Journal of Zhejiang University-SCIENCE C (Computers & Electronics)
ISSN 1869-1951 (Print); ISSN 1869-196X (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Modeling correlated samples via sparse matrix Gaussian graphical models

Yi-zhou HE[1], Xi CHEN[1], Hao WANG[‡2]

([1]*Lancaster University Management School, Lancaster University, Lancaster, LA1 4YU, UK*)

([2]*Department of Statistics, University of South Carolina, Columbia, South Carolina 29201, USA*)

E-mail: y.he2@lancs.ac.uk; x.chen14@lancs.ac.uk; haowang@sc.edu

Received Nov. 9, 2012; Revision accepted Jan. 14, 2013; Crosschecked Jan. 14, 2013

Abstract: A new procedure of learning in Gaussian graphical models is proposed under the assumption that samples are possibly dependent. This assumption, which is pragmatically applied in various areas of multivariate analysis ranging from bioinformatics to finance, makes standard Gaussian graphical models (GGMs) unsuitable. We demonstrate that the advantage of modeling dependence among samples is that the true discovery rate and positive predictive value are improved substantially than if standard GGMs are applied and the dependence among samples is ignored. The new method, called matrix-variate Gaussian graphical models (MGGMs), involves simultaneously modeling variable and sample dependencies with the matrix-normal distribution. The computation is carried out using a Markov chain Monte Carlo (MCMC) sampling scheme for graphical model determination and parameter estimation. Simulation studies and two real-world examples in biology and finance further illustrate the benefits of the new models.

Key words: Gaussian graphical models, Hyper-inverse Wishart distributions, Mutual fund evaluation, Network
doi:10.1631/jzus.C1200316 **Document code:** A **CLC number:** O212.8

1 Introduction

Many multivariate statistical methods assume that samples are independently distributed from a multivariate Gaussian distribution. Under this independence assumption, Gaussian graphical models (GGMs) are widely used for simultaneously discovering conditional independence structures and estimating covariance matrices (Lauritzen, 1996; Giudici and Green, 1999; Wang and Li, 2012). Very often, this assumption is violated in real applications where samples are naturally correlated owing to the data generating mechanisms. To accommodate sample dependence, existing methods often assume that the dependence structure is known (e.g., temporal correlations), and then model the structure explicitly (Carvalho and West, 2007; Zhang *et al.*, 2009; Wang, 2010; Wang *et al.*, 2011). When the sample dependence structure is unknown, this paper proposes a novel matrix graphical model for modeling the general dependence structure for better statistical inference.

One motivating example comes from building gene association networks from microarray data—an area where GGMs have been used extensively. When GGMs are used to generate graphs for genes, observations of microarrays are typically assumed to be independent. What if samples are dependent? In Efron (2009) and Allen and Tibshirani (2010; 2012), correlations among arrays are considered in the context of detecting significant genes in the two-class microarray. Correlations among samples can affect the observed pattern of correlations among variables. This 'leakage' effect can possibly lead to many false positive edges if standard GGMs are applied, and thus must be taken care of.

‡ Corresponding author

Another very meaningful motivation is to model panel data that arise frequently in economics and finance. Panel data contain repeated observations on the same cross-section of variables (e.g., firm's stock returns) over multiple time periods. The data thus will exhibit both cross-sectional dependence and time-series dependence. Naively assuming independence among samples will introduce severe bias into statistical analysis, and potentially lead to bad decision-making (Wooldridge, 2001; Petersen, 2009). Any good statistical procedure for panel data must carefully consider the correlation structures naturally inherent in the data. Our proposed matrix graphical models provide one such method to parsimoniously model correlation structures in panel data in order to produce reliable statistical inferences.

Our main contributions are the introduction of a matrix-variate Gaussian graphical model (MGGM) that simultaneously accounts for dependence among samples and among variables, and the efficient Markov chain Monte Carlo (MCMC) sampler for model fitting. The performance of the new methods is evaluated in a simulation and two real-world problems about gene expression network and mutual fund performance evaluation. New methods are found to improve parameter estimation and signal discovery over the standard models that consider no correlation among samples.

2 Gaussian graphical model for independent data

We first give a brief introduction of the standard Gaussian graphical models. Consider a p-dimensional vector which follows a multivariate normal distribution $N(0, \boldsymbol{U})$, where $\boldsymbol{U}$ is the covariance matrix. Let $\boldsymbol{\Omega} = (\omega_{ij}) \equiv \boldsymbol{U}^{-1}$ be the precision matrix. Let $G_{\boldsymbol{U}} = (V_{\boldsymbol{U}}, E_{\boldsymbol{U}})$ be an undirected graph with the node set $V_{\boldsymbol{U}} = \{1, 2, \cdots, p\}$ and the edge set $E_{\boldsymbol{U}} = \{(i, j) : \omega_{ij} \neq 0\}$ (i.e., edge (i, j) is missing whenever $\omega_{ij} = 0$).

The graph G then reflects the conditional independencies among components in $\boldsymbol{y}$. Observe independent samples $\boldsymbol{Y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n\}$ of size n from $N(0, \boldsymbol{U})$. The density of $\boldsymbol{Y}$ given a graph $G_{\boldsymbol{U}}$ can be represented as

$$p(\boldsymbol{Y} \mid \boldsymbol{U}, G_{\boldsymbol{U}}) = \frac{\prod_{P_{\boldsymbol{U}} \in \mathcal{P}_{\boldsymbol{U}}} N(\boldsymbol{Y}_{P_{\boldsymbol{U}}} \mid 0, \boldsymbol{U}_{P_{\boldsymbol{U}}})}{\prod_{S \in \mathcal{S}_{\boldsymbol{U}}} N(\boldsymbol{Y}_{S_{\boldsymbol{U}}} \mid 0, \boldsymbol{U}_{S_{\boldsymbol{U}}})}, \quad (1)$$

where $\mathcal{P}_{\boldsymbol{U}}$ and $\mathcal{S}_{\boldsymbol{U}}$ are sets of prime components and separators of $G_{\boldsymbol{U}}$ respectively, and $N(\boldsymbol{Y}_{P_{\boldsymbol{U}}} \mid 0, \boldsymbol{U}_{P_{\boldsymbol{U}}})$ is a multivariate Gaussian density with the covariance matrix of $\boldsymbol{U}_{P_{\boldsymbol{U}}}$ corresponding to the component $P_{\boldsymbol{U}}$.

For the constrained covariance matrix $\boldsymbol{U}$, the 'hyper-inverse Wishart' distributions (Dawid and Lauritzen, 1993) are the conjugate priors. Specifically, given $G_{\boldsymbol{U}}$, $\boldsymbol{U}$ follows a hyper-inverse Wishart, $\mathrm{HIW}_{G_{\boldsymbol{U}}}(b, \boldsymbol{B})$ prior, if its prior density is of the form

$$p(\boldsymbol{U} \mid G_{\boldsymbol{U}}) = \frac{\prod_{P_{\boldsymbol{U}} \in \mathcal{P}_{\boldsymbol{U}}} \mathrm{IW}(\boldsymbol{U}_{P_{\boldsymbol{U}}} \mid b, \boldsymbol{B}_{P_{\boldsymbol{U}}})}{\prod_{S_{\boldsymbol{U}} \in \mathcal{S}_{\boldsymbol{U}}} \mathrm{IW}(\boldsymbol{U}_{S_{\boldsymbol{U}}} \mid b, \boldsymbol{B}_{S_{\boldsymbol{U}}})}, \quad (2)$$

where the sub-matrix $\boldsymbol{U}_{P_{\boldsymbol{U}}}$ has an inverse Wishart $\mathrm{IW}(\boldsymbol{U}_{P_{\boldsymbol{U}}} \mid b, \boldsymbol{B}_{P_{\boldsymbol{U}}})$ prior for each complete prime component $P_{\boldsymbol{U}} \in \mathcal{P}_{\boldsymbol{U}}$. If the prime component is not complete, Roverato (2002) showed that the density on that prime component can be theoretically derived up to a normalizing constant. Bayesian graphical model determination requires the evaluation of the marginal likelihood at any graph $G_{\boldsymbol{U}}$. Under the hyper-inverse Wishart prior, $\boldsymbol{U} \sim \mathrm{HIW}_{G_{\boldsymbol{U}}}(b, \boldsymbol{B})$, this marginal likelihood density is calculated as

$$\begin{aligned} p(\boldsymbol{Y} \mid G_{\boldsymbol{U}}) &= \int_{\boldsymbol{\Omega} \in M(G_{\boldsymbol{U}})} p(\boldsymbol{Y} \mid \boldsymbol{U}, G_{\boldsymbol{U}}) p(\boldsymbol{U} \mid G_{\boldsymbol{U}}) \mathrm{d}\boldsymbol{U} \\ &= (2\pi)^{-\frac{np}{2}} \frac{H(b, \boldsymbol{B}, G_{\boldsymbol{U}})}{H(b+n, \boldsymbol{B}+\boldsymbol{S}_{\boldsymbol{Y}}, G_{\boldsymbol{U}})}, \quad (3) \end{aligned}$$

where $M(G_{\boldsymbol{U}})$ is the space of all positive-definite symmetric matrices that have fixed zeros for missing edges in $G_{\boldsymbol{U}}$, and $H(b, \boldsymbol{B}, G_{\boldsymbol{U}})$ and $H(b+n, \boldsymbol{B}+\boldsymbol{S}_{\boldsymbol{Y}}, G_{\boldsymbol{U}})$ are the prior and posterior normalizing constants respectively, where $\boldsymbol{S}_{\boldsymbol{Y}} = \boldsymbol{Y}\boldsymbol{Y}'$. For decomposable models, the normalizing constants are analytically available. For nondecomposable models, the normalizing constant is analytically intractable and requires numerical approximation (Atay-Kayis and Massam, 2005).

3 Gaussian graphical models for correlated samples

3.1 Matrix Gaussian graphical models

Now relax the assumption in Section 2 that the samples $\boldsymbol{y}_i$ are independent. We propose matrix Gaussian graphical models for accommodating the correlations among samples using the matrix-variate normal distribution. Consider a zero mean random

matrix $\boldsymbol{Y}$ of size $p \times n$. Then $\boldsymbol{Y}$ is matrix normal, $\boldsymbol{Y} \sim N(0, \boldsymbol{U}, \boldsymbol{V})$,

$$p(\boldsymbol{Y}) \equiv p(\boldsymbol{Y} \mid \boldsymbol{U}, \boldsymbol{V}) = k(\boldsymbol{U}, \boldsymbol{V}) \exp\left[-\mathrm{tr}\left(\frac{\boldsymbol{Y}'\boldsymbol{U}^{-1}\boldsymbol{Y}\boldsymbol{V}^{-1}}{2}\right)\right], \quad (4)$$

where $\boldsymbol{U} = (u_{ij})$ is the column covariance matrix of size $p \times p$, $\boldsymbol{V} = (v_{ij})$ is the row covariance matrix of size $n \times n$, and the constant $k(\boldsymbol{U}, \boldsymbol{V}) = (2\pi)^{-np/2}|\boldsymbol{U}|^{-n/2}|\boldsymbol{V}|^{-p/2}$ with $|\cdot|$ denoting the matrix determinant. The matrix-variate normal implies that the marginal distribution of one row $\boldsymbol{y}_{i,\star}$ is multivariate normal with covariance matrix $u_{ii}\boldsymbol{V}$ and precision matrix $\boldsymbol{\Lambda}/u_{ii} = \boldsymbol{V}^{-1}/u_{ii}$, and the marginal distribution of one column $\boldsymbol{y}_{\star,j}$ is multivariate normal with covariance matrix $v_{jj}\boldsymbol{U}$ and precision matrix $\boldsymbol{\Omega}/v_{jj} = \boldsymbol{U}^{-1}/v_{jj}$. Thus, the covariance matrix $\boldsymbol{V}$ captures the correlations among samples.

GGMs of Section 2 focus on $\boldsymbol{V} = \boldsymbol{I}_n$ where the columns $\boldsymbol{y}_{\star,j}$ are independent replicates. Correlations among samples imply a non-diagonal $\boldsymbol{V}$; however, conditional independencies among variables are still encoded in undirected graphical $\boldsymbol{\Omega}$. Zeros in $\boldsymbol{\Omega}$ define row-wise conditional independencies; that is, $\boldsymbol{y}_{i,\star}$ and $\boldsymbol{y}_{j,\star}$, conditional upon $\boldsymbol{y}_{-(ij,\star)}$, are independent if and only if $\omega_{ij} = 0$.

As in GGMs, we seek to determine the graph $G_{\boldsymbol{U}}$ for $\boldsymbol{U}$. Instead of assuming $\boldsymbol{V} = \boldsymbol{I}_n$, we incorporate the effect of $\boldsymbol{V}$ to $G_{\boldsymbol{U}}$ by imposing graphical models on $\boldsymbol{V}$ to parsimoniously estimate $\boldsymbol{V}$. Now, for a single data matrix $\boldsymbol{Y}$, two graphs are considered: a p-node graph $G_{\boldsymbol{U}}$ representing relationship among p variables, and an n-node graph $G_{\boldsymbol{V}}$ representing relationship among n samples. Here we assume $\boldsymbol{Y}$ has zero mean. This is not a restriction, because $\boldsymbol{Y}$ can be de-meaned by estimating $\boldsymbol{M}$ using methods in, for example, Allen and Tibshirani (2010).

Clearly, $\boldsymbol{U}$ and $\boldsymbol{V}$ are not identified since, for any $c \neq 0$ and $\boldsymbol{Y}$, the density $p(\boldsymbol{Y} \mid \boldsymbol{U}, \boldsymbol{V})$ is equal to the density $p(\boldsymbol{Y} \mid c\boldsymbol{U}, \boldsymbol{V}/c)$. We follow the idea of Wang and West (2009) to achieve identification by imposing the constraint $v_{11} = 1$. Furthermore, to specify priors for the restricted parameter of $\boldsymbol{V}$, we consider the conditioning approach, and assume independent priors for $\boldsymbol{U}$ and $\boldsymbol{V}$ with margins defined by

$$\begin{cases} \boldsymbol{U} \sim \mathrm{HIW}_{G_{\boldsymbol{U}}}(b, \boldsymbol{B}), \\ \boldsymbol{V} \sim \mathrm{HIW}_{G_{\boldsymbol{V}}}(d, \boldsymbol{D})1_{v_{11}=1}. \end{cases} \quad (5)$$

For the graphical model space priors on $(G_{\boldsymbol{U}}, G_{\boldsymbol{V}})$, we use independent Bernoulli distributions for edge inclusion indicators (Jones *et al.*, 2005). The prior on the graphical model space is $\pi(G_{\boldsymbol{U}}) = w_{\boldsymbol{U}}^{k_{\boldsymbol{U}}}(1 - w_{\boldsymbol{U}})^{m-k_{\boldsymbol{U}}}$, where $k_{\boldsymbol{U}}$ is the number of edges in $G_{\boldsymbol{U}}$ and $m = p(p-1)/2$. We use $w_{\boldsymbol{U}} = 2/(p-1)$ and $w_{\boldsymbol{V}} = 2/(n-1)$ in this paper.

3.2 Markov chain Monte Carlo sampling

To estimate MGGMs, we develop the following MCMC sampling scheme:

(a) Generate $G_{\boldsymbol{U}}$ from

$$(G_{\boldsymbol{U}} \mid \boldsymbol{V}, G_{\boldsymbol{V}}) \propto \frac{H(b, \boldsymbol{B}, G_{\boldsymbol{U}})\pi(G_{\boldsymbol{U}})}{H(b+n, \boldsymbol{B}+\boldsymbol{Y}\boldsymbol{\Lambda}\boldsymbol{Y}', G_{\boldsymbol{U}})}. \quad (6)$$

(b) Generate $\boldsymbol{U}$ from

$$(\boldsymbol{U} \mid \boldsymbol{V}, G_{\boldsymbol{U}}, G_{\boldsymbol{V}}) \sim \mathrm{HIW}_{G_{\boldsymbol{U}}}(b+n, \boldsymbol{B}+\boldsymbol{Y}\boldsymbol{\Lambda}\boldsymbol{Y}'). \quad (7)$$

(c) Generate $G_{\boldsymbol{V}}$ from

$$(G_{\boldsymbol{V}} \mid \boldsymbol{U}, G_{\boldsymbol{U}}) \propto p(\boldsymbol{Y} \mid G_{\boldsymbol{V}}, \boldsymbol{U}, G_{\boldsymbol{U}})\pi(G_{\boldsymbol{V}}), \quad (8)$$

where $p(\boldsymbol{Y} \mid G_{\boldsymbol{V}}, \boldsymbol{U}, G_{\boldsymbol{U}})$ is calculated below.

(d) Generate $\boldsymbol{V}$ from

$$(\boldsymbol{V} \mid \boldsymbol{U}, G_{\boldsymbol{V}}, G_{\boldsymbol{U}}) \sim \mathrm{HIW}_{G_{\boldsymbol{V}}}(d+p, \boldsymbol{D}+\boldsymbol{Y}'\boldsymbol{\Omega}\boldsymbol{Y})1_{v_{11}=1}. \quad (9)$$

In this sampling scheme, steps (b) and (d) involve sampling from the HIW, and the conditional HIW distributions. For decomposable graphs, Wang and West (2009) provided methods for generating these random variables. For nondecomposable graphs, sampling HIW is based on the block Gibbs sampler algorithm of Wang and Li (2012), while sampling the conditional HIW can be done by modifying the algorithm in Wang and Li (2012) to incorporate the constraint $v_{11} = 1$. In particular, in the block Gibbs sampler of Wang and Li (2012), whenever updating a block involving the first element v_{11}, we generate $\boldsymbol{V}$ from the conditional inverse-Wishart distribution under the constraint $v_{11} = 1$.

Step (c) requires the integration of a normal likelihood function with respect to the conditional HIW distributions to generate $G_{\boldsymbol{V}}$ in the sampling scheme. This integration can be theoretically derived

as follows:

$$
\begin{aligned}
&p(\boldsymbol{Y} \mid G_{\boldsymbol{V}}, \boldsymbol{U}, G_{\boldsymbol{U}}) \\
&= \int p(\boldsymbol{Y} \mid \boldsymbol{U}, \boldsymbol{V}) p(\boldsymbol{V} \mid \boldsymbol{U}, G_{\boldsymbol{U}}) \mathrm{d}\boldsymbol{V} \\
&\propto \int N(\boldsymbol{Y}; 0, \boldsymbol{U}, \boldsymbol{V}) \mathrm{HIW}_{G_{\boldsymbol{V}}}(\boldsymbol{V}; d, \boldsymbol{D}) 1_{v_{11}=1} \mathrm{d}\boldsymbol{V} \\
&\propto \frac{H(d, \boldsymbol{D}, G_{\boldsymbol{V}})}{\dfrac{p(v_{11}=1; d, \boldsymbol{D}, G_{\boldsymbol{V}})}{\dfrac{H(d+p, \boldsymbol{D}+\boldsymbol{Y}'\boldsymbol{\Omega}\boldsymbol{Y}, G_{\boldsymbol{V}})}{p(v_{11}=1; d+p, \boldsymbol{D}+\boldsymbol{Y}'\boldsymbol{\Omega}\boldsymbol{Y}, G_{\boldsymbol{V}})}}},
\end{aligned}
$$

where $p(v_{11}=1; d, \boldsymbol{D}, G_{\boldsymbol{V}})$ is the marginal density of v_{11} evaluated at $v_{11}=1$. Here the marginal density of v_{11} is implied from the joint distribution $\boldsymbol{V} \sim \mathrm{HIW}_{G_{\boldsymbol{V}}}(b, \boldsymbol{D})$. Note that Corollary 2 of Roverato (2002) implies that neither $p(v_{11}=1; d, \boldsymbol{D}, G_{\boldsymbol{V}})$ nor $p(v_{11}=1; d+p, \boldsymbol{D}+\boldsymbol{Y}'\boldsymbol{\Omega}\boldsymbol{Y}, G_{\boldsymbol{V}})$ depends on $G_{\boldsymbol{V}}$. Hence, the above integration is only proportional to $H(d, \boldsymbol{D}, G_{\boldsymbol{V}})/H(d+p, \boldsymbol{D}+\boldsymbol{Y}'\boldsymbol{\Omega}\boldsymbol{Y}, G_{\boldsymbol{V}})$ for a given graph $G_{\boldsymbol{V}}$.

Both steps (a) and (c) involve the evaluation of the ratio of the normalizing constant of an HIW distribution. For decomposable models, an analytical expression for the normalizing constant is available. For nondecomposable models, there is no analytical form for the normalizing constant. To estimate it, we use the Monte Carlo method proposed by Atay-Kayis and Massam (2005). Alternative methods to estimate this constant include those based on the importance sampler (Roverato, 2002; Dellaportas *et al.*, 2003).

The remaining substantial problem is how to effectively explore conditional posterior graphical model spaces $p(G_{\boldsymbol{U}} \mid \boldsymbol{V}, G_{\boldsymbol{V}})$ and $(G_{\boldsymbol{V}} \mid \boldsymbol{U}, G_{\boldsymbol{U}})$. In standard GGMs, MCMC is widely used to carry out model determination and there is no significant difference between specific MCMC samplers (Jones *et al.*, 2005). Although one can argue that alternative algorithms such as the shotgun stochastic search (Jones *et al.*, 2005) are more efficient than MCMC in identifying higher posterior graph regions, we focus our implementation on MCMC methods in all of our examples because of the following three critical reasons: First, stochastic search algorithms require the evaluation of the marginal likelihood density with $(\boldsymbol{U}, \boldsymbol{V})$ fully integrated out. Such an integration either requires Monte Carlo approximation in a decomposable case, or is simply computationally infeasible in a nondecomposable case for MGGM (Wang and West, 2009). In contrast, MCMC methods can avoid the complete integral over $(\boldsymbol{U}, \boldsymbol{V})$, and thus are feasible for unrestricted MGGM. Second, all existing stochastic search algorithms are developed for the stand-alone graphical model, that is, to learn the underlying graphs based on a sample of independent and identically distributed multivariate normal observations. Nevertheless, in many applications, GGM serves only as a modular structure for more elaborate models such as copula models, time-series models, mixture models, and the multivariate regression models developed in Section 6. MCMC is perhaps the only theoretically valid way to carry out model fitting in these situations. Third, we use the posterior median graph as our estimates of the true graph partly because the maximum a posteriori graph is impossible to obtain for MGGMs for reasons mentioned above. We find that multiple MCMC runs are able to provide similar posterior median graphs for both GGMs and MGGMs. One possible explanation is that MCMC is able to converge marginally on each individual edge inclusion probability, although it may not be able to find the overall highest posterior probability graphs. Thus, in the examples below, we use the simple local-move Metropolis-Hastings algorithm to sample from both $p(G_{\boldsymbol{U}} \mid \boldsymbol{V}, G_{\boldsymbol{V}})$ and $p(G_{\boldsymbol{V}} \mid \boldsymbol{U}, G_{\boldsymbol{U}})$.

4 Simulation

This simulation is concerned with the effect of a correlated sample on GGMs and the utility of modeling correlations among samples using MGGMs. It highlights the improvement in the quality of graphical model determination that can be obtained when correlations among samples are modeled rather than ignored if samples are truly correlated. It also shows the small loss of efficiency when modeling correlations that do not exist.

In all of these simulations, one single data matrix, $\boldsymbol{Y}$, is simulated from the matrix normal $N(0, \boldsymbol{U}, \boldsymbol{V})$, and is of size $p=100$ and $n=50$. Two structured covariances for $\boldsymbol{U}$ and four structured covariances for $\boldsymbol{V}$ are considered:

1. $\boldsymbol{U}_1$: autocorrelation matrix of the AR(2) model with AR parameters 0.7 and 0.2.

2. $\boldsymbol{U}_2$: block diagonal with block size 10. Within the block, $\boldsymbol{U}$ is set to be the constant correlation matrix $0.5\boldsymbol{J}_{10}+(1-0.5)\boldsymbol{I}_{10}$, where $\boldsymbol{J}_{10}$ is the 10×10

matrix of 1's and $\boldsymbol{I}_{10}$ is the identity matrix of order 10.

3. $\boldsymbol{V}_1$: identity matrix.

4. $\boldsymbol{V}_2$: autocorrelation matrix of the AR(1) model with AR parameter 0.5.

5. $\boldsymbol{V}_3$: block diagonal with block size 10. Within the block, $\boldsymbol{V}$ is set to be the constant correlation matrix $0.5\boldsymbol{J}_{10}+(1-0.5)\boldsymbol{I}_{10}$.

6. $\boldsymbol{V}_4$: block diagonal with block size 25. Within the block, $\boldsymbol{V}$ is set to be the constant correlation matrix $0.5\boldsymbol{J}_{25}+(1-0.5)\boldsymbol{I}_{25}$.

For each simulated data matrix $\boldsymbol{Y}$, we have applied unrestricted GGMs and MGGMs with several prior hyperparameter specifications and several initial graphs. In this paper, we report results for only one such specification, that is, $b=d=3$, $\boldsymbol{B}=3\boldsymbol{I}_{100}, \boldsymbol{D}=3\boldsymbol{I}_{50}$, and starting from empty graphs. The relative performance of MGGMs to GGMs is similar across the choices of initial graphs and hyperparameters. Results are obtained with 500 000 runs for GGMs and 500 000 sweeps after 100 000 burn-ins for MGGMs. Here one sweep consists of steps (a) and (b) in the algorithm of Section 3.2: one edge update in $G_{\boldsymbol{U}}$, one update of all elements in $\boldsymbol{U}$, one edge update in $G_{\boldsymbol{V}}$, and one update of all elements in $\boldsymbol{V}$.

To compare GGMs and MGGMs, we focus on their performances on the ability to recover existing edges of graph $G_{\boldsymbol{U}}$. This is a binary classification task where GGMs and MGGMs can be regarded as two classifiers. The results of a binary classification experiment can be summarized by four measures: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). According to the posterior distributions of $G_{\boldsymbol{U}}$, we compute the true positive rate (TPR), defined as $\mathrm{TPR}=\mathrm{TP}/(\mathrm{TP}+\mathrm{FN})$, and the positive predictive value (PPV), defined as $\mathrm{PPV}=\mathrm{TP}/(\mathrm{TP}+\mathrm{FP})$. TPR is the proportion of true discovered edges out of all true edges, and is equal to the power of the classifier. PPV is the proportion of true discovered edges out of all discovered edges, and is equal to one minus false discovery proportion. We use the posterior median probability graph as our estimate of the $G_{\boldsymbol{U}}$. A graph is called the posterior median probability graph, if each of its edges has a marginal edge inclusion probability greater than 0.5.

The entire process was repeated 50 times for each of $(\boldsymbol{U}_i, \boldsymbol{V}_j)$ $(i=1,2; j=1,2,3,4)$ and for both GGMs and MGGMs. Fig. 1 displays the boxplots of the TPR and PPV achieved by the different classifiers from GGMs and MGGMs. When correlations among samples truly exist, Figs. 1c–1h and Figs. 1k–1p show that allowing for the correlation structure among samples substantially and consistently improves the accuracy of the estimates of $G_{\boldsymbol{U}}$ in terms of TPR and PPV. In all cases where samples are correlated, MGGM is around 2% more efficient than GGM, as measured by TPR. For PPV, GGM fares particularly poorly compared with MGGM. For example, when $\boldsymbol{U}=\boldsymbol{U}_2$ and $\boldsymbol{V}=\boldsymbol{V}_3$, the median PPV is 78% for GGM and 90% for MGGM. The explanation is that when there actually exists sample dependence, 'leakage' between correlations in $\boldsymbol{U}$ and $\boldsymbol{V}$ can cause GGMs to signal false positive edges between variables if correlations among samples are not considered.

When correlations among samples do not exist, Figs. 1a and 1b and Figs. 1i and 1j show that there is a slight deterioration in the TPR for the MGGM classifier compared to the GGM procedure. This is expected because the samples are actually independent (i.e., $\boldsymbol{V}=\boldsymbol{I}_{50}$) and MGGMs also estimate $\boldsymbol{V}$. The loss in the efficiency in terms of TPR is about 3% for $\boldsymbol{U}_1$ (Fig. 1a) and 0.5% for $\boldsymbol{U}_2$ (Fig. 1i) for the median datasets.

5 Gene expression example

GGMs are frequently used to build networks from gene expression data. The majority of the modeling and computational techniques for exploring gene networks are developed under the assumption of independence among the samples. This independence assumption can be unrealistic in practice, as discussed in Efron (2009) and Allen and Tibshirani (2010; 2012), among others. This example illustrates that the correlations among arrays can exist and can be captured by MGGMs. It also shows that the graphs from MGGMs differ substantially from GGMs.

The breast cancer data (Jones *et al.*, 2005; Castelo and Roverato, 2006) contain $p=150$ genes related to the estrogen receptor pathway and $n=49$ samples. We first repeated the GGM analysis of Jones *et al.* (2005) under the assumption that there is no array correlation (i.e., $\boldsymbol{V}=\boldsymbol{I}_{49}$) and under the priors $b=3$ and $\boldsymbol{B}=4\boldsymbol{I}_{150}$. The

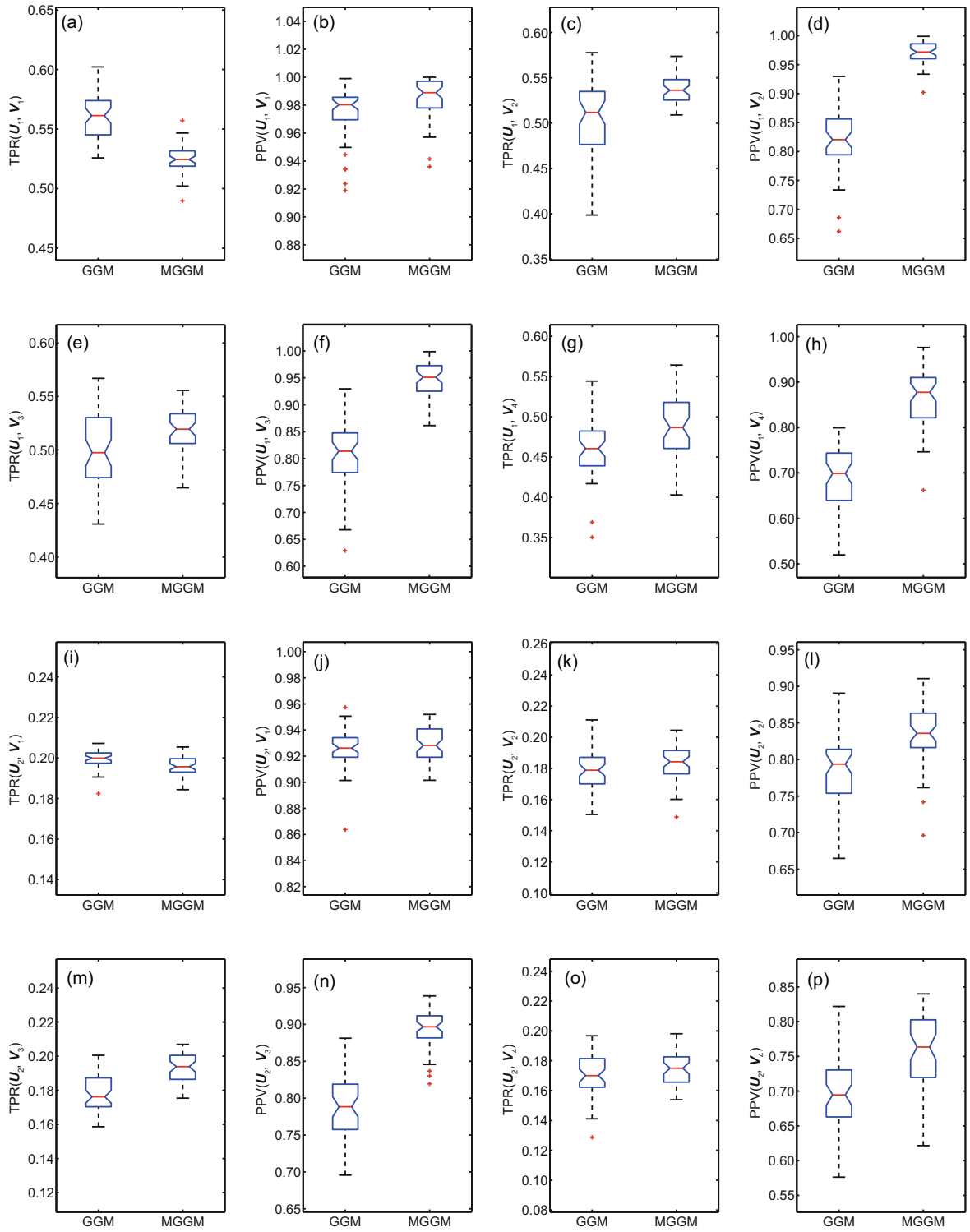

Fig. 1 Comparison of the Bayesian estimators of G_U for the TPR and PPV in the simulation study. Shown are box plots of the TPR and PPV over 100 simulations and for two estimators, GGM and MGGM, for different U and V matrices. The two U matrices are: U_1, autocorrelation matrix; U_2, block diagonal matrix. The four V matrices are: V_1, identity matrix; V_2, autocorrelation matrix; V_3, block diagonal with block size 10; V_4, block diagonal with block size 25

unrestricted Metropolis-Hastings stochastic search algorithm was run 10 million steps and the posterior median probability graph was used to estimate $G_{\boldsymbol{U}}$. The covariance matrix $\boldsymbol{U}$ was then estimated by $\hat{\boldsymbol{U}} = E_{G_{\boldsymbol{U}}}(\boldsymbol{U} \mid b+n, b+\boldsymbol{YY}')$ where the expectation was taken with respect to $\mathrm{HIW}_{G_{\boldsymbol{U}}}(b+n, b+\boldsymbol{YY}')$ for the posterior median probability graph $G_{\boldsymbol{U}}$ (Rajaratnam *et al.*, 2008). We then computed the column-wise correlations of the data matrix $\boldsymbol{Y}_{\mathrm{de}\boldsymbol{U}} = \hat{\boldsymbol{U}}^{-1/2}\boldsymbol{Y}$. Under the assumption $\boldsymbol{V} = \boldsymbol{I}_{49}$, $\boldsymbol{Y}_{\mathrm{de}\boldsymbol{U}}$ must be approximately distributed as $N(0, \boldsymbol{I}_{150}, \boldsymbol{I}_{49})$, and the distribution of the column-wise correlations of $\boldsymbol{Y}_{\mathrm{de}\boldsymbol{U}}$ must be similar to the distribution of the column-wise correlations from $N(0, \boldsymbol{I}_{150}, \boldsymbol{I}_{49})$. Fig. 2a shows the histogram of all 1176 column-wise correlations of $\boldsymbol{Y}_{\mathrm{de}\boldsymbol{U}}$, which is slightly over-dispersed than $N(0, \boldsymbol{I}_{150}, \boldsymbol{I}_{49})$ suggests: the standard deviation is estimated to be 0.11 rather than 0.8, which is numerically calculated from $N(0, \boldsymbol{I}_{150}, \boldsymbol{I}_{49})$.

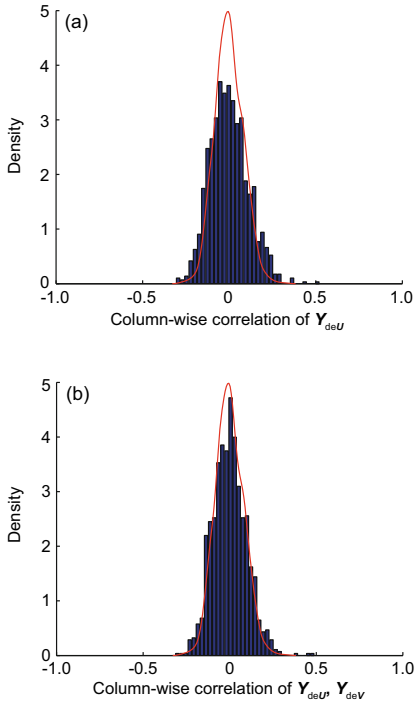

Fig. 2 Gene expression study. (a) Histogram of column-wise correlations of $\boldsymbol{Y}_{\mathrm{de}\boldsymbol{U}}$; (b) Histogram of column-wise correlations of $\boldsymbol{Y}_{\mathrm{de}\boldsymbol{U},\mathrm{de}\boldsymbol{V}}$. The solid curve represents numerically estimated density of the distribution of column-wise correlations from $\boldsymbol{N(0, I_{150}, I_{49})}$

Now we apply MGGMs with $b = d = 3$ and $\boldsymbol{B} = 4\boldsymbol{I}_{150}, \boldsymbol{D} = 4\boldsymbol{I}_{49}$ to the same data set. The model was estimated using 10 million sweeps for both burn-in and inference, and this gave agreement to two decimal places of the estimates of the edge inclusion probability of $G_{\boldsymbol{U}}$ when graphs were initialized from different starting positions. The code was written in C++. On a six-core 2.4 GHz CPU desktop computer, it took about 5 d to complete the computation. Fig. 3 summarizes the estimation results of $\boldsymbol{V}$. Fig. 3a shows the estimated posterior mean of correlations among arrays. Fig. 3b shows the histogram plot of all 1176 between-array correlations. A fair amount of correlations seem to exist between a few pairs of arrays. For example, array 1 is correlated with arrays 2, 3, and 5 with estimated correlations being 0.30, 0.33, and 0.19, respectively. Overall, 99% of pairs of all correlations are within 0.1, indicating that the majority of the arrays in this microarray data set are close to independence. The posterior mean estimates of $\hat{\boldsymbol{U}}$ and $\hat{\boldsymbol{V}}$ are then used to compute the de-correlated $\boldsymbol{Y}$ given by $\boldsymbol{Y}_{\mathrm{de}\boldsymbol{U},\mathrm{de}\boldsymbol{V}} = \hat{\boldsymbol{U}}^{-1/2}\boldsymbol{Y}\hat{\boldsymbol{V}}^{-1/2}$. The column-wise correlations of the $\boldsymbol{Y}_{\mathrm{de}\boldsymbol{U},\mathrm{de}\boldsymbol{V}}$ are compared with the distribution of the column-wise correlations from $N(0, \boldsymbol{I}_{150}, \boldsymbol{I}_{49})$ (Fig. 2b). $\hat{\boldsymbol{U}}$ and $\hat{\boldsymbol{V}}$ have de-correlated $\boldsymbol{Y}$ quite satisfactorily. In comparison with GGMs, the standard deviation of all 1176 column-wise correlations is now estimated to be 0.09, close to 0.08 which is numerically calculated from $N(0, \boldsymbol{I}_{150}, \boldsymbol{I}_{49})$.

To investigate the impact of array correlations, we compared the edge inclusion probabilities from GGMs with those from MGGMs. Overall, there are 160 and 104 edges that have probability 0.5 or higher in GGMs and MGGMs, respectively. Among the 160 edges in GGMs, 76 edges have probability 0.5 or higher in MGGMs. Fig. 4 displays the adjacency matrices of the graphs estimated from GGMs (Fig. 4a) and MGGMs (Fig. 4b). These adjacency matrices are computed from the median probability graph. These two estimated graphs differ substantially in terms of sparsity and overall pattern.

Finally, we conducted a series of robustness checks to verify that the difference in estimation performance between MGGM and GGM is not driven by the initial values of graphs or the model fitting algorithms. To do so, we first ran multiple MCMCs starting from different initial values of $G_{\boldsymbol{U}}$ for GGM (or $(G_{\boldsymbol{U}}, G_{\boldsymbol{V}})$ for MGGM), including the best decomposable graphs found from the initial 10 million

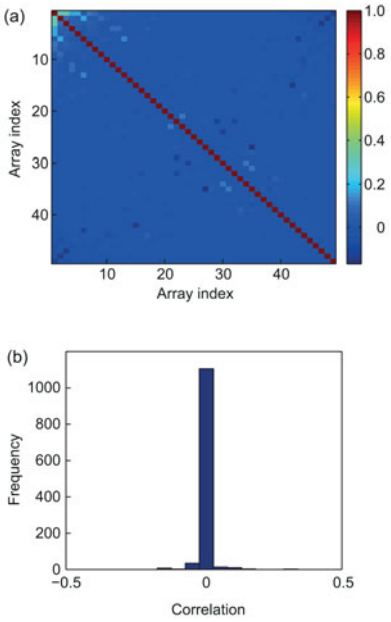

Fig. 3 The estimated correlation matrix (a) and the histogram of all correlations among arrays (b)

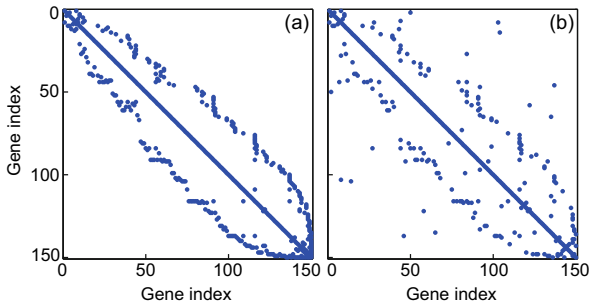

Fig. 4 Comparison of adjacency matrices of median probability graphs produced by GGMs (a) and MGGMs (b). Graph (a) has 160 edges and graph (b) has 104 edges

runs. Within either GGM or MGGM, the posterior median probability graphs resulting from multiple runs are relatively similar to each other—they differ only at fewer than 10 estimated edges. On the other hand, the number of differing edges between posterior median probability graphs of G_U estimated from GGM and from MGGM is always greater than 100. This indicates that the initial value cannot explain the main difference in the estimation performance between MGGM and GGM. Next, we show that the estimation difference between MGGM and GGM is not because of the difference in the modeling fitting algorithm. The unnormalized posterior mass of the posterior median probability graph G_U under GGM is around -9352. Jones *et al.* (2005) reported that the unnormalized posterior masses of the best decomposable graph and non-decomposable graphs found by their search algorithm are -9418 and -9228, respectively. Our result is comparable to theirs. In contrast, the posterior median probability graph G_U estimated by MGGM has an unnormalized posterior mass of around -9623 when it is evaluated under GGM. This indicates that GGM is unable to find the graph that is found by MGGM and it is indeed the model difference rather than the algorithm difference that generates the difference in the estimates.

6 Extensions to multivariate linear regression models with applications to mutual fund evaluation

6.1 Multivariate linear models with correlated errors

Multivariate linear regression (MLR) models provide a diverse and popular set of tools for modeling and predicting multivariate dependent variables that often arise in areas such as economics and finance (Wooldridge, 2001). Let $\boldsymbol{y}_t$ be a p-vector of dependent variables observed at time t, and let $\boldsymbol{X}_t$ be a $p \times k$ matrix of independent variables observed at time t. The multivariate linear model can be expressed as

$$\boldsymbol{y}_t = \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{e}_t, \quad t = 1, 2, \cdots, T, \tag{10}$$

where $\boldsymbol{\beta}$ is the k-vector of unknown regression coefficients, and $\boldsymbol{e}_t$ is the p-vector of unobserved random residuals. The assumptions about the correlation structure of the residual $\{\boldsymbol{e}_t\}$ are critical to the accuracy of statistical inferences about the $\boldsymbol{\beta}$. For example, the ordinary least squares (OLS) estimator of the standard errors is unbiased only when the residuals are independent and identically distributed. On the other hand, when the residuals are correlated, OLS can produce biased results. Many econometric methods have been proposed for adjusting the bias resulting from the correlated structure of $\{\boldsymbol{e}_t\}$. For a review of some of the popular methods, we refer to Petersen (2009). These methods often rely on asymptotic theory and may not be suitable for small-sample problems. Our basic model in Section 3 does not require any asymptotic theory, providing a new and attractive approach for estimating $\boldsymbol{\beta}$ when the residuals are both temporally correlated

and cross-sectionally correlated. The extension of the basic model in Section 3 to the MLR setting Eq. (10) is straightforward: Let the $p \times T$ residual matrix $(\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_T)$ follow the matrix-normal distribution $N(0, \boldsymbol{U}, \boldsymbol{V})$. The two covariance matrices $(\boldsymbol{U}, \boldsymbol{V})$ then automatically capture the time-series correlations and the cross-sectional correlations. We jointly estimate $(\boldsymbol{\beta}, \boldsymbol{U}, \boldsymbol{V})$ by extending the MCMC in Section 3.2 to incorporate an additional step for simulating $\boldsymbol{\beta}$. The posterior distribution of $\boldsymbol{\beta}$ then naturally takes account of the residual correlations.

To illustrate the utility of our methods, we consider an important finance application of Eq. (10) to evaluate mutual fund performance. Specifically, we will demonstrate how the new methods improve the estimation of a mutual fund's 'alpha', which is perhaps the most common measure of a mutual fund's historical performance. Traditionally, a mutual fund's 'alpha' is estimated via fitting a linear regression $y_{0,t} = \alpha_0 + \boldsymbol{x}_t'\boldsymbol{\beta}_0 + \epsilon_{0,t}$, for $t = 1, 2, \cdots, T$, where the vector $\boldsymbol{x}_t$ represents the benchmark returns, $y_{0,t}$ is the return of the mutual fund under evaluation, and α_0 is the fund's performance measure 'alpha'. The benchmark assets are often chosen according to asset pricing models. For example, the classical capital asset pricing model (CAPM) (Sharpe, 1964) indicates that the benchmark is the aggregated market return. A new method to estimate a fund's alpha was proposed by Pástor and Stambaugh (2002) based on the idea that nonbenchmark passive assets are also important in accurately estimating a fund's alpha. Their evaluation models can be expressed as the following multivariate linear regression:

$$\begin{cases} y_{0,t} = \alpha_0 + \boldsymbol{x}_t'\boldsymbol{\beta}_0 + \epsilon_{0,t}, \\ y_{i,t} = \alpha_i + \boldsymbol{x}_t'\boldsymbol{\beta}_i + \epsilon_{i,t}, \quad i = 1, 2, \cdots, p, \end{cases} \tag{11}$$

where $(y_{1,t}, y_{2,t}, \ldots, y_{p,t})$ are p nonbenchmark passive returns, $\boldsymbol{x}_t = (x_{1,t}, x_{2,t}, \cdots, x_{k,t})$ are k benchmark returns, and $\boldsymbol{\epsilon}_t = (\epsilon_{0,t}, \epsilon_{1,t}, \cdots, \epsilon_{p,t})$ is assumed by Pástor and Stambaugh (2002) to be correlated contemporaneously and not autocorrelated. That is, $\mathrm{cov}(\boldsymbol{\epsilon}) \equiv \mathrm{cov}(\boldsymbol{\epsilon}_1', \boldsymbol{\epsilon}_2', \cdots, \boldsymbol{\epsilon}_T')' = \boldsymbol{I}_{\mathrm{T}} \otimes \boldsymbol{U}$. But this temporal independence assumption is likely to be violated, as the monthly asset returns often exhibit short-term auto-correlations. Our matrix normal framework will remedy this issue by modeling both temporal and cross-sectional correlations in $\{\boldsymbol{\epsilon}_t\}$. That is, we assume the residuals follow $\mathrm{cov}(\boldsymbol{\epsilon}) \equiv \mathrm{cov}(\boldsymbol{\epsilon}_1', \boldsymbol{\epsilon}_2', \cdots, \boldsymbol{\epsilon}_T')' = \boldsymbol{V} \otimes \boldsymbol{U}$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are unknown covariance matrices characterizing the cross-sectional and the temporal correlations, respectively.

6.2 Vanguard managed funds

We applied our methods to 10 actively managed Vanguard mutual funds. We used their monthly returns available from the Center for Research in Security Prices (CRSP). The sample period is from Jan. 2004 through Dec. 2008, a total of five-year time span. The benchmark asset is the aggregated market returns (MKT). The nonbenchmark assets are: the Fama-French factors, i.e., the small-minus-big (SMB) and the high-minus-low (HML), the momentum factor (MOM), and the five industrial sector portfolios, IP1, IP2, IP3, IP4, and IP5, a total of eight nonbenchmark passive assets. An explanation of these variables is given in the caption of Fig. 5.

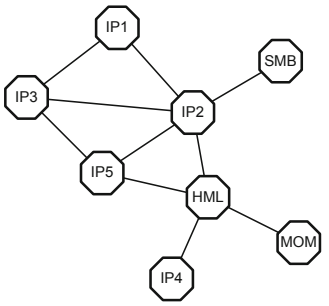

Fig. 5 Posterior median probability graph of errors of nonbenchmark assets from the analysis of Vanguard funds. The eight nonbenchmark assets are: the two Fama-French factors, i.e., the return spread on small and big market capitalization portfolios (SMB) and the return spread of high and low book-to-market ratio portfolios (HML), the momentum factor (MOM), and the five industrial portfolios which are: (1) durables, nondurables, wholesale, retail, and some services (IP1), (2) manufacturing, energy, and utilities (IP2), (3) business equipment, telephone and television transmission (IP3), (4) health care, medical equipment, and drugs (IP4), and (5) others such as mines, construction, hotels, entertainment, and finance (IP5)

For each of the 10 funds, we evaluated its performance over the five-year time span using the eight nonbenchmark assets. Thus, in the notation of Eq. (10), we have $T = 60$, $k = 1$, and $p = 8 + 1 = 9$. We chose standard non-informative priors for $\boldsymbol{\beta}$, $\boldsymbol{U}$, and $\boldsymbol{V}$ and ran the MCMC sampler for 20 000 steps. The posterior median probability graph of the eight

nonbenchmark assets is pictured in Fig. 5. As can be seen, this is not an empty graph, suggesting that there is a considerable amount of residual dependence structures among the eight nonbenchmark assets after controlling the benchmark asset. The benchmark does not explain all the cross-sectional correlation structures in the nonbenchmark assets.

Table 1 displays the estimates of α_0 for each of the 10 funds using OLS and our multivariate linear regression (MLR) method. First, we notice that the MLR estimates of α_0 differ substantially from the OLS estimates. This indicates that a mutual fund's performance can be estimated inconsistently across different models. Comparing the standard errors, we see that the MLR produces smaller standard errors than the OLS method, which is consistent with the founding in Pástor and Stambaugh (2002). Finally, we compare the standard error with the estimate of a fund's alpha to address the central question in mutual evaluation, i.e., whether the actively managed fund generates active returns that cannot be explained by the passive returns. If a manager indeed has the skill to outperform the benchmarks, then his/her fund's alpha must be significantly positive. Unfortunately, Table 1 shows that most α_0's are within two standard errors around 0, and thus these funds' performance is exclusively explained by the passive benchmark assets. Only the fund Dividend Growth and the fund Equity-Income appear to add value beyond the benchmark returns during this sample period.

Table 1 Monthly α_0 estimated using ordinary least squares (OLS) and multivariate linear regression (MLR) models with correlated errors from the analysis of Vanguard funds

Fund name	α_0		Standard error	
	OLS	MLR	OLS	MLR
Cap Opp	0.27	0.23	0.23	0.24
Dividend Growth	0.16	0.31*	0.12	0.15
Equity-Income	0.06	0.30*	0.14	0.12
Explorer	−0.03	−0.04	0.19	0.11
Growth & Income	−0.16	−0.12	0.09	0.12
Growth Equity	−0.25	−0.09	0.22	0.18
Mid Cap Growth	0.11	0.14	0.21	0.20
Morgan Growth	−0.07	−0.10	0.12	0.12
PRIMECAP	0.24	0.21	0.14	0.20
Selected Value	0.14	−0.01	0.20	0.25

* An estimated α_0 that is two standard errors away from 0

7 Discussion

This paper develops and applies modeling and computational methods for improved estimation of GGMs for correlated data using information of sample dependence. GGMs ignore information from possibly complex correlations between samples. The sample dependence can distort the variable dependence structure if sample dependence is not considered. As a method of solving these problems, we have introduced a matrix-variate Gaussian graphical model (MGGM) that can simultaneously model dependence among variables and samples.

Using a simulation, we demonstrate that modeling sample dependence can substantially improve the model estimation performance in terms of reducing the number of false positive and negative edges when samples are correlated. When samples are not correlated, the loss of efficiency for considering sample dependence is small. Using gene expression data analysis, we demonstrate that the assumption of array independence can be questionable in GGMs. In contrast, MGGMs can effectively capture dependence among both genes and arrays. The graph estimated using MGGM is significantly different from that estimated using GGMs. Using a mutual fund evaluation example, we demonstrate that the methods can be useful for better analyzing panel data that commonly occur in economics and finance.

Our new framework opens doors for further research in developing methods for correlated samples in more complicated situations. One interesting research direction is the generalization of the single matrix normal models to mixtures of matrix normals to allow heterogeneity in the samples. Another is to extend the basic model to more elaborate models as well as to incorporate prior information. The Bayesian structural vector autoregressive models of Kociecki *et al.* (2012) provide one interesting example that can be integrated with our proposed methods. We leave the exploration of these new ideas to future research.

Acknowledgements

The authors thank the anonymous referees and Dr. Rafal WOJAKOWSKI for very useful comments and suggestions.

References

Allen, G.I., Tibshirani, R., 2010. Transposable regularized covariance models with an application to missing data imputation. *Ann. Appl. Stat.*, **4**(2):764-790. [doi:10.1214/09-AOAS314]

Allen, G.I., Tibshirani, R., 2012. Inference with transposable data: modelling the effects of row and column correlations. *J. R. Stat. Soc. B*, **74**(4):721-743. [doi:10.1111/j.1467-9868.2011.01027.x]

Atay-Kayis, A., Massam, H., 2005. The marginal likelihood for decomposable and non-decomposable graphical Gaussian models. *Biometrika*, **92**(2):317-335. [doi:10.1093/biomet/92.2.317]

Carvalho, C.M., West, M., 2007. Dynamic matrix-variate graphical models. *Bayes. Anal.*, **2**(1):69-98. [doi:10.1214/07-BA204]

Castelo, R., Roverato, A., 2006. A robust procedure for Gaussian graphical model search from microarray data with p larger than n. *J. Mach. Learn. Res.*, **7**:2621-2650.

Dawid, A.P., Lauritzen, S.L., 1993. Hyper-Markov laws in the statistical analysis of decomposable graphical models. *Ann. Stat.*, **21**(3):1272-1317. [doi:10.1214/aos/1176349260]

Dellaportas, P., Giudici, P., Roberts, G., 2003. Bayesian inference for nondecomposable graphical Gaussian models. *Sankhya Ser. A*, **65**:43-55.

Efron, B., 2009. Are a set of microarrays independent of each other? *Ann. Appl. Stat.*, **3**(3):922-942.

Giudici, P., Green, P.J., 1999. Decomposable graphical Gaussian model determination. *Biometrika*, **86**(4):785-801. [doi:10.1093/biomet/86.4.785]

Jones, B., Carvalho, C., Dobra, A., Hans, C., Carter, C., West, M., 2005. Experiments in stochastic computation for high-dimensional graphical models. *Stat. Sci.*, **20**(4):388-400. [doi:10.1214/088342305000000304]

Kociecki, A., Rubaszek, M., Ca'Zorzi, M., 2012. Bayesian Analysis of Recursive SVAR Models with Overidentifying Restrictions. Eurosystem Working Paper Series, No. 1492.

Lauritzen, S.L., 1996. Graphical Models. Clarendon Press, Oxford.

Pástor, L., Stambaugh, R.F., 2002. Mutual fund performance and seemingly unrelated assets. *J. Financ. Econ.*, **63**(3):315-349. [doi:10.1016/S0304-405X(02)00064-8]

Petersen, M.A., 2009. Estimating standard errors in finance panel data sets: comparing approaches. *Rev. Financ. Stud.*, **22**(1):435-480. [doi:10.1093/rfs/hhn053]

Rajaratnam, B., Massam, H., Carvalho, C.M., 2008. Flexible covariance estimation in graphical Gaussian models. *Ann. Stat.*, **36**(6):2818-2849. [doi:10.1214/08-AOS619]

Roverato, A., 2002. Hyper-inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scand. J. Stat.*, **29**(3):391-411. [doi:10.1111/1467-9469.00297]

Sharpe, W.F., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. *J. Finance*, **19**(3):425-442. [doi:10.2307/2977928]

Wang, H., 2010. Sparse seemingly unrelated regression modelling: applications in finance and econometrics. *Comput. Stat. Data Anal.*, **54**(11):2866-2877. [doi:10.1016/j.csda.2010.03.028]

Wang, H., Li, S.Z.Z., 2012. Efficient Gaussian graphical model determination under G-Wishart prior distributions. *Electron. J. Stat.*, **6**:168-198. [doi:10.1214/12-EJS669]

Wang, H., West, M., 2009. Bayesian analysis of matrix normal graphical models. *Biometrika*, **96**(4):821-834. [doi:10.1093/biomet/asp049]

Wang, H., Reeson, C., Carvalho, C.M., 2011. Dynamic financial index models: modeling conditional dependencies via graphs. *Bayes. Anal.*, **6**:639-664. [doi:10.1214/11-BA624]

Wooldridge, J.M., 2001. Econometric Analysis of Cross Section and Panel Data. The MIT Press.

Zhang, G., Ferrari, S., Qian, M., 2009. An information roadmap method for robotic sensor path planning. *J. Intell. Robot. Syst.*, **56**(1-2):69-98. [doi:10.1007/s10846-009-9318-x]