W_t in terms of discount factors (West & Harrison, 1997) is also standard practice, typically using multiple discount factors related to components of the state vector and their expected degrees of random change in time, as illustrated in the example below. The innovations here concern graphical modelling and inference on (U, V). The key theory, conditional on U, V, concerns the conjugate sequential learning and forecasting as data is processed, as follows.

THEOREM 1. *Define $D_t = \{D_{t-1}, Y_t\}$ for $t = 1, 2, \ldots,$ with D_0 representing prior information. With initial prior $(\Theta_0 \mid U, V, D_0) \sim N(m_0, U \otimes C_0, V)$ we have, for all t:*

(i) *posterior at* $t-1$*:* $(\Theta_{t-1} \mid D_{t-1}, U, V) \sim N(m_{t-1}, U \otimes C_{t-1}, V)$;

(ii) *prior at* t*:* $(\Theta_t \mid D_{t-1}, U, V) \sim N(a_t, U \otimes R_t, V)$ *where* $a_t = (I_n \otimes G_t)m_{t-1}$ *and* $R_t = G_t C_{t-1} G_t' + W_t$;

(iii) *one-step forecast at* $t-1$*:* $(Y_t \mid D_{t-1}, U, V) \sim N(f_t, Uq_t, V)$ *with forecast mean matrix* $f_t = (I_n \otimes F_t' G_t)m_{t-1}$ *and scalar* $q_t = F_t' R_t F_t + 1$; *and*

(iv) *posterior at* t*:* $(\Theta_t \mid D_t, U, V) \sim N(m_t, U \otimes C_t, V)$ *with* $m_t = a_t + (I_q \otimes A_t)e_t$ *and* $C_t = R_t - A_t A_t' q_t$ *where* $A_t = R_t F_t / q_t$ *and* $e_t = Y_t - f_t$.

Proof. This stems from the theory of multivariate models applied to $\text{vec}(\text{Y}_\text{t})$ (West & Harrison, 1997). The main novelty here concerns the separability of covariance structures. That is: for all t, the distributions for state matrices have separable covariance structures; for example, $(\Theta_t \mid D_t, U, V)$ is such that $\text{cov}\{\text{vec}(\Theta_\text{t}) \mid \text{D}_\text{t}, \text{U}, \text{V}\} = \text{V} \otimes \text{U} \otimes \text{C}_\text{t}$; the sequential updating equations for the set of $qs \times p$ state matrices are implemented in parallel based on computations for the univariate component models, each of them involving the same scalar q_t, $s-$vector A_t and $s \times s$ matrices R_t, C_t at time t. □

Suppose now that U and V are constrained by graphs G_U and G_V, with priors as in equation (3) and sparsity priors over the graphs. Given data over $t = 1, \ldots, n$, the sequential updating analysis on (G_U, G_V) leads to the full joint density

$$p(Y_1, \ldots, Y_n \mid U, V) = \prod_{t=1}^{n} p(Y_t \mid U, V, D_{t-1}) = \prod_{t=1}^{n} N(e_t \mid 0, q_t U, V),$$

marginalized with respect to all state vectors. The one-step forecast error matrices e_t are conditionally independent matrix normal variates. Apart from the scalars q_t, this is essentially the framework of § 2. Thus, with a small change to insert the q_t, we are able to directly fit and explore dynamic graphical models using the analysis for random samples with embedded sequential updating computations.

7. A macro-economic example

An example concerns exploration of conditional dependence structures in macroeconomic time series related to US labour market employment. The data are Current Employment Statistics for the eight US states, New Jersey, New York, Massachusetts, Georgia, North Carolina, Virginia, Illinois and Ohio. We explore these data across nine industrial sectors: construction; manufacturing; trade, transportation and utilities; information; financial activities; professional and business services; education and health services; leisure and hospitality; and government. In our model framework, we have $q = 8$, $p = 9$ and monthly data over several years. Then U characterizes the residual conditional dependencies among states while V does the same for industrial sectors, in the context of an overall model that incorporates time-varying state parameters for underlying trend and annual seasonal structure in the series. Trend and seasonal elements are represented in standard form, the former as random walks and the latter as randomly varying seasonal effects. Specifically, in month t, the monthly employment change in state i and sector j is $Y_{t,ij}$, modelled as a first-order polynomial/seasonal effect model (West & Harrison, 1997) with the state vector comprising a local-level parameter and 12 seasonal factors, so that the state dimension is $s = 13$.

The univariate models of equation (4) have state vectors $\Theta_{t,ij} = (\mu_{t,ij}, \phi_{t,ij})'$, where $\mu_{t,ij}$ is the local level and $\phi_{t,ij} = (\phi_{t,ij,k}, \phi_{t,ij,k+1}, \ldots, \phi_{t,ij,11}, \phi_{t,ij,0}, \ldots, \phi_{t,ij,k-1})$ contains current monthly seasonal factors, subject to $1'\phi_{t,ij} = 0$ for all i, j and t. Further, $F_t = F$ (13×1) and $G_t = G$ (13×13) for all t, where $F' = (1, 1, 0, \ldots, 0)$. The state matrix G and the sequence of

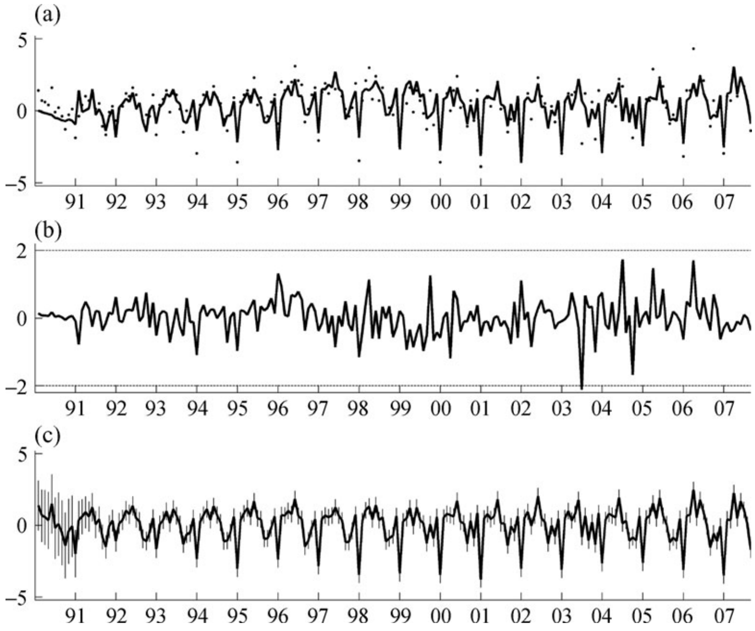

Fig. 3. One of the 72 time series in the econometric example, plotted over 1990–2007. (a): Dots are monthly changes in employment of North Carolina financial activities and the line joins the corresponding one-step ahead forecasts over time. (b): Corresponding standardized one-step ahead forecast errors $e_t/\surd q_t$. (c): Corresponding on-line estimated seasonal pattern with 95% pointwise credible intervals indicated by vertical bars at each month.

state evolution covariance matrices W_t (13×13) are

$$G = \begin{pmatrix} 1 & 0 \\ 0 & P \end{pmatrix}, \quad P = \begin{pmatrix} 0 & I_{11} \\ 1 & 0' \end{pmatrix}, \quad W_t = \begin{pmatrix} W_{t,\mu} & 0 \\ 0 & W_{t,\phi} \end{pmatrix},$$

with the latter having entries as follows. The univariate $W_{t,\mu}$ and 12×12 matrix $W_{t,\phi}$ are defined via discount factors δ_l and δ_s and the corresponding block components of C_t as $W_{t,\mu} = C_{t-1,\mu}(1-\delta_l)/\delta_l$ and $W_{t,\phi} = PC_{t-1,\phi}P'(1-\delta_s)/\delta_s$ for each t. The discount factor δ_l reflects the rate at which the levels $\mu_{t,ij}$ are expected to vary between months, with $100(\delta_l^{-1} - 1)$% of information on these parameters decaying each month. The factor δ_s plays the same role for seasonal parameters. We use $\delta_l = 0{\cdot}9$, $\delta_s = 0{\cdot}95$ to allow more adaptation to level changes than seasonal factors (West & Harrison, 1997); results, in terms of graphical model search and structure, are substantially similar using other values in appropriate ranges. In application, we can estimate discount factors and also extend the model to allow changes in discount factors to model change-points and other events impacting the series, based on monitoring and intervention methods (Pole et al., 1994; West & Harrison, 1997). Such considerations are secondary to our purposes in using this model for illustration of computational model search analysis for (U, V, G_U, G_V), but practically very germane. Model completion uses initial, vague priors with $m_0 = 0$, the 104×9 matrix, and $C_0 = 100I_{13}$. The constraint that $1'\phi_{t,ij} = 0$ is imposed by transforming m_0 and C_0 as discussed in West & Harrison (1997).

Model fitting estimates the movements in trend and seasonality, sequentially generating matrix series e_t whose row and column covariance patterns relate to (U, V). The North Carolina financial activities data and some aspects of the sequential model fit are graphed in Fig. 3. Priors for (U, V) use $B = 5I_8$, $D = 5I_9$ and $b = d = 3$, reflecting the range of residual variation,

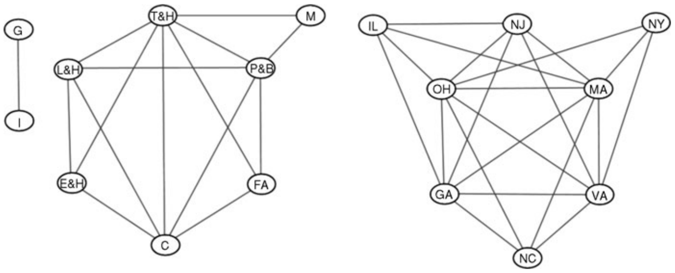

Fig. 4. Highest posterior probability graphs that illustrate aspects of inferred conditional dependencies among industrial sectors and among states in analysis of the econometric time series data.

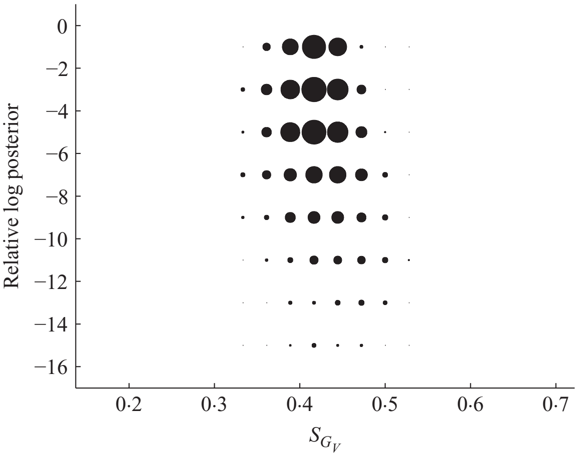

Fig. 5. Summary of posterior on sparsity of G_V in the econometric example. Circled areas are proportional to the fraction of posterior sampled graphs at several levels of posterior probability plotted against levels of sparsity S_{G_V} measured as the proportion of edges included.

sparsity-encouraging priors with prior edge inclusion probabilities $2/(q-1)$ for G_U and $2/(p-1)$ for G_V. The add/delete Metropolis-within-Gibbs sampler was run for 20 000 steps. Two chains were run: one starting at empty graphs and one at full graphs. The most probable model identified, $(\hat{G}_U, \hat{G}_V)$, is shown in Fig. 4. This, and the acceptance rates of graphs, were insensitive to the starting points. Beginning with empty graphs, the most probable model visited was found after 401 steps; its log posterior probability is −27 695·40, the sparsity of $(\hat{G}_U, \hat{G}_V)$ in terms of percentage of edges included is (72·4%, 42·1%) and the acceptance rates are (7·3%, 11·9%). Beginning with full graphs led to a most probable model with log posterior probability −27 695·43 after 2194 steps, sparsity (73·7%, 41·9%) and acceptance rates (7·7%, 12·2%). Posterior edge inclusion probabilities are also consistent between the two runs; see Table 1. Further, the most probable graphs sit in a region of graphs of similar sparsity and posterior probability and the posterior is dense around this mode; see Fig. 5.

Graphs with high probability in the region of the mode seem to reflect relevant dependencies in the econometric context. There are strongly evident conditional independencies particularly among subsets of the industrial sectors; see Table 1. Further, the posterior indicates overall

Table 1. *Posterior edge inclusion probabilities in graphical model analysis of the matrix econometric time series data*

	NJ	NY	MA	GA	NC	VA	IL	OH
NJ	1	0·05	1·00	0·55	0·01	1·00	1·00	1·00
NY		1	1·00	0·19	0·00	1·00	0·00	0·59
MA			1	0·98	0·96	1·00	1·00	1·00
GA				1	0·93	0·89	0·75	1·00
NC					1	1·00	0·06	1·00
VA						1	0·31	1·00
IL							1	1·00
OH								1

	C	M	T&U	I	FA	P&BS	E&H	L&H	G
C	1	0·02	1·00	0·16	0·75	1·00	0·99	1·00	0·06
M		1	1·00	0·28	0·02	0·98	0·01	0·03	0·01
T&U			1	0·02	1·00	1·00	0·93	1·00	0·02
I				1	0·06	0·34	0·02	0·01	0·55
FA					1	1·00	0·00	0·04	0·02
P&BS						1	0·02	1·00	0·00
E&H							1	0·75	0·02
L&H								1	0·01
G									1

US states: NJ, New Jersey; NY, New York; MA, Massachusetts; GA, Georgia; NC, North Carolina; VA, Virginia; IL, Illinois; OH, Ohio. *Industrial sectors*: C, industrial construction; M, manufacturing; T&U, trade, transportation & utilities; I, information; FA, financial activities; P&BS, professional & business services; E&H, education & health services; L&H, leisure & hospitality; G, government.

sparsity levels through posterior means of about 73% for the proportion of edges included in G_U and about 42% in G_V. Figure 5 further illustrates aspects of the posterior over sparsity for G_V.

8. Further comments

We have introduced Bayesian analysis of matrix-variate graphical models in random sampling and time series contexts. The main innovations include new priors for matrix normal graphical models, use of the parameter expansion approach, inference via Markov chain Monte Carlo for a specific graphical model, evaluation of marginal likelihoods over graphs using coupled candidate's formula approximations, and the extension of graphical modelling to matrix time series analysis.

On the use of parameter expansion, Roy & Hobert (2007) and Hobert & Marchev (2008) provide theoretical support for the method in Gibbs samplers; in our models, this approach induces tractable and computationally accessible posteriors, leads to good mixing of Markov chain simulations, and is theoretically fundamental to the new model/prior framework in addressing identification issues directly and naturally.

On model identification, an alternative approach might use unconstrained hyper-inverse Wishart priors for each of (U, V) and run the Markov chain Monte Carlo simulation on the unconstrained parameters, similar to a strategy sometimes used in multinomial probit models (McCulloch et al., 2000). It can be argued that this is computationally less demanding than using our explicitly constrained prior and that inferences can be constructed from the simulation output by transforming to constraint-compatible parameters $(Uv_{11}, V/v_{11})$. We had considered this, and note that posterior simulation analysis is marginally faster than under the explicitly identified model; in empirical studies, however, we find the computational benefit to be of negligible practical significance. Importantly, this approach relies on a proper prior for the effectively free, unidentified parameter v_{11}, and is sensitive to that choice. More importantly, the implied prior on $(Uv_{11}, V/v_{11})$ is nonstandard and difficult to interpret, and raises questions in prior elicitation and specification; for example, the implied margins for variances are those of ratios of inverse gamma variates and difficult to assess compared to the traditional inverse gamma, and there are now dependencies in priors on left and right covariance matrices. Perhaps most important are the resulting effects on approximate marginal likelihoods; in examples we have studied, the approach yields very different marginal likelihoods and the impact of the marginal prior on the unidentified v_{11} plays a key role in that. In contrast, and though very slightly more computationally demanding, the direct and explicitly constrained hyper-inverse Wishart prior is easy to interpret, specify and, with results from Carvalho et al. (2007), implement; synthetic examples have verified the resulting efficacy of the simulation and model search computations.

Our use of candidate's formula to provide different approximations to marginal likelihoods over graphs can be extended to multiple such approximations. We have explored other constructions, and found no obvious practical differences in the resulting estimates in simulated examples. This is an area open for theoretical investigation and in other model contexts. This also offers a route to extending the analysis here to nondecomposable graphical models.

Our examples are in modest dimensional problems where local move Metropolis–Hastings methods for the graphical model components of the analysis can be expected to be effective, building on experiences in multivariate models (Jones et al., 2005). To scale to higher dimensions, alternative computational strategies such as shotgun stochastic search over graphs (Dobra et al., 2004; Jones et al., 2005; Hans et al., 2007) become relevant. A critical perspective is to define analysis that will rapidly find regions of graphical model space supported by the data. It is far better to work with a small selection of high-probability models than a grossly incorrect model on full graphs, and as dimensions scale the latter quickly becomes infeasible. Shotgun stochastic search and related methods reflect this and offer a path towards faster, parallelizable model search. There is also potential for computationally faster approximations using expectation-maximization style and variational methods (Jordan et al., 1999).

An interesting class of matrix graphical structures arises under autoregressive correlation specifications for the two covariance matrices. This generates a class of Markov random field models that is of potential interest in applications such as texture image modelling. With the matrix data representing a spatial process on a rectangular grid, taking covariance matrices U and V as those of two stationary autoregressive processes provides flexibility in modelling patterns separately in horizontal and vertical directions. We have experimented with examples that suggest potential for this direction in applying the new theory and methods we have presented.

Acknowledgement

The authors are grateful for the constructive comments of the editor and two referees on the original version of this paper. Research reported here was partially supported under grants from the U.S. National Science Foundation and the National Institutes of Health. Any opinions, findings and conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of the NSF or NIH. Matlab code implementing the analysis and the GraphExplore visualization software are freely available at www.stat.duke.edu/research/software/west.

REFERENCES

BESAG, J. (1989). A candidate's formula: a curious result in Bayesian prediction. *Biometrika* **76**, 183.

CARVALHO, C. M., MASSAM, H. & WEST, M. (2007). Simulation of hyper-inverse Wishart distributions in graphical models. *Biometrika* **94**, 647–59.

CARVALHO, C. M. & WEST, M. (2007a). Dynamic matrix-variate graphical models. *Bayesian Anal.* **2**, 69–98.

CARVALHO, C. M. & WEST, M. (2007b). Dynamic matrix-variate graphical models—a synopsis. In *Bayesian Statistics* VIII, Ed. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, pp. 585–90. Oxford: Oxford University Press.

CHIB, S. (1995). Marginal likelihood from the Gibbs output. *J. Am. Statist. Assoc.* **90**, 1313–21.

DAWID, A. P. (1981). Some matrix-variate distribution theory: notational considerations and a Bayesian application. *Biometrika* **68**, 265–74.

DAWID, A. P. & LAURITZEN, S. L. (1993). Hyper-Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21**, 1272–317.

DOBRA, A., JONES, B., HANS, C., NEVINS, J. & WEST, M. (2004). Sparse graphical models for exploring gene expression data. *J. Mult. Anal.* **90**, 196–212.

DUTILLEUL, P. (1999). The MLE algorithm for the matrix normal distribution. *J. Statist. Comp. Simul.* **64**, 105–23.

FINN, J. D. (1974). *A General Model for Multivariate Analysis*. New York: Holt, Rinehart and Winston.

GALECKI, A. (1994). General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Commun. Statist.* A **23**, 3105–19.

GELMAN, A. (2004). Parameterization and Bayesian modeling. *J. Am. Statist. Assoc.* **99**, 537–45.

GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Anal.* **3**, 515–34.

GIUDICI, P. (1996). Learning in graphical Gaussian models. In *Bayesian Statistics* 5, Ed. J. M. Bernado, J. O. Berger, A. P. Dawid and A. F. M. Smith, pp. 621–28. Oxford: Oxford University Press.

GIUDICI, P. & GREEN, P. J. (1999). Decomposable graphical Gaussian model determination. *Biometrika* **86**, 785–801.

GUPTA, A. K. & NAGAR, D. K. (2000). *Matrix Variate Distributions* Monographs and Surveys in Pure & Applied Mathematics 104. London: Chapman & Hall.

HANS, C., DOBRA, A. & WEST, M. (2007). Shotgun stochastic search in regression with many predictors. *J. Am. Statist. Assoc.* **102**, 507–16.

HOBERT, J. P. & MARCHEV, D. (2008). A theoretical comparison of the data augmentation, marginal augmentation and PXDA algorithms. *Ann. Statist.* **2**, 532–54.

HUIZENGA, H. M., DE MUNCK, J. C., WALDORP, L. J. & GRASMAN, R. (2002). Spatiotemporal EEG/MEG source analysis based on a parametric noisecovariance model. *IEEE Trans. Biomed. Eng.* **49**, 533–9.

JONES, B., CARVALHO, C. M., DOBRA, A., HANS, C., CARTER, C. & WEST, M. (2005). Experiments in stochastic computation for high-dimensional graphical models. *Statist. Sci.* **20**, 388–400.

JORDAN, M., GHAHRAMANI, Z., JAAKKOLA, T. & SAUL, L. (1999). An introduction to variational methods for graphical models. *Mach. Learn.* **37**, 183–233.

LAURITZEN, S. L. (1996). *Graphical Models*. Oxford: Clarendon Press.

LIU, C., RUBIN, D. B. & WU, Y. N. (1998). Parameter expansion to accelerate EM: the PX-EM algorithm. *Biometrika* **85**, 755–70.

LIU, J. S. & WU, Y. N. (1999). Parameter expansion for data augmentation. *J. Am. Statist. Assoc.* **94**, 1264–74.

MARDIA, K. V. & GOODALL, C. R. (1993). Spatial-temporal analysis of multivariate environmental monitoring data. In *Multivariate Environmental Statistics*, Ed. G. P. Patil and C. R. Rao, pp. 347–85. Amsterdam: Elsevier.

MCCULLOCH, R. E., POLSON, N. G. & ROSSI, P. E. (2000). Bayesian analysis of the multinomial probit model with fully identified parameters. *J. Economet.* **99**, 173–93.

MITCHELL, M. W., GENTON, M. G. & GUMPERTZ, M. L. (2006). A likelihood ratio test for separability of covariances. *J. Mult. Anal.* **97**, 1025–43.

NAIK, D. N. & RAO, S. S. (2001). Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix. *J. Appl. Statist.* **29**, 91–105.

POLE, A., WEST, M. & HARRISON, P. J. (1994). *Applied Bayesian Forecasting and Time Series Analysis*. New York: Chapman-Hall.

QUINTANA, J. M. & WEST, M. (1987). Multivariate time series analysis: new techniques applied to international exchange rate data. *Statistician* **36**, 275–81.

ROY, V. & HOBERT, J. P. (2007). Convergence rates and asymptotic standard errors for MCMC algorithms for Bayesian probit regression. *J. R. Statist. Soc.* B **69**, 607–23.

THEOBALD, D. L. & WUTTKE, D. S. (2006). Empirical Bayes hierarchical models for regularizing maximum likelihood estimation in the matrix Gaussian procrustes problem. *Proc. Nat. Acad. Sci.* **103**, 18521–7.

WEST, M. & HARRISON, P. J. (1997). *Bayesian Forecasting and Dynamic Models*, 2nd ed. New York: Springer.

WHITTAKER, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Chichester, UK: John Wiley and Sons.

[*Received February* 2008. *Revised February* 2009]