Financial time series graphical modelling and portfolio analysis

By CRAIG REESON

Trinity College, Duke University, Durham, North Carolina 27708, U.S.A.

crr15@duke.edu
Summary

This paper considers the efficacy of the synthesis of dynamic linear models (DLMs) and Gaussian graphical models put forward in Carvalho & West (2007a) as a model for investment for an individual. The paper summarizes the background information and then applies the model to a set of 30 Vanguard mutual funds commonly found in 401k plans. I consider restrictions that effect individual investors, specifically no short sales and transaction costs. First, a local smoothing model is employed on 7 years of monthly data in an attempt to reduce the volatility of the realized returns. Then I apply a dynamic regression model to 5.5 years of daily data to both predict returns and reduce volatility. Furthermore, the implied regressions of the estimated covariance matrices are considered and used to examine the choice of index and managed funds. Following Rodriguez & Tower (2007), I test the hypothesis that each managed fund can be tracked by a basket of index funds and find that the higher the $R^2$ the lower the average returns, finding only minimal evidence. Finally, other results are considered, such as generating portfolios which are uncorrelated with a certain benchmark and consider its use as a metric for evaluating covariance models as well as simulating decomposable graphs.

Some key words: Gaussian graphical models; Hyper-inverse Wishart; Portfolio analysis; Dynamic linear models; Vanguard mutual funds; Dynamic regression models
1 Introduction

The question of how to design a portfolio among \( p \) securities has been an area of intense interest since Markowitz first studied the problem in 1952 (Markowitz, 1952). In his seminal work he laid the groundwork for both research and applied work that has changed the very way markets are thought about. His method consists of first estimating the future returns as well as the variance-covariance structure of these estimates, and then forming a set of efficient portfolios and then allowing the investor to choose his desired combination of risk and return.

Since Markowitz’s time much work has been done, in particular from a Bayesian perspective, to extend his results to consider different loss functions and especially to consider the computational and estimating efficiency of the model. However, little of this work has focused on the constraints that would face an ordinary investor.

The introduction of Bayesian dynamic linear models (DLMs) in West & Harrison (1997) has led to extensive work in one step ahead predictions and especially in trying to solve the problem of modeling the volatility structure through time.

In Carvalho & West (2007a) a new synthesis of matrix-variate dynamic linear models with Gaussian graphical models (Dawid & Lauritzen, 1993) was introduced. Using a matrix normal hyper-inverse Wishart prior allows for closed form sequential updating while also modeling conditional independence relationships. In the context of the model, this restricts entries in the precision matrix to be 0 and therefore lowers the parameter space being estimated. This model is further investigated in the following sections.

Section 2 develops the background finance and statistics necessary for analysis, section 3 explains the model, section 4 applies a simple version of the model to some monthly Vanguard mutual fund data, section 5 uses a state space model to analyze
daily Vanguard mutual fund data, section 6 compares managed and index funds, section 7 explores other related topics considered, and section 8 concludes.

2 Background

2.1 Portfolio Theory

In modern portfolio theory pioneered by Markowitz, given predicted returns for the next period \( r_{t+1} \), a target portfolio return \( m_{t+1} \), and the variance-covariance structure of the returns \( K_{t+1} \), the essential problem is to solve for weights \( w_{t+1} \) to minimize the portfolio variance \( w'_{t+1}K_{t+1}w_{t+1} \), subject to the constraints \( w'_{t+1}1 = 1 \) and \( w'_{t+1}r_{t+1} \geq m_{t+1} \). This problem can be solved easily using Lagrange multipliers (see for example Merton (1972)). The problem arises from the fact that both \( f_{t+1} \) and \( K_{t+1} \) are unknown and must be estimated by the investor. This problem has been widely considered (Sharpe, 1963) (Elton & Gruber, 1973) (Jobson & Korkie, 1980) (Jorion, 1986) (Quintana et al., 2003) and is traditionally addressed by assuming some model for the returns and applying this to estimate expected returns \( f_{t+1} \) and expected covariance of the returns \( \Sigma_{t+1} \) using historical data.

However, once the restriction that \( w_{t+1,i} \geq 0 \forall i \) is added this problem no longer has an analytical solution and requires numerical approximation to solve. For the analysis in this paper I use the quadratic programming function MATLAB®’s Optimization Toolbox.

The Markowitz formulation also ignores transaction costs. To include them in the model the constrain must be adjusted tot \( w'_{t}f \geq m_{t} \) to \( w'_{t}f - \kappa 1'\Delta w_{t} \geq m_{t} \) where \( \kappa \) accounts for proportional costs such as bid asks spreads or mutual fund fees. \( \Delta w_{t} = |w_{t} - w_{t-1}^{adjusted}| \) and \( w_{t-1,i}^{adjusted} = \frac{w_{t-1,i}(Y_{i-1}+1)}{\sum_{i}w_{t-1,i}(Y_{i-1}+1)} \), the weights at time \( t-1 \) adjusted for differences in the returns.
2.2 Matrix-Variate Dynamic Linear Models

The theory of matrix normal dynamic linear models (DLMs) is developed in chapter 16 of West & Harrison (1997). The theory allows for generalized, fully conjugate analysis of multivariate time series with cross-sectional covariances. Following the analysis in Carvalho & West (2007a), each of the individual time series, \( Y_{it} \), follows the DLM fully defined by the quadruple

\[
\{F_t, G_t, V_t, \sigma^2_{\text{it}}, W_t, \sigma^2_{\text{ti}}\}
\]

The notation, with \( t \) representing time and \( i \) indexing the series \( (i = 1, \ldots, p) \), represents the set of \( p \) DLMs:

\[
\text{Observation: } Y_{ti} = F'_t \theta_{ti} + \nu_{ti}, \quad \nu_{ti} \sim N(0, V_t \sigma^2_{\text{ti}}) \\
\text{Evolution: } \theta_{ti} = G_t \theta_{t-1,i} + \omega_{ti}, \quad \omega_{ti} \sim N(0, W_t \sigma^2_{\text{ti}})
\]

(1)

Where \( F_t \) is a known \( n \times 1 \) regression vector, \( V_t \) is a known signal to noise ratio at time \( t \), \( G_t \) is a known \( n \times n \) state evolution matrix, \( W_t \) is a known \( n \times n \) state innovation matrix, \( \theta_{ti} \) is a \( n \times 1 \) state vector specific to series \( i \), and \( \sigma^2_{\text{it}} \) is the variance of the unexplainable innovations.

Note that following Liu (2000) and Quintana et al. (2003) the variances, \( \sigma^2_{\text{ti}} \), are time varying. Moreover, past observations are discounted by some factor \( \delta \) such that \( 0 \ll \delta \leq 1 \). This allows for adaptation to changes in the volatility structure through time, the speed of adaptation controlled through \( \delta \). The formal model, developed in Uhlig (1994) involves subjecting \( \sigma^2_{\text{ti}} \) beta shocks at each time \( t \).

With the conditional independence assumptions \( \nu_{ti} \) and \( \omega_{ti} \) are independent across time and mutually independent at each time, and by letting \( \sigma_{t,ij} \) be the covariances between series \( i \) and \( j \) (so that \( \Sigma_t \) is the variance-covariance matrix), then these \( p \) univariate DLMs can be represented as the matrix DLM:
\[ Y_t' = F_t' \Theta_t + \nu_t, \quad \nu_t \sim N(0, V_t \Sigma_t) \]

\[ \Theta_t = G_t \Theta_{t-1} + \Omega_t, \quad \Omega_t \sim N(0, W_t, \Sigma_t) \]  

(2)

Where:

- \( Y_t \) is the \( p \times 1 \) observation vector of returns;
- \( F_t \) is the \( n \times 1 \) vector of predictors;
- \( \Theta_t \) is the \( n \times p \) matrix of regression coefficients;
- \( V_t \) are known scale factors that define the signal to noise ratio;
- \( \Sigma_t \) is the \( p \times p \) variance-covariance matrix of unexplained returns;
- \( G_t \) is a known \( n \times n \) matrix of state evolutions;
- \( W_t \) is a known \( n \times n \) left variance-covariance matrix for evolutions in the regression coefficients.

In general is defined as \( W_t = C_{t-1}^{(1-\delta)} \). This choice follows West & Harrison (1997) and Carvalho & West (2007a) as it is convenient for updating.

The updating equations for this model can be found in Carvalho & West (2007a). Note, however, that with \( D_t \) representing the data known at time \( t \), in order to allow for closed form updating the conditional priors are \( (\Sigma_0|D_0) \sim IW(b_0, S_0) \), the inverse Wishart distribution (the conjugate prior for variance matrices defined in Dawid (1981)) with degrees of freedom \( b_0 \) and location matrix \( S_0 \) given, and \( (\Theta_0|D_0, \Sigma_0) \sim N(m_0, C_0, \Sigma_0) \), the matrix normal distribution defined in Dawid (1981) with mean \( m_0 \), left covariance matrix \( C_0 \), and right covariance matrix \( \Sigma_0 \) given. Jointly, we say that \( (\Theta_0, \Sigma_0|D_0) \sim NIW(m_0, C_0, b_0, S_0) \), the normal inverse Wishart distribution.

### 2.3 Gaussian graphical models

Using graphical models to structure conditional independence relationships has a long history (c.f. Dempster (1972) Whittaker (1990), and Speed & Kiiveri (1986)).
Graphical models are especially powerful in high-dimensional problems because they allow for complex problems to be handled through a combination of simpler elements, generating computational efficiency and often reducing noise. Furthermore, especially relevant to the model considered in this paper, in multivariate normal distributions conditional independence restrictions are expressed as zeros in the precision matrix. The terminology for graphical models which can be found in Carvalho & West (2007a), Jones et al. (2005), and Lauritzen (1996) is defined below.

Consider a vector \( \mathbf{x} = (x_1, \ldots, x_p)' \) where \( \mathbf{x} \sim N(0, \Sigma) \) with precision \( \Omega = \Sigma^{-1} \), where the \( ij \)-th element is \( \omega_{ij} \). Then let \( G = (V, E) \) be the undirected graph defining the conditional independence relationships of \( \mathbf{x} \). Thus, the vertex set \( V \) corresponds to the \( p \) random variables in \( \mathbf{x} \), and \( E \) is the edge set which contains the element \((i, j)\) for all such elements \( i \) and \( j \) where \( i \) and \( j \) are not conditionally independent, i.e. when \( \omega_{ij} \neq 0 \). Thus \( \Omega \) is a positive definite matrix such that \( \omega_{ij} = 0 \ \forall (i, j) \notin E \). An overview of the terminology for graph theory can be found in appendix A.

Decomposable graphs are especially important because of the Hammersley-Clifford theorem (Hammersley & Clifford, 1971). This theorem implies that the likelihood factors over cliques and separators as follows, allowing for tractable analysis:

\[
p(X|G) = \frac{\prod_{C \in C} p(X_C)}{\prod_{S \in S} p(X_S)} \tag{3}
\]

The use of experts to specify a prior graphical structure which can be overridden by data suggests a Bayesian approach. Bayesian analysis of Gaussian graphical models, developed in Dawid & Lauritzen (1993) uses of the hyper-inverse Wishart (HIW) distribution. We say that \( \Sigma \sim HIW_G(b, \mathbf{D}) \) with degree of freedom parameter \( b \) and location matrix \( \mathbf{D} \) if each of the clique-marginals (and similarly for separators) are \( \Sigma_C \sim IW(b, D_C) \), where \( D_C \) is the positive-definite symmetric diagonal block of \( \mathbf{D} \).
corresponding to $\Sigma_C$. Note that $D^{-1} \in M(G)$, the set of all symmetric positive-definite matrices where the $ij$th element equals 0 if $(i, j) \notin V$. Then the distribution for $\Sigma$, constrained by $G$ factorizes over cliques and separators as:

$$p(\Sigma|b, D) = \frac{\prod_{C \in \mathcal{C}} p(\Sigma_C|b, D_C)}{\prod_{S \in \mathcal{S}} p(\Sigma_S|b, D_S)}$$

(4)

Finally, we define the matrix-normal/HIW distribution as in Carvalho & West (2007a). Let $\Sigma \sim HIW_G(b, D)$ on a given graph $G$, and $\mathbf{X}|\Sigma \sim N(m, W, \Sigma)$ for given $b$, $D$, $m$, and $W$. Then we denote this by $(\mathbf{X}, \Sigma) \sim NHIW_G(m, W, b, D)$ and $\mathbf{X} \sim HT_G(m, W, D, b)$ defined in Dawid & Lauritzen (1993).

3 The Model

3.1 Overview

The model in this paper forms a synthesis of matrix-variate DLMs and Gaussian graphical models. It incorporates a time varying stochastic volatility matrix making use of a discount factor to allow for changes in relationships over time. The essential components of the model follow, while the fully theory can be found in Carvalho & West (2007a) (summarized in Carvalho & West (2007b)).

If there are $p$ mutual funds and $n$ prediction variables known at time $t$ we model the returns of the funds at time $t$, $\mathbf{Y}_t' = (Y_{t1}, \ldots, Y_{tp})$, as the matrix-variate DLM

**Observation**: $\mathbf{Y}_t' = \mathbf{F}_t' \Theta_t + \nu_t, \quad \nu_t \sim N(0, V_t \Sigma_t)$

**Evolution**: $\Theta_t = \mathbf{G}_t \Theta_{t-1} + \Omega_t, \quad \Omega_t \sim N(0, W_t, \Sigma_t)$

(5)

With the variables as before, the only difference being that $\Sigma_t$ is now constrained by some graph $G$. 
3.2 Updating

One of the most appealing aspects of the model is the closed form updating that it allows. For proof of the updating equations see Carvalho & West (2007a). Letting $D_t$ stand for the data and information set known at time $t$, the updating is given by:

- Posterior at time $t - 1$
  $$(\Theta_{t-1}, \Sigma_{t-1}|D_{t-1}) \sim NHIW_G(m_{t-1}, C_{t-1}, b_{t-1}, S_{G,t-1})$$

- Prior at time $t$
  $$(\Theta_t, \Sigma_t|D_{t-1}) \sim NHIW_G(a_t, R_t, \delta b_{t-1}, \delta S_{G,t-1})$$
  With:
  $$a_t = G_t m_{t-1} \text{ and } R_t = G_t C_{t-1} G_t'/\delta$$

- Forecast for time $t$
  $$(Y_t|D_{t-1}) \sim HT_G(f_t, \delta Q_t S_{G,t-1}, \delta b_{t-1})$$
  With:
  $$f_t' = F_t' a_t \text{ and } Q_t = F_t' R_t F_t + V_t$$

- Posterior at time $t$
  $$(\Theta_t, \Sigma_t|D_t) \sim NHIW_G(m_t, C_t, b_t, S_{t,G})$$
  With:
  $$m_t = a_t + A_t e_t'$$
  $$C_t = R_t - A_t A_t' Q_t$$
  $$b_t = \delta b_{t-1} + 1$$
  $$S_t = S_{t-1} + e_t e_t'/Q_t$$

Where:
\[ \mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t / Q_t \text{ and } \mathbf{e}_t = \mathbf{Y}_t - \mathbf{f}_t \]

In order to move from \( \mathbf{S} \) to \( \mathbf{S}_G \) we employ use completion. This fixes the elements of \( \mathbf{S} \) for which \((i, j) \in E\) and then chooses the free parameters so that \( \mathbf{S}^{-1} \in M(G) \). A proof that such a matrix exists and is unique can be found in Dempster (1972) and the formulas used are from Massam & Neher (1998). This allows for efficiency in computation as we can run the DLM once to calculate \( \mathbf{S} \) and then only complete \( \mathbf{S} \) to get \( \mathbf{S}_G \) to calculate likelihoods and portfolios.

As this paper focuses on sequential, one step portfolio reallocation using the posterior distribution of \( \mathbf{\Sigma}_t \) at time \( t \), of key relevance in the forecasting is the point estimate for the mean, \( \mathbf{S}_G / (b_t - 2) \) (Dawid & Lauritzen, 1993).

To initialize the model we need a prior for \( \mathbf{\Theta} \) and \( \mathbf{\Sigma} \), \( (\mathbf{\Theta}_0, \mathbf{\Sigma}_0|D_0) \sim NHIW_G(\mathbf{m}_0, \mathbf{C}_0, b_0, \mathbf{S}_0) \) as well as a choice of discount factor \( \delta \) and the prior for the graph. Throughout this paper we use a prior for \( G \) where each edge has Bernoulli probability \( \beta \) of inclusion. This simplifies to \( \log \Pr(G) \propto |G| \log \frac{\beta}{1-\beta} \), a penalty for additional edges, and allows us to specify our desire for sparsity in the graph. In general all else equal sparser graphs are better as they lower the parameter space.

### 3.3 Retrospection

Some interest lies in retrospective analysis or smoothing of our results. Specifically, in model exploration, such as the choice of predictors, retrospective analysis can be used to create highest posterior density (HPD) regions and allow for accept and reject decisions based on Lindley’s method.

The retrospective estimate for \( \mathbf{\Theta}_{t-k} \) is:

\[
(\mathbf{\Theta}_{t-k}|\mathbf{\Sigma}_{t-k}, D_t) \sim N(\mathbf{a}_t(-k), \mathbf{R}_t(-k), \mathbf{\Sigma}_{t-k})
\]  

(6)
Where the parameters are calculated through the recurrences (where $a_t(0) = m_t$, $R_t(0) = C_t$):

\begin{align*}
B_{t-k} &= C_{t-k}C'_{t-k+1}R_{t-k} \\
\mathbf{a}_t(-k) &= \mathbf{m}_{t-k} + B_{t-k}[\mathbf{a}_t(-k + 1) - \mathbf{a}_{t-k+1}] \\
R_t(-k) &= C_{t-k} + B_{t-k}[R_t(-k + 1) - R_{t-k+1}]B'_{t-k}
\end{align*} (7)

The retrospective estimate for $\Sigma_{t-k}$ has no closed form solution. However, we can calculate the retrospective mean of $\Omega_{t-k}$. Following the methods in West & Harrison (1997) for multivariate inverse Wishart models, we have that letting $S_t(0) = S_t$, and $b_t(0) = b_t$ the retrospective estimate for each clique $\Omega_C$ can be calculated by:

\[ E[\Omega_{C,t-k}|D_t] = S_{C,t}^{-1}(-k)(b_t(-k) + |C| - 1) \] (8)

Where $S_{C,t}$ is the submatrix of $S_t$ corresponding to clique $C$ and $b_t(-k)$ and $S_{C,t}(-k)$ are:

\begin{align*}
S_{C,t}^{-1}(-k) &= (1 - \delta)S_{C,t}^{-1} + \delta S_{C,t}(-k + 1)^{-1} \\
b_t(-k) &= (1 - \delta)b_t(-k) + \delta b_t(-k + 1)
\end{align*} (9)

Each of these retrospective estimates can then be combined to form a retrospective estimate for $\Omega$. Using this, the harmonic mean of $\Sigma_{t-k}$, $\hat{\Sigma}_{t-k} = 1/E[\Omega_{t-k}]$, can be used as a reasonable estimate for $E[\Sigma_{t-k}]$, so that

\[ [\Theta_{t-k}]_{ij} \sim T_{b_t(-k)}[m_{ij}, [C_{t-k}]_{ii}E[\Sigma_{t-k}]_{jj}] \approx T_{b_t(-k)}[m_{ij}, [C_{t-k}]_{ii}][\hat{\Sigma}_{t-k}]_{jj} \] (10)
Central to our estimate for $\Sigma_t$ is the choice of graph, $G$, specifying the conditional dependence structure of the unpredictable returns. However, the search over graphs is not a trivial process. First, for $p > 10$ there are generally prohibitively many graphs, $\binom{2^p}{p}$, to calculate the probabilities and returns for all of them. Second, we can only compute the likelihood over decomposable graphs. With $p$ larger than 15 it rapidly becomes difficult to find decomposable graphs. Finally, the methods to compute the likelihood are computationally intensive, especially with large $p$ and $n$.

In order to deal with these issues, we employ an iterative hill climbing algorithm, a special case of the Shotgun Stochastic Search (SSS) algorithm developed in Hans (2005) and applied to graphical models in Jones et al. (2005). We begin with an initial graph, $G_0$, and explore its decomposable one edge neighbors (all graphs that can be generated by the addition or subtraction of a single edge). It was proved in Frydenberg & Lauritzen (1989) that all decomposable graphs can be reached through the iterated addition or deletion of single edges starting from a decomposable graph. After exploring all these one edge neighbors, we choose the neighbor with the highest result in the metric we are hill climbing with respect to (e.g. probability or Sharpe ratio) and then explore its neighbors. Iterating this process ensures that we eventually reach a mode in the graph space. Choosing a new initial graph and restarting the process allows us to find a number of modes in the graph space.

As it is impossible to explore all the decomposable graphs in most cases, the choice of initial graphs can be of considerable importance as different initial graphs will lead to different modes. In order to choose initial graphs that would allow for a large coverage of the graph space we chose initial graphs with a variety of edges. Beginning with the empty (full) graph and tested its neighbors with one additional (fewer) edge
for decomposability then chose one at random. Iterating this process until reaching
the desired number of edges leads to a psuedo-random decomposable graph with the
desired number of edges.

Two important questions addressed in our empirical examples are: Does using a
graph with high probability lead to choosing weights which produce better returns
(for example by Sharpe ratio), and if a graph is considered good by some metric (e.g.
has high probability, leads to weights with high returns or Sharpe ratios) over some
time period, will it be more likely to exhibit the same properties over the future time
periods than graphs which did not exhibit those properties?

3.5 Target

The overarching problem addressed in this paper is to solve for weights, $w_t$, to mini-
mize the portfolio variance $w_t'K_t w_t$, subject to the constraints $w_{ti} \geq 0, \forall i \in 1, \ldots, p$
,$w_t'1 = 1$, and $w_t'f \geq m_t$, where $m_t$ is the target at time $t$. In most literature (e.g.
Carvalho & West (2007a) and Polson & Tew (2000)) $m_t = m, \forall t$, where $m$ is some
constant such as 2% per month. This, however, can easily lead to problems when we
introduce the restriction that all the weights are positive. Under this regime if the
target is larger than any of the expected returns then there is no way to assign weights
which would achieve the target. This is especially a problem in the dynamic regression
model where the predicted returns, $f_t$, vary considerably through time. Therefore, in
this paper I introduce a novel moving target by letting $m_t = \min(\overline{f} + \rho \sigma_f, \max(f))$,
which ensures that the target is always attainable. The investor is able to choose $\rho$
for himself, though values of $\rho$ between 0 and 2 seem most reasonable.
3.6 Transaction Costs

Another concern related to the dynamic regression model is that the rapidly changing predicted returns, and thus target, can lead to large churn in the portfolio. In real world applications this would create transaction costs that eliminate any excess returns. Therefore, in the dynamic regression model, in addition to computing weights without transaction costs, we also consider a model with them. Specifically, we add the constraint $\kappa 1' \Delta w_t \geq m_t$ with $\Delta w_t$ defined as above. The choice of $\kappa$ is critical to the reduction in churn. Realistic values for $\kappa$ can range from $.01\% - 2\%$ depending on the liquidity of the security and any other fees attached. The solution to this optimization problem has no analytical solution, thus for the purposes of our analysis we use the constrained nonlinear optimization function in MATLAB's Optimization Toolbox.

4 Case Study: Monthly Vanguard Mutual Funds

4.1 The Data

Recent research has suggested that firm-level volatility is increasing and correlations among firms has decreased. Specifically, this means that while it once took approximately 20 securities to be diversified (Bloomfield et al., 1977), to achieve the same level of diversification now requires 50 securities (Campbell et al., 2001). However, to the individual investor, researching and buying so many securities individually is not practical. One option for the investor to avoid assuming diversifiable risk is to invest in one or more mutual funds. Almost all retirement accounts today offer a variety of mutual funds for their investors. In order to evaluate the efficacy of the model, it is applied to a collection of 30 Vanguard mutual funds using monthly data from July
2000 through June 2007 ($n = 84$), available from the Center for Research in Security Prices (CRSP) database. The names of the funds are available in appendix B. A time series plot of the data can be seen in figure 2. As can be seen from the figure the returns of the funds are highly correlated and can be anywhere from +20% to −20% in a given month. For the purposes of analysis the data were divided into two periods: July 2000-June 2005 ($n = 60, \bar{r} = 3.5\%, \bar{\sigma} = 17.9\%$) and July 2005-June 2007 ($n = 24, \bar{r} = 18.4\%, \bar{\sigma} = 9.2\%$) to first fit the model and then to test it.

4.2 Local Smoothing Model

The model considered here is a special case of the model defined earlier. Specifically, we let $F'_t$ and $G_t$ be the identity matrices (we do not use any predictors), $V_t = 1$, and let $W_t$ be defined by the discount factor $\delta$ as described above.

$$
\text{Observation : } Y_t = \theta_t + \nu_t, \quad \nu_t \sim N(0, \Sigma_t)
$$

$$
\text{Evolution : } \theta_t = \theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, W_t \Sigma_t)
$$

This model essentially assumes no predictive power and simply hopes to better estimate the structure in the volatility of the returns. Therefore, the only fair way to compare the performance of graphs is to compare the variance of returns in the minimum variance portfolios induced by the implied volatility structure of the graph. However, for the sake of interest graphs which generated the highest return over the period were also considered.

4.3 Initializing Model

To initialize the model priors are needed for $\theta$ and $\Sigma$, which implies the necessary choice of $m_0, C_0, b_0$, and $S_0$ as well as a choice of discount factor $\delta$ and the probability
of edge inclusion, $\beta$. As there is so little data compared to the number of parameters to be estimated I choose informative priors. Specifically, $m_0$ is chosen as $.01\mathbf{1}$, where $\mathbf{1}$ is a vector of ones. This implies an average yearly return of roughly 12%, though I choose $C_0 = 4^2$ to show significant uncertainty about the prior. $S_0$ is chosen to imply a prior location on $\Sigma$ where the monthly standard deviation of the funds of 4% with a correlation of .5. Specifically I choose $b_0 = 1$ to allow for fast adaptation and $S_0 = (4\%)^2/2I_{30} + (4\%)^2/2J_{30}$, where $I_{30}$ is the $30 \times 30$ identity matrix and $J_{30}$ is the $30 \times 30$ matrix of ones. This prior is in line with Giudici & Green (1999). $\delta = .98$ is chosen to allow for some small changes in the volatility structure across time. Finally, I set $\beta = .05$ to induce sparsity in the graphical structure. This amounts to a penalty of roughly $-3$ in the log probability for each edge.

4.4 Results

As there are $2^{435}$ graphs, it is impossible to search over even a small percentage of this graph space. Therefore, to efficiently find interesting graphs, we employ the hill climbing method of section 3.4, choosing the one edge neighbor with the highest posterior probability. This provides similar results to searching for least volatile returns, and in a model which does not predict future returns is more valuable than searching for high return graphs. Following this method, over three million graphs were searched from initial graphs separated by 15 edges. The highest posterior probability graph is shown in figure 3.

To examine the efficacy of using graphical models to model the volatility structure we examine the relationship between sparsity, probability, and risk (as measured in standard deviations). Figure 4 demonstrates a clear relationship between high probability graphs and low risk. The relationship between sparsity and risk is less
clear (figure 5); however we can see that there exists a small range of edges with high probability, while graphs with fewer than 20 are incredibly unlikely to be the true model.

One of the questions I set out to answer is how the effectiveness of graphs change through time. As there is such a large parameter space of graphical models, it is important that graphs retain their properties over a long period of time because it would not be feasible to regularly search the graph space. This is addressed in this section by dividing the data into two subsets, as described above. We then examine whether the likelihood of the graphs and other key properties are correlated across time periods. Figure 6 demonstrates that the likelihood of the graphs are remarkably highly correlated through time. From figure 7 we can see that the prior has a stronger effect on period two as there is less data. As expected, in figure 8 we see that the variance of the returns, which is highly correlated with the probability of the graph is similarly related through time. Finally, even returns show positive trends through time as can be seen in figure 9. Therefore, we can conclude that at least in this example estimating the volatility structure using graphical models produces lasting results.

5 Case Study: Daily Vanguard Mutual Funds

5.1 The Data

In order to more fully evaluate the efficacy of our model, I abandon the local smoothing model and consider a prediction model using prior returns for both the funds as well as other predictor variables. Considerable academic research has shown that there is daily autocorrelation in mutual fund returns (Goriaev et al., 2004), (Busse, 2001). Therefore, to allow for some prediction and to better estimate the covariance
matrix, we move from monthly data to daily data using the same funds for the period July 2002 through December 2006, also available from the Center for Research in Security Prices (CRSP) database. The 15 separate predicting variables over the time period are available via DataStream. The names of the predictors are available in appendix C. Plots of some of the data can be seen in figure 10. For purposes of analysis we divided the data into three sets: July 2002-June 2004 ($n = 504, \bar{r} = 12.6\%, \bar{\sigma} = 20.5\%$), July 2004-December 2006 ($n = 631, \bar{r} = 15.4\%, \bar{\sigma} = 12.3\%$), January 2007-December 2007 ($n = 250, \bar{r} = 7.4\%, \bar{\sigma} = 17.2\%$) in order to first fit the model, search over the graph space, and then to test the graphs.

5.2 Prediction Model

The model considered here is another special case of the model defined earlier. Specifically, I let $G_t$ be the identity matrix (this assumes a random walk for state space changes), and let $W_t$ be defined by the discount factor $\delta$ as described above. Then returns are modeled as:

$$
\begin{align*}
\text{Observation:} & \quad Y_t' = F_t' \Theta_t + \nu_t, \quad \nu_t \sim N(0, \Sigma_t) \\
\text{Evolution:} & \quad \Theta_t = \Theta_{t-1} + \Omega_t, \quad \Omega_t \sim N(0, W_t, \Sigma_t)
\end{align*}
$$

(12)

I choose $\delta = .997$ for the analysis which corresponds to a monthly discount rate of about .94 and a yearly discount rate of roughly .5. Therefore, an observation today is worth about two observations from a year ago. Furthermore, $\beta = .05$, to encourage sparser models.

In order to predict returns at time $t$, the returns of the funds and predicting variables at time $t - 1$ and a constant are used. This means that in the notation above $Y_t'$ is a $1 \times 30$ vector of returns, $\Sigma_t$ is a $30 \times 30$ matrix of the unexplained variations, $\Theta_t$ is a $46 \times 30$ state-space matrix, and $F_t'$ is a $1 \times 46$ matrix of the
predictor variables.

The model requires initial parameters $m_0$, $C_0$, $S_0$, and $b_0$. Once again we use informative priors, however, as this dataset is much larger the choices are in general overwhelmed by the data. First we consider $S_0/(b_0 - 2)$, the expectation of $\Sigma_0$ when $b_0 > 2$. Specifically, if we consider that daily returns generally have a standard deviation of roughly 2% a natural choice to induce sparsity is to have $S_0/(b_0 - 2) = .02^2 I_{30}$. Furthermore, to display our vague prior beliefs we let $b_0 = 3$. The choice of $C_0$ can then be thought of in the context of $\theta_{ij} \sim T_{b_t}[m_{ij}, c_{ii} s_{jj}/(b_t - 2)]$. As we believe there to be no good a priori beliefs about the dynamic regression coefficient $m_{ij}$, we can express this by noting if $s_{jj}/(b_0 - 2) = (2/100)^2$, then letting $c_{ii} = 100^2$ implies a scale factor of 2 which allows for appropriate adjustments. This gives $C_0 = 100^2 I_{46}$.

Finally, we are left to choose $m_0$. As we can see in figure 11, the choice of initial parameters does not matter much after a few months given our large initial uncertainty. A reasonable choice is to let $m_0 = 0$, the $46 \times 30$ matrix of zeros in order to induce sparsity, while this has the minor drawback of implying $f_1$ equals the zero vector.

5.3 Choosing Predictors

An attempt to fit the prediction model following the updating equations in section 3.2 soon reveals a problem. As mentioned above, it is impossible to generate sparsity in the structure of $\Theta_t$ and maintain a closed form solution. Therefore, in fitting the model we only have 1386 data points to fit $46 \times 30 = 1380$ parameters in the state space matrix. Clearly we cannot get a good estimate for the state space with so many predictors. An example of how this effects the outcome of the model can be seen in figure 12. Using too many predictors suppresses the returns on the portfolio - having
fewer predictors allows for better estimates of their effects and thus higher portfolio returns with lower risk. The optimal number of predictors for this data set seems to be five or six.

Two methods for determining predictors were tried. First, using the relationship $\theta_{ij} \sim T_b[m_{ij}, c_{ii}s_{jj}/(b_t - 2)]$ one can follow Lindley’s method (c.f. section 4.3 of Lee (2004)) and construct 95% HPD regions and compare how often predictors are therefore significant in predicting the returns of the funds over the first two years of data. See for example figure 13. I do this for the full graph to get a sense of the general predictive power of the variables using $m_0$ as the $46 \times 30$ zero matrix in order to ensure the greatest sparsity. By considering which coefficients do not have HPD intervals surrounding zeros at the end of our data set, we are able to eliminate unnecessary predictors. As can be seen in figure 14, almost all of the predictors can be rejected by Lindley’s method. The two predictors which proved to be useful for many funds were the returns at $t - 1$ of the Mid Cap Index and the Euro to Dollar exchange rate.

Another method considered is to consider which predicting variables, when used as the sole predictor along with a constant (and $\rho = 2$), give the highest Sharpe ratios over the two years in the hopes that they will retain their predictive power in the future. The three highest funds in terms of predictive power, with Sharpe ratios of over 3.7 were Capital Opportunities, Growth Index, and Growth Equity. Interestingly, when considering the entire period from July 2002 through December 2006 these predictors were still all close to the top in terms of predictors that generated the best return.

Combining the predictors from both methods results in a much more manageable seven predictors to estimate instead of 45. The results can clearly be seen in terms of portfolio return in figure 15. One important question for anyone interested in
applying this model may be which method gave predictors that worked best with future Sharpe ratios. The answer, illustrated in figure 16, seems to be that those that are statistically significant resulted in better returns, which is surprising given how the other predictors are chosen. This may imply that choosing them one at a time is not a good indication for how well they will do as a group. For the purposes of our analysis, our combination of 5 predictors did slightly better than either of these so a combination may be best.

5.4 Graph Search

Once the predictors are chosen and the model is initialized, the final aspect of the model is the graph specifying the conditional dependence structure of the unpredictable returns. Here there are two important questions to answer: Does using a graph with high probability lead to choosing weights which produce better returns (by Sharpe ratio), and if a graph has a high probability (or leads to weights with high returns or Sharpe ratios) over some time period, will it be more likely to exhibit the same properties over the future time periods?

We search for graphs using the hill climbing method described above. The top probability graph is found in figure 17 and the top return graph found is in figure 18. The cumulative returns of the portfolio generated from the prediction model specified above is found in figure 19. This plot is interesting as the top probability graph does not deviate much from the complete graph. One thing to note is that the full graph actually has a higher likelihood than the top probability graph and the top probability graph has almost half the edges (207).

Figures 20 and 21 show the relationship between probability, risk, sparsity, and return in our graphical models. One consistent observation across datasets is that
higher probability graphs display lower risk. Also, probability increases sharply as edges are added as we would expect with the highly correlated nature of the data. Finally, the highest return graphs are lower in probability and sparser than the highest probability graphs.

5.5 Graphs Through Time

The results here are similar to the results from the local smoothing model. Specifically, we first examine whether graphs which are high probability over one period are high probability over the next period. In figure 22 we find that graphs which are high in probability from July 2004-December 2006 continue to be high probability graphs during January 2007-December 2007, which is interesting as there is a marked increase in volatility over this time period as seen in the original data plots. However, unlike the previous case study here there is no marked difference between the likelihood and the probability. There is enough data here to override the effect of our prior on graphs in the analysis.

Also, unlike the previous study, here we find no correlation between return in the first period and the second period. This is seen in figure 23. This however changes when we remove the restriction that all the weights be positive. In figure 23 this correlation through time can clearly be seen. The effect here is that an individual investor should choose a graph with high probability which will minimize his or her volatility since he cannot hope to improve his expected return through the choice of graph, whereas an institutional investor can consider searching for graphs for ones which generate high returns as there is some hope that the effect will carry over into the next period.
5.6 Transaction Costs

One striking difference between this prediction model and earlier local smoothing models is how variable the weights are. It is clear from looking at figure 25 that without some constraints in the changing weights, the transaction costs will eat up any returns generated the model. To control for this, in figure 26 we solve for the weights over the period July 2002 to December 2006 on a sample graph following the transaction cost optimization presented in section 3.6. We let $\kappa = .01$ to represent a 1% charge for every trade which clearly results in a stabilizing effect. After a burn in period of 6 months the expected returns, $f$, stabilize near zero and the weights stabilize with the exception of a few large shocks.

6 Managed vs. Index Funds

6.1 Overview

A central problem in the choice of investors is which fund to choose. Investors are offered a wide array of choices including the choice between passively managed index funds which try to track a general market and actively managed funds which attempt to beat a given index while charging a larger fee for their effort. The choice between these two types of funds is a central problem for investors and is considered extensively in Rodriguez & Tower (2007). Here we consider this question from the perspective of graphical models. Our data consists of 12 index funds and 18 managed funds.

6.2 Theory of Marginalizing

An interesting application of marginalizing graphs (example in figure 27) arises from the consideration of the relationships between a single managed fund and the index
funds. For example, if an expert really think that a single managed fund should be conditionally dependent on only three or four index funds, what does this mean for the structure of the 30 node graph? We can attempt to get at this through considerations of our prior on graph $G_0$ which says that every edge is in or out with probability $\beta_0$. From this, we can then say that after the marginalization of one node that the remaining edges are in or out of the new graph $G_1$ with some probability $\beta_1$ which should depend on $\beta_0$.

The solution is tractable at least for low dimensions. For example, with just one marginalization $\beta_1$ can be found as follows. Consider nodes $i$ and $j$ in the original graph. $Pr(i \sim j|G_0) = \beta_0$, which is read “The probability that $i$ and $j$ are neighbors in graph $G_0$ is $\beta_0$.” Now assume another node, node $k$, is marginalized to generate graph $G_1$. Then $Pr(i \sim j|G_1) = \beta_1 = Pr(i \sim j|G_0) + (1 - Pr(i \sim j|G_0))(Pr(i \sim k; j \sim k|G_0)) = \beta_0 + (1 - \beta_0)(\beta_0^2)$.

Solutions to $\beta$’s of larger $k$ can be found in appendix D, along with a method for computing the values. Solutions involve enumerating large numbers of possible graphical structures. In general, decent approximations can be arrived at using $\beta_k = \beta_{k-1} + (1 - \beta_{k-1})(2\beta_{k-1}\beta_0 - \beta_0^2)$, or by large simulation. However, with simulation one has to be careful as it is very difficult to simulate decomposable graphs of dimension greater than fifteen, and empirical evidence suggests that decomposable graphs marginalize differently than non-decomposable graphs (see for example figure 28). Returning to our example, if we believed that each managed fund should depend on about 4 index funds, this would imply that $\beta_{17} = \frac{4}{12}$ and, using our approximation, this would imply a $\beta_0$ of about 0.075, alluding to the idea sparsity is undone in marginalization.
6.3 Implied Regressions

In Rodriguez & Tower (2007), managed funds are compared to a basket of tracking index funds. Specifically, each managed fund is regressed on all of the index funds while restricting the regression coefficients on the index funds to be positive and sum to one. This results in a basket of index funds for each managed fund which best tracks the return of the managed fund. Rodriguez & Tower (2007) find that each managed fund is tracked by $2 - 7$ index funds ($1 - 3$ of which are usually significant), which allows them to rate the performance of managed funds relative to their basket of index funds.

As shown in Stevens (1998), the inverse of the covariance matrix, $V^{-1} = K$ for the $p$ vector of returns determines a linear regression relating these terms to each other. Specifically, take the first row of $K$, $k_1 = (k_{11}, \ldots, k_{1p})$. Then we have that the regression coefficient for the return of asset $i > 1$ on asset 1 is, $\beta_{1i} = -k_{1i}/k_{11}$ and that $k_{11} = \frac{1}{\sigma_1^2(1 - R_1^2)}$, where $R_1^2$ is the proportion of the variability in the returns of fund 1 explained by the other $p - 1$ funds. In matrix form:

\[
V^{-1} = K = \begin{pmatrix}
\frac{1}{\sigma_1^2(1 - R_1^2)} & \frac{\beta_{12}}{\sigma_2^2(1 - R_2^2)} & \cdots & \frac{\beta_{1p}}{\sigma_p^2(1 - R_p^2)} \\
\frac{\beta_{21}}{\sigma_1^2(1 - R_1^2)} & \frac{1}{\sigma_2^2(1 - R_2^2)} & \cdots & \frac{\beta_{2p}}{\sigma_p^2(1 - R_p^2)} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\beta_{p1}}{\sigma_1^2(1 - R_1^2)} & \frac{\beta_{p2}}{\sigma_2^2(1 - R_2^2)} & \cdots & \frac{1}{\sigma_p^2(1 - R_p^2)}
\end{pmatrix}
\]

Therefore, we can use our estimated covariance matrix, $\hat{\Sigma}_t = Q_t S_t / (b_t - 2)$ (with $\Sigma$ constrained by a graph $G$), to incorporate a similar analysis to Rodriguez & Tower (2007). An important property that makes this analysis efficient is that if we choose a subset of funds $A \subset \{1, \ldots, p\}$, then the submatrix $\hat{\Sigma}_A$ is consistent with the graph $G_A$ formed by marginalizing the nodes $A^C$ from graph $G$. 

25
In this section we employ this analysis to the top probability graph from section 4.4. The results of the regression are found in table 8. Figure 29 and figure 30 plot smoothed \( R^2 \) and regression coefficients respectively through time. These plots imply a constancy in the weights and explanatory power though time, implying little change in overall strategy in the managed funds. In Rodriguez & Tower (2007) analysis of \( R^2 \) to excess returns provides a clear relationship between excess return and the explanatory power of the index funds. However, in our analysis the trend is small at best (see figure 31) and without more index funds to compare with managed funds it is hard to interpret the reason for such a relationship. Another difference includes the weight on the total stock index funds, which in Rodriguez & Tower (2007) is almost never included in the regressions, whereas in our analysis is the largest component.

One surprising observation is that there is no clear relationship between \( R^2 \) and standard deviation in returns for the managed portfolios shown in figure 32, perhaps implying that none of these index funds adequately represents the entire market. In general our results provide are consistent Vanguard’s descriptions of the funds and economic intuition. One object of interest is that the only growth indices are small cap, explaining the value placed on the Total Stock Index as a large cap blend from the large cap growth funds.

Finally, we note that as there is large uncertainty about which graphical model is the true model, one computationally intensive option is to simulate from the posterior distribution (Carvalho et al., 2007) to get a more robust estimate for the implied regression.
7.1 Portfolios Uncorrelated in Expectation with \( j \)th Fund

One area of interest for institutional investors in particular is having the returns of their portfolio be uncorrelated in expectation with a certain benchmark, for example the S&P 500. This example results in portfolio which is “market neutral.”

Consider the portfolio returns on a portfolio, \( R_P = w_1R_1 + w_2R_2 + \cdots + w_pR_p \). Then the variance of the portfolio is \( \text{var}(R_P) = \text{cov}(R_P, R_P) = \text{cov}(R_P, w_1R_1 + w_2R_2 + \cdots + w_pR_p) = w_1\text{cov}(R_P, R_1) + w_2\text{cov}(R_P, R_2) + \cdots + w_p\text{cov}(R_P, R_p) \). Note that this is the weighted sum of the covariances of the portfolio with the returns of the individual securities. Further, the covariances can be expressed in matrix form as \( Kw \). Much interest lies in the minimum variance portfolio, where \( w \propto K^{-1}1 \). This implies that the covariances of the return of the portfolio with respect to the individual securities is \( Kw \propto 1 \), therefore the covariance is the same for all securities.

In the current problem we want to set one of these covariances equal to zero. If the benchmark is one of the securities the investor is allowed to invest in, this problem becomes minimizing \( w'Kw \), subject to the constraints \( w'1 = 1 \), and \( w'k_j = 0 \), where \( j \) is the benchmark and the covariance matrix \( K = [k_1, \ldots, k_p] \). Another possible constraint is \( w'f \geq m_t \).

Solving this optimization problem using Lagrange Multipliers results in the following solution where the vector of weights \( w \) satisfies:

\[
\begin{align*}
w &= K^{-1}(\lambda_11 + \lambda_2k + \lambda_3f) \\
\end{align*}
\]

And \( \lambda_1, \lambda_2, \) and \( \lambda_3 \) are given by:
\[
\begin{pmatrix}
1'K^{-1} & 1'K^{-1}k_j & 1'K^{-1}f \\
1'K^{-1}k_j & k_j'K^{-1}k_j & k_j'K^{-1}f \\
1'K^{-1}f & k_j'K^{-1}f & f'K^{-1}f
\end{pmatrix}
\begin{pmatrix}
\lambda_1 \\
\lambda_2 \\
\lambda_3
\end{pmatrix}
= 
\begin{pmatrix}
1 \\
0 \\
m
\end{pmatrix}
\]  
(14)

Code to solve for these weights was developed in the MATLAB® environment. Implementing this code over a collection of five securities provides interesting new results, seen in figure 33. Specifically, we see a trend whereby graphs with higher probabilities produce a correlation coefficient closer to zero. This provides another metric for testing the efficacy of one step ahead estimates for the covariance matrix, again showing that high probability graphs produce better estimates for the covariance matrix.

Plots of the weights are shown in figure 34 and the leverage ratio in figure 35. It is clear that these portfolios would not be practical for implementation in practice with their high leverage ratio and large variability in weights. However, future investigation of the leverage ratio required as the number of securities increases may be of interest.

7.2 Simulating Decomposable Graphical Models from our Prior

The prior used over graphs is that each edge is independent with Bernoulli edge inclusion probability \( \beta \). One area of interest is being able to simulate decomposable graphs uniformly from the prior, which could be used in prior inference. Simulation was considered from the perspective of the junction tree. In order to simulate from this we need to determine several things: the distribution of the number of cliques implied by the specific \( \beta \), the distribution on the size of each clique implied by the \( \beta \) and number of cliques, and the implied distribution on the size of the separators
implied by the above. If we can determine these distributions, we can generate a
junction tree randomly and then uniformly assign our variables to these nodes.

Another possible method would be to generate a random graph by considering
each edge one at a time and assigning it with probability $\beta$ and then find its nearest
decomposable graph using minimum triangulation (Berry et al., 2004). This method,
however, has not been attempted.

7.3 Counting Decomposable Graphs

As we saw in the above section, simulation of decomposable graphs relies on an
understanding of the underlying distributions. Therefore, of interest is the number of
decomposable graphs with vertex set $p$, as well as the breakdown by number of edges
and cliques. The only effort in literature to calculate the number of decomposable
graphs is Wormald (1985), where a generating function is derived which allows for
the determination of the number of $n$-vertex labelled chordal graphs with a given
connectivity. I considered counting from the junction tree perspective. Specifically,
the goal was to enumerate all possible junction trees and then can determine the
possible permutations of graphs belonging to each junction tree. However, there
proved too many difficulties to overcome.

8 Discussion

In Carvalho & West (2007a) a new, rich class of matrix DLMs that incorporate
conditional independence structure in the cross-sectional precision matrix of a set of
time series. This model is especially useful in portfolio theory where a central problem
is the estimation of the volatility matrix. Reducing the parameter space allows for
better estimation and reduction of noise. As the examples in this paper show, this
model improves upon the standard DLM model and allows for improved portfolio performance even for individual investors. This paper further advances the previous work by hill climbing in graphs in metrics other than probability, including return and Sharpe Ratios. Analysis of other datasets, including a set of 19 stocks in the Dow Jones Industrial Average over a 10 year period and an analysis of 30 Fidelity mutual funds provide similar results to the dataset used in this paper.

Future advances include new priors for the graph, implementing graphs in factor models, and comparisons with other models. Other fruitful areas of interest include the simulation of decomposable graphs and theoretical understanding of the marginalization of graphs. Code and data used in this paper are available upon request to the author.
References


Consider a vector $\mathbf{x} = (x_1, \ldots, x_p)'$ where $\mathbf{x} \sim N(\mathbf{0}, \Sigma)$. Then let $G = (V, E)$ be the undirected graph defining the conditional independence relationships of $\mathbf{x}$. If $(i, j) \in E$, then we say that $j$ is a neighbor of $i$, or conversely, $i$ is a neighbor of $j$. We represent this as $i \sim j$. Consider a subset $V_A \subseteq V = \{1, \ldots, p\}$. Then the graph $G_A = (V_A, E_A)$, where $E_A$ consists of all $(i, j) \in E$ such that $i, j \in V_A$, is called a subgraph of $G$. Furthermore, let $G_A, G_B,$ and $G_S$ be three subgraphs of $G$ such that $V_A \cup V_B \cup V_S = V$. Then we say that $G_A, G_B,$ and $G_S$ decompose $G$.

Moreover, if for any $a \in V_A$ and $b \in V_B$, $a \sim b \Rightarrow a$ or $b \in V_S$ then we say that $G_S$ separates $G_A$ and $G_B$. If $\forall i, j \in V_S i \sim j$, we say that $G_S$ is complete. A subgraph $G_S$ which is complete and separates two other subgraphs is called a separator. If the separator is chosen to be minimal, then a proper (the separated subgraphs are nonempty) decomposition exists which results in the prime components of the graph, a collection of subgraphs which cannot be further decomposed.

All graphs can be represented as a tree of its prime components, a junction tree, an example of which can be seen in figure 1. If all the prime components of a graph, $G$, do not contain any cycles of length 4 or more, then there exists a perfect ordering $P_1; S_2, P_2; S_3, P_3; \ldots$ of the prime components and separators such that $S_n = P_n \cap (P_1 \cup \cdots \cup P_{n-1})$. Note that the perfect ordering need not be unique. If all the prime components are complete then we say that the graph $G$ is decomposable.
### Vanguard Mutual Funds

#### Domestic Funds

<table>
<thead>
<tr>
<th>Style</th>
<th>Blend</th>
<th>Growth</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>Blend</td>
<td>Growth</td>
<td>Size</td>
</tr>
<tr>
<td>Equity Income</td>
<td>500 Index</td>
<td>Growth Equity</td>
<td>Large</td>
</tr>
<tr>
<td>U.S. Value</td>
<td>Dividend Growth</td>
<td>Morgan Growth</td>
<td></td>
</tr>
<tr>
<td>Value Index</td>
<td>Growth and Income</td>
<td>PRIMECAP</td>
<td></td>
</tr>
<tr>
<td>Windsor</td>
<td>Total Stock Index</td>
<td>US Growth</td>
<td></td>
</tr>
<tr>
<td>Windsor II</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selected Value</td>
<td>Extended Market Index</td>
<td>Capital Opportunities</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>Mid Cap Index</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Strategic Equity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small Cap Value Index</td>
<td>Small Cap Index</td>
<td>Explorer</td>
<td>Small</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Growth Index</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Small Cap Growth Index</td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>North America</td>
<td>Europe</td>
<td>Pacific</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------------------</td>
<td>-----------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td></td>
<td>Global Equity (42%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>International Value (59%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>European Stock Index (100%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Global Equity (31%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>International Value (18%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pacific Stock Index (100%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## C Predictors

<table>
<thead>
<tr>
<th>Futures</th>
<th>Exchange Rates</th>
<th>Options</th>
<th>Foreign Indices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corn</td>
<td>Euro to U.S. Dollar</td>
<td>Implied S&amp;P Volatility</td>
<td>BOVESPA</td>
</tr>
<tr>
<td>Crude</td>
<td>U.S. Dollar Index</td>
<td></td>
<td>BSE 100 (India)</td>
</tr>
<tr>
<td>Gold</td>
<td></td>
<td></td>
<td>DAX 30 Performance</td>
</tr>
<tr>
<td>LIBOR</td>
<td></td>
<td></td>
<td>FTSE 100</td>
</tr>
<tr>
<td>NIKKEI</td>
<td></td>
<td></td>
<td>Shanghai SE A Share</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Futures were continuous settlement price
Label the graphs $G_0 \ldots G_p$ where the subscript defines how many nodes have been marginalized. Let $|G|$ represent the number of edges in graph $G$ and let $N_k^G$ represent the expected number of neighbors per node in graph $G$ where $k$ denotes the number of nodes marginalized. Then the formulas for $\beta_k$ can be computed by exhaustion using the following steps:

1. All possible graphs on $p$ nodes are generated. As there are $2^{p(p-1)/2}$ possible graphs this step is computationally intensive and can only be done practically for $p \leq 6$.

2. Marginalize each graph one node at a time from node 1 to node $p$, keeping track of the number of edges in the graph at each time point. This results in a $p+1 \times 2^{p(p-1)/2}$ matrix where the $m$th row is the number of edges after marginalizing $m-1$ nodes and the columns are for each possible initial graph. Note that the last row is unnecessary as it will always be 0 as there will be no nodes left.

3. From this we can compute the average number of neighbors per node for each graph. The equation for this is $\frac{2|G|}{(p-k)}$.

4. We aggregate these average values of neighbors per node over graphs with the same number of starting edges. Specifically, in the matrix considered above, we find all columns with the same number in the first row and average these columns together to form a $p+1 \times \binom{p}{2}$ matrix where the row represents the number of nodes marginalized from 0 \ldots $p$ and the column represents the number of edges in $G_0$. This gives $E[N_k||G_0| = e] = \frac{1}{\binom{p}{2}} \sum_{|G_0|=e} N_k^{G_0}$.

5. As $|G_0| \sim Bin(\binom{p}{2}, \beta_0)$, for specific values of $\beta_0$ we can use the binomial pdf to determine the expected number of neighbors after $k$ marginalizations. Specifi-
cally, $E[N_k|\beta_0] = \sum_{i=1}^{(\frac{p}{2})} \Pr(|G| = i)E[N_k||G_0| = i]$. This is easy to calculate given $\beta_0$ and $E[N_k||G_0| = i]$.

6. Given $E[N_k|\beta_0]$, we can calculate $\beta_k$ as the $\beta_k$ which results in that expected number of neighbors. Specifically, $\beta_k = E[N_k|\beta_0]/(p - k - 1)$.

7. Finally, once we calculate $\beta_k$ over a series of $\beta_0$’s, we can determine the formula using simple linear regression. Specifically, we solve for the $a$’s in $\beta_k = a_1\beta_0 + a_2\beta_0^2 + \cdots + a_{(\frac{k+2}{2})}\beta_0^{(\frac{k+2}{2})}$.

This gives the following formulas for $\beta_k$:

$\beta_0 = \beta_0$

$\beta_1 = \beta_0 + (1 - \beta_0)(\beta_0^2)$

$\beta_2 = \beta_0 + (1 - \beta_0)(\beta_0^2)(-2 - 2\beta_0 + 5\beta_0^2 - 2\beta_0^3)$

$\beta_3 = \beta_0 + (1 - \beta_0)(\beta_0^2)(-3 - 6\beta_0 + 9\beta_0^2 + 36\beta_0^3 - 91\beta_0^4 + 84\beta_0^5 - 36\beta_0^6 + 6\beta_0^7)$

(15)
<table>
<thead>
<tr>
<th>Managed Funds</th>
<th>Vanguard 500</th>
<th>Emerging Market</th>
<th>European Stock</th>
<th>Extended Market</th>
<th>Growth</th>
<th>Mid Cap</th>
<th>Pacific Stock</th>
<th>Small Cap Growth</th>
<th>Small Cap</th>
<th>Small Cap Value</th>
<th>Total Stock</th>
<th>Value</th>
<th>R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cap Opp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.29</td>
<td>0.81</td>
</tr>
<tr>
<td>Dividend Growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.77</td>
<td>0.71</td>
</tr>
<tr>
<td>Equity-Income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.77</td>
<td>0.88</td>
</tr>
<tr>
<td>Explorer</td>
<td>0.02</td>
<td>−0.02</td>
<td>0.22</td>
<td>0.18</td>
<td>0.00</td>
<td>0.29</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>Global Equity</td>
<td>0.14</td>
<td>0.30</td>
<td>−0.14</td>
<td>0.34</td>
<td>0.10</td>
<td>−0.10</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>Growth &amp; Income</td>
<td>0.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.44</td>
<td>0.97</td>
</tr>
<tr>
<td>Growth Equity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.82</td>
<td>0.87</td>
</tr>
<tr>
<td>International Explorer</td>
<td>0.19</td>
<td>0.37</td>
<td>0.18</td>
<td>0.04</td>
<td>0.23</td>
<td>−0.17</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>International Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Mid Cap Growth</td>
<td>0.21</td>
<td>0.40</td>
<td>−0.08</td>
<td>0.22</td>
<td>0.13</td>
<td>−0.08</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.77</td>
<td>0.95</td>
</tr>
<tr>
<td>Morgan Growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.12</td>
<td></td>
</tr>
<tr>
<td>PRIMECAP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.12</td>
<td>0.87</td>
</tr>
<tr>
<td>Selected Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.80</td>
<td>0.66</td>
</tr>
<tr>
<td>Strategic Equity</td>
<td>0.04</td>
<td>0.04</td>
<td>−0.19</td>
<td>0.56</td>
<td>0.01</td>
<td>0.20</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>US Growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>US Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.87</td>
<td>0.83</td>
</tr>
<tr>
<td>Windsor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.99</td>
<td>0.88</td>
</tr>
<tr>
<td>Windsor II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.82</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Table 1: Results of the implied regression using the top probability graph on monthly data through December 2007.
Figure 1: Example of a junction tree. The graph on the top right is decomposed into its cliques and separators.

$C_1 = \{1, 2, 5\}$

$S_2 = \{2, 5\}$

$C_2 = \{2, 3, 4, 5\}$
Figure 2: Time series plots of the monthly data.
Figure 3: Highest posterior probability graph found. Funds resorted to maximize concentration along diagonal. Has 37 edges. Logically places growth funds together.
Figure 4: Created from a random subset of 3000 graphs explored. High probability graphs are generally sparse and produce less volatile returns.
Figure 5: Top graph demonstrates that between 10 and 20% sparsity seems to maximize probability. Lower graph shows general trend towards higher probability graphs producing less volatile returns.

Figure 6: Likelihoods from the first 60 months predict likelihoods for the final 24 months very well.
Figure 7: Although the likelihoods are very similar, the posterior probabilities can be construed as quite different because the prior distribution has more effect with fewer data points.

Figure 8: Low volatility in the first period predicts low volatility in the second period. The difference in scale is due to a lower overall volatility in the second period. Specifically, $\bar{\sigma}_1 = 17.9\%$ $\bar{\sigma}_2 = 9.2\%$
Figure 9: Slight evidence that higher returns in the first period predict higher returns in the second period. Note the banding in the returns is due to the positive weight restrictions. Many graphs produce the same or very similar weights because of this restriction.

Figure 10: Time series plots of the daily data.
Figure 11: Example of a regression coefficient with different initial values converging through time. Note that after a few months the prior value has a minimal effect.

Figure 12: This figure demonstrates that too many predictors can hold back the performance of the portfolio. Five or six predictors seem optimal. The graph is generated by randomly sampling \( k \) predictors 200 times and then calculating the Sharpe ratio of the portfolio generated by these predictors on the full graph. Initial values include \( \rho = 2 \) and \( m_0 \) is a \( k \times 30 \) matrix where \( m_{ij} \sim N(0, .001^2) \).
Figure 13: Examples of regression coefficients through time. $MorganGrowth_{t-1}$ on $Explorer_t$. Bands represent 95% confidence intervals.
Figure 14: Spy of matrix showing whether Coefficient passes Lindley’s method
Figure 15: This figure demonstrates how fewer predictors can create greater returns. Vertical line represents the point up to which the returns were used to choose the predictors. Weights generated with $\rho = 2$, without restricted to positive weights, without transaction costs, and using the full graph.

Figure 16: This figure demonstrates that the predictors chosen through Lindley’s method perform better in terms of return than those chosen by Sharpe ratios. Weights generated with $\rho = 2$ without being restricted to positive weights, without transaction costs, and using the full graph.
Figure 17: Highest probability graph found so far hill climbing by probability with initial values and predictors as specified in this paper. This graph is much busier than the graph found from the monthly data. This is probably due to the large number of data points overruling our prior. 207 edges.
Figure 18: Highest return graph found so far hill climbing by probability with initial values and predictors as specified in this paper. 84 edges.
Figure 19: Demonstrates that the predictors chosen through Lindley’s method perform better in terms of return than those chosen by Sharpe ratios. Weights generated with $\rho = 2$ without being restricted to positive weights, without transaction costs, and using the full graph.
Figure 20: This figure shows the relationship between log probability, risk (measured in standard deviations of returns), sparsity, and return in the daily data. Note that in all figures return and risk are given in annualized figures.
Figure 21: This figure shows the relationship between log probability, risk (measured in standard deviations of returns), sparsity, and return in the daily data.
Figure 22: This figure shows that probability and likelihood are highly correlated through time. High probability and likelihood graphs remain high over the immediate future. The major difference between the daily and monthly results is that here the prior is dominated by the data.

Figure 23: This figure shows that the returns of a portfolio generated by a graph over one time period does not do a good job of predicting the returns over the next period. This is the case where weights are restricted to be positive.
Figure 24: This figure shows that when the positive weight restriction is relaxed there is a reasonable correlation between past returns and future returns for portfolios generated by graphs.

Figure 25: Weights of full prediction model over time using the full graph when restricted to be positive. Note that the weights are very volatile compared those that result from a local smoothing model.
Figure 26: Weights of full prediction model over time using the full graph when restricted to be positive and subjected to transaction costs where $\kappa = .01$. Note that the weights are sticky to where they initially start, which implies it will be good to initialize the model.
Figure 27: Example of the marginalization of a graph. In this example node 5 is marginalized. All of its neighbors are connected in the marginalized graph.
Figure 28: Example results from uniformly simulating graphs with $p = 10$ and marginalizing 4 nodes from the graphs. The horizontal axis is the number of edges in the initial graph and the vertical axis shows the average number of neighbors for each node after the marginalization.

Figure 29: Smoothed $R^2$ through time for all 18 managed funds regressed on the 12 index funds. While most of the funds have similar $R^2$’s, two clearly lag: Dividend Growth and Selected Value. Selected Value, as the only one in its style box is unsurprising, while Dividend Growth is less expected.
Figure 30: Regression coefficients for Selected Value through time. After the first 24 data points the weights settle down.
Figure 31: Annualized return of the managed funds graphed against the implied $R^2$. There appears to be only a slight relationship between return and $R^2$, a relationship which is hard to interpret without more index funds.

Figure 32: Annualized standard deviation of the managed funds graphed against the implied $R^2$. There does not appear to be any clear relationship.
Figure 33: This figure demonstrates that higher probability graphs on average produce portfolios which have lower sample correlations with the benchmark. This is consistent with higher probability graphs forming a better estimate of the one step ahead covariance matrix.
Figure 34: In order to achieve an expected covariance of zero with such correlated funds the weights are very large.

Figure 35: The large weights necessary to achieve an expected covariance of zero lead to large leverage ratios. The leverage ratio is computed as the sum of the absolute values of the weights. Large leverage ratios lead to high risk and high costs for borrowing the funds necessary.