J. R. Statist. Soc. B (2005)
67, *Part* 3, *pp.* 321–341

Structural learning with time-varying components: tracking the cross-section of financial time series

Makram Talih

Hunter College of the City University of New York, USA

and Nicolas Hengartner

Los Alamos National Laboratory, USA

[Received June 2003. Revised November 2004]

Summary. When modelling multivariate financial data, the problem of structural learning is compounded by the fact that the covariance structure changes with time. Previous work has focused on modelling those changes by using multivariate stochastic volatility models. We present an alternative to these models that focuses instead on the latent graphical structure that is related to the precision matrix. We develop a graphical model for sequences of Gaussian random vectors when changes in the underlying graph occur at random times, and a new block of data is created with the addition or deletion of an edge. We show how a Bayesian hierarchical model incorporates both the uncertainty about that graph and the time variation thereof.

Keywords: Covariance selection; Dynamic graphical model; Markov chain Monte Carlo methods; Markov random field; Multivariate stochastic volatility; Precision matrix

1. Introduction

Estimating the covariance structure of multivariate data is of fundamental importance in their statistical analysis. For instance, in empirical finance, the estimation of the variance–covariance matrix is central to asset pricing, portfolio selection and investment strategies based on historical indicators and market data. However, when modelling multivariate financial data, the problem of structural learning becomes compounded by the fact that the covariance structure changes with time. Previous work has focused on modelling those changes by using the so-called multivariate stochastic volatility models of Jacquier *et al.* (1994). In this paper, we present an alternative to those models that focuses instead on the hidden graphical structure that is related to the precision matrix. The latter is the inverse of the variance–covariance matrix, a natural parameter in the multivariate normal model, which is intimately related to the coefficients in the simultaneous regression of each variable on all the remaining variables.

This paper develops a graphical model for sequences of Gaussian random vectors when changes in the underlying graph, which is specified by zeros in the precision matrix, occur at random times, and a new block of data is created with the addition or deletion of an edge. We show how a Bayesian hierarchical model incorporates both the uncertainty about that graph and the time variation thereof. Our main objective is to learn the graph underlying the most current block of data. In fact, our framework allows us to make inference about the whole history

Address for correspondence: Makram Talih, Department of Mathematics and Statistics, Hunter College of the City University of New York, 695 Park Avenue, New York, NY 10021, USA.
E-mail: talih@math.hunter.cuny.edu

 1369–7412/05/67321

up to and including the last block. Dahlhaus and Eichler (2003) have developed a related class of models, termed time series graphs. For stationary time series, conditional independence is characterized in the frequency domain by zeros in the inverse spectral matrix. Using Gibbs potentials, Guyon and Hardouin (2002) developed another class of dynamic spatiotemporal models, which they called Markov chain Markov fields.

The plan of this paper is as follows. Section 2 gives some background. In Section 3, we develop the statistical model. Section 4 presents our posterior sampling strategy. In Section 5, we use a simulated data set as a test-case, and, in Section 6, consider a small data set consisting of five US industry portfolios. Section 7 discusses extensions to our methodology.

2. Covariance selection

As in Koster (1996), we define an *undirected graph* as a pair $g=(V, E)$ where V is a set of *vertices* $V=\{1,2,\ldots,d\}$ and E is a set of *(undirected) edges*

$$E \subset \{\{i, j\} : i \neq j, i \in V, j \in V\},$$

connecting (unordered) pairs of distinct vertices. Our graphs do not allow for loops from a vertex to itself, nor for multiple edges. Dempster (1972) has pioneered an approach to modelling the variance–covariance matrix in which some off-diagonal entries in its inverse are set to 0, thereby exploiting conditional (near) independence relations in the data. The resulting models are summarized by an undirected graph, whose vertices represent the variables, and whose edges indicate non-zero off-diagonal entries in the precision matrix.

2.1. Graphical Gaussian models

Let $X=(X_i)_{i\in V}$ be a d-dimensional normal random vector, indexed by the vertices of an undirected graph $g=(V, E)$. For subsets $U, W \subset V$, X_U denotes the subvector $X_U=(X_i)_{i\in U}$, and $U \setminus W$ is shorthand for $U \cap W^c$. The notation $X_i \perp\!\!\!\perp X_j | X_{V\setminus\{i,j\}}$ means that X_i and X_j are conditionally independent given $X_{V\setminus\{i,j\}}$. The undirected graphs that we consider are so-called *conditional independence graphs* (Whittaker (1990), page 60), in that

$$\{i, j\} \notin E \qquad \text{if and only if } X_i \perp\!\!\!\perp X_j | X_{V\setminus\{i,j\}}.$$

Let Σ be the variance–covariance matrix of X and $K=\Sigma^{-1}$ its precision matrix. Let the mean vector μ be zero. In Gaussian graphical models, the absence of an edge between two nodes is signalled by a zero off-diagonal entry in the precision matrix. The conditional independence relation $X_i \perp\!\!\!\perp X_j | X_{V\setminus\{i,j\}}$ is read off the corresponding entry k_{ij} of K, since the partial correlation coefficient between X_i and X_j given the remaining variables is simply

$$\rho_{ij} = -k_{ij}/\sqrt{(k_{ii}k_{jj})}, \tag{1}$$

which is 0 if and only if $k_{ij}=0$. Thus, for graphical Gaussian models,

$$\{i, j\} \notin E \qquad \text{if and only if } k_{ij}=0.$$

The log-likelihood for X is given by

$$\log\{f(x|K, g)\} = -\frac{d}{2}\log(2\pi) + \frac{1}{2}\log|K| - \frac{1}{2}\sum_{i\in V}\sum_{j\in V} x_i k_{ij} x_j.$$

Isolating the terms involving variable X_i in the quadratic form in the log-likelihood shows that the conditional distribution of X_i given the

$$X_{-i} \stackrel{\text{def}}{=} \{X_j, j \neq i\}$$

is Gaussian, with

$$\begin{aligned}\mathbf{E}(X_i|X_{-i}) &= -\sum_{j\neq i} \frac{k_{ij}}{k_{ii}} X_j = -\sum_{j:\{i,j\}\in E} \frac{k_{ij}}{k_{ii}} X_j \\ &= \mathbf{E}(X_i|X_{\{j:\{i,j\}\in E\}}),\end{aligned} \tag{2}$$

and

$$\text{var}(X_i|X_{-i}) = \frac{1}{k_{ii}}. \tag{3}$$

Relation (2) corresponds to the regression of X_i onto the X_{-i}, where the coefficients are

$$\beta_{ij} = -\frac{k_{ij}}{k_{ii}} \times \mathbf{1}_{\{\{i,j\}\in E, i\neq j\}}. \tag{4}$$

The multiple-regression coefficient of this regression is R_i^2, and the residual variance is

$$\frac{1}{k_{ii}} = \sigma_i^2(1 - R_i^2). \tag{5}$$

3. Model proposed

In this section, we set out to specify the data-generating mechanism, as well as the priors that are involved in developing a hierarchical Bayesian model.

3.1. A slowly varying sequence of graphs

Consider the sequence of graphs which arises in blocks as follows:

$$\underbrace{G_1, \ldots, G_1}_{N_1}, \underbrace{G_2, \ldots, G_2}_{N_2}, \ldots, \underbrace{G_B, \ldots, G_B}_{N_B}.$$

Each graph G_b is an undirected graph with fixed vertex set $V = \{1, \ldots, d\}$ and edge set E_b, and is repeated N_b times. In our model, the graphs G_b and G_{b+1} in successive blocks differ only by the addition or deletion of at most one edge. Let the corresponding edge sets be E_b and E_{b+1} respectively. The Hamming distance between G_b and G_{b+1} is the size of the symmetric difference

$$E_b \triangle E_{b+1} = (E_b \backslash E_{b+1}) \cup (E_{b+1} \backslash E_b)$$

between their edge sets:

$$d(G_b, G_{b+1}) = \#(E_b \triangle E_{b+1}) \leqslant 1.$$

To model time variation, we define Markov transition probabilities

$$\mathbf{P}(G_{b+1} = g | G_b = h) = \frac{1}{\#D(h)} \mathbf{1}_{\{g\in D(h)\}},$$

where

$$D(h) = \{g : d(h, g) \leqslant 1\},$$

and $\mathbf{1}_U$ is the indicator function of a set U. Note that $g \in D(h)$ if and only if $h \in D(g)$, and that $h \in D(h)$ for all h. The joint probability of a sequence

$$\mathbf{g} \stackrel{\text{def}}{=} (g_1, \ldots, g_B)$$

is thus

$$p(\mathbf{g}) = \frac{p_1(g_1)}{\#D(g_1)\ldots\#D(g_{B-1})} \times \prod_{b=1}^{B-1} \mathbf{1}_{\{g_{b+1} \in D(g_b)\}},$$

for some initial distribution $p_1(\cdot)$ for the graph G_1 in the first block. If we allow all graphs, then $\#D(g) = 1 + d(d-1)/2$ for all g, which does not depend on g. Further, we can take the initial distribution $p_1(\cdot)$ for the Markov chain to be the uniform distribution $p_1(\cdot) = 2^{-d(d-1)/2}$. This is clearly also the stationary distribution, since the marginals $p_b(\cdot)$ remain uniform. Note that the uniform distribution $p_1(\cdot)$ puts more mass on graphs with medium-sized edge sets, as opposed to very sparse or nearly saturated graphs. It can be thought of as the distribution of an Erdös–Rényi random graph with each of the $d(d-1)/2$ edges drawn (their indicators set to 1) independently according to a Bernoulli($\frac{1}{2}$) distribution.

The resulting joint prior $p(\mathbf{g})$ becomes a uniform distribution on

$$\{\mathbf{g} = (g_1, \ldots, g_B) : g_{b+1} \in D(g_b), b = 1, \ldots, B-1\}.$$

3.2. Data-generating mechanism

We view the data as a stream of d-dimensional observations

$$\underbrace{X_{1,1}, \ldots, X_{1,N_1}}_{N_1}, \underbrace{X_{2,1}, \ldots, X_{2,N_2}}_{N_2}, \ldots, \underbrace{X_{b,1}, \ldots, X_{b,N_b}}_{N_b}, \ldots,$$

where the block lengths $N_1, \ldots, N_b, \ldots$ are random. Conditionally on the number of blocks B, and a fixed sample size n, the last block observed has only

$$M_B \stackrel{\text{def}}{=} n - \sum_{b=1}^{B-1} N_b$$

observations, whence $M_B \leqslant N_B$. Henceforth, define $M_b \equiv N_b$ for $b = 1, \ldots, B-1$.

A sufficient representation of the full data is as follows, where $\mathbb{X} = (\mathbb{X}_1, \ldots, \mathbb{X}_B)$, $\mathbf{G} = (G_1, \ldots, G_B)$ and $\mathbf{M} = (M_1, \ldots, M_B)$:

$$(\mathbb{X}, \mathbf{G}, \mathbf{M}) = \{(\mathbb{X}_1, G_1, M_1), \ldots, (\mathbb{X}_B, G_B, M_B)\}.$$

$\mathbb{X}_b$ is a $d \times M_b$ matrix with column vectors $X_{b,1}, \ldots, X_{b,M_b}$. Each graph $G_b = (V, E_b)$ is repeated M_b times. Data in block b are assumed independent and identically distributed from $N_d(0, \Sigma_b)$, with the precision matrix $K_b = \Sigma_b^{-1} \in \mathcal{M}^+(G_b)$. Here, for any undirected graph $g = (V, E)$, the set $\mathcal{M}^+(g)$ is defined as the set of all symmetric and positive definite matrices A such that $A_{ij} = 0$ whenever $\{i, j\} \notin E$, $i \neq j$.

We condition on B throughout this paper, regarding it as a tuning parameter for the maximum number of blocks required. Our approach contrasts with that of Barry and Hartigan (1993), who ascertained the number of blocks B in a sequence of observations *a posteriori*.

3.2.1. Distribution of block lengths

The length of a block is the interarrival time between structural changes. We model the interarrival times $\mathbf{N} = (N_1, \ldots, N_B)$ as independent and identically distributed geometric random

variables. Conditionally on the sample size n and the number of blocks B, the vector of block lengths $\mathbf{M} = (M_1, \ldots, M_B)$, where $M_b = N_b$, for $b = 1, \ldots, B-1$, and $M_B = n - \Sigma_{b=1}^{B-1} M_b$, follows a uniform distribution $u(\mathbf{m})$ over the space of vectors

$$\left\{ \mathbf{m} = (m_1, \ldots, m_B) : 0 \leqslant m_b \leqslant n, \sum_{b=1}^{B} m_b = n \right\}.$$

Note that empty blocks are not ruled out.

3.3. Parameterization

For a precision matrix $K_b \in \mathcal{M}^+(G_b)$ and undirected graph $G_b = (V, E_b)$, we use two parameters θ_b and σ_b^2, and set regression coefficients (4) and conditional variances (3) to

$$\begin{aligned} \beta_{ij}^b = -\frac{k_{ij}^b}{k_{ii}^b} = \frac{\theta_b}{\nu_i^b} \mathbf{1}_{\{\{i,j\} \in E_b, i \neq j\}}, \\ \frac{1}{k_{ii}^b} = \frac{\sigma_b^2}{\nu_i^b} \end{aligned} \tag{6}$$

respectively. Here,

$$\nu_i^b \stackrel{\text{def}}{=} \max(1, \#\{j : \{i,j\} \in E_b\})$$

is the degree of vertex i in the graph when i is not a disconnected singleton and is equal to 1 when i is disconnected.

As a result of expression (6), the precision matrix K_b has off-diagonal elements

$$k_{ij}^b = -\frac{\theta_b}{\sigma_b^2} \mathbf{1}_{\{\{i,j\} \in E_b, i \neq j\}}. \tag{7}$$

Restricting $\theta_b \in (-1, 1)$ in equation (7) is sufficient to ensure that the resulting precision matrix K_b is positive definite, regardless of dimension d, as it makes K_b strictly diagonally dominant:

$$\sum_{j \neq i} \frac{|k_{ij}^b|}{k_{ii}^b} = \frac{|\theta_b|}{\nu_i^b} \sum_{j \neq i} \mathbf{1}_{\{\{i,j\} \in E_b\}} \leqslant |\theta_b| < 1.$$

However, the partial correlation coefficients for this model are given in equation (1) by

$$\begin{aligned} \rho_{ij}^b &= -\frac{k_{ij}^b}{\sqrt{(k_{ii}^b k_{jj}^b)}} \\ &= \frac{\theta_b}{\sqrt{(\nu_i^b \nu_j^b)}} \mathbf{1}_{\{\{i,j\} \in E_b, i \neq j\}}. \end{aligned} \tag{8}$$

Since both ν_i^b and ν_j^b are at least 1, it follows that the choice of $\theta_b \in (-1, 1)$ is also necessary.

3.3.1. Prior specification in the non-hierarchical model

For tying together the parameters in consecutive blocks, we consider first the case when $\theta_b \equiv \theta_1$ and $\sigma_b^2 \equiv \sigma_1^2$ for all b, and we choose the following convenience priors for θ_1 and σ_1^2:

$$\pi_1(\theta_1) \propto \frac{\pi/2}{\cos^2(\theta_1 \pi/2)}, \qquad \pi_1(\sigma_1^2) \propto \frac{1}{\sigma_1^2}. \tag{9}$$

The improper prior $\pi_1(\theta_1)$ corresponds to a uniform prior for $\tan(\theta_1\pi/2)$ and is U shaped.

3.3.2. Prior specification in the hierarchical model

A more realistic model allows the strength (8) of persistent edges, as well as volatility at the nodes, to vary 'smoothly' from one block to the next. We let the parameters (θ_b, σ_b^2) themselves be governed by appropriately defined autoregressive processes, in the spirit of the stochastic volatility models of Jacquier *et al.* (1994). For this, we define the initial distributions $\pi_1(\theta_1)$ and $\pi_1(\sigma_1^2)$ as in expression (9), and, for $b = 2, \ldots, B$, let

$$\begin{aligned} \tan(\theta_{b+1}\pi/2) &= \phi_\theta \ \tan(\theta_b \pi/2) + \varepsilon_{b+1}, \\ \log(\sigma_{b+1}^2) &= \phi_\sigma \ \log(\sigma_b^2) + \varepsilon'_{b+1}, \end{aligned} \tag{10}$$

where the ε_b are independent and identically distributed according to an $N\{(1-\phi_\theta)\alpha_\theta, \eta_\theta^2\}$ distribution and the ε'_b according to an $N\{(1-\phi_\sigma)\alpha_\sigma, \eta_\sigma^2\}$ distribution, for some hyperparameters

$$\psi_\theta \stackrel{\text{def}}{=} (\alpha_\theta, \phi_\theta, \eta_\theta^2), \qquad \psi_\sigma \stackrel{\text{def}}{=} (\alpha_\sigma, \phi_\sigma, \eta_\sigma^2).$$

Thus, our choice of joint prior for

$$\boldsymbol{\theta} \stackrel{\text{def}}{=} (\theta_1, \ldots, \theta_B)$$

is given by

$$\pi(\boldsymbol{\theta}|\psi_\theta) \propto \left\{ \prod_{b=1}^{B} \frac{\pi/2}{\cos^2(\theta_b \pi/2)} \right\} \exp\left[-\sum_{b=1}^{B-1} \frac{\{\tan(\theta_{b+1}\pi/2) - \phi_\theta \tan(\theta_b \pi/2) - (1-\phi_\theta)\alpha_\theta\}^2}{2\eta_\theta^2} \right], \tag{11}$$

whereas that for

$$\boldsymbol{\sigma}^2 \stackrel{\text{def}}{=} (\sigma_1^2, \ldots, \sigma_B^2)$$

is given by

$$\pi(\boldsymbol{\sigma}^2|\psi_\sigma) \propto \left(\prod_{b=1}^{B} \sigma_b^{-2} \right) \exp\left[-\sum_{b=1}^{B-1} \frac{\{\log(\sigma_{b+1}^2) - \phi_\sigma \log(\sigma_b^2) - (1-\phi_\sigma)\alpha_\sigma\}^2}{2\eta_\sigma^2} \right]. \tag{12}$$

The priors in expressions (11) and (12) encourage parameter values in successive blocks to be similar. The prior of independence, which is represented by the Jacobian (product) terms in each expression, is penalized by a 'smoothness' or similarity penalty (the exponential term) which assigns a higher weight to successive values of the parameters that are close.

3.3.3. Hyperparameters

In the hierarchical prior specification, an extra layer of parameters is added, to determine the autoregressive behaviour of the θs and the σ^2s in expression (10). For that matter, we choose a triangle function (with height 2 at 1, and 0 at –1) for the hyperparameters ϕ_θ and ϕ_σ, as we expect

a typically strong positive autoregressive coefficient. For the trend parameters α_θ and α_σ, we assume an $N(0, a^2)$ prior for some choice of a^2 (e.g. $a^2 = 1$). For the variance parameters η_θ^2 and η_σ^2, we simply take as prior a χ_1^2-distribution, which controls the persistence. Since we only allow changes in the parameters at the end of each block, and typically consider a small number of blocks, it is clear that estimation of the hyperparameters $\psi_\theta = (\phi_\theta, \alpha_\theta, \eta_\theta^2)$ and $\psi_\sigma = (\phi_\sigma, \alpha_\sigma, \eta_\sigma^2)$ is of no particular relevance, as we effectively have $B \ll n$ 'observed' parameters (θ_b, σ_b^2). We shall thus marginalize over ψ_θ and ψ_σ from the posterior.

3.3.4. Heterogeneous variance parameters

Within each block, model (6) forces the residual variances $1/k_{ii}^b$ to be the same across the diagonal, adjusted only by the degree of the vertex i in the graph G_b. However, from equation (5), the residual variance from the regression of X_i onto the X_{-i} is given by

$$\frac{1}{k_{ii}^b} = \sigma_{b,i}^2 (1 - R_{b,i}^2),$$

where $R_{b,i}^2$ is the multiple-regression coefficient from that regression. Thus, it is quite natural to parameterize the residual variances $1/k_{ii}^b$ by

$$\frac{1}{k_{ii}^b} = \frac{\sigma_{b,i}^2}{\nu_i^b} \tag{13}$$

instead, while modifying the parameterization of the regression coefficients as follows:

$$\beta_{ij}^b = -\frac{k_{ij}^b}{k_{ii}^b} = \frac{\theta_b}{\nu_i^b} \frac{\sigma_{b,i}}{\sigma_{b,j}} \mathbf{1}_{\{\{i,j\} \in E_b, i \neq j\}}. \tag{14}$$

This is especially relevant for the industry portfolio returns data that we analyse in Section 6, since some industries tend to be more volatile, or more sensitive to market shocks, than others. As a result of equation (14), the precision matrix K_b has off-diagonal elements

$$k_{ij}^b = -\frac{\theta_b}{\sigma_{b,i} \sigma_{b,j}} \mathbf{1}_{\{\{i,j\} \in E_b, i \neq j\}}.$$

4. Posterior inference and sampling strategy

Let $f(\mathbb{X}|\mathbf{g}, \mathbf{m}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2)$ denote the likelihood function. The joint posterior distribution is

$$q(\mathbf{g}, \mathbf{m}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2 | \mathbb{X}) \propto f(\mathbb{X}|\mathbf{g}, \mathbf{m}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2)\, p(\mathbf{g})\, u(\mathbf{m}) \int \pi(\boldsymbol{\theta}|\psi_\theta)\, \xi(\psi_\theta)\, \mathrm{d}\psi_\theta \int \pi(\boldsymbol{\sigma}^2|\psi_\sigma)\, \xi(\psi_\sigma)\, \mathrm{d}\psi_\sigma.$$

Our aim is to draw a sample from this joint posterior distribution by using the Metropolis–Hastings algorithm. We start at an arbitrary point in the state space and perturb the state variable one component at a time, marginalizing over the hyperparameters ψ_θ and ψ_σ. We refer the reader to Robert and Casella (1999) for more on Markov chain Monte Carlo methods. We describe how to generate proposals next, one component of the state variable at a time.

4.1. Perturbing the graph sequence

4.1.1. The e-toggle

Let $g = (V, E)$ be an undirected graph with vertex set V and edge set E. Let E^c denote the edge set of the graph $g^c = (V, E^c)$ given by

$$\{i, j\} \in E^c, i \neq j \qquad \text{if and only if } \{i, j\} \notin E, i \neq j.$$

We define the set $\mathcal{E}$ of all possible edges as the set of subsets of $E \cup E^c$ of cardinality 0 or 1. Thus, the empty set (of size 0) corresponds to the null edge, which we denote by o. We include explicitly the null edge here for computational convenience. It follows that $\#\mathcal{E} = 1 + d(d-1)/2$, regardless of g.

Starting with a graph $g = (V, E)$, select an edge $e \in \mathcal{E}$, and define a new graph $h = (V, E')$ with vertex set V and edge set E' given by

$$E' = \begin{cases} E \cup \{e\}, & \text{if } e \in E^c, \\ E \backslash \{e\}, & \text{if } e \in E, \\ E, & \text{if } e = o. \end{cases}$$

We call this operation a *toggle* and call the graph h that is thus obtained the *e-toggle* of the graph g (relative to the edge e). We denote by $h = g \pm e$ the *e-toggle* of g. By a slight abuse of notation, we also denote the new edge set E' by $E' = E \pm e$.

4.1.2. Legalese

If all graphs are allowed, then

$$\begin{aligned} D(g) &= \{h : d(g,h) \leqslant 1\} \\ &= \{h : h = g \pm e, \text{for some } e \in \mathcal{E}\}. \end{aligned}$$

The subset $\mathcal{E}(g)$ of edges in $\mathcal{E}$ for which $g \pm e$ is in $D(g)$ are called the legal edges. Here, $\mathcal{E}(g) \equiv \mathcal{E}$ for all g; hence $\#D(g) = \#\mathcal{E}(g) \equiv \#\mathcal{E} = 1 + d(d-1)/2$. However, when only some subclass $\mathcal{G}$ of the graphs is allowed, then $\mathcal{E}(g)$ is usually a proper subset of $\mathcal{E}$, defined in such a way as to ensure that $\mathcal{G}$ is closed under the e-toggle operation:

$$h \pm e = g \in \mathcal{G}, e \in \mathcal{E}(g) \qquad \text{if and only if } g \pm e = h \in \mathcal{G}, e \in \mathcal{E}(h).$$

Of historical interest is the subclass $\mathcal{G}$ of *triangulated graphs*; see Dirac (1961) and Gavril (1972). An important result is that the subclass $\mathcal{G}$ of triangulated graphs can be traversed via the toggle operation (Frydenberg and Lauritzen, 1989). We shall not be concerned with triangulated graphs any further here; however, we point out that our modelling and subsequent inference strategy are well adapted to such graphs.

4.1.3. The toggle move on a sequence of graphs

Let the current state of the Metropolis chain be a sequence of graphs $\mathbf{g} = (g_1, \ldots, g_B)$, whose common vertex set is V, and whose edge sets $E_1, \ldots, E_B$ satisfy

$$E_{b+1} = E_b \pm e_{b+1}, \qquad b = 1, \ldots, B-1,$$

where $\{e_{b+1}\} = E_{b+1} \triangle E_b$. We now describe a strategy for generating a new sequence of graphs $\mathbf{h} = (h_1, \ldots, h_B)$ which will form a proposed state for the Metropolis chain. We seek to maintain the property that consecutive graphs h_b and h_{b+1} differ by at most one edge. For that, we simply choose an edge $e \in \mathcal{E}$ such that $e \in \mathcal{E}(g_b)$ for all b, and, as illustrated in Figs 1(a) and 1(b), define the sequence $\mathbf{h} = (h_1, \ldots, h_B)$ by

$$h_b = g_b \pm e, \qquad b = 1, \ldots, B. \tag{15}$$

The edge sets E'_b of the h_b are such that the relative changes of the E_b are conserved:

$$\begin{aligned} E'_{b+1} \triangle E'_b &= (E_{b+1} \pm e) \triangle (E_b \pm e) \\ &= E_{b+1} \triangle E_b = \{e_{b+1}\}. \end{aligned}$$

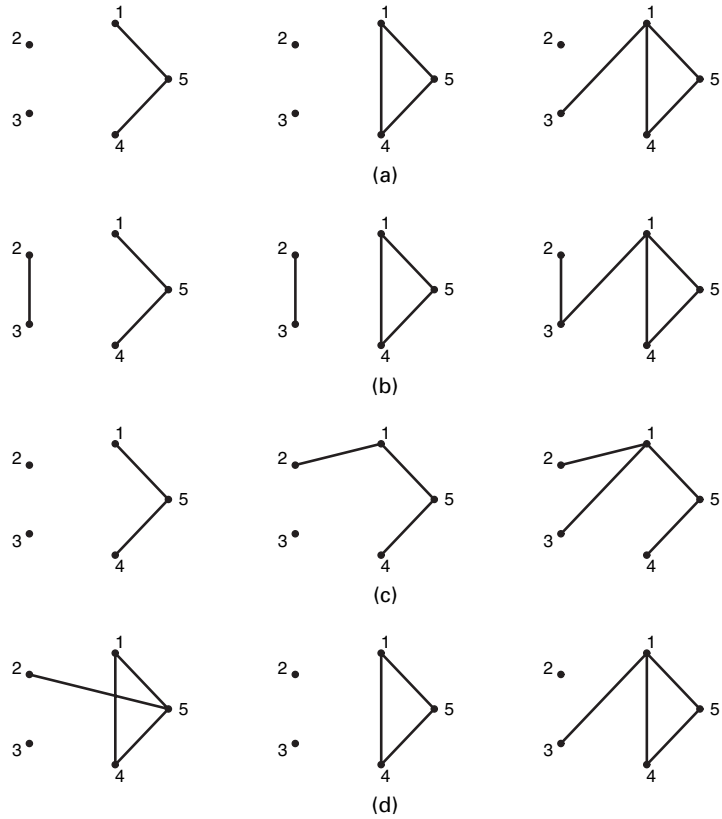

Fig. 1. (a) Current sequence, (b) toggle move, (c) fork move, blocks 2–3, and (d) fork move, block 1: the toggle move in (b) adds the edge $\{2,3\}$ across the whole sequence; the fork move in (c), operating from left to right, keeps the first graph anchored, toggling both $\{1,2\}$ and $\{1,4\}$; the fork move in (d), operating from right to left, keeps graphs 3 and 2 anchored, toggling edges $\{2,5\}$ and $\{1,4\}$

Note that, conditionally on the current state $\mathbf{g}$ of the Metropolis–Hasting chain, once the edge $e \in \cap_{b=1}^{B} \mathcal{E}(g_b)$ has been chosen, the proposed state $\mathbf{h}$ is entirely determined by equation (15). Hence, the acceptance rate for the toggle move is given by

$$\alpha(\mathbf{g},\mathbf{h}) = \min\left\{1, \frac{f(\mathbb{X}|\mathbf{h},\mathbf{m},\boldsymbol{\theta},\boldsymbol{\sigma}^2)\ p(h)}{f(\mathbb{X}|\mathbf{g},\mathbf{m},\boldsymbol{\theta},\boldsymbol{\sigma}^2)\ p(g)} \frac{\#\cap_{b=1}^{B}\ \mathcal{E}(g_b)}{\#\cap_{b=1}^{B}\ \mathcal{E}(h_b)} \mathbf{1}_{\{h_b = g_b \pm e, e \in \mathcal{E}(g_b) \cap \mathcal{E}(h_b), b=1,\ldots,B\}}\right\}.$$

When all graphs are allowed, i.e. when $\mathcal{E}(g) \equiv \mathcal{E}$ for all g, the toggle move is symmetric: hence the proposal ratio is equal to 1. This, together with $p(\mathbf{g})$ being uniform, implies that only the likelihood ratio enters the calculation of the acceptance probability $\alpha(\mathbf{g},\mathbf{h})$.

4.1.4. The fork move on a sequence of graphs

The toggle move that we have just described can easily be extended. For $b_0 > 1$, suppose that we want to leave the first $b_0 - 1$ blocks unchanged and to modify the remaining blocks $b_0, \ldots, B$. Now, apply the toggle move on the subsequence $(g_{b_0}, \ldots, g_B)$ of $\mathbf{g}$. This defines a new sequence $\mathbf{l}$ with $l_b = g_b$ for $b = 1, \ldots, b_0 - 1$ and $l_b = g_b \pm e$, with $e \in \mathcal{E}(g_b)$ for $b = b_0, \ldots, B$. But, by toggling only a subsequence of $\mathbf{g}$, the edge sets of the graphs l_{b_0-1} and l_{b_0} now differ by two edges: $\{e_{b_0}\} \equiv E_{b_0} \triangle E_{b_0-1}$ and $\{e\} \equiv (E_{b_0} \pm e) \triangle E_{b_0}$. As a remedy, we (un)toggle the edge $e' \equiv e_{b_0}$.

Note that, conditionally on $\mathbf{g}$ and on the edge e, the edge e' is determined: $e' \equiv e_{b_0}$, and e' must lie in $\mathcal{E}(l_b)$ for all $b = b_0, \ldots, B$. Thus, we obtain the acceptance rate

$$\alpha(\mathbf{g}, \mathbf{h}) = \min\left\{1, \frac{f(\mathbb{X}|\mathbf{h}, \mathbf{m}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2)\ p(\mathbf{h})}{f(\mathbb{X}|\mathbf{g}, \mathbf{m}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2)\ p(\mathbf{g})} \frac{\#\cap_{b=b_0}^{B}\ \mathcal{E}(g_b)}{\#\cap_{b=b_0}^{B}\ \mathcal{E}(h_b)} \mathbf{1}_{\{h_b = (g_b \pm e) \pm e', e' \in \mathcal{E}(g_b \pm e), e \in \mathcal{E}(g_b), b \geqslant b_0\}}\right\}.$$

As noted earlier, when all graphs are allowed, only the likelihood ratio enters the calculation of $\alpha(\mathbf{g}, \mathbf{h})$, which is appealing. We name the toggle–toggle move just described a *fork* move.

4.2. Perturbing the times between changes

Proposals on the vector of block lengths can be made locally in the sense that only two components of $\mathbf{m}$ are changed at any given time. Indeed, suppose that some $b \in \{1 \ldots B-1\}$ is drawn at random. Then, a symmetric proposal $\mathbf{m}'$ can be constrained only by

$$m'_b + m'_{b+1} = m_b + m_{b+1}. \tag{16}$$

We generate proposals by selecting a range of values (m'_b, m'_{b+1}) with the constraint (16), and sampling one such proposal uniformly from that range. We have found that taking too small a range of possible new values (m'_b, m'_{b+1}) results in almost sure acceptance of the proposal, whereas taking too large a range results in unduly low acceptance probabilities. Thus, the range of proposals that are considered is adjusted to have a moderate acceptance rate, and thence satisfactory mixing. We call this move on the vector of interarrival times a 'shift' move. Indeed, it consists of shifting forwards or backwards one of the arrival times (poles), while keeping the arrival time just before and the arrival time just after unchanged.

4.3. Perturbing the parameter matrix

With the parameterization (6) of K_b in terms of (θ_b, σ_b^2), there are two cases to consider.

4.3.1. Non-hierarchical case

In the simplest case, $\theta_b \equiv \theta$ and $\sigma_b^2 \equiv \sigma^2$ for all b. We make symmetric proposals on the tangent scale for θ according to

$$w(\theta_{\text{new}}|\theta_{\text{old}}) = (2\pi\zeta_\theta^2)^{-1/2} \frac{\pi/2}{\cos^2(\theta_{\text{new}}\pi/2)} \exp\left[-\frac{\{\tan(\theta_{\text{new}}\pi/2) - \tan(\theta_{\text{old}}\pi/2)\}^2}{2\zeta_\theta^2}\right],$$

and on the logarithmic scale for σ^2 according to

$$w(\sigma^2_{\text{new}}|\sigma^2_{\text{old}}) = (2\pi\zeta_\sigma^2)^{-1/2} \frac{1}{\sigma^2_{\text{new}}} \exp\left[-\frac{\{\log(\sigma^2_{\text{new}}) - \log(\sigma^2_{\text{old}})\}^2}{2\zeta_\sigma^2}\right],$$

where the proposal parameters ζ_θ and ζ_σ are tuning parameters. In the application that is explored in this paper, we have found that the values $\zeta_\theta = 0.9$ and $\zeta_\sigma = 0.25$ lead to acceptable convergence rates for the Metropolis chain.

4.3.2. Hierarchical case

In the more flexible hierarchical case, the parameters θ_b and σ_b^2 are allowed to vary from block to block, as described in Section 3.3. We make proposals for the components of the sequences $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_B)$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_B^2)$ that are symmetric on the tangent scale and logarithmic scale respectively. As in the previously described fork move, we can choose to perturb only part of the parameter sequence, leaving the other part unchanged.

4.3.3. Perturbing the hyperparameters

For the autoregressive coefficients ϕ_θ and ϕ_σ, we make symmetric proposals on the tangent scale. For the variance parameters η_θ^2 and η_σ^2, we work with symmetric proposals on the logarithmic scale. For the trend parameters α_θ and α_σ, we simply make symmetric proposals on the natural scale. The variances of such proposal densities are adjusted to have a fast drop in the autocorrelation function of the resulting Metropolis chain. We have found it more efficient to update those hyperparameters jointly as $\psi_\theta = (\phi_\theta, \alpha_\theta, \eta_\theta^2)$ and $\psi_\sigma = (\phi_\sigma, \alpha_\sigma, \eta_\sigma^2)$, since they tend to be correlated *a posteriori*, and, for the application at hand, have found it adequate to use an $N(0, \varsigma_\theta^2)$ proposal distribution with $\varsigma_\theta = 1.3$, for ψ_θ, and an $N(0, \varsigma_\sigma^2)$ proposal distribution with $\varsigma_\sigma = 1.1$, for ψ_σ.

5. Simulation example

To test the performance of our methodology, we simulate a data set according to the hierarchical prior specification, allowing for parameters to change at the end of each block.

We generate the data by using the graphs in Fig. 2, and the following parameters:

$$(\theta_1, \sigma_1^2) = (0.85, 3),$$
$$(\theta_2, \sigma_2^2) \approx (0.87, 7.4),$$
$$(\theta_3, \sigma_3^2) \approx (0.85, 5.3).$$

These values are obtained by starting with $(\theta_1, \sigma_1^2) = (0.85, 3)$ and drawing subsequent parameters according to the autoregression in expression (10), with hyperparameters set to

$$\psi_\theta = (\alpha_\theta, \phi_\theta, \eta_\theta^2) = (4, 0.75, 0.25),$$
$$\psi_\sigma = (\alpha_\sigma, \phi_\sigma, \eta_\sigma^2) = (1, 0.75, 0.25).$$

There are three blocks in our data here, of lengths $M_1 = 400$, $M_2 = 200$ and $M_3 = 400$.

We set $B = 4$ and study how our methodology recovers the structural changes in the data, with one more block than necessary. Talih (2003) contains further test cases which the reader can consult. After 600 burn-in loops, which correspond to 100 iterations in each component of the state vector $(\mathbf{g}, \mathbf{m}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2, \psi_\theta, \psi_\sigma)$, we carry out 120 000 sampling loops, to have marginal posterior samples of size 20 000 in each component. Convergence and mixing of the Metropolis chain were diagnosed by using trace and autocorrelation function plots. The acceptance rates for our fork, shift and parameter moves ranged from 10% to 22%, with the exception of an acceptance rate of 7% for the variance parameters $\boldsymbol{\sigma}^2$.

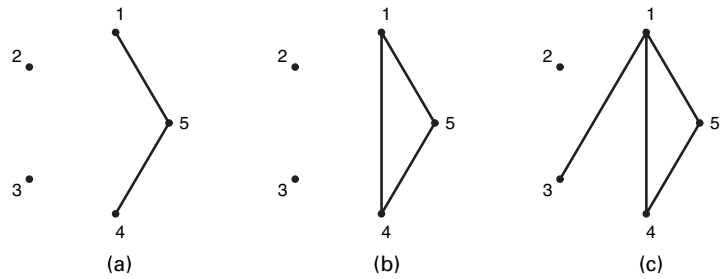

Fig. 2. Test case—1000 independent observations are drawn from a multivariate normal distribution with mean vector zero and precision matrices in each block constrained by the graphs shown: (a) observations 1–400; (b) observations 401–600; (c) observations 601–1000

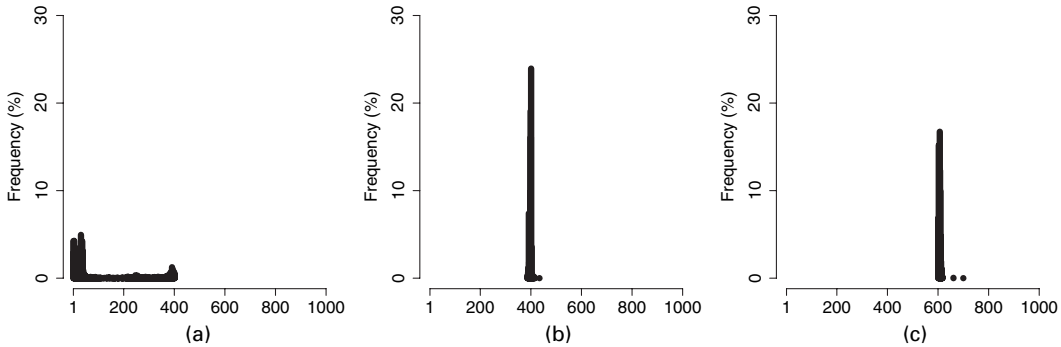

Fig. 3. Test case—marginal posterior distributions for the first three changepoints: (a) posterior position of pole 1 (highest posterior density, 30; mean, 80); (b) posterior position of pole 2 (highest posterior density, 400; mean, 398); (c) posterior position of pole 3 (highest posterior density, 606; mean, 606)

Fig. 3 shows the marginal posterior distributions for the position of each pole. The posterior mode for the first changepoint is at 30, matched closely by modes near 0, which seem to imply a very small, if not empty, first block. The second changepoint is detected at 400, which is where the edge $\{1, 4\}$ is added to form the complete subgraph $\{1, 4, 5\}$. The third changepoint is detected at 606, which is almost exactly where the second structural change occurs in the data with the addition of the edge $\{1, 3\}$.

Histograms of the marginal posterior densities of the parameters are not shown here. However, both the first change in volatility from 3 to 7.4 and that from 7.4 to 5.3 are detected. The 95% highest posterior density regions for σ_4^2, σ_3^2, σ_2^2 and σ_1^2 are $[5.05, 5.75]$, $[6.23, 7.89]$, $[2.68, 4.23]$ and $[1.25, 6.05]$ respectively, with posterior modes at 5.25, 6.81, 3.03 and 1.95 respectively. In fact, the posterior density for σ_1^2 turns out to be bimodal, with a second mode around 3, and skewed to the right, which explains the low acceptance rate of 7% that was noted earlier. Also, the correlation parameters are well approximated *a posteriori*. The 95% highest posterior density regions for θ_4, θ_3, θ_2 and θ_1 are $[0.83, 0.88]$, $[0.86, 0.91]$, $[0.82, 0.88]$ and $[0.81, 0.95]$ respectively, with posterior modes at 0.86, 0.89, 0.85 and 0.91. The posterior density for θ_1 is also bimodal, with a second mode at 0.85.

Fig. 4 shows *a posteriori* most probable sequences of graphs $\mathbf{g} = (g_1, g_2, g_3, g_4)$. For each one of the sequences presented, the graphs in the last three blocks are indeed the true graphs. In fact, for the first block, all 11 possible neighbours of the graph in the second block were sampled; they are not all shown here. Thus, we can marginalize over the first block, according to the probabilities that are quoted, to obtain, with posterior probability 1, the true graph sequence. Note that the 4-sequence with highest posterior probability remains the sequence where the last three graphs are the true graphs, and the first graph is identical to the second. This is evidence towards merging blocks 1 and 2.

6. US industry portfolios data

For this paper, we choose to work at the portfolio level using five industries, with annually rebalanced value-weighted monthly returns from July 1926 to December 2001, providing 906 months of data. The five industry portfolios are given in Table 1. This data set was obtained from the virtual data library of Kenneth French at Dartmouth: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. We work with log-returns so the assumption of normality is not grossly violated. Log-returns have, in theory,

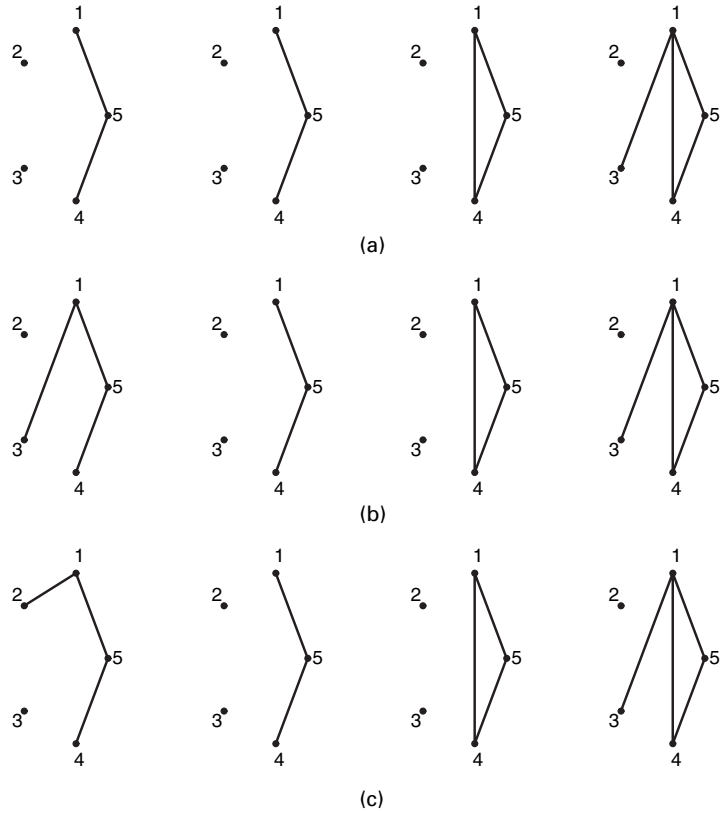

Fig. 4. Test case—*a posteriori* most probable sequences of graphs, with matching probabilities: (a) probability 0.863; (b) probability 0.029; (c) probability 0.024

Table 1. Five industry portfolios, with corresponding main industry groups

Portfolio	*Industry group*†
F	Finance (60–69)
M	Manufacturing (20–39)
S	Shops: wholesale and retail (50–59); some services (70–79)
U	Utilities (49)
O	Other: agriculture, mines, oil, construction and transportation, telecommunications, health and legal services

†The standard industrial classifications are given in parentheses.

mean 0. In fact, the 5-vector of industry-specific sample means, using the full sample from July 1926 to December 2001, had entries ranging from 0.69% to 0.84%, which are practically 0. None-the-less, we centre the observations around their sample mean to focus on the precision matrix.

We shall not address, here, the issues that are involved in modelling both the mean process and the precision matrices jointly. Finance theory, in essence, relates the first and the second moments of asset returns; therefore one should not assume stochastic volatility while assuming constant, or zero, means. Campbell *et al.* (1997), section 12.2.3, reviewed some of the most com-

mon models of changing volatility that link first and second moments. Aguilar and West (2000) developed a class of dynamic linear factor models to capture both the multivariate stochastic volatility and the factor structure of financial time series. In Talih (2004), we account for such modelling advances in the context of structural learning through precision matrices.

6.1. Preliminary findings

Linear factor models and clustering are the most commonly used statistical techniques for covariance modelling in the context of stock returns; see for example Brown *et al.* (1997). Graphical modelling is an alternative that is most suited to the study of the precision matrix.

A naïve approach to determining the graph structure underlying multivariate stock returns, during a fixed time period, is as follows. The inverse of the sample covariance matrix estimates $K = \Sigma^{-1}$, and sample partial correlations whose absolute values fall below an arbitrary cut-off level are set to 0. For instance, we may choose a threshold of 20% to account for multiple testing, as well as to avoid obtaining an overly complex graph. Note that a Bonferroni-type correction may yield a different threshold. However, the determination of the graph structure from mere truncation of the sample precision matrix is not without serious drawbacks. One such difficulty is the correlation between sample partial correlation coefficients. Indeed, as discussed in Cox and Wermuth (1990),

$$\operatorname{cov}(\hat{k}_{i_1,j_1}, \hat{k}_{i_2,j_2}) = k_{i_1,i_2} k_{j_1,j_2} + k_{i_1,j_2} k_{i_2,j_1}.$$

Another problem is that partial correlation coefficients, indeed the whole graph, change with time. For instance, take a moving window of 151 months, shifting forwards every month. The observations within each period are assumed independent. Within each window, the underlying graph is determined by setting to 0 any sample partial correlation whose absolute value falls

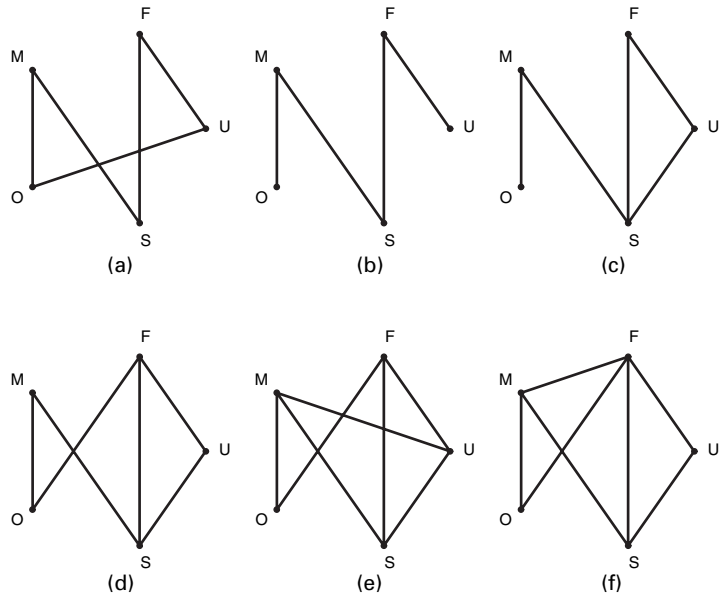

Fig. 5. Naïve determination of the time-varying sequence of graphs (only a subsequence of the graphs that were obtained is shown here): (a) February 1982–May 1986; (b) June 1986–July 1987; (c) August 1987–October 1991; (d) November 1991–October 1992; (e) November 1992–July 1997; (f) August 1997–April 1998

below the threshold of 20%. Consecutive graphs that are identical are grouped into blocks, and the duration of each block is indicated. Fig. 5 presents a subsequence of the graphs thus obtained, demonstrating a slowly varying graphical structure.

In Section 3 of this paper, we have developed a hierarchical Bayesian framework for dealing with both the uncertainty about and the time variation of the graph structure that is evident in this industry portfolio returns data. Our model resolves the logical inconsistency of using moving window sample estimates to determine the associated graphs.

6.2. Data analysis

We fix the maximum number of desired blocks at $B = 9$ here, to have roughly one structural change for every decade between July 1926 and December 2001. In contrast with the simulation example of Section 5, we let each industry carry its own variance parameter, since industries tend to react differently to shocks and changes to the economy. Indeed, using the full sample from July 1926 and December 2001 reveals heterogeneous sample variances, which are 0.47%, 0.33%, 0.26%, 0.39% and 0.34% respectively for portfolios F, M, O, S and U. Thus, we adopt the parameterization that is described by equations (13) and (14).

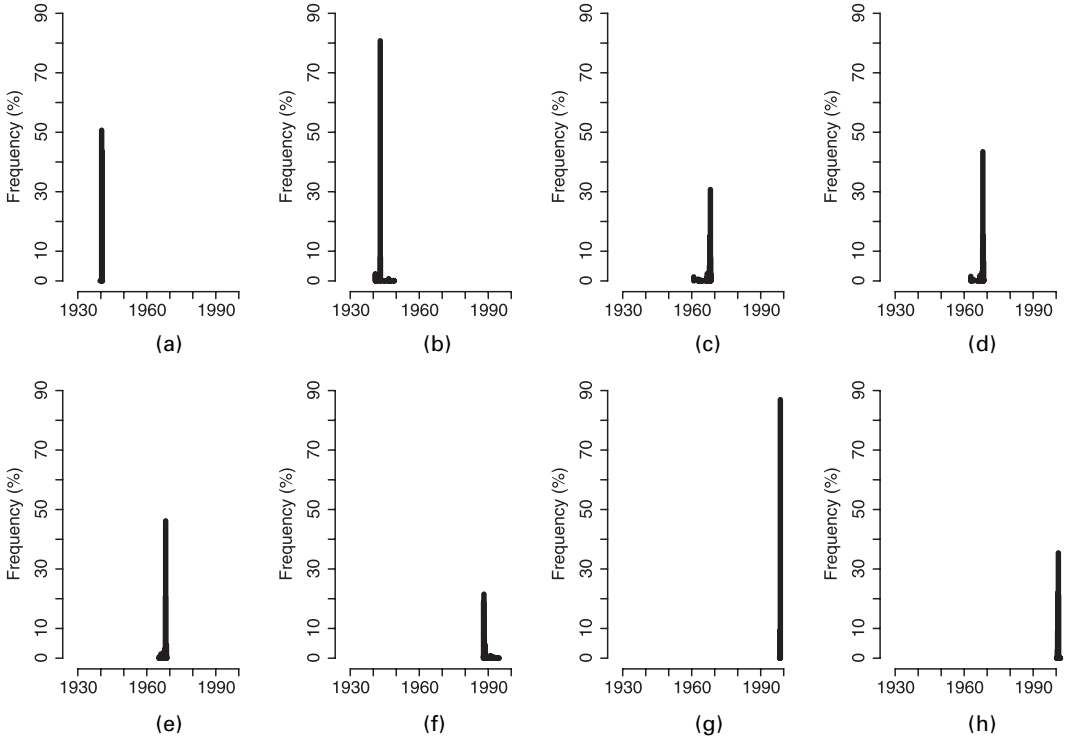

Fig. 6. Industry portfolios data—marginal posterior distributions for each of the eight changepoints; (a) posterior position of pole 1 (highest posterior density, May 1940; mean, June 1940); (b) posterior position of pole 2 (highest posterior density, January 1943; mean, December 1942); (c) posterior position of pole 3 (highest posterior density, February 1968; mean, October 1967); (d) posterior position of pole 4 (highest posterior density, February 1968; mean, December 1967); (e) posterior position of pole 5 (highest posterior density, March 1968; mean, January 1968); (f) posterior position of pole 6 (highest posterior density, February 1988; mean, July 1988); (g) posterior position of pole 7 (highest posterior density, July 1998; mean, July 1998); (h) posterior position of pole 8 (highest posterior density, December 2000; mean, November 2000)

After 600 burn-in loops, corresponding to 100 iterations in each component of the state variable $(\mathbf{g}, \mathbf{m}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2, \psi_\theta, \psi_\sigma)$, we carry out 300 000 sampling loops, to obtain marginal posterior samples of size 50 000 in each component. The acceptance rates for our fork, shift and parameter moves ranged from 7% to 24%, with the exception of the acceptance rates of 5% and 1% for the hyperparameters ψ_θ and ψ_σ respectively.

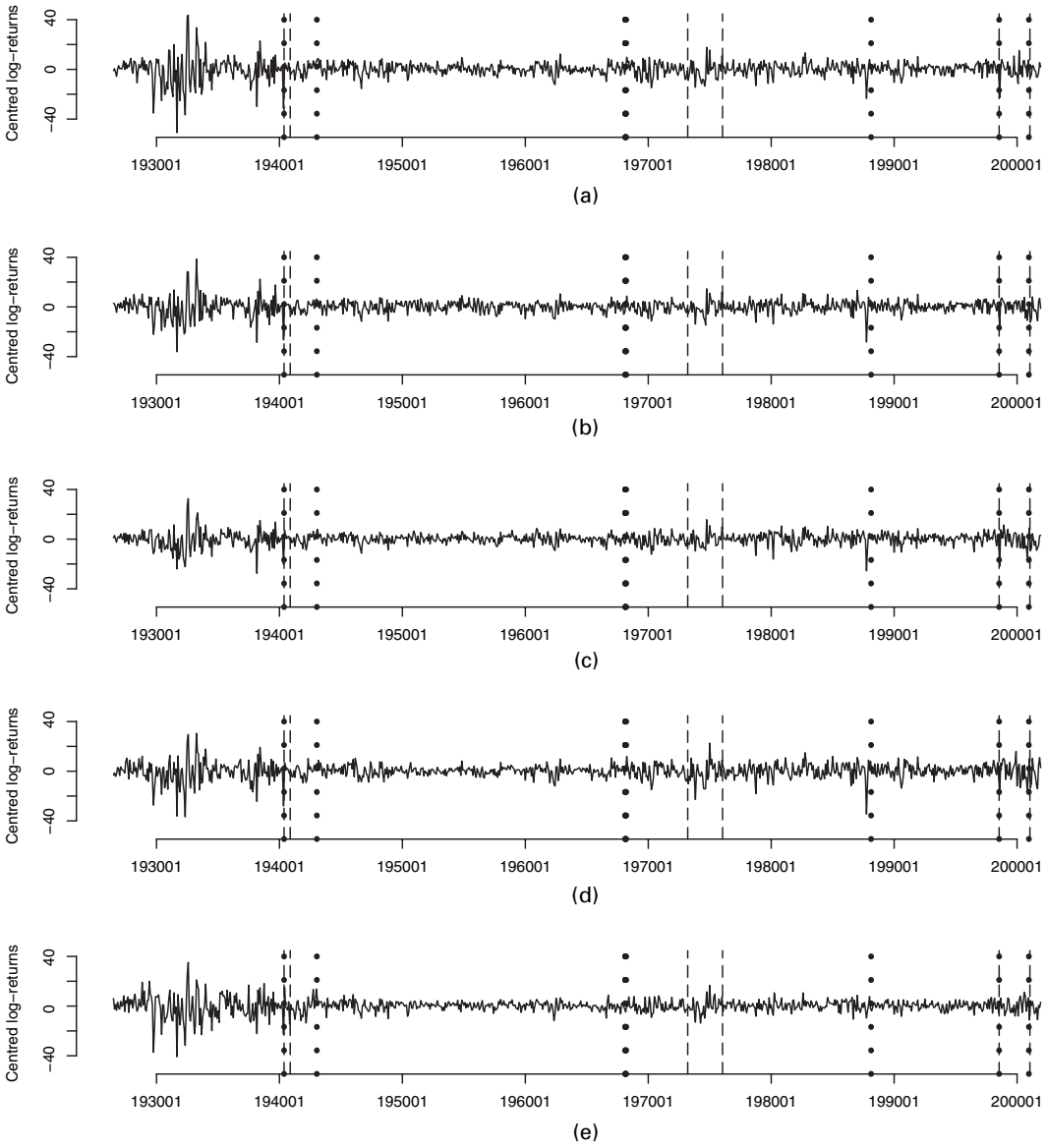

Fig. 7. Time series of centred log-returns by industry portfolio (when $B = 9$ (⋮), the posterior modes of the dates at which the changepoints occur are May 1940, January 1943, February 1968, March 1968, February 1988, July 1998 and December 2000; when $B = 7$ (⋮) they are at May 1940, November 1940, March 1973, January 1996, July 1998 and January 2001; the log-returns are on a scale of 1/100): (a) F; (b) M; (c) O; (d) S; (e) U

Fig. 6 shows the marginal posterior distributions for the position of each changepoint, having conditioned on a maximum number of $B=9$ blocks. The posterior modes of the dates at which the changes occur are as follows: May 1940, January 1943, February 1968, February 1968, March 1968, February 1988, July 1998 and December 2000. Note that the third, fourth and fifth poles lie on top of each other, which suggests merging them into one changepoint, thus possibly using seven blocks.

The highest posterior density blocks resulting from $B=9$ are shown in Fig. 7, which plots the actual series of centred log-returns in time. As seen from Fig. 7, some of the changepoints that have been detected mark shifts in volatility across the whole market.

However, it is evident from Fig. 7 that the different industries react in varying degrees to such shifts of regime, thereby affecting the covariance structure as a whole. This has been our viewpoint, which we can further bolster by performing simple likelihood ratio tests for the null hypothesis that the covariance matrices Σ_b and Σ_{b+1} in consecutive blocks are equal. Let Σ be their common value under the null hypothesis. Let m_b and m_{b+1} be the (highest posterior density) lengths of blocks b and $b+1$. The sample covariance matrices for the centred log-returns are given by $S_b = \mathbb{X}_b \mathbb{X}_b^{\mathrm{T}}/(m_b-1)$ and $S_{b+1} = \mathbb{X}_{b+1}\mathbb{X}_{b+1}^{\mathrm{T}}/(m_{b+1}-1)$. Ignoring the graphical modelling for now, the statistic

$$-2\log(M) = (m_b + m_{b+1} - 2)\log|S_b + S_{b+1}| - (m_b-1)\log|S_b| - (m_{b+1}-1)\log|S_{b+1}|$$

is approximately distributed as $(1-c)^{-1}\chi^2_{d(d+1)/2}$, where

$$c = \frac{2d^2+3d-1}{6(d+1)}\left(\frac{1}{m_b-1} + \frac{1}{m_{b+1}-1} - \frac{1}{m_b+m_{b+1}-2}\right).$$

This is the modified Wilks likelihood ratio test and is reviewed in Seber (1984). The results of this test lend support to our viewpoint, as seen from Table 2.

Running the same analysis with $B=7$ instead of $B=9$, the highest posterior density block lengths for the last two blocks become 30 and 11, instead of 29 and 12 respectively. This change makes the last test statistic in Table 2 equal to 36 instead of 23, which has an approximate P-value of 0% instead of 8%. The main effect of going from $B=9$ to $B=7$ (and indeed to $B=6$) is that the changepoint that is detected in the 1980s when $B=9$ is now shifted back to the mid-1970s, as shown in Fig. 7.

Interestingly, our procedure detects some historical events that have had an influence on the

Table 2. Results of the modified Wilks likelihood ratio tests for equality of Σ_b and Σ_{b+1}, with $B=9$

b	$\chi^2_{d(d-1)/2}$	*P-value (%)*
1	80	0
2	113	0
3†	202	0
6	62	0
7	202	0
8	23	8

†Block 3 is compared directly with block 6.

US economy, such as the American involvement in World War II (1940) and the Asian financial crisis (1997–1998). A more detailed account can be found in Talih (2003).

As shown in Fig. 7, most changepoints percolate across the whole market, typically indicating wide changes in volatility, although, on inspection of the estimated posterior densities of the industry-specific variance parameters, which are not shown here, some shifts in volatility are especially salient for industries such as finance (F) or manufacturing (M). Such uneven changes in volatility are accompanied by structural changes in the underlying graph, as seen in the *a posteriori* most likely sequences of graphs in Figs 8 and 9 which, together, account for 92% posterior probability.

In closing, some remarks are in order.

The grouping of companies into five industry portfolios is the coarsest classification that we can work with. A finer industry classification, into perhaps 17 or more industry portfolios, can bring about more specifically those industries which are most affected by shocks and changes to the economy.

Data sampled at a higher frequency than monthly, perhaps daily, or even intradaily data, will offer a more relevant time frame for the analysis that we propose, at least from the practitioner's viewpoint.

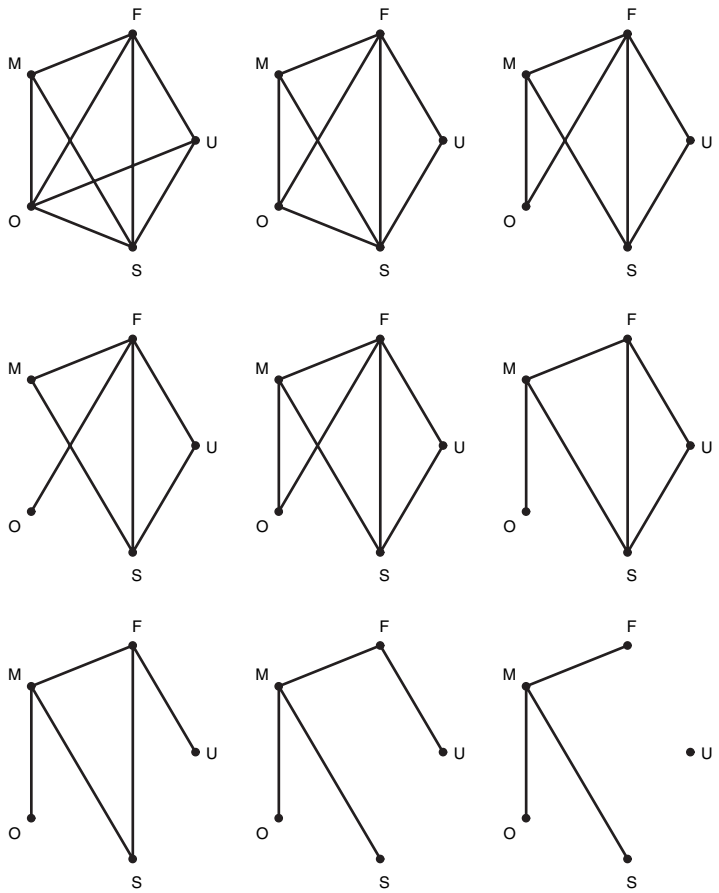

Fig. 8. Industry portfolios data: the *a posteriori* most probable sequence of graphs $(g_1, \ldots, g_9)$ read from top left to bottom right, with corresponding posterior probability 0.802

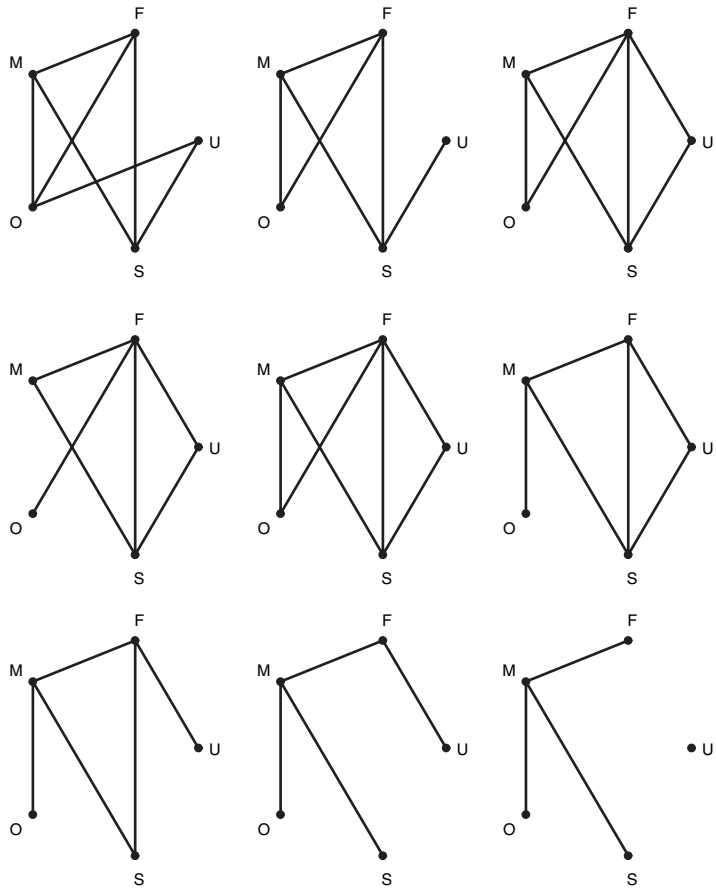

Fig. 9. Industry portfolios data: the second *a posteriori* most probable sequence of graphs $(g_1, \ldots, g_9)$ read from top left to bottom right, with corresponding posterior probability 0.121

7. Conclusions and discussion

In this paper, we have developed a graphical model for sequences of Gaussian random vectors when changes in the underlying graph occur at random times, and a new block of data is created with the addition or deletion of an edge. We have shown how a Bayesian hierarchical model incorporates both the uncertainty about that graph and the time variation thereof.

Our main objective was to learn the graph underlying the last, most current, block of data. In fact, our Bayesian framework has allowed us to make inference about the whole history up to and including the last block. By carefully designing the proposal distributions, we have ensured a relatively expeditious exploration of the posterior in each component of our state variable, via the Metropolis–Hastings algorithm, and have demonstrated our methodology on a simulated data set. Also, we could recover historically meaningful structural changes for the US industries data.

An important practical goal is that of on-line learning and prediction of the precision matrix, updated continually on a daily, or even intraday, basis. The Markov chain Monte Carlo framework is sufficiently flexible to allow for sampling from the posterior predictive distribution of a future graph or precision matrix given data up to and including the present. Development

of efficient, easy-to-use computer code implementing our modelling and posterior inference strategies would be especially useful to an investor who seeks to allocate and re-evaluate the composition of her portfolio dynamically. The R code that we have written for this paper, which will be released separately as soon as it is optimized, is a first step in that direction.

Next, treating the number of blocks B in the data as a random variable to be learned *a posteriori*, instead of a tuning parameter fixed *a priori*, would make inference and especially prediction more automatic. However, the changepoint process is not easily separated from that which generates successive graphs and parameter values; hence great care must be exercised, especially when attempting to merge neighbouring blocks. In Talih (2004), we address the issue of a random B and extend our model to allow for more than one structural change in the graph at the end of each block.

Finally, it would be of great value to adopt a more general parameterization of the precision matrix. For structural learning, our current parameterization, although appealing for its simplicity, is rather restrictive. Indeed, as seen from equation (8), up to the connectivity indices in the denominator, the strength of all connections in a graph G_b is determined by one parameter θ_b. Our model is biased towards strong associations, and weak associations might be classified as non-existent. Therefore, the resulting graphs might not reflect the true interaction structure, and it becomes difficult to interpret such graphs as conditional independence graphs. However, even a simplistic interaction structure can be useful, and indeed lucrative, in the context of portfolio allocation. In fact, the market is such that it continually updates itself with the investors' expectations and (sometimes untrue) beliefs.

Nevertheless, in the case of triangulated graphs, for which a unique decomposition of the likelihood is available, Dawid and Lauritzen (1993) have developed a class of 'compatible' prior distributions on precision matrices. These priors can be used to localize, and ultimately to speed up, the calculation of the posterior likelihood ratios that are involved in the Markov chain Monte Carlo proposal evaluations, as in Giudici and Green (1999). However, in the time-varying graphical model framework that we have put forth in this paper, the requirement of compatibility is difficult to meet, since we lose the appealing localization property.

Acknowledgements

This work has been an integral part of the first author's doctoral dissertation at Yale University, Department of Statistics, under the supervision of Nicolas Hengartner. The first author has also greatly benefited from the guidance that he has received from John Hartigan and Joseph Chang, who have both served on his doctoral committee. The referees' outstanding feed-back has much improved the quality of this paper.

References

Aguilar, O. and West, M. (2000) Bayesian dynamic factor models and portfolio allocation. *J. Bus. Econ. Statist.*, **18**, 338–357.

Barry, D. and Hartigan, J. A. (1993) A Bayesian analysis for change point problems. *J. Am. Statist. Ass.*, **88**, 309–319.

Brown, S. J., Goetzmann, W. N. and Grinblatt, M. (1997) Positive portfolio factors. *Working Paper 97-01*. Yale International Center for Finance, New Haven.

Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997) *The Econometrics of Financial Markets*. Princeton: Princeton University Press.

Cox, D. R. and Wermuth, N. (1990) An approximation to maximum likelihood estimates in reduced models. *Biometrika*, **77**, 747–761.

Dahlhaus, R. and Eichler, M. (2003) Causality and graphical models in time series analysis. In *Highly Structured Stochastic Systems* (eds P. Green, N. Hjort and S. Richardson). Oxford: Oxford University Press.

Dawid, A. P. and Lauritzen, S. L. (1993) Hyper Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.*, **21**, 1272–1317.
Dempster, A. P. (1972) Covariance selection. *Biometrics*, **28**, 157–175.
Dirac, G. A. (1961) On rigid circuit graphs *Abh. Math. Sem. Univ. Hambrg*, **25**, 71–76.
Frydenberg, M. and Lauritzen, S. L. (1989) Decomposition of maximum likelihood in mixed graphical interaction models. *Biometrika*, **76**, 539–555.
Gavril, F. (1972) Algorithms for minimum coloring, maximum clique, minimum covering by cliques, and maximum independent set of chordal graph. *SIAM J. Computn*, **1**, 180–187.
Giudici, P. and Green, P. J. (1999) Decomposable graphical model determination. *Biometrika*, **86**, 785–801.
Guyon, X. and Hardouin, C. (2002) Markov chain Markov field dynamics: models and statistics. *Statistics*, **13**, 339–363.
Jacquier, E., Polson, N. G. and Rossi, P. E. (1994) Bayesian analysis of stochastic volatility models (with discussion). *J. Bus. Econ. Statist.*, **12**, 371–417.
Koster, J. T. A. (1996) Markov properties of nonrecursive causal models. *Ann. Statist.*, **24**, 2148–2177.
Robert, C. P. and Casella, G. (1999) *Monte Carlo Statistical Methods*. New York: Springer.
Seber, G. A. F. (1984) *Multivariate Observations*, pp. 102–105, 448–451. New York: Wiley.
Talih, M. (2003) Markov random fields on time-varying graphs, with an application to portfolio selection. *PhD Thesis*. Department of Statistics, Yale University, New Haven.
Talih, M. (2004) Online learning of dynamic graphical models. *Working Paper*. Department of Mathematics and Statistics, Hunter College of the City University of New York, New York.
Whittaker, J. (1990) *Graphical Models in Applied Multivariate Statistics*. Chichester: Wiley.