Proc. Valencia / ISBA 8th World Meeting on Bayesian Statistics
Benidorm (Alicante, Spain), June 1st–6th, 2006

Dynamic Matrix-Variate Graphical Models – A Synopsis[1] –

Carlos M. Carvalho & Mike West
ISDS, Duke University, USA
carlos@stat.duke.edu

Summary

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 models are then applied in the context of financial time series for predictive portfolio analysis providing a significant improvement in performance of optimal investment decisions.

Keywords and Phrases: Dynamic Linear Models, Gaussian Graphical Models, Portfolio Analysis.

1. INTRODUCTION

Bayesian dynamic linear models (DLMs) (West and Harrison, 1997) are used for analysis and prediction of time series of increasing dimension and complexity in many applied fields. 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 models of increasingly 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 increasing prevalence of high-dimensional multivariate time series and the consequent needs to scale and more highly structure analysis methods. Contexts of high-dimensional and rapidly sampled financial time series are central examples, though similar needs are emerging in many areas of science, social science and engineering. This paper introduces a broad new class of time series models to address this: the framework synthesises 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.

The presentation outlines the framework of matrix-variate DLMs and Gaussian graphical models for structured, parameter constrained covariance matrices based

[1]To appear in *Bayesian Statistics 8,* the Proceedings of the Valencia/ISBA 8th World Meeting on Bayesian Statistics. This is a 6-page synopsis of: *Dynamic Matrix-Variate Graphical Models,* by C.M. Carvalho & M. West, *Bayesian Analysis*, **2** (to appear).

on the use of the family of hyper-inverse Wishart distributions. We then discuss formal model specification and details of the resulting methodology for both constant and, of more practical relevance, time-varying structured covariance matrices in the new models. We summarise the theory that extends DLM sequential updating, forecasting and retrospective analysis to this new model class.

Our applied examples combine these flexible new Bayesian models with Bayesian decision analysis in financial portfolio prediction studies. 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, as well as illustrating the real practical utility, and benefits over existing models, of the new methodology.

2. BACKGROUND

2.1. *Matrix-Variate Dynamic Linear Models*

The class of Matrix Normal DLMs (Quintana and West, 1987) represents a general, fully-conjugate framework for multivariate time series analysis and dynamic regression with estimation of cross-sectional covariance structures.

The model for the $p \times 1$ vector $\mathbf{Y}_t$ is defined by

$$\begin{aligned} \mathbf{Y}_t' &= \mathbf{F}_t'\mathbf{\Theta}_t + \boldsymbol{\nu}_t', & \boldsymbol{\nu}_t &\sim N(\mathbf{0}, V_t\mathbf{\Sigma}), & (1)\\ \mathbf{\Theta}_t &= \mathbf{G}_t\mathbf{\Theta}_{t-1} + \mathbf{\Omega}_t & \mathbf{\Omega}_t &\sim N(\mathbf{0}, \mathbf{W}_t, \mathbf{\Sigma}), & (2) \end{aligned}$$

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

2.2. *Gaussian Graphical Models*

Graphical model structuring for multivariate models characterizes conditional independencies via graphs (Dawid and Lauritzen, 1993; Jones *et al*, 2005) and provides methodologically useful decompositions of the sample space into subsets of variables 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 normal distributions, conditional independence restrictions are simply expressed through zeros in the off-diagonal elements of the precision 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 fully conjugate Bayesian analysis of decomposable Gaussian graphical models (David 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

$$\mathbf{\Sigma} \sim HIW_G(b, \mathrm{D}) \qquad (3)$$

has a degree-of-freedom parameter b and location matrix $\mathrm{D} \in M(G)$ implying that for each clique $P \in \mathcal{P}$, $\mathbf{\Sigma}_P \sim IW(b, \mathrm{D}_P)$ where D_P is the positive-definite symmetric diagonal block of D corresponding to $\mathbf{\Sigma}_P$. The full HIW is conjugate to the likelihood from a Gaussian sample with variance $\mathbf{\Sigma}$ on G.

3. SPARSITY IN DLMS: GENERALIZATION TO HIW

Graphical structuring can be incorporated in matrix normal DLMs to provide models for $\mathbf{\Sigma}$ that allow structure, induce parsimony and lead to statistical and computational efficiencies as a result. For a given decomposable graph, take the hyper-inverse Wishart as a conjugate prior for $\mathbf{\Sigma}$; it turns out that the closed-form, sequential updating theory of DLMs can be generalized (Theorem 1 of Carvalho and West, 2006) to this richer model class.

Consider the matrix normal DLM above and suppose $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$ is constrained by a graph G. If D_t is the data and information set conditioned upon at any time t, assume the NHIW initial prior of the form

$$(\mathbf{\Theta}_0, \mathbf{\Sigma} \mid D_0) \sim NHIW_G(\mathbf{m}_0, \mathbf{C}_0, b_0, \mathbf{S}_0). \tag{4}$$

In components,

$$(\mathbf{\Theta}_0 \mid \mathbf{\Sigma}, D_0) \sim N(\mathbf{m}_0, \mathbf{C}_0, \mathbf{\Sigma}) \qquad \text{and} \qquad (\mathbf{\Sigma} \mid D_0) \sim HIW_G(b_0, \mathbf{S}_0), \tag{5}$$

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, (detailed in Carvalho and West, 2006), generating a sequence of NHIW priors and posteriors for $(\mathbf{\Theta}, \mathbf{\Sigma})$.

4. TIME-VARYING $\mathbf{\Sigma}_T$

Importantly, the above development extends to the practically critical context of time-varying $\mathbf{\Sigma} \to \mathbf{\Sigma}$ to induce a novel class of *graphical multivariate volatility models.* 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, as well as areas in engineering and the natural sciences.

Based on a specified discount factor δ, $(0 < \delta \leq 1)$, beginning at $t-1$ with current posterior

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

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

$$(\mathbf{\Sigma}_t \mid D_{t-1}) \sim HIW_G(\delta b_{t-1}, \delta \mathbf{S}_{t-1}). \tag{6}$$

The time-evolution maintains the inverse-Wishart form for the prior of $\mathbf{\Sigma}_t$, while increasing the spread of the HIW distribution by reducing the degrees-of-freedom and maintaining the location at $\mathbf{S}_{t-1}/b_{t-1}$. 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,

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

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

5. LARGE-SCALE DYNAMIC PORTFOLIO ALLOCATION

Bayesian forecasting models and Bayesian decision analysis in asset allocation problems has been routine for a number of years, from the seminal paper of Quintana (1992) to more recent work (e.g., Aguilar and West, 2000, Quintana *et al*, 2003) with increasingly large problems. Forecast 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.

At 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 optimal portfolio weights can be expressed in terms of the precision matrix $\mathbf{K}_t = \mathbf{Q}_t^{-1}$, via

$$w_{ti}^{(m)} = \lambda \frac{f_{ti} - \sum_{j \neq i} (k_{tij}/k_{tii}) f_{tj}}{k_{tii}^{-1}} \tag{7}$$

(Stevens, 1998), where λ is a Lagrange multiplier. In normal models, 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. Hence the investment in asset i depends on the ratio of the expected return that cannot be explained by the linear combination of assets to the *unhedgeable* risk.

In higher-dimensional portfolios, the optimal weights can be very volatile over time due to the uncertainty in the estimation of covariance matrices. Structured variance models should help: the above equations suggest that conditional independence assumptions can directly influence 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 noise is added to the problem.

Two examples bear this out: "sparse" (with graphical modelling constraints) models lead to portfolios that are both less risky and more profitable than under the standard construction with "full" (unconstrained) variance matrices.

5.1. *Example: International Exchange Rates*

This first application is a dynamic version of the example in Carvalho, Massam and West (2005) where portfolios for $p = 11$ international currency exchange rates relative to the US dollar were compared. The study here uses the graph displayed in Figure 1. For each model at each time point, mean-variance and minimum-variance portfolios were computed based on the one-step ahead forecasts. In comparing the impact on portfolio predictions and decisions of the proposed structured model *vis-a-vis* the unconstrained DLM, the overall conclusion is 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 higher realized cumulative returns, lower risk portfolios in terms of one-step ahead predictive variances and lower volatility of the optimal portfolio weights as they are sequentially revised, consistent with the idea of more stable portfolios. This example demonstrates the relevance of appropriate model structuring: the graphical model DLM generates more accurate predictions and optimal portfolio decisions, with lower

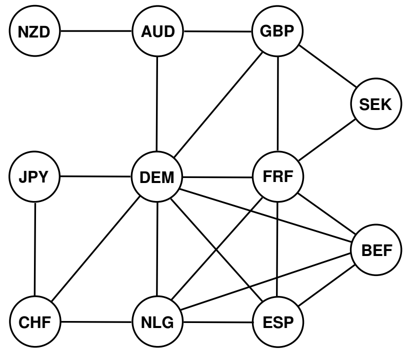

Figure 1: *Graph determining the conditional independence structure in the exchange rate/portfolio investment example.*

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.

5.2. *Example: Portfolio Allocation in the S&P 500*

A higher-dimensional application involves $p = 346$ securities forming part of the S&P500 stock index. In this higher-dimensional setting, graphical models to induce structure are particularly key. The example also addresses graphical model structure selection, evaluating models using Metropolis stochastic search (Jones *et al*, 2005) to explore the full space of graphical DLMs using only the first 1,200 observations in the data set. On the remaining data, we sequentially updated portfolios use a few graphical DLMs selected based on the posterior at $t = 1,200$. The key conclusions (see Figure 2 here, and details in Carvalho and West, 2006) are that (a) 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; and that (b) higher realized returns are coupled with lower risk and lower volatility of time trajectories of portfolio weights.

6. SUMMARY COMMENTS

The marriage of DLMs with graphical models defines a new, rich class of matrix DLMs that allow for the incorporation of conditional independence structure in the stochastically varying, cross-sectional structure of a set of time series. Our examples focus on Bayesian decision analysis for sequential portfolio allocation, where the utility and benefits of the new models are sharply illuminated. The value of data-consistent structuring and constraints on variance matrices across series is evident; the implied parsimony in parametrization, statistical efficiency in estimation and reduced uncertainty translates into more accurate predictions and decisions. Our

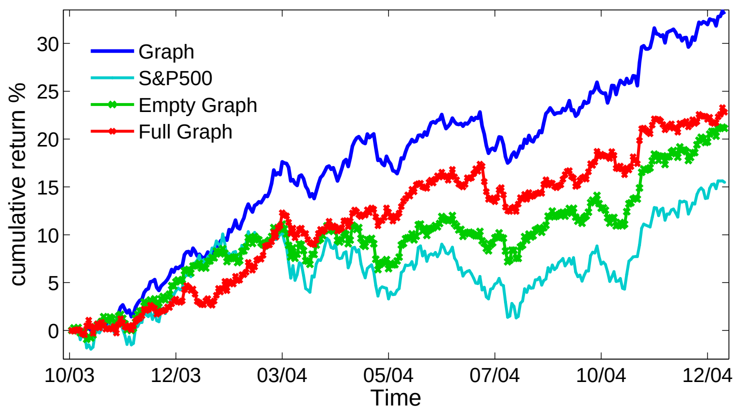

Figure 2: *S&P 500 portfolios: comparison of cumulative returns under different models* ($\delta = 0.98$).

second example 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 on a given graph. The models open up a rich new area for application of Bayesian modelling, as well as research directions in the class of dynamic graphical models.

ACKNOWLEDGMENTS

The authors acknowledge the support of grants from the NSF (DMS 0102227 and 0342172) and the NIH (U54-CA-112952-01).

REFERENCES

Aguilar, O. and West, M. (2000). Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *J. Business Econ. Studies* **18**, 338–357.

Carvalho, C.M. and West, M. (2006). Dynamic matrix-variate graphical models. *Bayesian Analysis* (to appear)

Carvalho, C.M., Massam, H. and West, M. (2005) Simulation of hyper inverse-Wishart distributions in graphical models. *ISDS discussion paper*, (submitted).

Dawid, A.P. and Lauritzen, S.L. (1993). Hyper-Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **3**, 1272–1317.

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

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

Quintana, J.M. (1992). Optimal portfolios of forward currency contracts. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) Oxford: University Press, 753–762.

Quintana, J.M., Lourdes, V., Aguilar, O. and Liu, J. (2003). Global gambling (with discussion). *Bayesian Statistics 7* (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, eds.) Oxford: University Press, 349-368.

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

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. New York: Springer