

Dynamic Financial Index Models: Modeling Conditional Dependencies via Graphs

BY HAO WANG

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

CRAIG REESON

*The Goldman Sachs Group, Inc.,
New York, New York 10004, U.S.A.
craig.reeson@gs.com*

AND CARLOS M. CARVALHO

*Booth School of Business, The University of Chicago,
Chicago, Illinois 60637, U.S.A.
carlos.carvalho@chicagobooth.edu*

SUMMARY

We discuss the development and application of dynamic graphical models for multivariate financial time series in the context of Financial Index Models. The use of graphs generalizes the independence residual variation assumption of index models with a more complex yet still parsimonious model. Working with the dynamic matrix-variate graphical model framework, we develop general time-varying index models that are analytically tractable. In terms of methodology, we carefully explore strategies to deal with graph uncertainty and discuss the implementation of a novel computational tool to sequentially learn about the conditional independence relationships defining the model. Additionally, motivated by our applied context, we extend the DGM framework to accommodate random regressors. Finally, in a case study involving 100 stocks, we show that our proposed methodology is able to generate improvements in covariance forecasting and portfolio optimization problems.

Some key words: Bayesian forecasting; Covariance matrix forecasting; Dynamic matrix-variate graphical models; Index models, Factor models; Gaussian graphical models; Portfolio selection

1 INTRODUCTION

Since the seminal work of [Sharpe \(1964\)](#), Financial Index Models are in the core of asset pricing and portfolio allocation problems. These models assume that all systematic variation in the returns of financial securities can be explained by one, or a set of market indices (factors). The central empirical implication of this assumption is a highly structured covariance matrix for the distribution of returns as, after conditioning on the chosen set of market indices, the residual covariance matrix is diagonal. The attractiveness of this approach is immediate as it offers a very simple, economically justifiable and stable way to estimate potentially very large covariance matrices. A vast body of literature is dedicated to testing the validity of Index Models and the selection of indices – we refer the reader to [Cochrane \(2001\)](#) for a detailed account of the area.

The covariance matrix of returns is a key input in building optimal portfolios and its estimation is often challenging as the number of parameters grows exponentially with the number of assets considered. It is necessary, therefore, to work with structured models that reduce the dimensionality of the problem and deliver more effective estimates and, in turn, better investment decisions. In this paper, we explore a generalization of Financial Index Models with more complex patterns of covariation between returns by allowing conditional dependencies via the introduction of graphical constraints. We work with the matrix-variate dynamic graphical model (DGM) framework of [Carvalho & West \(2007a,b\)](#) but, unlike their original work, graphs are used to increase complexity and not to reduce it.

We take the view that, given its popularity in empirical finance, Index Models such as the *Capital Asset Pricing Model* (CAPM) and the *Fama-French* (FF) are appropriate for the purpose of asset allocation. The central idea of our work is to show that it is possible to improve upon traditional estimates from Index Models and provide more flexible, efficient and still parsimonious strategies for estimating covariances. In addition, we provide two extensions to DGMs: *(i)* we consider the problem of sequential inference about the graphical structure and, *(ii)* define the sequential updating process in the presence of stochastic regressors.

The proposed forecasting model is tested on stock returns data in a portfolio selection exercise. Using 100 NYSE monthly stock returns from 1989 through 2008, we find that our strategy yields better out-of-sample forecast of realized covariance matrix and lower portfolio variance than the two traditional implementations of index models, the capital asset pricing model (CAPM) and the Fama-French (FF) model.

We start by describing Index Models in Section 2 along with its use in the dynamic linear model context. Section 3 and 4 present the necessary background of dynamic matrix-variate graphical models. In Section 5 we discuss issues of dealing with graph (model) uncertainty through time and a simulation study is presented in Section 6. Section 7 expands the DGM context to allow for random regressors. Finally, in Section 8 we explore the use of DGMs as a tool to improve the implementation of Financial Index Models.

2 FINANCIAL INDEX MODELS

A k -dim Index Model assumes that stock returns are generated by

$$Y_{it} = \alpha_i + \sum_{j=1}^k \theta_{ij} f_{jt} + \nu_{it}$$

where f_{jt} is the j th common factor at time t , and residuals ν_{it} are uncorrelated to index f_{jt} and to one another. This implies that the covariance matrix of returns can be written as:

$$\mathbf{V}_t = \mathbf{\Theta}'_t \mathbf{\Psi}_t \mathbf{\Theta}_t + \mathbf{\Sigma}_t$$

where $\mathbf{\Theta}_t$ is the matrix of factor loadings of stocks, $\mathbf{\Psi}_t$ is the covariance matrix of the factors, and $\mathbf{\Sigma}_t$ is a diagonal matrix containing the residual return variances.

Some interesting Index Models include the single index model and three index model. The single index uses the excess return of the market as the single index. This model corresponds to the standard Capital Asset Pricing Model of [Sharpe \(1964\)](#). More recently, and perhaps the most commonly used approach is the three index model proposed by [Fama & French \(1993\)](#) where two new factors (besides the market) are added: value-weighted market index with size and book-to-market factors.

These models are usually estimated by running a set of independent regressions where the excess return of each stock is regressed against the indices for a certain window of time. Call $\hat{\theta}_i$ the estimates of the regression coefficients for stock i and the $\hat{\sigma}_{ii}$ the residual variance estimate. This yields the following estimator for the covariance matrix of stock returns:

$$\hat{\mathbf{V}} = \hat{\mathbf{\Theta}} \hat{\mathbf{\Psi}} \hat{\mathbf{\Theta}}' + \hat{\mathbf{\Sigma}},$$

where $\hat{\mathbf{\Psi}}$ is the sample covariance matrix of indices, $\hat{\mathbf{\Theta}} = [\hat{\theta}_1, \dots, \hat{\theta}_p]$ is the matrix of regression coefficients for all p assets and $\hat{\mathbf{\Sigma}}$ is the diagonal matrix of residual variances. This strategy usually defines the one-step forecast of the covariance matrix to be the current estimate of the covariance matrix $\hat{\mathbf{V}}$.

In our work, we recast the above strategy in a natural model based on a state-space or dynamic linear model (DLM) ([West & Harrison, 1997](#)) representation. This follows the work of [Zellner & Chetty \(1965\)](#); [Quintana & West \(1987\)](#); [Carvalho & West \(2007a\)](#), to cite a few. We use a dynamic regression framework where, in its full generality, a $p \times 1$ vector time series of returns \mathbf{Y}_t follows the dynamic linear model

$$\mathbf{Y}'_t = \mathbf{F}'_t \mathbf{\Theta}_t + \boldsymbol{\nu}'_t, \quad \boldsymbol{\nu}_t \sim N(\mathbf{0}, \mathbf{\Sigma}_t), \quad (1)$$

$$\mathbf{\Theta}_t = \mathbf{\Theta}_{t-1} + \mathbf{\Omega}_t \quad \mathbf{\Omega}_t \sim N(\mathbf{0}, \mathbf{W}_t, \mathbf{\Sigma}_t), \quad (2)$$

for $t = 1, 2, \dots$, where (a) $\mathbf{Y}_t = (Y_{ti})$, the $p \times 1$ observation vector; (b) $\mathbf{\Theta}_t = (\theta_{ti})$, the $n \times p$ matrix of states; (c) $\mathbf{\Omega}_t = (\omega_{ti})$, the $n \times p$ matrix of evolution innovations; (d) $\boldsymbol{\nu}_t = (\nu_{ti})$, the $p \times 1$ vector of observational innovations; (e) for all t , the $n \times 1$ regressor vector \mathbf{F}_t , is known. Also, $\mathbf{\Omega}_t$ follows a matrix-variate normal with mean 0, left and

right covariance matrices \mathbf{W}_t and Σ_t , respectively. In terms of scalar elements, we have p univariate models with individual n -vector state parameters, namely

$$\text{Observation: } Y_{ti} = \mathbf{F}'_t \boldsymbol{\theta}_{ti} + \nu_{ti}, \quad \nu_{ti} \sim N(0, \sigma_{ii,t}^2), \quad (3)$$

$$\text{Evolution: } \boldsymbol{\theta}_{ti} = \boldsymbol{\theta}_{t-1,i} + \boldsymbol{\omega}_{ti}, \quad \boldsymbol{\omega}_{ti} \sim N(0, \mathbf{W}_t \sigma_{ii,t}^2), \quad (4)$$

for each i, t . Each of the scalar series shares the same \mathbf{F}_t elements, and the reference to the model as one of exchangeable time series reflects these symmetries. This is a standard specification in which the correlation structures induced by Σ_t affect both the observation and evolution errors; for example, if $\sigma_{ij,t}$ is large and positive, vector series i and j will show concordant behavior in movement of their state vectors and in observational variation about their levels. Specification of the entire sequence of W_t in terms of discount factors (West & Harrison, 1997) is also standard practice, typically using discount factors related to the state vector and their expected degrees of random change in time.

The above representation provides sequential, closed-form analytical updates of the one-step ahead forecast distributions of future returns and posterior distributions for states and parameters defining the model. This allows for proper accounting of the uncertainty associated with all necessary inputs in sequential investment decisions.

According to traditional Index Models, Σ_t is a diagonal matrix as all common variation between returns should be captured by the elements in Θ_t . We will depart from this standard assumption and allow for a more flexible representation of the residual covariance matrix leading to potentially more complex forms of \mathbf{V} . This is done via the introduction of conditional independencies determined by graphical constraints in Σ_t . The use of these models in sequential portfolio problems is originally proposed by Carvalho & West (2007a) and further analyzed by Quintana et al. (2009). In both references however, graphs were used to reduce the dimensionality of an otherwise fully unstructured covariance matrix of returns. Here, we come from a different direction and show that graphs can be successfully used to increase the complexity of an otherwise highly structured covariance matrix. Before continuing, we need to define the necessary notation for the introduction of graphical models in DLMS.

3 GAUSSIAN GRAPHICAL MODEL

3.1 Basic structure

Graphical model structuring characterizes conditional independencies via graphs (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. Define a p -vector \mathbf{x} with elements x_i and zero-mean multivariate normal distribution with $p \times p$ variance matrix Σ and precision $\Omega = \Sigma^{-1}$ with elements ω_{ij} . Write $G = (V, E)$ as the undirected graph whose vertex set V corresponds to the set of p random variables in \mathbf{x} , and 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 Ω 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

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

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 Σ_P , subject to the consistency condition that sub-matrices in the separating components are identical (Dawid & Lauritzen, 1993). That is, if $S = P_1 \cap P_2$ the elements of Σ_S are common in Σ_{P_1} and Σ_{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 refer to all prime components (as well as their separators) as cliques of the graph. Due to its mathematical and computational convenience, we will only consider decomposable graphs. In this context, the sequential updating and model assessment procedures remain tractable, especially in the high-dimensional settings. It is also our experience and belief that this restriction is not severe as the space of decomposable graphs is very large allowing for the necessary flexibility to our modeling goals. We now briefly review the theory of hyper-inverse Wishart distributions and its extensions to DLMS.

3.2 Hyper-inverse Wishart distributions

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

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

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

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

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

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

where each component in the products of both numerator and denominator is IW as in equation (7). Finally, both the expected value of Σ and Ω can be obtained in close form following the results in Rajaratnam et al. (2008) and Jones et al. (2005) respectively.

4 DYNAMIC MATRIX-VARIATE GRAPHICAL MODEL

The matrix-variate graphical model framework combines 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 $n \times p$ random matrix \mathbf{X} and $p \times p$ random variance matrix Σ have a joint matrix-normal, hyper-inverse Wishart (NHIW) distribution if $\Sigma \sim HIW_G(b, \mathbf{D})$ on G and $(\mathbf{X}|\Sigma) \sim N(\mathbf{m}, \mathbf{W}, \Sigma)$ for some $b, \mathbf{D}, \mathbf{m}, \mathbf{W}$. We denote this by $(\mathbf{X}, \Sigma) \sim NHIW_G(\mathbf{m}, \mathbf{W}, b, \mathbf{D})$ with \mathbf{X} marginally following a matrix hyper-T (as defined in Dawid & Lauritzen, 1993) denoted by $HT_G(\mathbf{m}, \mathbf{W}, \mathbf{D}, b)$.

Back to the DGM context and given Σ_t constrained by any decomposable graph G , Carvalho & West (2007a,b) define the details of the full sequential and conjugate updating, filtering and forecasting for the dynamic regressions and time-varying Σ_t . This approach incorporate graphical structuring into the traditional matrix-variate DLM context and provides a parsimonious yet tractable model for Σ_t . Consider the matrix normal DLM described in equation (1) and (2). With the usual notation that $D_t = \{D_{t-1}, \mathbf{Y}_t\}$ is the data and information set upon any time t , assume the NHIW initial prior of the form

$$(\Theta_0, \Sigma_0 | D_0) \sim NHIW_G(\mathbf{m}_0, \mathbf{C}_0, b_0, \mathbf{S}_0). \quad (8)$$

In components, $(\Theta_0 | \Sigma_0, D_0) \sim N(\mathbf{m}_0, \mathbf{C}_0, \Sigma_0)$ and $(\Sigma_0 | D_0) \sim HIW_G(b_0, \mathbf{S}_0)$, which incorporates the conditional independence relationships from G into the prior. For now assume full knowledge of G defining the conditional independence relationships in \mathbf{Y} . Full sequential updating can be summarized in the following Theorem 1.

Theorem 1. (Carvalho & West, 2007a,b) *Under the initial prior of equation (8) and with data observed sequentially to update information sets D_t the sequential updating for the matrix normal DGM on G is given as follows:*

- (i) *Posterior at $t - 1$: $(\Theta_{t-1}, \Sigma_{t-1} | D_{t-1}) \sim NHIW_G(\mathbf{m}_{t-1}, \mathbf{C}_{t-1}, b_{t-1}, \mathbf{S}_{t-1})$*
- (ii) *Prior at t : $(\Theta_t, \Sigma_t | D_{t-1}) \sim NHIW_G(\mathbf{a}_t, \mathbf{R}_t, \delta b_{t-1}, \delta \mathbf{S}_{t-1})$ where $\mathbf{a}_t = \mathbf{m}_{t-1}$ and $\mathbf{R}_t = \mathbf{C}_{t-1} + \mathbf{W}_t$*

- (iii) *One-step forecast:* $(\mathbf{Y}_t \mid D_{t-1}) \sim HT_G(\mathbf{f}_t, q_t \delta \mathbf{S}_{t-1}, \delta b_{t-1})$ where $\mathbf{f}'_t = \mathbf{F}'_t \mathbf{a}_t$ and $q_t = \mathbf{F}'_t \mathbf{R}_t \mathbf{F}_t + 1$
- (iv) *Posterior at t:* $(\Theta_t, \Sigma_t \mid D_t) \sim NHIW_G(\mathbf{m}_t, \mathbf{C}_t, b_t, \mathbf{S}_t)$ with $\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t \mathbf{e}'_t$, $\mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{A}'_t / q_t$, $b_t = \delta b_{t-1} + 1$, $\mathbf{S}_t = \delta \mathbf{S}_{t-1} + \mathbf{e}_t \mathbf{e}'_t / q_t$ where $\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t / q_t$ and $\mathbf{e}_t = \mathbf{Y}_t - \mathbf{f}_t$.

The above derivation uses a “locally smooth” discount factor-based model to allow Σ_t to vary stochastically. This is a common approach in dynamic linear models (Quintana et al., 2003) where information is discounted through time by a pre-specified discount factor δ . This provides sequential estimates of Σ_t that keep adapting to new data while further discounting past observations. This is easily seen in the representation of the posterior harmonic mean that has the form of a exponentially weighted moving average estimate define as

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

In practical terms the choice of δ represents a similar problem as the choice of the data window in the usual estimation of index models. Extensive discussion of choice of δ in dynamic variance models appears in Chapter 16 of West & Harrison (1997).

So far, G was assumed known and held fixed for all t . This is clearly a limitation of the framework of Carvalho & West (2007a) as it is not necessarily the case that the same set of conditional independence constraints remain fixed across time. Moreover, it is rarely the case that knowledge about G is available and data driven approaches to determine G are required which represents a non-trivial question in empirical applications. Carvalho & West (2007a) present one example where graphs were selected via the computationally intensive stochastic search ideas of Jones et al. (2005). Quintana et al. (2009) consider similar strategies and briefly explore the issue of time variation in G when modeling currencies. It is clear that the use of dynamic matrix-variate graphical models requires proper accounting of the uncertainty associated with G . Before continuing in our exploration of the use of graphs in index models, we add to this discussion and consider alternatives to learn about the conditional independence relationships defining the models.

5 GRAPHICAL MODEL UNCERTAINTY AND SEARCH

5.1 Marginal likelihood over Graphs

In the standard static context, from a Bayesian perspective, model selection involves the posterior distribution of graphs, given by:

$$p(G|\mathbf{x}) \propto p(\mathbf{x}|G)p(G)$$

where $p(\mathbf{x}|G)$ is the marginal likelihood of G . The marginal likelihood function for any graph G is computed by integrating out the covariance matrix with respect to the prior

$$p(\mathbf{x}|G) = \int_{\Sigma^{-1} \in M(G)} p(\mathbf{x}|\Sigma, G)p(\Sigma|G)d\Sigma$$

where $M(G)$, as before, indicates the set of all positive-definite symmetric matrices constrained by G .

Under a hyper-inverse Wishart prior for Σ and observed data \mathbf{x} of sample size n , the above integration for decomposable graph becomes a simple function of the prior and posterior normalizing constants, $H(b, \mathbf{D}, G)$ and $H(b + n, \mathbf{D} + \mathbf{S}_x, G)$:

$$p(\mathbf{x}|G) = (2\pi)^{-np/2} \frac{H(b, \mathbf{D}, G)}{H(b + n, \mathbf{D} + \mathbf{S}_x, G)}$$

where the normalizing constant $H(b, \mathbf{D}, G)$ is given by

$$H(b, \mathbf{D}, G) = \frac{\prod_{P \in \mathcal{P}} |\frac{\mathbf{D}_P}{2}|^{\binom{b+|P|-1}{2}} \Gamma_{|P|}(\frac{b+|P|-1}{2})^{-1}}{\prod_{S \in \mathcal{S}} |\frac{\mathbf{D}_S}{2}|^{\binom{b+|S|-1}{2}} \Gamma_{|S|}(\frac{b+|S|-1}{2})^{-1}}, \quad (9)$$

with $\Gamma_k(a)$ the multivariate gamma function.

In the dynamic set up, a fully Bayesian analysis will consider the graph predictive probability of $\pi(G | D_{t-1})$ over \mathcal{G} , the set of all decomposable graphs, and specify the unconditional predictive distribution $p(\mathbf{Y}_t | D_{t-1})$ as $E_G\{p(\mathbf{Y}_t | D_{t-1}, G)\}$ with the expectation taken with respect to $p(G | D_{t-1})$, namely,

$$(\mathbf{Y}_t | D_{t-1}) \sim \sum_{G \in \mathcal{G}} \pi(G | D_{t-1}) p(\mathbf{Y}_t | D_{t-1}, G). \quad (10)$$

Equation (10) indicates that the predictive probability $\pi(G | D_{t-1})$ is central to evaluating the predictive distribution $p(\mathbf{Y}_t | D_{t-1})$. The two possibilities for consideration of predicting G are as follows: (i) fixed graph for all t , that is for some $G \in \mathcal{G}$, DLM(G) holds for all t ; (ii) time varying graphs where for some possible sequence of graphs $G_t \in \mathcal{G}, (t = 1, 2, \dots)$, DLM(G_t) holds at time t .

For (i), the predictive probability of graphs for time t is defined as

$$\pi(G | D_{t-1}) = p(G | D_{t-1}) \propto p(G) p(\mathbf{Y}_{1:t-1} | G) \quad (11)$$

where the marginal likelihood of a DLM on any graph G is

$$p(\mathbf{Y}_{1:t-1}|G) = p(\mathbf{Y}_{t-1}|D_{t-2}, G)p(\mathbf{Y}_{t-2}|D_{t-3}, G) \dots p(\mathbf{Y}_1|D_0, G),$$

with 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.

For (ii), the time dependence is made explicit with time subscripts, so that a graph G_i at time t is $G_{t,i}$. Denote $\pi(G_{t,i} | D_{t-1})$ as the predictive probability at time $t - 1$ for

graph G_i . It is natural to dynamic modeling that, as time progresses, what occurred in the past becomes less and less relevant to inference made for the future. Applying this notion to graphs, past data loses relevance to current graphs as t increases. Once again, one practical possibility is to use a discount factor to reduce the impact of past information to current inferences, similarly to the discounting ideas used in modeling Σ_t . The predicted probability of $G_{t,i}$ for time t at time $t - 1$ could be written as

$$\pi(G_{t,i} | D_{t-1}) \propto \frac{H(b_0, \mathbf{S}_0, G_{t,i})}{H(\delta b_{t-1}, q_{t-1} \delta \mathbf{S}_{t-1}, G_{t,i})} \pi_0(G_{t,i}). \quad (12)$$

This model implies that, the most recent exponentially weighted residual covariance matrix \mathbf{S}_{t-1} could predict both the one-step ahead residual graphical structure and the residual covariance matrix.

Given any particular graph predicting model M_G and discount factor δ , predicting \mathbf{Y}_t in $\alpha = (M_G, \delta)$ is based on the predictive density

$$p(\mathbf{Y}_t | D_{t-1}, \alpha) = \sum_{G_{t,i} \in \mathcal{G}} p(\mathbf{Y}_t | D_{t-1}, \alpha, G_{t,i}) p(G_{t,i} | D_{t-1}, \alpha). \quad (13)$$

5.2 Sequential stochastic search

Regardless of the choice of how to model G in time the model selection problem gets further complicated by the explosive combinatorial nature of the space of possible graphs. Without the restriction of decomposability there are $2^{\binom{p}{2}}$ elements in graph space, where p represents the number of vertices. Decomposability accounts for approximately 10% of this number which is still impossible to enumerate for moderate size p . Any attempt to deal with these models requires the development of efficient computational tools to explore the model space. Here, we propose an extension to the shotgun stochastic search (SSS) of Jones et al. (2005) to sequentially learn $(G_{t,i} | D_{t-1})$. In a nutshell, our analysis generates multiple graphs at each time t from the predictive probability $\pi(G_{t,i} | D_{t-1})$, using SSS.

Suppose that, at time $t - 1$, we have saved a sample of the top N graphs $G_{t-1,i}, i = 1, \dots, N$ with highest predictive probabilities $\pi(G_{t-1,i} | D_{t-1})$. Proceeding to time t , we adopt the following search algorithm.

1. Evaluate new predictive probability $\pi(G_{t+1,i} | D_t)$ of these N graphs from time $t - 1$;
2. From among the N graphs, propose the i th graph as a new starting graph with probability proportional to $\pi(G_{t+1,i} | D_t)^c$, where c is an annealing parameter;
3. Start with $G_{t+1,i}$ and apply SSS. After each stage of SSS, compute the Bayesian model average (BMA) estimator of a predicted quantity of interest, e.g. predictive covariance matrix, using current top N graphs;

4. Stop search when certain distance between the last two BMA estimates is below a small number, set $t = t + 1$ and return to (1).

The evaluation and resample steps of (1) and (2) are important because top graphs from the previous step still represent the majority of our knowledge and should be good starting points for a new SSS once a new data sample becomes available. This scheme can also be seen as a “particle filter” for G under the resample-propagate particle learning framework of [Lopes et al. \(2011\)](#).

6 A SIMPLE EXAMPLE

To focus the idea of sequential learning in dynamic graphical model, we first consider a simple local trend DLM, namely

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\theta}_t + \boldsymbol{\nu}_t, & \boldsymbol{\nu}_t &\sim N(0, \boldsymbol{\Sigma}_t), \\ \boldsymbol{\theta}_t &= \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, & \boldsymbol{\omega}_t &\sim N(0, W_t \boldsymbol{\Sigma}_t). \end{aligned}$$

This is a special case of the general DLMs presented in previous sections. We extend the example in [Carvalho & West \(2007a\)](#) where data from $p = 11$ international currency exchange rates relative to the US dollar is analyzed. In all models, we use fairly diffuse priors, and annealing parameter $c = 1$.

We ran a set of parallel analysis differing through the value δ , with δ values of 0.93, 0.95, 0.97 and 0.99, and graph predicting models, with $M_G = M_F$ as is described by equation 11, and $M_G = M_C$ as is described by equation 12. At each of the eight pairs of (M_G, δ) , and time t , the predictive density of equation 13 is approximated by summing over top 1000 graphs at each time t , resulting in a full marginal likelihood function of (M_G, δ) . Figure 1 displays the plots over time of the eight models’ relative cumulative log predictive density under baseline model $(M_F, 0.95)$. When comparing predictive density within each δ ’s, Figure 1 shows that all four time-varying graphs generate smaller predictive density than their fixed graph peers. Figure 2 highlights the change of the relative cumulative log predictive density of the top two models. Overall, the chosen MLE from such analysis is $(M_F, 0.97)$ over the period up to the end of 08/1992 and $(M_F, 0.95)$ over the period from then until the end of data at 06/1996. The change from $\delta = 0.97$ to $\delta = 0.95$ at the end of 08/1992 reflects a more adaptive model being favored since then. The occurrence of one or two rather marked changes of relative predictive density may be due to major economic changes and events. A key such event was Britain’s withdrawal from the EU exchange rate agreement (ERM) in the September 1992 and into 1993, and this led to the deviation from the steady behaviour anticipated under a model with relative high discount factor 0.97 to the more adaptive 0.95. A second period of change of structures occurred in early 1995 with major changes in Japanese interest rate policies as a response to a weakening Yen and a move toward financial restructuring in Japan.

Figure 3 displays four snapshots of predicted adjacency matrices at four different time points. These adjacency matrices are from the median probability graph

(Scott & Carvalho, 2008), which is defined as graph consisting of those edges which have overall edge inclusion probability greater than or equal to 0.5 of being in a graph. At each time t , the edge inclusion probabilities are predicted from the outputs of the graphical model search under top identified models: $(M_F, 0.97)$ before 09/1992 and $(M_F, 0.95)$ after 09/1992. As can be seen, these best predicted graphs have several persistent signals as well as the similar overall pattern of graphs over time.

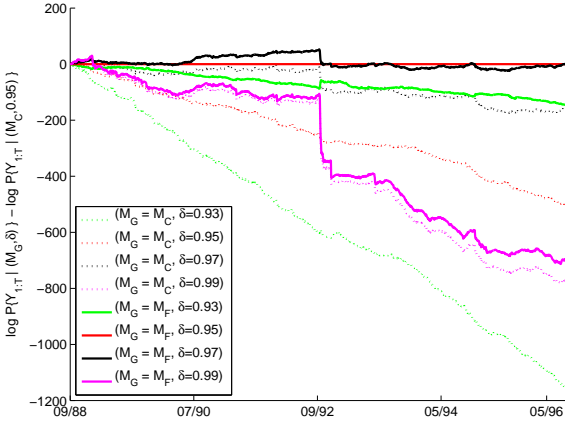


Figure 1: Relative cumulative log predictive density over time under the baseline model $(M_F, 0.95)$. The figure illustrates that the predictive density of time-varying graphs (dashed lines) are generally smaller than those of fixed graphs (solid lines)



Figure 2: Relative cumulative log predictive density of model $(M_F, 0.97)$ under the baseline model $(M_F, 0.95)$

This example serves to illustrate some features of inference with dynamic graphical models. In each of the DGMs (M_G, δ) , and for any specified sequence of graphs $\{G_t\}$, the prior, posterior, and forecast distributions are all standard distributions of well-understood forms, whether they be hyper-inverse Wishart or hyper T. Forecasts that

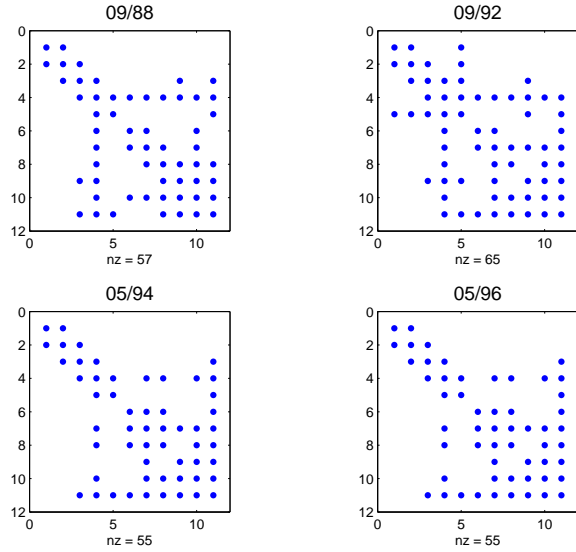


Figure 3: Four snapshots of adjacency matrices of median probability graphs predicted by using the output the stochastic search under corresponding top models: $(M_F, 0.97)$ (upper two panels), and model $(M_F, 0.95)$ (lower two panels).

take into account of graph uncertainties are easily calculated from the finite mixture of hyper T distributions of equation 13. If one is concerned about which are the best graph predicting models or which discount factors to use, their corresponding cumulative marginal density may be used to choose these specifications.

The two proposed graph predicting models together with the covariance matrix discount factors allows us to separately infer the dynamics of graphs and the dynamics of covariance matrices. In this particular example the marginal likelihoods favor static models M_F for all values of δ 's. This suggests that time-varying graphs inferred by a moving window may not produce consistently better predicting results than fixed graphs with signals detected sequentially using all historical data. On the other hand, covariance matrices Σ_t seem to be time-varying, since the marginal likelihoods of δ favor values of 0.95 or 0.97. Further exploration of these issues are necessary as should be regarded as an applied question to be dealt within a particular application. Our hope was to demonstrate that the methods presented in Section 5 provide a computationally attractive way to address these modeling questions.

7 RANDOM REGRESSION VECTOR DLM

Our applied interests are motivated by models where we attempt to predict \mathbf{Y}_t with a regression vector \mathbf{F}_t that is random and unknown before time t . Now, let $I_t = \{\mathbf{Y}_1, \dots, \mathbf{Y}_t, \mathbf{F}_1, \dots, \mathbf{F}_t\}$ denote the data and information set. Assume \mathbf{F}_t has a prior $p(\mathbf{F}_t | I_{t-1})$ at time t . Then under the assumption that the priors of (Θ_t, Σ_t) and \mathbf{F}_t

are conditionally independent given I_{t-1} , namely, $(\Theta_t, \Sigma_t) \perp\!\!\!\perp \mathbf{F}_t \mid I_{t-1}$, the following results apply.

Theorem 2. *Under the initial prior of equation (8) and with data observed sequentially to update information sets I_t the sequential updating for the matrix normal DLM on G is given as follows:*

- (i) *Posterior at $t - 1$: $(\Theta_{t-1}, \Sigma_{t-1} \mid I_{t-1}) \sim NHIW_G(\mathbf{m}_{t-1}, \mathbf{C}_{t-1}, b_{t-1}, \mathbf{S}_{t-1})$*
- (ii) *Prior at t : $(\Theta_t, \Sigma_t \mid I_{t-1}) \sim NHIW_G(\mathbf{a}_t, \mathbf{R}_t, \delta b_{t-1}, \delta \mathbf{S}_{t-1})$ where $\mathbf{a}_t = \mathbf{m}_{t-1}$ and $\mathbf{R}_t = \mathbf{C}_{t-1} + \mathbf{W}_t$*
- (iii) *One-step forecast: $p(\mathbf{Y}_t \mid I_{t-1}) = \int HT_G(\mathbf{f}_t, q_t, \delta \mathbf{S}_{t-1}, \delta b_{t-1}) p(\mathbf{F}_t \mid I_{t-1}) d\mathbf{F}_t$ with first two moments:*

$$\begin{aligned} \mathbf{r}_t &\equiv E(\mathbf{Y}_t \mid I_{t-1}) = \mathbf{a}'_t \mu_{\mathbf{F}_t} \\ \mathbf{Q}_t &\equiv \text{cov}(\mathbf{Y}_t \mid I_{t-1}) = \mathbf{a}'_t \Sigma_{\mathbf{F}_t} \mathbf{a}_t + \{V_t + \mu'_{\mathbf{F}_t} \mathbf{R}_t \mu_{\mathbf{F}_t} + \text{tr}(\mathbf{R}_t \Sigma_{\mathbf{F}_t})\} E(\Sigma_t \mid I_{t-1}) \end{aligned}$$

where $\mathbf{f}'_t = \mathbf{F}'_t \mathbf{a}_t$ and $q_t = \mathbf{F}'_t \mathbf{R}_t \mathbf{F}_t + V_t$, the first and second moments of the predictive regression vector, $\mu_{\mathbf{F}_t} = E(\mathbf{F}_t \mid I_{t-1})$ and $\Sigma_{\mathbf{F}_t} = \text{cov}(\mathbf{F}_t \mid I_{t-1})$.

- (iv) *Posterior at t : $(\Theta_t, \Sigma_t \mid I_t) \sim NHIW_G(\mathbf{m}_t, \mathbf{C}_t, b_t, \mathbf{S}_t)$ with $\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t \mathbf{e}'_t$, $\mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{A}'_t q_t$, $b_t = \delta b_{t-1} + 1$, $\mathbf{S}_t = \delta \mathbf{S}_{t-1} + \mathbf{e}_t \mathbf{e}'_t / q_t$ where $\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t / q_t$ and $\mathbf{e}_t = \mathbf{Y}_t - \mathbf{f}_t$.*

Proof. (i)(ii)(iv) follow directly from Theorem 1. (iii) results from the properties of conditional expectations applied to $p(\mathbf{Y}_t \mid I_{t-1})$, $E(\mathbf{Y}_t \mid I_{t-1})$ and $\text{cov}(\mathbf{Y}_t \mid I_{t-1})$. \square

The above theorem suggests a two stage model on the vector time series $\{\mathbf{Y}_t\}$: first, a model is fitted on low dimensional regression vectors $\{\mathbf{F}_t\}$; second, the fitted model provides the necessary quantities for the dynamic graphical DLMS. Some specific contexts of $\{\mathbf{F}_t\}$ include:

- Pre-fixed regression vector in which the \mathbf{F}_t values are specified in advance by design. This is the assumption made by the standard dynamic linear model, which yields a degenerated prior distribution $p(\mathbf{F}_t \mid I_{t-1})$ with $\mu_{\mathbf{F}_t} = \mathbf{F}_t$ and $\Sigma_{\mathbf{F}_t} = 0$. In such cases, Theorem 1. applies as a special case of Theorem 2.
- Independent and identically distributed regression vector in which the n - vector \mathbf{F}_t are commonly assumed to be independent and identically distributed from a multivariate normal distribution with mean vector $\mu_{\mathbf{F}}$ and covariance matrix $\Sigma_{\mathbf{F}}$.
- Dynamic regression vector in which another dynamic model structure could be imposed on vector process $\{\mathbf{F}_t\}$. For example, in asset pricing models, if \mathbf{F}_t is the market excessive return, a AR-GARCH type of model could be applied.

| Model | Mean | Std | Min | 25th | Median | 75th | Max |
|--------|-------|-------|--------|--------|--------|-------|-------|
| Sample | 0.158 | 0.166 | -0.530 | 0.046 | 0.159 | 0.269 | 0.836 |
| CAPM | 0.040 | 0.170 | -0.594 | -0.075 | 0.036 | 0.150 | 0.825 |
| FF | 0.014 | 0.154 | -0.557 | -0.092 | 0.011 | 0.114 | 0.816 |

Table 1: Summary statistics of correlations among sampled stocks. First row, summary of sample correlations; second and third row report summary of residual correlations after fitting CAPM and FF models respectively. For each case, at the end of April of each year from 1994 to 2008, pairwise correlations are calculated based on the monthly excess returns over the prior 60 months. Summary statistics are based on the estimated values pooled over all years.

8 EXAMPLE: PORTFOLIO ALLOCATION IN STOCKS

To demonstrate the use of DGMs in the Index Model context we work with 100 stocks randomly selected from the population of domestic commonly traded stocks in the New York Stock Exchange. By selecting a random sample of 100 we hope to reduce potential selection biases. The sample period is from January 1989 to December 2008 in a total of 240 monthly returns. Monthly returns of a one-month Treasury bill is used as the risk-free rate in the computation of the excess returns. Excess returns from the a market weighted basket of all stocks in the AMEX, NYSE and NASDAQ were used as the *market* returns. This index along with the Fama-French three factor return data were obtained from the data library of Professor Kenneth R. French ¹. Summary statistics for the excess returns series are given in the first row in Table 1. The median pairwise correlation is 0.159, indicating that there were potentially large payoffs to portfolio diversification.

In an initial exploration of the data we fitted OLS regressions to the returns using capital asset pricing model(CAPM) and Fama-French(FF) models. The second and third row in Table 1 shows summary statistics of cross-sectional residual correlations. The generally lower correlations compared with sample correlation suggest that the indexes capture most of the common variation among the securities under consideration. However, there are remaining signals in the residuals as indicated by the maximum and minimum correlations, and these are in precisely the quantities are aiming to explore relaxing the independence assumption with the inclusion of graphs.

To appreciate the importance and contribution of the use of graphical models, we consider the following alternatives: (1) sample covariance model; (2) Standard dynamic CAPM; (3) Dynamic CAPM with graphs; (4) Standard dynamic FF; (5) Dynamic FF with graphs, and (6) mixtures of (3) and (5).

In model (1), at each month t , the one-step ahead covariance matrix is based on the data from the preceding 60 months as the *in-sample* period. For model (2)-(6), we use weak priors, $m_0 = 0, C_0 = 10000I, b_0 = 3, S_0 = 0.0003I_{100}$, and $\delta = 0.983$

¹see, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

corresponding to a rolling window of about 60 months. For the random regression vector \mathbf{F}_t , we use sample mean and covariance matrix of the past 60 months as forecasts of the first and second predictive moments, $\mu_{\mathbf{F}_t}$ and $\Sigma_{\mathbf{F}_t}$. Furthermore, based on simulation experiments in Section 6, we chose model graph uncertainty with the predictive model of equation 11 for alternatives (3) and (5). In (6), CAPM and FF models are compared with each other and then averaged based upon their conditional marginal likelihood $p(\mathbf{Y}_{1:t} | \mathbf{F}_{1:t})$. The resulting posterior probabilities of FF model reaches 1 after a short period time. This should not be surprising as most of the current literature points to the use of a multi-factor model as oppose to the traditional single factor CAPM. Due to this fact, the overall performances of model (5) and (6) are close so we only report results from model (5) hereafter.

Figure 4 displays the estimated expected number of edges over time starting from January 1994 under model (3) and (5). Three results are worth noting here. First, all graphs are sparse relative to the total 4950 possible edges. The inclusion of graphs provides the necessary flexibility to capture the remaining signals from the residual covariance matrix and the data is responsible to inform which of these non-zero entries are relevant. Second, when comparing with each other, the CAPM model has more edges than FF – once again no surprises here: FF imposes a richer structure for Σ so we should expect more non-zero elements in the residual covariation of assets when the market returns are the only covariate. Third, as more information becomes available, more signals in the residuals are detected.

We now evaluate these forecasting models in two ways: forecasting ability of future correlation matrices and in the construction of optimal portfolios. This is a predictive test, in the sense that our investment strategy does not require any hindsight.

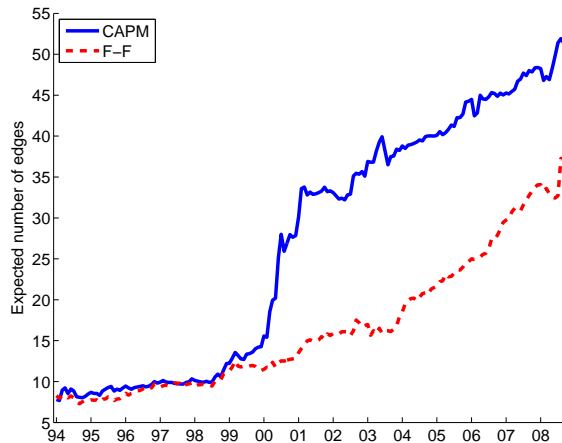


Figure 4: Estimated expectation of numbers of edge across each month.

8.1 Out-of-sample covariance forecasts

At the end of every month, the correlations forecasts from each model are compared to the sample correlations realized over a subsequent 12 months period, in the first experiment, and 36 months in the second experiment. Forecast performance is evaluated in terms of the absolute difference between the realized and forecasted values. Table 2 provide summary statistics on the absolute differences from these two experiments. When the performances evaluated using subsequent 12 months data are compared with those from subsequent 36 months, the average absolute forecast errors are reduced. The drop in forecast errors suggests that there is a lot of noise in covariance matrices measured over a period as short as 12 months. Nevertheless, as in both experiments, the relative performance of each model are generally the same.

The full sample covariance model, which is the most complex model in terms of number of free parameters, has the highest median absolute error and root mean square error. All other models are better than full covariance model. More complex model does not necessarily offer smaller forecast errors. This message is consistent with many empirical studies on correlation matrix forecast of stock returns.

Comparing the empty and the graphical models within either CAPM or FF family, we see that models with graphs are better than their empty graph peers. This is more evident in the CAPM family. Model (3) has reduced the median of the absolute differences and the root mean square errors relative to model (2), while model (5) has almost the same absolute differences as model (4). The clearer advantage in CAPM family is because there are more structures in the residuals left unexplained by only the market index than by the FF three indexes. In general, the improvement of out-of-sample covariance forecasts is minor. This is actually as expected, since the signals are very sparse which indicates the covariance matrix based on the graphical models do not differ much from those based on the traditional index models. However, as the experiments in the following section will show, these signals, though sparse, are influential when the forecasted covariance matrices are used to build optimal portfolios.

8.2 Portfolio optimization

From a practical point of view, the optimization experiments provide perhaps more important metrics for evaluating forecasting models. The setup of our portfolio optimization experiments is as follows. To highlight the role of the second predictive moment, we first form the global minimum variance portfolio. At the end of April of each year starting from 1994, we use the different models to predict the one-step ahead covariance matrix for the 100 stocks. These predictions are the input to a quadratic programming routine that defines the minimum variance portfolio (Markowitz, 1959). Short sales are allowed so that the weights are only required to be summed up to 1. These weights are then applied to buy-and-hold portfolio returns until the next April, when the forecasting and optimization procedures are repeated. The resulting time series of monthly returns of portfolios allow us to characterize the performance of opti-

| Model | 12 month | | | | 36 month | | | |
|---------------------|----------|-------|-------|---------------------|----------|-------|-------|---------------------|
| | Median | Std | 95th | $\sqrt{\text{MSE}}$ | Median | Std | 95th | $\sqrt{\text{MSE}}$ |
| (1) Full covariance | 0.238 | 0.200 | 0.654 | 0.340 | 0.160 | 0.141 | 0.460 | 0.235 |
| (2) CAPM Empty | 0.234 | 0.186 | 0.612 | 0.323 | 0.146 | 0.127 | 0.413 | 0.212 |
| (3) CAPM Graph | 0.230 | 0.184 | 0.605 | 0.319 | 0.143 | 0.123 | 0.402 | 0.207 |
| (4) FF Empty | 0.230 | 0.185 | 0.609 | 0.321 | 0.143 | 0.124 | 0.405 | 0.208 |
| (5) FF Graph | 0.230 | 0.185 | 0.607 | 0.320 | 0.143 | 0.123 | 0.404 | 0.208 |

Table 2: Performance of correlation forecasting models. Forecasts of monthly return correlation matrices are generated from different models, based on the prior 60 months of data for model (1) and based on discount factor $\delta = 0.983$ for model (2)-(5). Forecasts are then compared against the realized sample covariance estimated over the subsequent 12 months (first four) columns and 36 months (last four columns). The last estimation period ends in December 2005. Summary statistics are provided for both the distribution of the absolute difference between realized and forecasted value of pairwise correlations: Std, standard deviation of absolute differences; 95th, 95th quantile of absolute difference, and $\sqrt{\text{MSE}}$, the root mean square errors of forecasts.

mized portfolio based on each model. We also form a mean-variance portfolio using the first two moment forecast $\{\mathbf{r}_t, \mathbf{Q}_t\}$ from Theorem 2 with a target annualized excessive mean return of 15%.

Table 3 summarizes these optimization results. These are all expressed on an annualized basis. In comparison within each group, it is clear that the introduction of the graphical structure helps. The annualized standard deviation of the optimized portfolio based on the graphical CAPM model is 10.7%, yielding a Sharpe ratio of 0.688, compared to a Sharpe ratio of 0.533 for the standard CAPM portfolio. The same advantage of using graphs can be found in two models within FF class. The conclusion from this example is simple: it pays to allow for a more flexible residual covariance structure in the implementation of Index Models.

9 FURTHER COMMENTS

By allowing more flexible models for the residual covariance matrix, Financial Index models can be improved in their abilities to build more effective optimal portfolios. In this paper we take advantage of the DGMS framework of [Carvalho & West \(2007a\)](#) and show that graphical models can also be used to identify sparse signals in the residual covariance matrices and thereby obtain a more complex representation of the distribution of asset returns. Unlike [Carvalho & West \(2007a\)](#) and [Quintana et al. \(2009\)](#), in the Index Model framework, graphs are used as a parsimonious way to increase the complexity of an otherwise very restrictive model. In that sense, it is our hope that our work complements the widely used tool box of dynamic linear models for the analysis asset returns. Our first example helps illustrate the model implementation

| Model | Minimum variance portfolio | | | Mean variance portfolio | | |
|---------------------|----------------------------|-------|--------|-------------------------|-------|--------|
| | Rate | Std | Sharpe | Rate | Std | Sharpe |
| (1) Full covariance | - | - | - | - | - | - |
| (2) CAPM Empty | 0.064 | 0.120 | 0.533 | 0.064 | 0.119 | 0.535 |
| (3) CAPM Graph | 0.074 | 0.107 | 0.688 | 0.075 | 0.107 | 0.700 |
| (4) FF Empty | 0.062 | 0.109 | 0.569 | 0.069 | 0.109 | 0.627 |
| (5) FF Graph | 0.070 | 0.105 | 0.661 | 0.072 | 0.106 | 0.678 |

Table 3: Performance of portfolios based on forecasting models. At the end of April of each year from 1994 through 2007, forecasts of covariance matrices of monthly excessive returns are generated from different models. Since $T = 60 < p = 100$, the sample covariance is close to singular, we omit its results. Based upon each model's forecasts of covariance matrices, a quadratic programming procedure is used to find the global minimum variance portfolio (panel A first three column), and mean variance portfolio with target excess annual return 15% (panel B last three column). Short sales are allowed so that the weights are only constrained to sum up to 1. These weights are then applied to form portfolio returns for the next 12 months until next April, at the end of which forecasting and optimization steps are repeated and the portfolios are formed. Summary statistics are presented: Rate, the annualized excessive returns $r - r_T$, where the annualized portfolio return r is determined by $(1 + r)^{14} = \prod_{i=1}^{168} (1 + r_i)$, and annualized risk-free return r_T is determined by $(1 + r_T)^{14} = \prod_{i=1}^{168} (1 + r_{T,i})$ with r_i and $r_{T,i}$ denoting the monthly return of portfolio and risk-free asset; Std, the annualized standard deviation of excess returns $r_i - r_{T,i}$; and Sharpe ratio, the annualized excessive return divided by the annualized standard deviation.

and highlight the issue of specifying discount factors and graph predicting models. The second example discusses and explores aspects of random regression vectors and variable selection. This analysis confirmed that the CAPM and FF model generally do well in explaining the variation of stock returns, but identifying relevant non-zero entries in the unexplained covariation are of real practical value: the resulting covariance matrix forecast has lower out-of-sample forecast errors, and the corresponding portfolios achieve lower level of realized risk in terms of variance and higher realized returns.

In addition to our case studies, we have also provided a fully Bayesian framework of two-stage forecast of covariance matrices, a mechanism of graph evolution, and the use of sequential stochastic search for high-dimensional graphical model space.

In regards to the modeling of graphical structure through time, alternative approaches include the use of first-order Markov probabilities in which the graph obtained at time t depends on which of the graphs obtained at time $t - 1$, but not on what happened prior to $t - 1$, and higher-order Markov probabilities that extend the dependence to graphs at time $t - 2, t - 2, \dots$, etc. These alternatives require the learning of a higher-dimensional transition matrix between graphs. Even a sparse representation of the transition matrix, such as each graph only moves to its neighbors between two time points is limited in a sense that the sparse pattern would restrict the evolution of graphs between time.

The sequential stochastic search algorithm combines the sequential Monte Carlo idea and shotgun stochastic search algorithm. Exploration of a static model space to find high posterior probability graphs can be successfully carried out using direct search such as shotgun stochastic search method, certainly up to 100 vertices or so while traditional MCMC is competitive only for relatively small graphs (Jones et al., 2005). However, fast searching a sequence of large model space is more challenging. This problem can be eased by noticing that from one step to the next we do not expect large changes in the mass of the distribution. Therefore, we could use the high probability graphs from the previous step as starting points to initiate a new search and rapidly traverse the graphical model space around these promising models.

REFERENCES

- CARVALHO, C. M. & WEST, M. (2007a). Dynamic matrix-variate graphical models. *Bayesian Analysis* **2**, 69–98.
- CARVALHO, C. M. & WEST, M. (2007b). Dynamic matrix-variate graphical models - A synopsis. In *Bayesian Statistics VIII*, J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith & M. West, eds. Oxford University Press.
- COCHRANE, J. (2001). *Asset Pricing*. Princeton University Press.
- DAWID, A. P. & LAURITZEN, S. L. (1993). Hyper-Markov laws in the statistical analysis of decomposable graphical models. *Annals of Statistics* **21**, 1272–317.

- FAMA, E. F. & FRENCH, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* **33**, 3–56.
- JONES, B., CARVALHO, C., DOBRA, A., HANS, C., CARTER, C. & WEST, M. (2005). Experiments in stochastic computation for high-dimensional graphical models. *Statistical Science* **20**, 388–400.
- LAURITZEN, S. L. (1996). *Graphical Models*. Oxford: Clarendon Press.
- LOPES, H., CARVALHO, C., JOHANNES, M. & POLSON, N. (2011). Particle learning. In *Bayesian Statistics IX*, J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith & M. West, eds. Oxford University Press. To appear.
- MARKOWITZ, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. New York: Wiley.
- QUINTANA, J., LOURDES, V., AGUILAR, O. & LIU, J. (2003). Global gambling. In *Bayesian Statistics VII*, J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith & M. West, eds. Oxford University Press.
- QUINTANA, J. M., CARVALHO, C. M., SCOTT, J. & COSTIGLIOLA, T. (2009). Futures markets, bayesian forecasting and risk modeling. In *The Handbook of Applied Bayesian Analysis*, T. O’Hagan & M. West, eds. Oxford University Press.
- QUINTANA, J. M. & WEST, M. (1987). Multivariate time series analysis: New techniques applied to international exchange rate data. *The Statistician* **36**, 275–281.
- RAJARATNAM, B., MASSAM, H. & CARVALHO, C. M. (2008). Flexible covariance estimation in graphical gaussian models. *Annals of Statistics* **36**, 2818–49.
- SCOTT, J. G. & CARVALHO, C. M. (2008). Feature-inclusion stochastic search for gaussian graphical models. *Journal of Computational and Graphical Statistics* **17**, 790–808.
- SHARPE, W. F. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance* **19**, 425–442.
- WEST, M. & HARRISON, P. (1997). *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag.
- ZELLNER, A. & CHETTY, V. K. (1965). Prediction and decision problems in regression models from the bayesian point of view. *Journal of the American Statistical Association* **60**, 608–616.