Bayesian Analysis (2007) **2**, Number 1, pp. 69–98

Dynamic Matrix-Variate Graphical Models

Carlos M. Carvalho* and Mike West†

Abstract. This paper introduces a novel class of Bayesian models for multivariate time series analysis based on a synthesis of dynamic linear models and graphical models. The synthesis uses sparse graphical modelling ideas to introduce structured, conditional independence relationships in the time-varying, cross-sectional covariance matrices of multiple time series. We define this new class of models and their theoretical structure involving novel matrix-normal/hyper-inverse Wishart distributions. We then describe the resulting Bayesian methodology and computational strategies for model fitting and prediction. This includes novel stochastic evolution theory for time-varying, structured variance matrices, and the full sequential and conjugate updating, filtering and forecasting analysis. The models are then applied in the context of financial time series for predictive portfolio analysis. The improvements defined in optimal Bayesian decision analysis in this example context vividly illustrate the practical benefits of the parsimony induced via appropriate graphical model structuring in multivariate dynamic modelling. We discuss theoretical and empirical aspects of the conditional independence structures in such models, issues of model uncertainty and search, and the relevance of this new framework as a key step towards scaling multivariate dynamic Bayesian modelling methodology to time series of increasing dimension and complexity.

Keywords: Bayesian Forecasting, Dynamic Linear Models, Gaussian Graphical Models, Graphical Model Uncertainty, Hyper-Inverse Wishart Distribution, Portfolio Analysis.

1 Introduction

Bayesian dynamic linear models (DLMs) (West and Harrison 1997) are used extensively for analysis and prediction of time series of increasing dimension and complexity in finance (Aguilar and West 2000; Quintana et al. 2003), engineering (Godsill and Rayner 1998; Fong et al. 2002; Godsill et al. 2004), ecology (Calder et al. 2003), medicine (West et al. 1999) and other areas. The time-varying regression structure, or state-space structure, and the sequential nature of DLM analysis flexibly allows for the creation and routine use of interpretable forecasting models of realistic complexity. The inherent Bayesian framework naturally allows and encourages the integration of data, expert information and systematic interventions in model fitting and assessment, and thus in forecasting and decision making.

The current work responds to the increasingly pressing need to scale multivariate time series analysis methodology to higher-dimensional problems. Many application areas are generating data of increasing dimension and complexity, and modellers must respond with increasing attention to structure and parameter parsimony in statistical

*ISDS, Duke University, Durham, NC, http://www.isds.duke.edu/~carlos/
†ISDS, Duke University, Durham, NC, http://www.isds.duke.edu/~mw/

models. Increasing sparsity of parameters in higher-dimensions is a pre-requisite for scalability of methods in time series as in other areas. We address this by introducing a synthesis of multi- and matrix-variate DLMs with graphical modelling to induce sparsity and structure in the covariance matrices of such models, including time-varying matrices in multivariate time series.

Section 2 outlines the framework of matrix-variate DLMs, a natural framework for evaluation of inter-connections among several or many series and of the changes in dependency structures over time. These models are routinely used in financial applications, in particular. Section 3 outlines the structure of Gaussian graphical models, and Bayesian models for structured, parameter constrained covariance matrices based on the use of the family of hyper-inverse Wishart distributions. Section 4 then defines the new modelling framework, including the formal model specification and details of the resulting methodology for both constant and, of more practical relevance, time-varying covariance matrices in matrix-DLMs. This includes extensions of the standard DLM sequential updating, forecasting and retrospective analysis theory. Section 5 then describes the use of formal models inducing variance matrix discounting into the new models for structured, time-varying covariance matrices. Section 6 develops a study in a key motivating application context, that of financial portfolio prediction and decision analysis (Quintana and West 1987; Quintana 1992; Quintana et al. 2003; Aguilar and West 2000). We discuss theoretical and empirical findings in the context of an initial example using 11 exchange rate time series, and then a more extensive and practical study of 346 securities from the S&P Index. This latter application also develops and applies graphical model search and selection ideas, based on existing MCMC and stochastic search methods now translated to the DLM context. Section 7 provides a brief overview and summary comments, and pointers to near-term research including broader questions of model uncertainty.

2 Matrix-Variate Dynamic Linear Models

The class of Matrix Normal DLMs
(Quintana 1987; Quintana and West 1987; West and Harrison 1997) represents a general, fully-conjugate framework for multivariate time series analysis and dynamic regression with estimation of cross-sectional covariance structures. The framework involves common structure for each of the univariate series, thus making these models particularly well-suited for the analysis of time series of similar, related items, such as stock prices, bond prices, temporal gene expression data, and so forth.

We begin with development of models with constant but unknown observational variances and cross-series covariances. This is developed in this and the following section, and then we extend to the key practical case of time-varying variance matrices in Section 4.

Consider p univariate time series Y_{ti} following individual DLMs

$$\left\{\mathbf{F}_t, \mathbf{G}_t, V_t\sigma_i^2, \mathbf{W}_t\sigma_i^2\right\}.$$

Here t is the time index and i indexes the individual series ($i = 1, \ldots, p$). The notation above represents the set of p DLMs

$$\begin{aligned} \text{Observation:} \quad Y_{ti} &= \mathbf{F}_t'\boldsymbol{\theta}_{ti} + \nu_{ti}, & \nu_{ti} &\sim N(0, V_t\sigma_i^2), \quad (1) \\ \text{Evolution:} \quad \boldsymbol{\theta}_{ti} &= \mathbf{G}_t\boldsymbol{\theta}_{t-1,i} + \boldsymbol{\omega}_{ti}, & \boldsymbol{\omega}_{ti} &\sim N(0, \mathbf{W}_t\sigma_i^2), \quad (2) \end{aligned}$$

where: $\mathbf{F}_t$ is a known $n \times 1$ regression vector, $\mathbf{G}_t$ is a known $n \times n$ state evolution matrix, $\mathbf{W}_t$ is a known $n \times n$ evolution innovation variance matrix, V_t are known scale factors, $\boldsymbol{\theta}_{ti}$ is the series-i specific $n \times 1$ state vector, and the σ_i are unknown scale factors. Standard conditional independence assumptions are that the observation error terms ν_{ti} and state evolution innovations $\boldsymbol{\omega}_{ti}$ are independent across time and mutually independent at each time. The multivariate model is completed with a cross-sectional covariance structure that impacts on both observation and evolution terms. Let $\boldsymbol{\Sigma}$ be a $p \times p$ covariance matrix with diagonal elements $\sigma_{ii} = \sigma_i^2$ and off-diagonals σ_{ij}, $(i \neq j)$.

Combine the model components as follows:

- $\mathbf{Y}_t = (Y_{t1}, \ldots, Y_{tp})'$, the $p \times 1$ observation vector;
- $\boldsymbol{\Theta}_t = (\boldsymbol{\theta}_{t1}, \ldots, \boldsymbol{\theta}_{tp})$, the $n \times p$ matrix of states;
- $\boldsymbol{\Omega}_t = (\boldsymbol{\omega}_{t1}, \ldots, \boldsymbol{\omega}_{tp})$, the $n \times p$ matrix of evolution innovations; and
- $\boldsymbol{\nu}_t = (\nu_{t1}, \ldots, \nu_{tp})'$, the $p \times 1$ vector of observational innovations.

Then the model is

$$\begin{aligned} \mathbf{Y}_t' &= \mathbf{F}_t'\boldsymbol{\Theta}_t + \boldsymbol{\nu}_t', & \boldsymbol{\nu}_t &\sim N(\mathbf{0}, V_t\boldsymbol{\Sigma}), \quad (3) \\ \boldsymbol{\Theta}_t &= \mathbf{G}_t\boldsymbol{\Theta}_{t-1} + \boldsymbol{\Omega}_t & \boldsymbol{\Omega}_t &\sim N(\mathbf{0}, \mathbf{W}_t, \boldsymbol{\Sigma}), \quad (4) \end{aligned}$$

where the evolution innovation matrix $\boldsymbol{\Omega}_t$ follows a *matrix-variate normal* with mean $\mathbf{0}$ (a $n \times p$ matrix), left covariance matrix $\mathbf{W}_t$ and right covariance matrix $\boldsymbol{\Sigma}$; see Dawid (1981) and Appendix A below.

The cross-sectional structure comes in via the elements σ_{ij} $(i, j = 1, \ldots, p)$ of the $(p \times p)$ covariance matrix $\boldsymbol{\Sigma}$. The model of (3) and (4) implies that, for all $i, j = 1, \ldots, p$,

$$\begin{aligned} Cov(\nu_{ti}, \nu_{tj}) &= V_t\sigma_{ij}, \\ Cov(\boldsymbol{\omega}_{ti}, \boldsymbol{\omega}_{tj}) &= \mathbf{W}_t\sigma_{ij}. \end{aligned}$$

The correlation structure induced by $\boldsymbol{\Sigma}$ affects both the observational and evolution errors; thus, if σ_{ij} is large and positive, series i and j will show a similar behavior in both their underlying state evolution and in the observational variation about their level.

3 Gaussian Graphical Models

3.1 Basic Structure

Graphical model structuring for multivariate models characterizes conditional independencies via graphs (Whittaker 1990; Lauritzen 1996; Jones et al. 2005), and provides methodologically useful decompositions of the sample space into subsets of variables (graph vertices) so that complex problems can be handled through the combination of simpler elements. In high-dimensional problems, graphical model structuring is a key approach to parameter dimension reduction and, hence, to scientific parsimony and statistical efficiency when appropriate graphical structures are identified.

In the context of a multivariate normal distribution, conditional independence restrictions are simply expressed through zeros in the off-diagonal elements of the precision (or concentration) matrix. A $p-$vector $\mathbf{x}$ with elements x_i has a zero-mean multivariate normal distribution with $p \times p$ variance matrix $\mathbf{\Sigma}$ and precision $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$ with elements ω_{ij}. Write $G = (V, E)$ for the undirected graph whose vertex set V corresponds to the set of p random variables in $\mathbf{x}$, and whose edge set E contains elements (i, j) for only those pairs of vertices $i, j \in V$ for which $\omega_{ij} \neq 0$. The canonical parameter $\mathbf{\Omega}$ belongs to $M(G)$, the set of all positive-definite symmetric matrices with elements equal to zero for all $(i, j) \notin E$.

The density of $\mathbf{x}$ factorizes as

$$p(\mathbf{x}|\mathbf{\Sigma}, G) = \frac{\prod_{P \in \mathcal{P}} p(\mathbf{x}_P|\mathbf{\Sigma}_P)}{\prod_{S \in \mathcal{S}} p(\mathbf{x}_S|\mathbf{\Sigma}_S)}, \tag{5}$$

a ratio of products of densities where $\mathbf{x}_P$ and $\mathbf{x}_S$ indicate subsets of variables in the prime components (P) and separators (S) of G, respectively. Given G, this distribution is defined completely by the component-marginal covariance matrices $\mathbf{\Sigma}_P$, subject to the consistency condition that sub-matrices in the separating components are identical (Dawid and Lauritzen 1993). That is, if $S = P_1 \cap P_2$ the elements of $\mathbf{\Sigma}_S$ are common in $\mathbf{\Sigma}_{P_1}$ and $\mathbf{\Sigma}_{P_2}$.

A graph is said to be decomposable when all of its prime components are complete subgraphs of G, implying no conditional independence constraints within a prime component; we also then refer to all prime components (as well as their separators) as cliques of the graph. We develop our theory for decomposable graphical models, now briefly reviewing and then extending the use of hyper-inverse Wishart distributions.

3.2 Hyper-Inverse Wishart Distributions

The fully conjugate Bayesian analysis of decomposable Gaussian graphical models (Dawid and Lauritzen 1993) is based on the family of *hyper-inverse Wishart* (HIW) distributions for structured variance matrices. If $\mathbf{\Omega} = \mathbf{\Sigma}^{-1} \in M(G)$, the hyper-inverse Wishart

$$\mathbf{\Sigma} \sim HIW_G(b, \mathbf{D}) \tag{6}$$

has a degree-of-freedom parameter b and location matrix $\mathbf{D} \in M(G)$. This distribution is the unique hyper-Markov distribution for $\mathbf{\Sigma}$ with consistent clique-marginals that are inverse Wishart. Specifically, for each clique $P \in \mathcal{P}$, $\mathbf{\Sigma}_P \sim IW(b, \mathbf{D}_P)$ with density

$$p(\mathbf{\Sigma}_P|b, \mathbf{D}_P) \propto |\mathbf{\Sigma}_P|^{-(b+2|P|)/2} \exp\left(-\frac{1}{2} tr(\mathbf{\Sigma}_P^{-1}\mathbf{D}_P)\right) \tag{7}$$

where $\mathbf{D}_P$ is the positive-definite symmetric diagonal block of $\mathbf{D}$ corresponding to $\mathbf{\Sigma}_P$. The full HIW is conjugate to the likelihood from a Gaussian sample with variance $\mathbf{\Sigma}$ on G, and the full HIW joint density factorizes over cliques and separators in the same way as (5); that is,

$$p(\mathbf{\Sigma}|b, \mathbf{D}) = \frac{\prod_{P\in\mathcal{P}} p(\mathbf{\Sigma}_P|b, \mathbf{D}_P)}{\prod_{S\in\mathcal{S}} p(\mathbf{\Sigma}_S|b, \mathbf{D}_S)}, \tag{8}$$

where each component in the products of both numerator and denominator is IW as in equation (7).

Definition: Matrix-Normal/HIW Distributions

Our new models utilise HIW distributions together with matrix and multivariate normal distributions, in a direct and simple extension of the usual normal, inverse Wishart distribution theory to the general framework of graphical models. The setup and notation is as follows: The $n \times p$ random matrix $\mathbf{X}$ and $p \times p$ random variance matrix $\mathbf{\Sigma}$ have a joint matrix-normal, hyper-inverse Wishart (NHIW) distribution if $\mathbf{\Sigma} \sim HIW_G(b, \mathbf{D})$ on G and $(\mathbf{X}|\mathbf{\Sigma}) \sim N(\mathbf{m}, \mathbf{W}, \mathbf{\Sigma})$ for some $b, \mathbf{D}, \mathbf{m}, \mathbf{W}$. We denote this by $(\mathbf{X}, \mathbf{\Sigma}) \sim NHIW_G(\mathbf{m}, \mathbf{W}, b, \mathbf{D})$ with $\mathbf{X}$ marginally following a matrix hyper-T (as defined in Dawid and Lauritzen 1993) denoted by $HT_G(\mathbf{m}, \mathbf{W}, \mathbf{D}, b)$.

4 Sparsity in DLMs: Generalization to HIW

4.1 Framework

As discussed above, Gaussian graphical models are a representation of conditional independence structure in multivariate distributions where decompositions of the joint distribution provide computational efficiencies and a reduction in the space of parameters. Taking advantage of the latter, we now show how graphical structuring can be incorporated in matrix normal DLMs providing a parsimonious model for $\mathbf{\Sigma}$. For a given decomposable graph, the hyper-inverse Wishart is used as a conjugate prior for $\mathbf{\Sigma}$ and the analytical, closed-form, sequential updating procedure can be generalized. The methodological developments in this section assume the choice of a particular decomposable graph G for all time points. In practical settings we face two situations: either G is specified based on a combination of substantive reasoning and prior data, or G is drawn from a set of (possible many) candidate graphs to allow for uncertainty about the graphical structure. In the latter case we may then apply the following analysis on each of the graphs in parallel and structure assessment follows by embedding within a

model mixture context (West and Harrison, 1997, chapter 12). The two examples of section 6 below speak to each of these two situations.

Consider the matrix normal DLM described in Equations (3) and (4), and suppose $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$ is constrained by a graph G. With the usual notation that D_t is the data and information set conditioned upon at any time t, assume the NHIW initial prior of the form

$$(\mathbf{\Theta}_0, \mathbf{\Sigma}|D_0) \sim NHIW_G(\mathbf{m}_0, \mathbf{C}_0, b_0, \mathbf{S}_0). \tag{9}$$

In components,

$$(\mathbf{\Theta}_0|\mathbf{\Sigma}, D_0) \sim N(\mathbf{m}_0, \mathbf{C}_0, \mathbf{\Sigma}) \qquad \text{and} \qquad (\mathbf{\Sigma}|D_0) \sim HIW_G(b_0, \mathbf{S}_0), \tag{10}$$

which incorporates the conditional independence relationships from G into the prior. This is in fact the form of the conjugate prior for sequential updating at all times t, as is now detailed.

4.2 Sequential Updating and Forecasting

Theorem 1. *Under the initial prior of equation (9) and with data observed sequentially to update information sets as $D_t = \{D_{t-1}, \mathbf{Y}_t\}$, the sequential updating for the matrix normal DLM on G is given as follows:*

(i) *Posterior at $t-1$:*

$$(\mathbf{\Theta}_{t-1}, \mathbf{\Sigma}|D_{t-1}) \sim NHIW_G(\mathbf{m}_{t-1}, \mathbf{C}_{t-1}, b_{t-1}, \mathbf{S}_{t-1})$$

(ii) *Prior at t:*

$$(\mathbf{\Theta}_t, \mathbf{\Sigma}|D_{t-1}) \sim NHIW_G(\mathbf{a}_t, \mathbf{R}_t, b_{t-1}, \mathbf{S}_{t-1})$$

where

$$\mathbf{a}_t = \mathbf{G}_t\mathbf{m}_{t-1} \quad \textit{and} \quad \mathbf{R}_t = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t' + \mathbf{W}_t$$

(iii) *One-step forecast:*

$$(\mathbf{Y}_t|D_{t-1}) \sim HT_G(\mathbf{f}_t, Q_t\mathbf{S_{t-1}}, b_{t-1})$$

where

$$\mathbf{f}_t' = \mathbf{F}_t'\mathbf{a}_t \quad \textit{and} \quad Q_t = \mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t + V_t$$

(iv) *Posterior at t:*

$$(\mathbf{\Theta}_t, \mathbf{\Sigma}|D_t) \sim NHIW_G(\mathbf{m}_t, \mathbf{C}_t, b_t, \mathbf{S}_t)$$

with

$$\begin{aligned} \mathbf{m}_t &= \mathbf{a}_t + \mathbf{A}_t\mathbf{e}_t' \\ \mathbf{C}_t &= \mathbf{R}_t - \mathbf{A}_t\mathbf{A}'Q_t \\ b_t &= b_{t-1} + 1 \\ \mathbf{S}_t &= \mathbf{S}_{t-1} + \mathbf{e}_t\mathbf{e}_t'/Q_t \end{aligned}$$

where

$$\mathbf{A}_t = \mathbf{R}_t\mathbf{F}_t/Q_t \quad \textit{and} \quad \mathbf{e}_t = \mathbf{Y}_t - \mathbf{f}_t$$

Proof. This theorem is a direct extension of the theory for matrix DLMs using inverse Wishart distributions for constant variance matrices, as described in Quintana (1987), Quintana and West (1987) and West and Harrison (1997), to the more general framework of graphical models and HIW distributions. The components of the theorem that are novel and require discussion here are *(iii)* and the updating related to $\mathbf{\Sigma}$ in *(iv)*.

- *Proof of (iii)*: It is clear that

$$(\mathbf{Y}_t|\mathbf{\Sigma}, D_{t-1}) \sim N(\mathbf{f}_t, Q_t\mathbf{\Sigma}),$$

  with $(\mathbf{\Sigma}|D_{t-1}) \sim HIW_G(b_{t-1}, S_{t-1})$ so, for each clique C, the marginal distribution of $\mathbf{Y}_t^C$ is simply a $T(\mathbf{f}_t, Q_t\mathbf{S}_{t-1}^C, b_{t-1})$. The overall marginal distribution of $\mathbf{Y}_t$ is then a hyper-T distribution given by the Markov combination (consistent with G) of T-distributions over cliques and separators, as defined in Dawid and Lauritzen (1993), and denoted here by HT_G.

- *Proof of (iv)*: The updating for $\mathbf{\Sigma}$ follows directly the conjugacy of the HIW described in Dawid and Lauritzen (1993). Here we are simply exploiting this conjugacy for repeated sequential updates based on the likelihood contributions from the realized forecast errors $\mathbf{e}_t$ that factorize on the graph in the conjugate form.

□

4.3 Retrospection

Interest often also lies in retrospective estimation. At any time T with data D_T, the so-called k-step filtered distribution for the state matrix $p(\mathbf{\Theta}_{T-k}|D_T)$, $(1 \leq k \leq T)$, is then available as a direct byproduct of conditional independencies of DLMs. This result generalizes the retrospective cascade of filtering distributions to the graphical context. Given that $\mathbf{\Sigma}$ is a fixed parameter (not a state), the results developed in West and Harrison (1997) extend to the matrix DLMs with graphical structure. In summary, the filtered distribution of the state matrix $\mathbf{\Theta}_{T-k}$ and $\mathbf{\Sigma}$ is most easily obtained recursively as follows (details in West and Harrison 1997):

$$(\mathbf{\Theta}_{t-k}, \mathbf{\Sigma}|D_t) \sim NHIW_G(\mathbf{a}(-k)_t, \mathbf{R}(-k)_t, b_t, \mathbf{S}_t) \tag{11}$$

where the parameters are calculated through the following recurrences:

$$\begin{aligned}
\mathbf{B}_{t-k} &= \mathbf{C}_{t-k}\mathbf{G}'_{t-k+1}\mathbf{R}^{-1}_{t-k+1} \\
\mathbf{a}_t(-k) &= \mathbf{m}_{t-k} + \mathbf{B}_{t-k}[\mathbf{a}_t(-k+1) - \mathbf{a}_{t-k+1}] \\
\mathbf{R}_t(-k) &= \mathbf{C}_{t-k} + \mathbf{B}_{t-k}[\mathbf{R}_t(-k+1) - \mathbf{R}_{t-k+1}]\mathbf{B}'_{t-k},
\end{aligned}$$

with starting values $\mathbf{a}_t(0) = \mathbf{m}_t$ and $\mathbf{R}_t(0) = \mathbf{C}_t$.

5 Time-Varying Σ_t

The above development is now extended to the practically critical context of time-varying variances and covariances across series, modifying $\mathbf{\Sigma}$ to $\mathbf{\Sigma}_t$ and developing an initial class of stochastic evolution models for these dynamic multivariate matrices. Models of $\mathbf{\Sigma}_t$ varying stochastically over time are key in areas such as finance, where univariate and multivariate volatility models have been center-stage in both research and front-line financial applications for over two decades (Quintana and West 1987; Bollerslev et al. 1992; Quintana 1992; Jacquier et al. 1994; Kim et al. 1998; Aguilar and West 2000; Quintana et al. 2003). It is important to point out that, once again, we assume that G is given and constant across time points.

Our stochastic model for time variation is a "locally smooth", discount factor-based model that extends models for full, unconstrained $\mathbf{\Sigma}_t$ matrices introduced in Liu (2000) and Quintana et al. (2003). These references developed a general and flexible framework and a multivariate volatility model that provided a more general foundation for earlier methods of Uhlig (1994), Quintana et al. (1995) and West and Harrison (1997).

The model involves constructing a Markov process in which transition distributions $p(\mathbf{\Sigma}_t|\mathbf{\Sigma}_{t-1})$ are defined based on matrix-Beta random innovations applied to elements of the Bartlett decomposition of $\mathbf{\Sigma}_{t-1}$. The details extend this Beta-Bartlett evolution from its original application in models for full, unconstrained variance matrices to the context here of models constrained on graphs. Full details of the construction and theory are given in Appendix B below; here we note the basic ideas and operational results.

Based on a specified discount factor δ, $(0 << \delta \leq 1)$, the matrix Beta-Bartlett stochastic evolution model has the following key implications and features:

- Beginning at time $t-1$ with the current posterior

$$(\mathbf{\Sigma}_{t-1}|D_{t-1}) \sim HIW_G(b_{t-1}, \mathbf{S}_{t-1}),$$

  the stochastic evolution of $\mathbf{\Sigma}_{t-1}$ to $\mathbf{\Sigma}_t$ implies the time t prior

$$(\mathbf{\Sigma}_t|D_{t-1}) \sim HIW_G(\delta b_{t-1}, \delta \mathbf{S}_{t-1}). \tag{12}$$

- Notice how the time-evolution maintains the inverse-Wishart form for the prior of $\mathbf{\Sigma}_t$. The key additional two features are that the evolution increases the spread of the HIW distribution by reducing the degrees-of-freedom via multiplication with the chosen discount factor, while maintaining the location of the distribution at $\mathbf{S}_{t-1}/b_{t-1}$.

Otherwise, the theory and updating analysis as earlier presented are essentially unchanged. Observing $\mathbf{Y}_t$ generates the realized forecast error $\mathbf{e}_t$ and the time t prior is updated as before, with the discount factor modification implied by the modified time t prior; that is,

$$(\mathbf{\Sigma}_t|D_t) \sim HIW_G(b_t, \mathbf{S}_t)$$

with

$$b_t = \delta b_{t-1} + 1 \quad \text{and} \quad \mathbf{S}_t = \delta \mathbf{S}_{t-1} + \mathbf{e}_t \mathbf{e}_t'. \tag{13}$$

Some additional comments and details are relevant to both interpretation and practical application:

- Notice how the past information is now discounted prior to updating the summary parameters of the HIW distributions. To exemplify this, consider the posterior harmonic mean of $\mathbf{\Sigma}_t$, namely $\hat{\mathbf{\Sigma}}_t = E(\mathbf{\Sigma}_t^{-1}|D_t)^{-1} = \mathbf{S}_t/b_t$. For large t, this has the form of an exponentially weighted moving average estimate of the error variances. In detail, $b_t \longrightarrow (1-\delta)^{-1}$ and

$$\hat{\mathbf{\Sigma}}_t \approx (1-\delta) \sum_{l=0}^{t-1} \delta^l \mathbf{e}_{t-l} \mathbf{e}_{t-l}'.$$

  This provides a framework where forward estimates of $\mathbf{\Sigma}_t$ keep adapting to new data while further discounting past observations. The rate of information decay implied by δ is a modelling choice with standard analysis using relatively high value of δ, typically between 0.9 and 1. Extensive discussion of choice of discount factors, and the broader role of discount concepts in dynamic modelling, appears in West and Harrison (1997) (with discussion specific to dynamic variance models in chapter 16). Modifications in the discount factor may also be used to allow for more abrupt changes in volatility to be incorporated into the model. Standard approaches in applied work will evaluate models across a crude grid of values and assess sensitivity of resulting inferences and predictions across those models, perhaps even embedding that approach within a model mixture framework West and Harrison (1997) (especially examples in chapters 3,10,12).

- As in scalar and multivariate IW models (West and Harrison 1997), the mean of the retrospective or filtered distributions of the $\mathbf{\Sigma}_t^{-1}$ sequence can be recursively calculated by

$$E(\mathbf{\Sigma}_{t-k}^{-1}|D_t) = \mathbf{S}_t^{-1}(-k)(b_t(-k) + p - 1) \tag{14}$$

  where,

$$\mathbf{S}_t^{-1}(-k) = (1-\delta)\mathbf{S}_{t-k}^{-1} + \delta \mathbf{S}_t(-k+1)^{-1} \tag{15}$$

$$b_t(-k) = (1-\delta)b_{t-k} + \delta b_t(-k+1) \tag{16}$$

  with starting values $\mathbf{S}_t(0) = \mathbf{S}_t$ and $b_t(0) = b_t$. So, for every clique of the graph G, retrospective estimates of $\mathbf{\Omega}_{t-k}^C$ can be computed using equation (14) and combined to form a retrospective estimate of $\mathbf{\Omega}_{t-k}$, while direct inversion will yield the harmonic posterior mean of $(\mathbf{\Sigma}_t|D_T)$.

6 Large-Scale Dynamic Portfolio Allocation

6.1 Predictive Portfolio Decisions

The use of dynamic Bayesian forecasting models and Bayesian decision analysis in asset allocation problems has been routine for a number of years. The one-step ahead forecast distributions of future returns are the key components of mean-variance portfolio optimization methods that allow for parameter change and uncertainty to be taken into account in sequential investment decisions. Quintana (1992), Putnam and Quintana (1994), Quintana and Putnam (1996), Aguilar and West (2000) and Quintana et al. (2003) are examples of carefully developed DLMs that implement portfolio rules in fixed income and currency markets; aspects of Bayesian portfolio selection are also discussed in detail in Polson and Tew (2000).

The static (non-time series) portfolio example of Carvalho et al. (2005) is a first illustration of how structuring the covariance matrices of assets can reduce the uncertainty about optimal portfolio weights and so induce less volatile investment opportunities. DLMs with structured covariance matrices, as developed in Theorem 1, are now developed to properly explore such matters in a realistic dynamic portfolio decision/allocation context.

In the examples here we base derivation of optimal portfolios on the quadratic programming procedures developed by Markowitz (1959). At any time t, given the first two moments $(\mathbf{f}_t, \mathbf{Q}_t)$ of the predictive distribution of a vector of next-period returns and a fixed scalar return target m, the investor decision problem reduces to choosing the vector of portfolio weights $\mathbf{w}_t$ to minimize the one-step ahead portfolio variance $\mathbf{w}_t'\mathbf{Q}_t\mathbf{w}_t$ subject to constraints $\mathbf{w}_t'\mathbf{f}_t = m$ and $\mathbf{w}_t'\mathbf{1} = 1$. The general solution for the above optimization through Lagrange multipliers creates the so called *efficient frontier* where the mean-variance efficient portfolio is given by

$$\mathbf{w}_t^{(m)} = \mathbf{Q}_t^{-1}(a\mathbf{f}_t + b\mathbf{1}) \tag{17}$$

where $a = \mathbf{1}'\mathbf{Q}_t^{-1}\mathbf{e}$ and $b = -\mathbf{f}_t'\mathbf{Q}_t^{-1}\mathbf{e}$, and where $\mathbf{e} = (\mathbf{1}m - \mathbf{f}_t)/d$ with $d = (\mathbf{1}'\mathbf{Q}_t^{-1}\mathbf{1})(\mathbf{f}_t'\mathbf{Q}_t^{-1}\mathbf{f}_t) - (\mathbf{1}'\mathbf{Q}_t^{-1}\mathbf{f}_t)^2$. An interesting alternative portfolio that involves only estimates of $\mathbf{Q}_t$ is the minimum-variance portfolio that arises from equation (17) when $a = 0$ and $b^{-1} = \mathbf{1}'\mathbf{Q}_t^{-1}\mathbf{1}$. This latter strategy isolates the effects of $\mathbf{Q}_t$ on investment decisions and is of interest when competing models for covariance estimation are considered.

Perold (1988), Polson and Tew (2000) and Ledoit and Wolf (2004) point out that building high-dimensional portfolios tend to result in extreme and very unstable weights assigned to each asset. This is due to the large amount of uncertainty in the estimation of covariance matrices, especially when the number of historical observations is relatively small if compared to the number of assets considered. From (17) it is clear that the solution for optimal portfolios is a direct function of the precision matrix $\mathbf{K}_t = \mathbf{Q}_t^{-1}$. A

nice representation appears in Stevens (1998), namely

$$w_{ti}^{(m)} = \lambda \frac{f_{ti} - \sum_{j \neq i} (k_{tij}/k_{tii}) f_{tj}}{k_{tii}^{-1}} \tag{18}$$

with λ being the Lagrange multiplier. If it is assumed that the returns are normally distributed, expression (18) shows that the weight assigned to asset i depends on the ratio of the intercept of its regression on all other assets relative to the conditional variance of the regression. In other words, the amount of money invested in asset i depends on the ratio of the expected return that cannot be explained by the linear combination of assets to the *unhedgeable* (or *nondiversifiable*) risk.

Note that the numerator is a function of the off-diagonal elements of $\mathbf{K}_t$ hence it is not surprising that conditional independence assumptions have a direct influence over the uncertainty about $\mathbf{w}_t$. If, in fact, the unhedgeable risk can be obtained by a regression involving a smaller number of regressors (i.e. having some of the k_{tij}'s equal to zero) this has to be taken into account; failing to do so implies that unnecessary parameters are being estimated and nothing but uncertainty is added to the problem. In the following two applications, we show how imposing conditional independence constraints help create portfolios that not only are less risky but also turn out to be more profitable. First we revisit the exchange rate example of Carvalho et al. (2005) followed by an example involving a large set of securities in the S&P 500 stock index. The goal is to simply compare the performance of dynamic portfolios built from both a "full" (unconstrained) and a "sparse" (with graphical modelling constraints) DLM. In all examples a simple DLM with time-varying covariance structure (as in Section 5) is used; the form is a local trend model, namely

$$\mathbf{Y}_t = \boldsymbol{\theta}_t + \boldsymbol{\nu}_t, \qquad \boldsymbol{\nu}_t \sim N(0, \boldsymbol{\Sigma}_t), \tag{19}$$

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(0, W_t \boldsymbol{\Sigma}_t). \tag{20}$$

This is special case of the general DLMs presented in previous sections and includes time-varying structured variance matrices.

6.2 Example: International Exchange Rates

This first example is a dynamic version of the example in Carvalho et al. (2005) where portfolios for $p = 11$ international currency exchange rates relative to the US dollar were compared; the data appear in Figure 1. This example serves to illustrate the methodology and to demonstrate that graphical model structure in a dynamic model can lead to substantial improvements in model fit and, importantly, resulting Bayesian decision analysis in a practical context. In this study, the graph G specified is the same graph used in Carvalho et al. (2005) (Figure 2) which was identified as the posterior mode by the Metropolis stochastic search described in Jones et al. (2005), assuming the returns were zero mean, independent normal with a constant variance matrix. That prior analysis thus provides statistical support for the relevance of this specific graph, while the structure is evidently interpretable in the financial econometric setting: it groups as

a clique the core European Union currencies, and defines other links and cliques that structurally reflect international trading relationships with the other countries represented.

We use this graphical model structure in a DLM with a time-varying variance matrix, exploring comparisons with the traditional DLM with an unconstrained variance matrix that include evaluations of the impact on portfolio predictions and decisions. In all models here, we use fairly diffuse initial priors (small b_0 and with identity $\mathbf{D}_0$). We discuss results from four analyses that differ only in the chosen value of the variance matrix discount factor: δ takes the values 0.90, 0.95, 0.97 and 0.99. Apart from comparisons, much of the discussion that illuminates the practical relevance of graphical model structuring is focused on the model with $\delta = 0.97$.

For each model at each time point, mean-variance and minimum-variance portfolios were computed based on the one-step ahead forecasts. This leads to the evaluation of a number of key quantities of interest in the model comparison:

- the empirical accuracy of portfolio predictions and resulting decisions, in terms of realized returns;
- the projected risk measures in terms of the one-step ahead variances of the optimal portfolios;
- uncertainties about the cumulative returns over any specified period of time, and
- the time-trajectories of the optimal portfolio weights themselves as they are sequentially updated.

Cumulative returns for mean-target portfolios in the analysis using the usual unconstrained DLM and those from the described graphical model structured DLM are displayed in Figure 3 (for different values of the discount factor δ). The target mean return chosen is $m = 0.01\%$, corresponding to a (20 day) monthly target return of approximately 2%. Portfolio weights were computed from the moments of the one-step ahead forecast distributions and sequentially revised over the full time period. The optimal portfolio weights then computed are used to reallocate investments across currencies (assuming no transaction or other costs). This leads to evaluation of the cumulative returns. Evidently, the graphical DLM dominates in terms of risk/return, fact that is reinforced in Figure 4 where the same comparison using the baseline minimum-variance portfolio is presented.

Figure 5 displays the ratio of optimal portfolio risk (using the baseline minimum variance portfolio) under the full model relative to the graphical model. This example comes from the DLMs with $\delta = 0.97$, and the results are similar across values of δ and also under the mean-target portfolios. The risk is simply the *ex ante* standard deviation of the one-step ahead optimal portfolio, derived from the corresponding variance $\mathbf{w}_t'\mathbf{Q}_t\mathbf{w}_t$ at the optimizing weight vector $\mathbf{w}_t$. This shows that the estimated portfolio variance is likely to be smaller under the model with graphical structure, indicating that

the parsimonious model is able to help reduce predicted investment risk. This is also highlighted in Figure 7.

As discussed before one of the reasons behind the good performance of the structured models is possibly the smaller variation of portfolio weights. We explore this notion in Figure 6 where the estimated variance of the weight of each asset in the mean-variance portfolio is presented. These plots are given for, as before, the analysis using the usual unconstrained variance matrix DLM and for the graphical model structured DLM. Again we use the model with $\delta = 0.97$ and the results are typical of those of all models. Notice that throughout the entire period of time, the variances are substantially higher under the unconstrained model than under the graphical model, implying that the optimal portfolio in the former moves money between currencies at higher volumes on a daily basis. Thus the full model has both more uncertainty about the optimal decisions and increased costs due to portfolio rebalancing fees were this to be implemented live.

The overall conclusions of this example, as evidenced in the figures, are that the DLM graphical model uniformly outperforms the unconstrained, full variance matrix DLM across the full time period of portfolio decisions. The uniform dominance is reflected in several practically critical elements:

- higher realized cumulative returns,
- lower risk portfolios in terms of both one-step ahead predictive variances of optimal portfolios and variances of cumulative returns, and
- lower volatility of the optimal portfolio weights as they are sequentially revised, consistent with more stable portfolios.

This example demonstrates the relevance of appropriate model structuring and the fact that the resulting parameter parsimony can indeed help in portfolio allocation processes. The graphical model DLM generates more accurate predictions and optimal portfolio decisions, with lower risk in terms of both the nominal predicted portfolio risk and in terms of realized outcomes. In addition, the more stable portfolios add to the practical benefits since they would imply, in a live context, reduced costs in terms of transaction fees for moving money between currencies period-to-period.

6.3 Example: Portfolio Allocation in the S&P 500

A higher-dimensional application involves $p = 346$ securities forming part of the S&P500 stock index. These series are all the companies that remained in the index from January 1999 until December 2004, a period of $t = 1,508$ daily observations. Again, the goal of the application is to compare the performance of dynamic portfolios built from models with graphical structure relative to the full/unconstrained model. This problem illustrates one focal point of research in portfolio allocation theory where there is great interest in the development of models to efficiently deal with increasingly large sets of assets within a common analysis framework. Here the utility of graphical models to

induce sparsity in the inverse variance matrix - sparsity that increases in its impact on parameter estimation uncertainty as dimension grows - is particularly key.

This application also addresses the question of graphical model structure selection, evaluating models using the graphical model Metropolis stochastic search method introduced in Jones et al. (2005). This method was applied to explore the full space of graphical DLMs using only the first 1,200 observations times in the data set. Across the remaining time period we then evaluate the sequentially updated portfolios using a few graphical DLMs selected based on the posterior at $t = 1,200$. This serves to illustrate both statistical and practical relevance of graphical model structuring.

For any set of observations $\mathbf{Y}_{1:T}$ over times $t = 1 : T$ the sequential updating of a DLM on any graph G leads to direct and efficient computation of the marginal likelihood as

$$p(\mathbf{Y}_{1:T}|G) = p(\mathbf{Y}_T|D_{T-1}, G)p(\mathbf{Y}_{T-1}|D_{T-2}, G)\dots p(\mathbf{Y}_1|D_0, G), \tag{21}$$

where for each element in the product, $(\mathbf{Y}_t|D_{t-1}, G) \sim HT_G(\mathbf{f}_t, \mathbf{S_{t-1}}, b_{t-1})$ as defined in Theorem 1. Hence, taking $T = 1,200$ we can compute this marginal likelihood value trivially on G, and then embed this in the local model search method of Jones et al. (2005) to generate candidate graphs G' in neighborhoods of G to evaluate as Metropolis-Hastings candidates. The only difference between our use of this stochastic search and that of Jones et al. (2005) is the differing model context that leads to a different computation of this marginal likelihood. A fully specified model over graphs includes the assumption of independent edge inclusion indicators for all pairs of nodes (variables) and a common prior probability of edge inclusion, taken here as $\beta = 0.5$ for richer graphs and also as a benchmark for the example. This prior structure is also precisely as described and used in Jones et al. (2005).

This analysis explored hundreds of millions of graphs and we saved the 100,000 with highest posterior probability; across this set of graphs, both the log marginal likelihood function over G and the log posterior probabilities over G vary by over 1000 units, so that a relatively small number of "top graphs" – those with higher posterior probability – really dominate the posterior based on the data up to $T = 1,200$. On any one of these graphs G, the posterior at $t = 1,200$ then provides the initial prior representing D_0 in a new sequential DLM analysis for the remaining 308 observations. In each model a discount factor of 0.98 was used, in line with values used by Quintana et al. (1995) and Quintana et al. (2003).

Figure 8 displays the cumulative returns for baseline minimum-variance portfolios built on the final 308 time periods. For comparison purposes we show results from the "top" graphical DLM (the posterior mode from the stochastic search, labelled "graph"), the "full graph" representing the usual unconstrained variance matrix DLM as in the exchange rate example, and also now the "empty graph" in which the series are unrelated. Cumulative returns of the S&P500 provide a benchmark for the portfolios created in the example and are also plotted. The top graphical DLM G has exactly 29,181 edges (of the total possible of 59,685). Evidently, selection of this model based on statistical considerations alone leads to practical dominance over the usual full model as well as the other two (albeit straw men) indices. The relative improvements in realized portfolio

performance are practically very substantial.

To emphasize the increased (desired) stability of time trajectories of optimal portfolio weights in structured models, Figure 9 displays the weights of 4 selected companies based on the DLM analysis using the full, unconstrained model and the top graphical model.

Finally, Figure 10 shows cumulative returns from 5 different graphical DLMs selected from the 100,000 models (details in figure's caption). The fact that the cumulative returns are quite similar for all 5 graphs indicates that, for the purpose of investment opportunities, a small number of edges is really relevant. This suggests that exploring subgroups of graphs that generate "equivalent" cumulative returns may generate insights about portfolio allocation theory in connection with covariance selection models.

Some of the key conclusions are as follows:

- Selection of graphs G according to high posterior probability on the training data of 1,200 observations leads to graphical model DLMs that generate substantial improvements in realized portfolio applications of the subsequent data relative to the usual full model and other competing methods. Again it is clear that the imposition of constraints in the covariance matrix - constraints that are defined by the training data as consistent with the time-varying structure of associations across series - generate more profitable investment opportunities.

- Higher realized returns are coupled with lower risk and lower volatility of time trajectories of portfolio weights.

- Graphical models with higher posterior probabilities on the training data analysis are seen to yield higher realized portfolio returns (and more stable portfolios, of course); that is, both statistical and practical significance support the view of parsimony and structure in problems of increasing dimension.

- The example shows the ease with which the general methodology can be implemented with series of modest dimension ($p = 346$) and build on the existing methods for exploring uncertainty about graphical model structures as well as efficient estimation of time-varying states and cross-series variance matrices within graphical models.

7 Summary Comments

By combining dynamic linear models with decomposable graphical models, we have defined a new, rich class of matrix DLMs that allow for the incorporation of conditional independence structure in the cross-sectional precision matrix of a set of time series. The use of the hyper-inverse Wishart distributions on Gaussian graphical models provides a conjugate framework that allows for sequential updating, on-line predictions and decisions, and retrospective filtering analyses that extend the existing theory and methods of Bayesian forecasting with dynamic models. The new framework includes

the use of time-varying covariance structures generated with extensions of the Beta-Bartlett construction of Markovian evolutions of variance matrices from the case of full, unconstrained matrices (inverse Wishart models) to the full class of graphical models (hyper-inverse Wishart models).

Our examples focus on questions of practical utility in financial time series and Bayesian decision analysis for sequential portfolio allocation and investment. The real practical value of data-consistent structuring and constraints on variance matrices across series is evident in these examples, where the analyses lead to parsimony in parametrization, statistical efficiency in estimation and reduced uncertainty as a result, and this translates into more accurate and more stable portfolio decisions: the resulting portfolios achieve higher realized returns than in models with no graphical structuring, and have lower levels of predicted and realized risk in terms of both variances of returns and volatility of portfolio weights. High-dimensional portfolios are regarded as one of the most challenging problems in financial theory (Polson and Tew 2000) and theoretical developments for a precise understanding of the connections between conditional independence assumptions and optimal investment strategies is of key importance for further advances in that area, including comparisons based on other performance measures such as Value-at-Risk and that take into account transaction costs and other economic relevant variables.

Our second example also discusses and explores aspects of graphical model uncertainty and model selection, evaluating posterior distributions over graphical structures G as well as time-varying state and variance parameters within any given graphical structure. This analysis used the ideas and methods of Jones et al. (2005), which is recommended reading for readers interested in the broader questions of graphical model choice, and of the computational and prior specification issues in graphical modelling. Looking ahead, some next steps will involve the development of model mixing and averaging, where we will naturally begin to explore combinations of DLMs in a variety of possible multi-process modelling contexts (West and Harrison 1997). This development of sequential model selection procedures that address uncertainty about graphs while allowing for efficient on-line updates is an open research area and one of key importance in further applications of DLMs in real forecasting problems.

For readers interested in developing analyses of this kind, we will make available custom software implementing the analyses reported and exemplified here.

Acknowledgments

The authors are grateful for the constructive comments of the editor, associate editor and two anonymous referees on the original version of this paper. The authors acknowledge the support of grants from the NSF (DMS 0102227 and 0342172) and the NIH (U54-CA-112952-01).

Appendix A: Matrix-Normal Distributions

For an $n \times p$ random matrix $\mathbf{X}$, the notation $\mathbf{X} \sim N(\mathbf{m}, \mathbf{W}, \mathbf{\Sigma})$ denotes the matrix-normal (or matrix-variate normal) distribution of $\mathbf{X}$. Here the $n \times p$ matrix $\mathbf{m}$ is the mean with $E(X_{ij}) = m_{ij}$; the $n \times n$ variance matrix $\mathbf{W}$ is the left covariance matrix, and the $p \times p$ variance matrix $\mathbf{\Sigma}$ is the right covariance matrix.

Denote the elements of $\mathbf{\Sigma}$ by σ_{ij} with diagonals $\sigma_{ii} = \sigma_i^2$, and those of $\mathbf{W}$ by w_{ij} with diagonals $w_{ii} = w_i^2$. Then any column $\mathbf{x}_i$ of $\mathbf{X}$ has a multivariate normal distribution $N(\mathbf{m}_i, \mathbf{W}\sigma_i^2)$ where $\mathbf{m}_i$ is the corresponding column of the mean matrix $\mathbf{m}$. Any pair of columns $\mathbf{x}_i$ and $\mathbf{x}_j$ has covariance matrix $\mathbf{W}\sigma_{ij}$.

Correspondingly, the i^{th} row of $\mathbf{X}$ has a multivariate normal distribution with mean vector the corresponding row of $\mathbf{m}$, and variance matrix $\mathbf{\Sigma}w_i^2$. Any two rows i and j have covariance matrix $\mathbf{\Sigma}w_{ij}$.

See Dawid (1981) for primary reference and development of theory, and Quintana (1987), Quintana (1992) and Quintana and West (1987) for early applications.

Appendix D: Matrix Beta-Bartlett HIW Evolutions

Begin with the time $t-1$ posterior

$$(\mathbf{\Sigma}_{t-1}|D_{t-1}) \sim HIW_G(b_{t-1}, \mathbf{S}_{t-1}).$$

The stochastic map from $\mathbf{\Sigma}_{t-1}$ to $\mathbf{\Sigma}_t$ is established through multiplicative beta shocks applied to the diagonal of the Bartlett decomposition of $\mathbf{\Sigma}_{t-1}^{-1}$. Using the HIW extension of the Bartlett decomposition (Roverato 2000), we have $\mathbf{\Sigma}_{t-1} = \mathbf{\Phi}^{-1'}\mathbf{\Phi}^{-1}$ where: $\mathbf{S}_{t-1}^{-1} = \mathbf{T}'\mathbf{T}$ and $\mathbf{\Phi} = \mathbf{\Psi}\mathbf{T}$ where $\mathbf{\Psi} \in M^t(G)$ has elements $(\psi_{ii})^2 \sim \chi^2_{b_{t-1}+\nu_i}$ and $\psi_{ij} \sim N(0,1)$ for $(i,j) \in V$. Here, $M^t(G)$ is the set of all upper triangular matrices with positive diagonal elements such that the entries $(i,j) \notin E$ are zero, and ν_i is the number of non-zero elements in row i of $\mathbf{\Psi}$.

Evolving to time t follows a transformation of the diagonal elements of $\mathbf{\Psi}$ to $\mathbf{\Psi}^* \in M^t(G)$ defined by elements:

$$\psi^*_{ii} = \psi_{ii}\sqrt{r_i} \quad \text{and} \quad \psi^*_{ij} = \psi_{ij} \quad \text{for} \quad (i,j) \in V,$$

where the r_i are independent random "shocks" distributed as

$$r_i \sim Beta\left[\delta_i(b_{t-1}+\nu_i)/2, (1-\delta_i)(b_{t-1}+\nu_i)/2\right]$$

with $\delta_i = (\delta b_{t-1} + \nu_i)/(b_{t-1}+\nu_i)$.

It then follows that $(\psi^*_{ii})^2 \sim \chi^2_{\delta b_{t-1}+\nu_i}$ and $\psi^*_{ij} \sim N(0,1)$ for all pairs $(i,j) \in V$; hence, by setting $\mathbf{T}^* = \sqrt{\delta}\mathbf{T}$, $\mathbf{\Phi}^* = \mathbf{\Psi}^*\mathbf{T}^*$ and then generating $\mathbf{\Sigma}_t = \mathbf{\Phi}^{*-1'}\mathbf{\Phi}^{*-1}$, it follows that $\mathbf{\Sigma}_t$ has the marginal "evolved" HIW distribution

$$(\mathbf{\Sigma}_t|D_{t-1}) \sim HIW_G(\delta b_{t-1}, \delta\mathbf{S}_{t-1}). \tag{22}$$

References

Aguilar, O. and West, M. (2000). "Bayesian dynamic factor models and variance matrix discounting for portfolio allocation." *Journal of Business and Economic Statistics*, 18: 338–357. 69, 70, 76, 78

Bollerslev, T., Chou, R., and Kroner, K. (1992). "ARCH modeling in finance." *Journal of Econometrics*, 52: 5–59. 76

Calder, C., Lavine, M., Muller, P., and Clark, J. (2003). "Incorporating multiple sources of stochasticity into dynamic population models." *Ecology*, 84: 1395–1402. 69

Carvalho, C., Massam, H., and West, M. (2005). "Simulation of hyper-inverse Wishart distributions in graphical models." ISDS Discussion Paper. 78, 79

Dawid, A. P. (1981). "Some matrix-variate distribution theory: Notational considerations and a Bayesian application." *Biometrika*, 68: 265–274. 71, 85

Dawid, A. P. and Lauritzen, S. L. (1993). "Hyper-Markov laws in the statistical analysis of decomposable graphical models." *The Annals of Statistics*, 3: 1272–1317. 72, 73, 75

Fong, W., Godsill, S. J., Doucet, A., and West, M. (2002). "Monte Carlo smoothing with application to speech enhancement." *IEEE Trans. Signal Processing*, 50: 438–449. 69

Godsill, S. J., Doucet, A., and West, M. (2004). "Monte Carlo smoothing for non-linear time series." *Journal of the American Statistical Association*, 99: 156–168. 69

Godsill, S. J. and Rayner, P. J. W. (1998). *Digital Audio Restoration: A Statistical Model-Based Approach*. Springer-Verlag. 69

Jacquier, E., Polson, N., and Rossi, P. (1994). "Bayesian Analysis of stochastic volatility models." *Journal of Business and Economic Statistics*, 12: 371–417. 76

Jones, B., Carvalho, C., Dobra, A., Hans, C., Carter, C., and West, M. (2005). "Experiments in stochastic computation for high-dimensional graphical models." *Statistical Science*, 20: 388–400. 72, 79, 82, 84

Kim, S., Shephard, N., and Chib, S. (1998). "Stochastic volatility: likelihood inference and comparison with ARCH model." *Review of Economic Studies*, 65: 361–393. 76

Lauritzen, S. L. (1996). *Graphical Models*. Clarendon Press, Oxford. 72

Ledoit, O. and Wolf, M. (2004). "Honey, I shrunk the sample covariance matrix." *Journal of Portfolio Management*, 30: 110–119. 78

Liu, J. (2000). *Bayesian Time Series Analysis: Methods Using Simulation-Based Computation*. Duke University: PhD. Thesis. 76

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. New York, USA: John Wiley and Sons. 78

Perold, A. (1988). "Large-scale portfolio optimization." *Management Science*, 30: 1143–1160. 78

Polson, N. and Tew, B. (2000). "Bayesian portfolio selection: An empirical analysis of the S&P500 index 1970 - 1996." *Journal of Business and Economic Statistics*, 18: 164–173. 78, 84

Putnam, B. and Quintana, J. (1994). "New Bayesian statistical approaches to estimating and evaluating models of exchange rates determination." In *Proceedings of the ASA Section on Bayesian Statistical Science*. American Statistical Association. 78

Quintana, J. (1987). *Multivariate Bayesian Forecasting Models*. University of Warwick: PhD. Thesis. 70, 75, 85

— (1992). "Optimal Portfolios of forward currency contracts." In Berger, J., Bernardo, J., Dawid, A., and Smith, A. (eds.), *Bayesian Statistics IV*, 753–762. Oxford University Press. 70, 76, 78, 85

Quintana, J., Chopra, V., and Putnam, B. (1995). "Global Asset Alocation: Stretching returns by shrinking forecasts." In *Proceedings of the ASA Section on Bayesian Statistical Science*. American Statistical Association. 76, 82

Quintana, J., Lourdes, V., Aguilar, O., and Liu, J. (2003). "Global gambling." In Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Heckerman, D., Smith, A., and West, M. (eds.), *Bayesian Statistics VII*, 349–368. Oxford University Press. 69, 70, 76, 78, 82

Quintana, J. and Putnam, B. (1996). "Debating currency markets efficiency using multiple-factor models." In *Proceedings of the ASA Section on Bayesian Statistical Science*. American Statistical Association. 78

Quintana, J. and West, M. (1987). "Multivariate time series analysis: New techniques applied to international exchange rate data." *The Statistician*, 36: 275–281. 70, 75, 76, 85

Roverato, A. (2000). "Cholesky decomposition of a hyper-inverse Wishart matrix." *Biometrika*, 87: 99–112. 85

Stevens, G. (1998). "On the inverse of the covariance matrix in portfolio analysis." *The Journal of Finance*, 53: 1821–1827. 79

Uhlig, H. (1994). "On singular Wishart and singular multivariate beta distributions." *Annals of Statistics*, 22: 395–405. 76

West, M. and Harrison, P. (1997). *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag. 69, 70, 75, 76, 77, 84

West, M., Prado, R., and Krystal, A. (1999). "Evaluation and comparison of EEG traces: Latent Structure in Non-Stationary Time Series." *Journal of the American Statistical Association*, 94: 1083–1095. 69

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Chichester, United Kingdom: John Wiley and Sons. 72

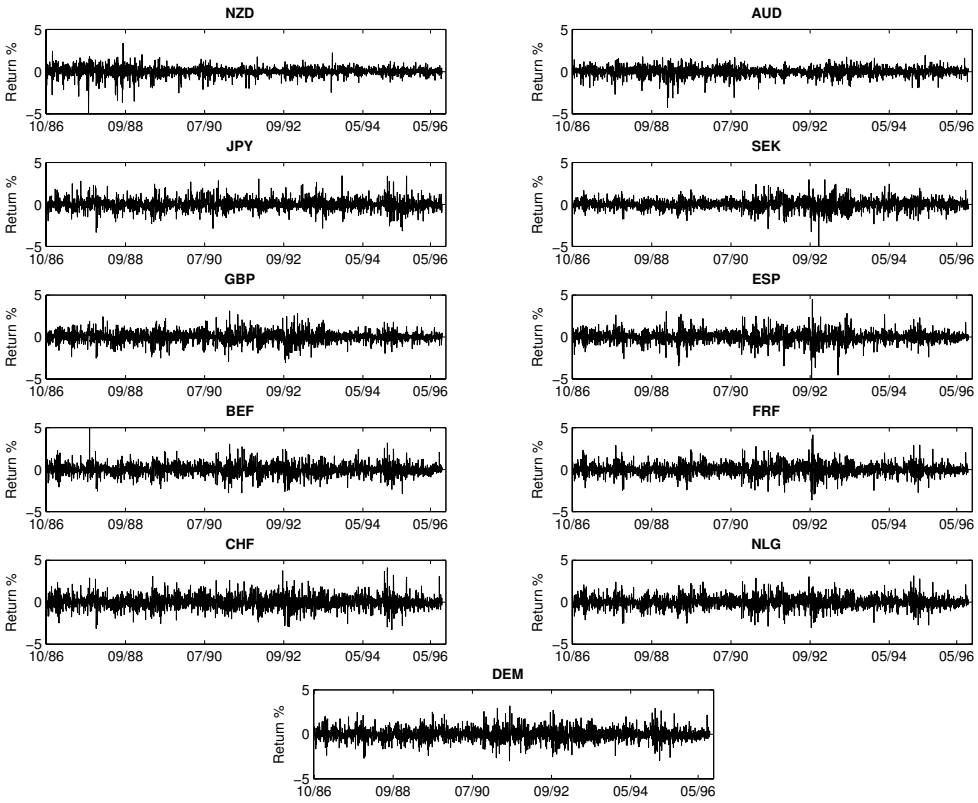

Figure 1: Daily exchange rate returns data. The data set consists of 2,566 daily returns for each of these 11 national currencies, over the period of about 10 years: 10/9/86 to 8/9/96. Daily returns are computed as $Y_{ti} = 100(P_{ti}/P_{t-1,i} - 1)$ for currency i on day t, where P_{ti} is the daily closing spot exchange rate on day t.

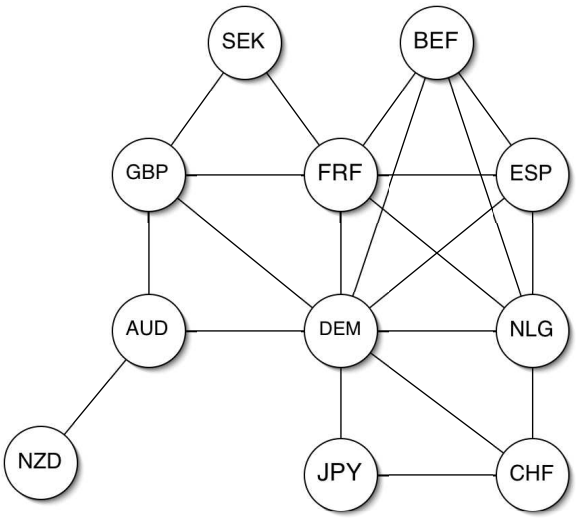

Figure 2: Graph determining the conditional independence structure in the exchange rate/portfolio investment example.

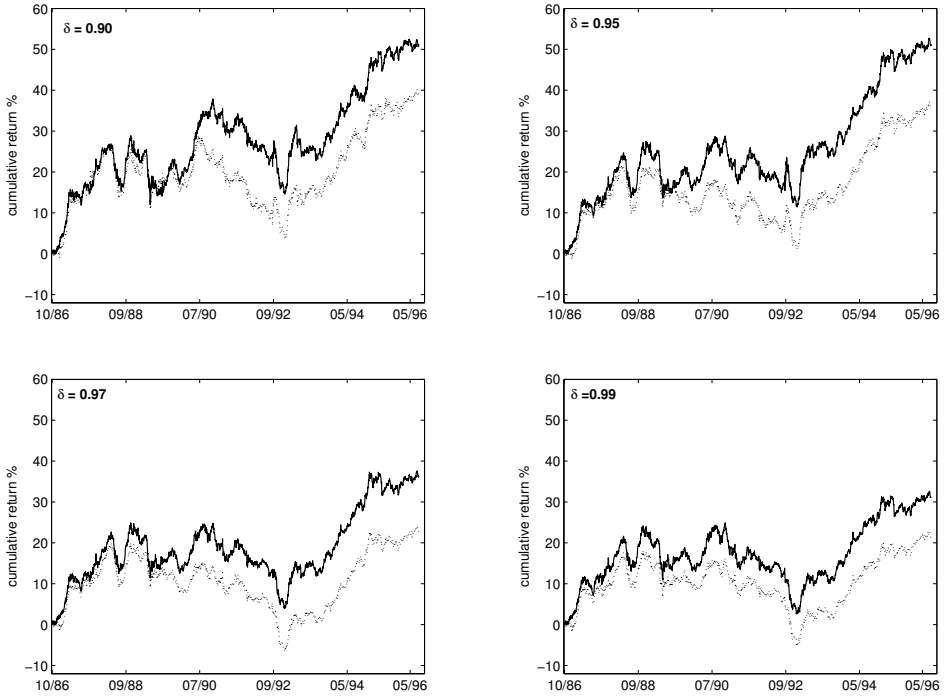

Figure 3: Cumulative returns for mean-target portfolios in the analysis using the usual unconstrained DLM and those from the described graphical model structured DLM. The panels differ only in terms of the choice of δ.

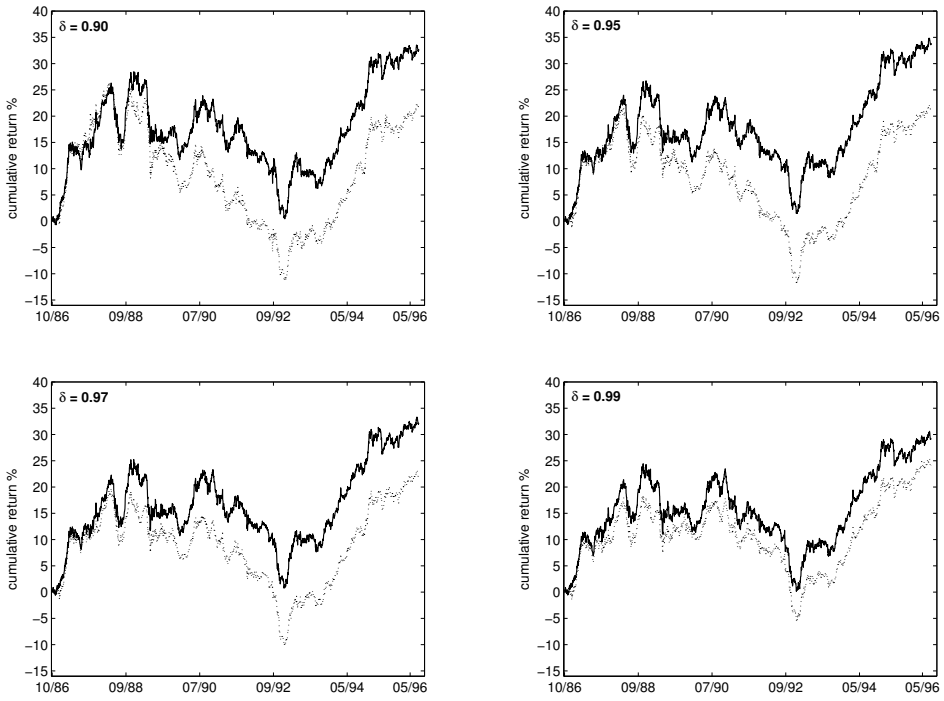

Figure 4: Cumulative returns as in Figure 3, but now using the baseline minimum-variance portfolio construction. Color coding and display format is precisely as in Figure 3.

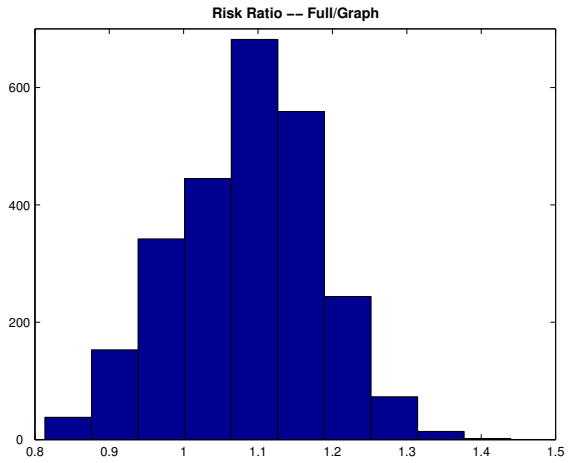

Figure 5: *Ex ante* (at time t given D_{t-1}) ratio of optimal portfolio risk using the baseline minimum variance portfolio under the full model relative to the graphical model. Here, $\delta = 0.97$.

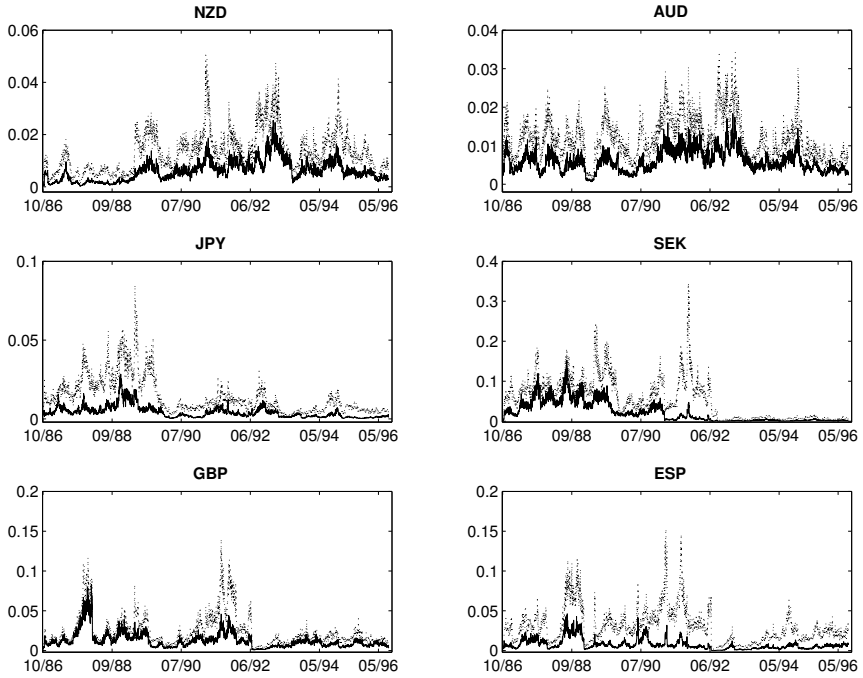

Figure 6: Each panel displays the estimated variance of the weight of each asset in the mean-variance portfolio. Repeat simulation of the prior $p(\mathbf{\Sigma}_t|D_{t-1})$ at each time t allows for computation of the one-step ahead predictive distributions of the full vector of portfolio weights $p(\mathbf{w}_t|D_{t-1})$, and these simulated portfolios can then be summarised. This figure simply computes the corresponding Monte Carlo estimates of variances of each of the weights.

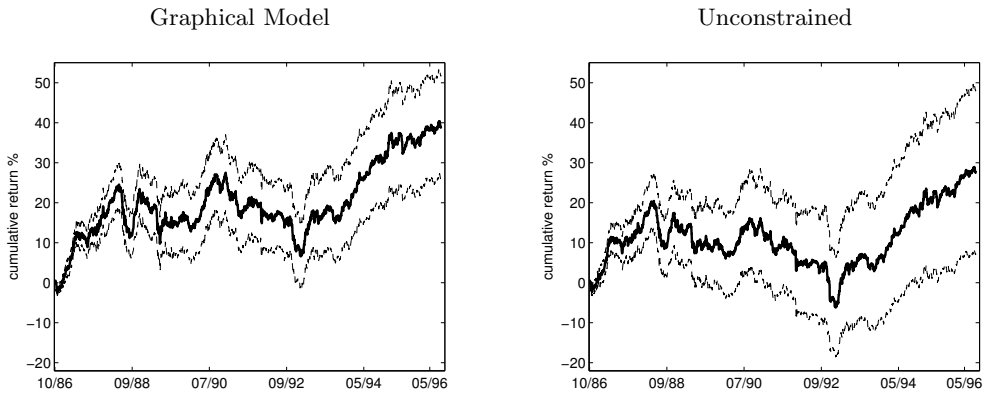

Figure 7: To highlight the issues illustrated in Figure 6, 2-S.D. limit using the optimal portfolio variances are plotted about the realized cumulative returns over time. The resulting intervals are much wider for the full model than for the graphical model.

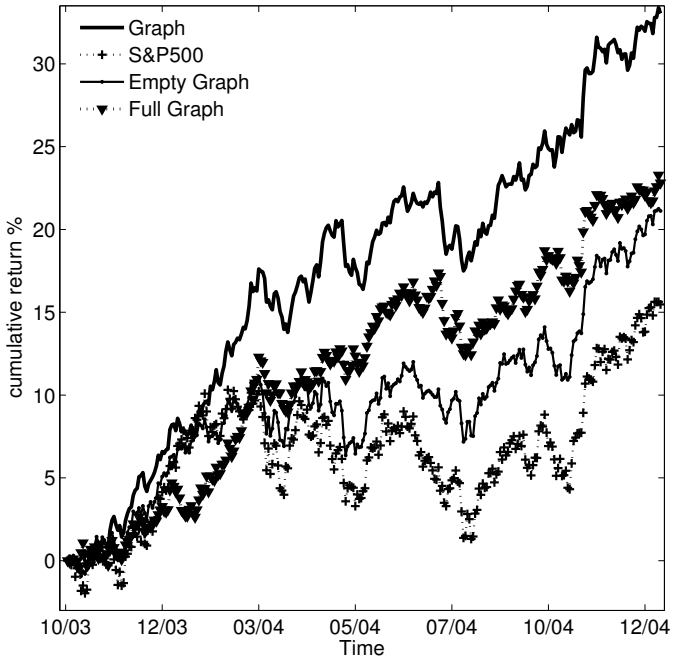

Figure 8: S&P 500 portfolios: Cumulative returns for baseline minimum-variance portfolios. Here, $\delta = 0.98$.

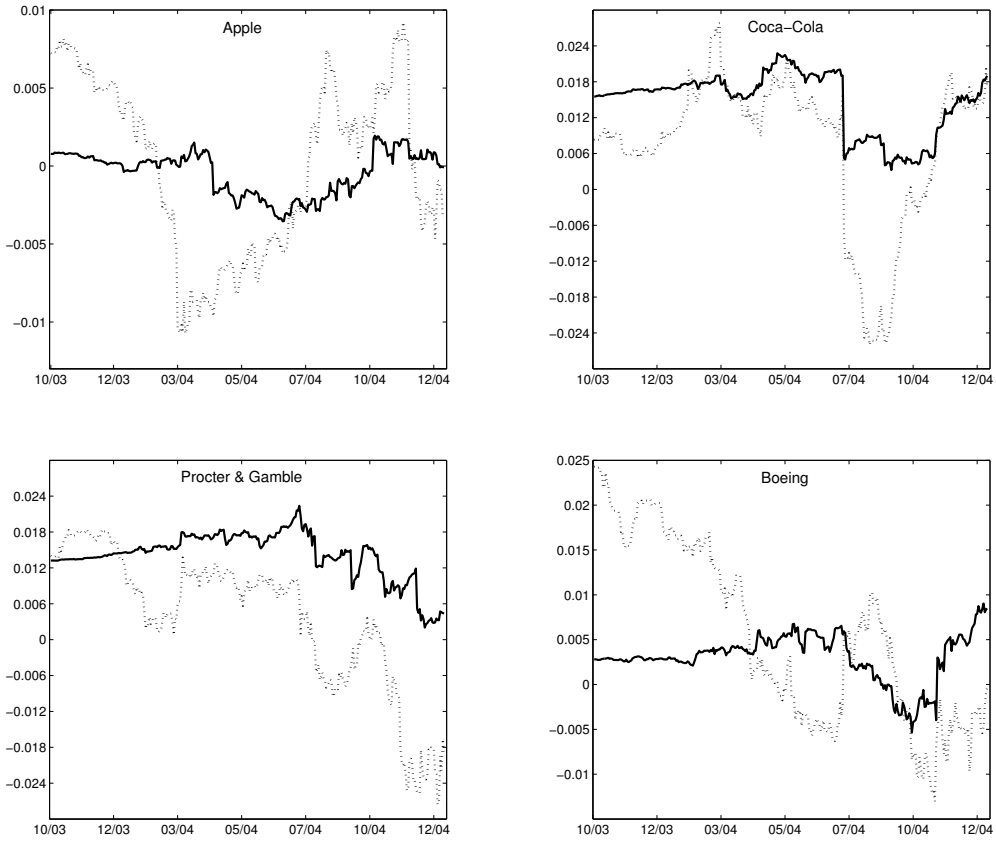

Figure 9: Optimal portfolio weights for 4 selected companies, based on DLM analysis using the full, unconstrained model and the top graphical model. Here, $\delta = 0.98$.

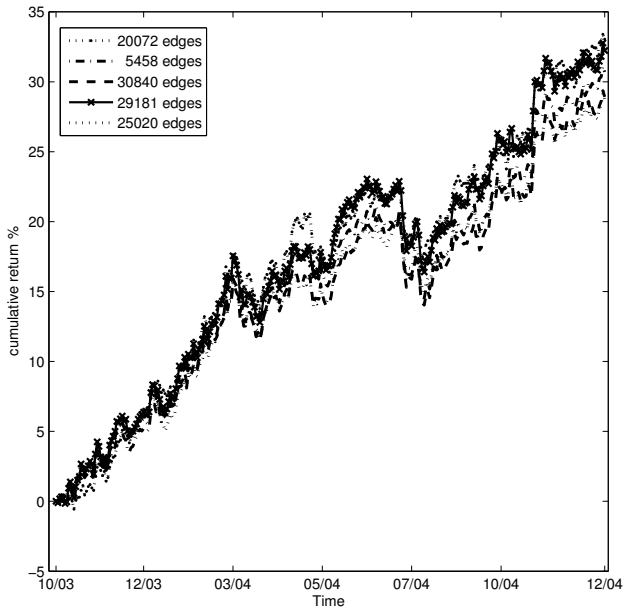

Figure 10: S&P 500 portfolios: Cumulative returns from 5 different graphical DLMs selected from the 100,000 models: these are top graph (posterior mode) that has 29,181 edges; the 75% percentile model that has 30,840 edges; the median model with 25,020 edges; the 25% model with 20,072 edges; and the sparsest model of the 100,000, one that has just 5,458 edges. Here, $\delta = 0.98$.