

Copyright © 2003 by Merrill Windous Liechty
All rights reserved

COVARIANCE MATRICES AND SKEWNESS:
MODELING AND APPLICATIONS IN FINANCE

by

Merrill Windous Liechty

Institute of Statistics and Decision Sciences
Duke University

Date: _____

Approved:

Dr. Peter Müller, Supervisor

Dr. James O. Berger

Dr. Michael L. Lavine

Dr. Campbell R. Harvey

Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Institute of Statistics and Decision Sciences
in the Graduate School of
Duke University

2003

ABSTRACT

(Bayesian correlation estimation, shadow prior, portfolio selection
with higher moments)

COVARIANCE MATRICES AND SKEWNESS: MODELING AND APPLICATIONS IN FINANCE

by

Merrill Windous Liechty

Institute of Statistics and Decision Sciences
Duke University

Date: _____

Approved:

Dr. Peter Müller, Supervisor

Dr. James O. Berger

Dr. Michael L. Lavine

Dr. Campbell R. Harvey

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the
Institute of Statistics and Decision Sciences in the Graduate School of
Duke University

2003

Abstract

This Ph.D. dissertation is concerned with using model based and computation intensive statistical approaches to gain insight into substantive issues in finance related topics. It addresses the use of posterior (predictive) simulation for Bayesian inference in high dimensional and analytically intractable models. It consists of three studies.

One study focuses on the issue of covariance estimation. I propose prior probability models for variance-covariance matrices. The proposed models address two important issues. First, the models allow a researcher to represent substantive prior information about the strength of correlations among a set of variables. Second, even in the absence of such substantive prior information, it provides increased flexibility. This is achieved by including a clustering mechanism in the prior probability model. Clustering is with respect to variables or with respect to pairs of variables. This leads to shrinkage towards a mixture structure implied by the clustering, instead of towards a diagonal structure as is commonly done. With departure from the standard MLE or inverse-Wishart prior there are many computational difficulties that arise in modeling covariances.

A second study follows from the computational challenges that arise in the covariance estimation study. Because of the requirement that the correlation matrix be positive (semi-) definite, the high dimensional space of the correlation matrix is analytically intractable. Calculating the normalizing constant is then left to numerical techniques. I examine a method for estimating the constant, and also an argument for ignoring it in certain cases. An additional strategy called the “shadow prior” is explored by introducing an additional level of hierarchy into the model.

A third study follows the foundational work of Markowitz’s portfolio selection in

1952. Issues that he pointed out but didn't address include parameter uncertainty, utility optimization, and inclusion of higher moments. I build upon his foundation by incorporating a skew-normal model. Using Bayesian methods, higher moments and parameter uncertainty can be accommodated in a natural way in the portfolio selection process. Preference over portfolios is framed in terms of utility function maximization, which are optimized using Bayesian methods. Findings suggest that skewness is important for distinguishing between 'good' and 'bad' variance.

Contents

Abstract	iv
List of Tables	ix
List of Figures	xv
Acknowledgements	xviii
1 Bayesian Methods in Finance	1
2 Bayesian Correlation Estimation	4
2.1 Introduction	4
2.2 Models	8
2.2.1 Motivating example	9
2.2.2 Common correlation	10
2.2.3 Grouped correlation	12
2.2.4 Grouped variables	13
2.3 Implementation and posterior simulation	14
2.3.1 Sampling the full conditional of r_{ij}	15
2.3.2 Sampling the full conditionals of μ and σ^2	15
2.3.3 Sampling the full conditional of ϑ	18
2.4 Extensions and applications	19
2.4.1 Random effects distributions in hierarchical models	19
2.4.2 ARCH models and time-varying correlations	20
2.4.3 Probit models	22
2.5 Examples	23

2.5.1	Simulated data	23
2.5.2	Structure in the stock market	25
2.5.3	Population genetics	26
2.6	Conclusion	28
3	Shadow Prior	31
3.1	Introduction	31
3.2	Example: Correlation matrices	34
3.3	Simulation study	35
3.4	Conclusion	38
4	Portfolio Selection with Higher Moments	39
4.1	Introduction	39
4.2	Higher moments and Bayesian probability models	42
4.2.1	Economic importance	44
4.2.2	Probability models for higher moments	47
4.2.3	Model choice	50
4.3	Optimization	50
4.3.1	Simplifications made in practice	51
4.3.2	Bayesian optimization methods	57
4.4	Optimal portfolios in practice	59
4.4.1	Data description	59
4.4.2	Model choice and select summary statistics	60
4.4.3	Expected utility for competing methods	62
4.5	Conclusion	65
5	Conclusions	74

Appendixes	77
A Skew Normal Probability Model	77
A.1 Density and moment generating function	77
A.2 First three moments of linear combination	78
A.3 Model specification	79
A.3.1 Likelihood and priors	79
A.3.2 Full conditionals	79
A.4 Estimation using the slice sampler	80
B Symbols	82
References	89
Biography	102

List of Tables

2.1 Empirical correlation matrix for monthly stock returns for nine equity securities from April, 1996 to May, 2002. The list includes energy companies: British Petroleum, Chevron, Exxon and Reliant; And financial services companies: Merrill Lynch, Bank of America, Citi-Bank, and Lehman Brothers. Enron could potentially be in either group based on different criterion. 10

2.2 Simulation study with (8×8) correlation matrix consisting of two groups of correlations and 500 observations. The $(8 \times 7)/2 = 28$ correlations are evenly divided between two groups that have parameters values $\mu_1 = -0.54$, $\sigma_1^2 = 0.05$, $\mu_2 = 0.37$, and $\sigma_2^2 = 0.08$. Let ϑ_0 denote the true simulation model. Let $\hat{\vartheta}$ denote the posterior mean estimates. The matrix on the left shows the group that each of the correlations is in (ϑ_0), and the matrix on the right shows the posterior probability that the correlations are from group 1 ($\hat{\vartheta}$). 24

2.3 Result of reversible jump Markov chain Monte Carlo exercise to determine which model best fits monthly stock returns for nine equity securities from April, 1996 to May, 2002. The list includes energy companies: British Petroleum, Chevron, Exxon and Reliant; And financial services companies: Merrill Lynch, Bank of America, Citi-Bank, and Lehman Brothers. Enron could potentially be in either group based on different criterion. The reversible jump Markov chain Monte Carlo results in 50% posterior probability on the grouped variables model with two groups. 26

2.4 Posterior estimates of μ and σ^2 in grouped variables model with two groups for monthly stock returns for nine equity securities from April, 1996 to May, 2002. The list includes energy companies: British Petroleum, Chevron, Exxon and Reliant; And financial services companies: Merrill Lynch, Bank of America, Citi-Bank, and Lehman Brothers. Enron could potentially be in either group based on different criterion. 26

2.5	Posterior estimates of probability that variable i is in a particular group, in grouped variables model with two groups for monthly stock returns for nine equity securities from April, 1996 to May, 2002. The list includes energy companies: British Petroleum, Chevron, Exxon and Reliant; And financial services companies: Merrill Lynch, Bank of America, Citi-Bank, and Lehman Brothers. According to this model, Enron is classified as an energy stock.	27
2.6	Posterior estimates of R , the correlation matrix, in grouped variables model with two groups for monthly stock returns for nine equity securities from April, 1996 to May, 2002. The list includes energy companies: British Petroleum, Chevron, Exxon and Reliant; And financial services companies: Merrill Lynch, Bank of America, Citi-Bank, and Lehman Brothers. According to this model, Enron is classified as an energy stock.	28
2.7	Posterior estimates of the probability that variable i is a particular group, in grouped variables model with two groups of 11 variables on 40 broccoli (<i>brassica</i>) plants. The variables include: days to leaf, days to flower, days to harvest, leaf length, fruit number, leaf width, petiole length, leaf number, height, leaf biomass, and stem biomass.	29
4.1	Evaluating the distributional representation of four equity securities: Model choice results for analysis of the daily stock returns of General Electric, Lucent Technologies, Cisco Systems, and Sun Microsystems from April 1996 to March 2002. The four models that are used are the multivariate normal (MV-Normal), the multivariate skew-normal of Azzalini and Dalla Valle (1996) with a diagonal Δ matrix (MVS-Normal D- Δ), and the multivariate skew-normal of Sahu <i>et. al.</i> (2002) with both a diagonal and full Δ matrix (MVS-Normal F- Δ). Maximum log likelihood values are used to compute Bayes factors between the multivariate normal model and all of the other models and is reported on the log scale. The model with the highest Bayes factor best fits the data. The two models with diagonal Δ provide the best fit suggesting that there is little co-skewness. Sahu <i>et. al.</i> (2002) diagonal Δ model fits best overall.	61

4.2	Evaluating the distributional representation of global asset allocation benchmarks: Model choice results for weekly benchmark indices from January 1989 to June 2002 (Lehman Brothers government bonds, LB corporate bonds, and LB mortgage bonds, MSCI EAFE (non-U.S. developed market equity), MSCI EMF (emerging market free investments), Russell 1000 (large cap), and Russell 2000 (small cap)). The four models that are used are the multivariate normal (MV-Normal), the multivariate skew-normal of Azzalini and Dalla Valle (1996) with a diagonal Δ matrix (MVS-Normal D- Δ), and the multivariate skew-normal of Sahu <i>et. al.</i> (2002) with both a diagonal and full Δ matrix (MVS-Normal F- Δ). Maximum log likelihood values are used to compute Bayes factors between the multivariate normal model and all of the other models and is reported on the log scale. The model with the highest Bayes factor best fits the data. The Sahu <i>et. al.</i> (2002) full Δ model fits the data best.	62
4.3	Parameter estimates for diagonal Δ skew normal on four securities: Parameter estimates for the diagonal Δ model of Sahu <i>et. al.</i> (2002) used to fit the daily stock returns of General Electric, Lucent Technologies, Cisco Systems, and Sun Microsystems from April 1996 to March 2002. These estimates are the result of a Bayesian Markov chain Monte Carlo iterative sampling routine. These parameters combine to give the mean $(\mu + \sqrt{2/\pi}\Delta\mathbf{1})$, variance $(\Sigma + (1 - 2/\pi)\Delta\Delta')$, and skewness (see Appendix A for formula).	63
4.4	Parameter estimates for full Δ skew normal on global asset allocation benchmark: Parameter estimates for full Δ model of Sahu <i>et. al.</i> (2002) used to fit the weekly benchmark indices Lehman Brothers government bonds, LB corporate bonds, and LB mortgage bonds, MSCI EAFE (non-U.S. developed market equity), MSCI EMF (emerging market free investments), Russell 1000 (large cap), and Russell 2000 (small cap) from January 1989 to June 2002. These estimates are the result of a Bayesian Markov chain Monte Carlo iterative sampling routine. These parameters combine to give the mean $(\mu + (2/\pi)^{1/2}\Delta\mathbf{1})$, variance $(\Sigma + (1 - 2/\pi)\Delta\Delta')$, and skewness (see Appendix A for formula).	68

4.5 Two moment optimization for four equity securities: This table contains predictive utilities for the weights that maximize utility as a linear function of the two moments of the multivariate normal model by three different methods for daily stock returns of General Electric, Lucent Technologies, Cisco Systems, and Sun Microsystems from April 1996 to March 2002. The first method is based on predictive or future values of the portfolio (results in $\omega_{2,pred}$ where the 2 represents the number of moments in the model), the second is based on the posterior parameter estimates ($\omega_{2,param}$), and the third is the method proposed by Michaud ($\omega_{2,Michaud}$). The weights that are found by each method are ranked by the three moment predictive utility they produce (*i.e.* $E[u_{3,pred}(\omega)] = \omega' m_p - \lambda \omega' V_p \omega + \gamma \omega' S_p \omega \otimes \omega$, where the 3 signifies that the utility function is linear in the first three moments of the skew normal model, and m_p , V_p , and S_p are the predictive mean, variance and skewness) for varying values of λ and γ . The highest utility obtained signifies the method that is best for portfolio selection according to the investor's preferences. For each combination of λ and γ , $\omega_{2,pred}$ gives the highest expected utility. . . 69

4.6 Three moment optimization for four equity securities: Predictive utilities for the weights that maximize utility as a linear function of the three moments of the multivariate skew normal model by three different methods for daily stock returns of General Electric, Lucent Technologies, Cisco Systems, and Sun Microsystems from April 1996 to March 2002. The first method is based on predictive or future values of the portfolio (results in $\omega_{3,pred}$ where the 3 represents the number of moments in the model), the second is based on the posterior parameter estimates ($\omega_{3,param}$), and the third is the method proposed by Michaud ($\omega_{3,Michaud}$). The weights that are found by each method are ranked by the three moment predictive utility they produce (*i.e.* $E[u_{3,pred}(\omega)] = \omega' m_p - \lambda \omega' V_p \omega + \gamma \omega' S_p \omega \otimes \omega$, where the 3 signifies that the utility function is linear in the first three moments of the skew normal model, and m_p , V_p , and S_p are the predictive mean, variance and skewness) for varying values of λ and γ . The highest utility obtained signifies the method that is best for portfolio selection according to the investor's preferences. For each combination of λ and γ , $\omega_{3,pred}$ gives the highest expected utility. . . 69

4.7 Two moment optimization for global asset allocation benchmark indices: Predictive utilities for the weights that maximize utility as a linear function of the two moments of the multivariate normal model by three different methods for weekly benchmark indices Lehman Brothers government bonds, LB corporate bonds, and LB mortgage bonds, MSCI EAFE (non-U.S. developed market equity), MSCI EMF (emerging market free investments), Russell 1000 (large cap), and Russell 2000 (small cap) from January 1989 to June 2002. The first method is based on predictive or future values of the portfolio (results in $\omega_{2,pred}$ where the 2 represents the number of moments in the model), the second is based on the posterior parameter estimates ($\omega_{2,param}$), and the third is the method proposed by Michaud ($\omega_{2,Michaud}$). The weights that are found by each method are ranked by the three moment predictive utility they produce (*i.e.* $E[u_{3,pred}(\omega)] = \omega' m_p - \lambda \omega' V_p \omega + \gamma \omega' S_p \omega \otimes \omega$, where the 3 signifies that the utility function is linear in the first three moments of the skew normal, and m_p , V_p , and S_p are the predictive mean, variance and skewness) for varying values of λ and γ . The highest utility obtained signifies the method that is best for portfolio selection according to the investor's preferences. For each combination of λ and γ , $\omega_{3,pred}$ gives the highest expected utility. 70

4.8 Three moment optimization for global asset allocation benchmark indices: Predictive utilities for the weights that maximize utility as a linear function of the three moments of the multivariate skew normal model by three different methods for weekly benchmark indices Lehman Brothers government bonds, LB corporate bonds, and LB mortgage bonds, MSCI EAFE (non-U.S. developed market equity), MSCI EMF (emerging market free investments), Russell 1000 (large cap), and Russell 2000 (small cap) from January 1989 to June 2002. The first method is based on predictive or future values of the portfolio (results in $\omega_{3,pred}$ where the 3 represents the number of moments in the model), the second is based on the posterior parameter estimates ($\omega_{3,param}$), and the third is the method proposed by Michaud ($\omega_{3,Michaud}$). The weights that are found by each method are ranked by the three moment predictive utility they produce (*i.e.* $E[u_{3,pred}(\omega)] = \omega' m_p - \lambda \omega' V_p \omega + \gamma \omega' S_p \omega \otimes \omega$, where the 3 signifies that the utility function is linear in the first three moments of the skew normal, and m_p , V_p , and S_p are the predictive mean, variance and skewness) for varying values of λ and γ . The highest utility obtained signifies the method that is best for portfolio selection according to the investor's preferences. For all combinations of λ and γ except one ($\lambda = 0$, $\gamma = 0.5$), $\omega_{2,pred}$ gives the highest expected utility. 71

4.9 Portfolio weights: Four equity securities: Three moment (skew normal) utility based portfolio weights for daily stock returns of General Electric, Lucent Technologies, Cisco Systems, and Sun Microsystems from April 1996 to March 2002. The weights maximize the expected utility function $E[u_{3,pred}(\omega)] = \omega' m_p - \lambda \omega' V_p \omega + \gamma \omega' S_p \omega \otimes \omega$, (where the 3 signifies that the utility function is linear in the first three moments, and m_p , V_p , and S_p are the predictive mean, variance and skewness) for varying values of λ and γ . Three different methods of maximization are used. The first is based on predictive or future values of the portfolio (results in $\omega_{3,pred}$ where the 3 represents the number of moments in the model), the second is based on the posterior parameter estimates ($\omega_{3,param}$), and the third is the method proposed by Michaud ($\omega_{3,Michaud}$). 72

4.10 Portfolio weights: Global asset allocation benchmark indices: Two moment (normal) utility based portfolio weights for weekly benchmark indices Lehman Brothers government bonds, LB corporate bonds, and LB mortgage bonds, MSCI EAFE (non-U.S. developed market equity), MSCI EMF (emerging market free investments), Russell 1000 (large cap), and Russell 2000 (small cap), from January 1989 to June 2002. The weights maximize the expected utility function $E[u_{2,pred}(\omega)] = \omega' m_p - \lambda \omega' V_p \omega$, (where the 2 signifies that the utility function is linear in the two moments of the normal model, and m_p and V_p are the predictive mean and variance) for varying values of λ and γ . Three different methods of maximization are used. The first is based on predictive or future values of the portfolio (results in $\omega_{2,pred}$ where the 2 represents the number of moments in the model), the second is based on the posterior parameter estimates ($\omega_{2,param}$), and the third is the method proposed by Michaud ($\omega_{2,Michaud}$). . . 73

List of Figures

3.1	Using a common correlation model with 8 variables and varying values for ν^2 , estimates of the ratio of the normalizing constant $C(\cdot, \nu^2)$ are found. As ν^2 becomes small the average ratio becomes one and the variance of this ratio approaches zero. This suggests, that for small values of ν^2 it is reasonable to assume that $C(\delta^*, \nu^2)/C(\delta, \nu^2) = 1$.	36
3.2	Autocorrelation plots for posterior draws of μ and σ^2 using the shadow prior.	37
3.3	Kernel density estimates for posterior samples of μ and σ^2 using a common correlation model with 8 variables both with (\cdots) and without ($—$) the shadow prior. The resulting posterior distributions are similar enough to conclude that the modification of the model that comes from introducing the shadow prior is negligible.	37
4.1	This figure contains univariate estimates for Cisco Systems and General Electric daily stock returns from April 1996 to March 2002. The solid lines represents the kernel density estimate, while the dotted lines are the normal density with sample mean and variance. In one dimension the normal distribution closely matches the returns for these two stocks.	43
4.2	This figure contains a bivariate normal estimate for Cisco Systems and General Electric daily stock returns from April 1996 to March 2002. The plot is a bivariate normal with sample mean and covariance. The scatter points are the actual data. In two dimensions the bivariate normal distribution is a poor model to use for these joint returns. The actual returns exhibit co-skewness and much fatter tails than the normal approximation.	44

- 4.3 This figure contains plots of the mean, variance and skewness of portfolios consisting of two assets. Daily returns from April 1996 to March 2002 for General Electric and Lucent Technologies, Sun Microsystems and Cisco Systems, and General Electric and Cisco Systems are considered. The top row has the mean of the portfolio (equal to the linear combination of the asset means) as the weight of the first asset varies from 0 to 1. The solid line in the plots in the second row represents the linear combination of the variances of the assets, while the dotted line represents the variance of portfolios (variance of linear combination). The variance of the portfolio is always less than or equal to the linear combination of the variances of the underlying assets. The solid line in the third row of plots is the linear combination of the skewness of the two assets in the portfolio, and the dotted line is the skewness of the portfolio. The skewness of the portfolio does not dominate, nor is dominated by the linear combination of the skewness. Picking a portfolio based solely on minimum variance could lead to a portfolio with minimum skewness as well (see GE vs. Cisco). 46
- 4.4 This figure shows the space of possible portfolios based on historical parameter estimates from the daily returns of General Electric, Lucent Technologies, Cisco Systems, and Sun Microsystems from April 1996 to March 2002. The top left plot is the mean-standard deviation space, the top right plot is the mean verses the cubed-root of skewness. The bottom left plot is the standard deviation verses the cubed-root of skewness, and the bottom right plot is a three dimensional plot of the mean, standard deviation and cubed-root of skewness. In all plots that contain the skewness there is a sparse region where the skewness is zero. 56
- 4.5 This figure shows the mean-variance space of possible portfolios based on historical parameter estimates from the daily returns of General Electric, Lucent Technologies, Cisco Systems, and Sun Microsystems from April 1996 to March 2002. The portfolios are shaded according to the utility associated with each. In the left plot the utility function is $E[u_{pred}(\omega)] = \omega' m_p - 0.5 \omega' V_p \omega$, which is a linear function of the first two moments. The maximum utility is obtained by a portfolio on the frontier and is marked by a '+'. The plot on the right is shaded according to the utility function $E[u_{pred}(\omega)] = \omega' m_p - 0.5 \omega' V_p \omega + 0.5 \omega' S_p \omega \otimes \omega$, which is a linear function of the first three moments. The maximum utility is obtained by a portfolio on the frontier and is marked by a '+'. 57

4.6 This figure contains a bivariate skew normal estimate for Cisco Systems and General Electric daily stock returns from April 1996 to March 2002. The plot is a bivariate skew normal distribution with parameters estimated using the described MCMC algorithm and the model of Sahu *et. al.* (2002) with a diagonal Δ . The scatter points are the actual data. In two dimensions the bivariate skew normal distribution is a better model to use for these joint returns than the normal.

Acknowledgements

I thank my colleagues, friends and family for help and support in completing this work. My advisor Peter Müller allowed me to pick my topic and gave me unwavering support in my exploration of the subject matter. Peter has been the ideal advisor. I am very thankful for having him along on the journey.

My brother John Liechty played a major role in my life as I was growing up. He has pointed me in the right direction in many aspects. Many of the ideas presented herein result directly from him as he has once again pointed me in the right direction. John has worked in the capacity of co-advisor with Peter. I greatly appreciate his efforts and insights.

My interest in financial applications of Bayesian statistics has been spurred on by my interactions with Campbell Harvey. He has shown interest in my work from the first time I met him, and has been very patient in helping me learn more about finance. I greatly appreciate the countless meetings he has made time for. Cam has also been instrumental in the topic of portfolio selection in this work. He posed the problem to me, and supported me as a research assistant while we worked on it. I could not have asked for a better external committee member than Cam has been.

I would like to thank the two other members of my committee, Jim Berger and Michael Lavine. Both have given me additional insight into the topics discussed herein. In addition, Mike West has given me several useful comments on this thesis.

Most of all I would like to thank my wife and best friend Sara and my daughter Annabelle both of whom this work is dedicated. Without their loving support this would not have been possible.

My parents Jay and Suzanne Liechty have instilled in me a desire to continue to learn, and I thank them for the loving environment they provided for me to grow

up in. I would like to thank my brothers and sisters aswell.

Finally I would like to thank the ISDS staff for logistical and computing support over the years. These include Cheryl McGee, Sean O'Connell, Eric van Gyzen, Nicole Scott, Pat Johnson, Marsha Harrison, Tonya Rambert, and Krista Moyle.

Chapter 1

Bayesian Methods in Finance

Assumptions underly every mathematical model of reality. Models for second and higher moments have been neglected in favor of simplifying assumptions. The assumptions are necessary because they simplify reality to the extent that the model can be useful for establishing a fundamental understanding. Complexity of models has been dictated by computational capabilities. With the increased speed and availability of computers, underlying assumptions can be relaxed to allow for models that more closely fit reality. Markov chain Monte Carlo (MCMC) methods allow for analytically intractable models to be used in an efficient manner. Because of this relatively recent development some issues that were once impossible to address are now readily addressable. Covariance matrices, models that include higher moments, and analytically intractable normalizing constants are examined in this work.

I propose prior probability models for variance-covariance matrices. The proposed models address two important issues. First, the models allow a researcher to represent substantive prior information about the strength of correlations among a set of variables. Second, even in the absence of such substantive prior information, the increased flexibility of the proposed models is important. It mitigates critical dependence on strict parametric assumptions in standard prior models. For exam-

ple, the model allows *a posteriori* different levels of uncertainty about correlations among different subsets of variables. This is achieved by including a clustering mechanism in the prior probability model. Clustering is with respect to variables and pairs of variables.

Many current probability models for variance-covariance matrices shrink towards a diagonal structure. In contrast, this approach leads to *a posteriori* shrinkage towards a mixture structure implied by the clustering. I discuss appropriate posterior simulation schemes to implement posterior inference in the proposed models based on simulated data, a stock return data set and a population genetics data set. Part of this discussion focuses on computational strategies for evaluating normalizing constants that are a function of parameters of interest. The normalizing constants result from the restriction that the correlation matrix be positive definite. By introducing an additional level of hierarchy, it is found that the issue of integration over the constraint space for the correlation matrix can be side stepped. The cost involved is extra variance in the posterior estimates of the correlations. However this extra level of variance can be determined by the researcher, hence this method is very unrestrictive.

Following the argument that higher order moments are present and important in financial data, I examine the role of third moments by building on the Markowitz portfolio selection process. I incorporate higher order moments of the assets in utility functions based on predictive asset returns. I propose the use of the skew normal distribution as a sampling distribution of the asset returns. I show that this distribution has many attractive features when it comes to modeling multivariate returns. Preference over portfolios is presented in terms of expected utility maximization. I discuss estimation and optimal portfolio selection using Bayesian decision theoretic methods. These methods allow for a comparison to other optimization approaches

where parameter uncertainty is either ignored or accommodated in a non-traditional manner. The results suggest that it is important to incorporate higher order moments in portfolio selection. Further, I show that this approach leads to higher expected utility than the resampling methods common in the practice of finance. I also set up the framework for determining the markets preference over skewness in assets allocation.

Chapter 2

Bayesian Correlation Estimation

2.1 Introduction

Of the many probability models and parameterizations that exist for covariance matrices few allow for easy interpretation and prior elicitation. I propose a collection of models for covariance matrices where correlations are put together based on similarities among the correlations or based on groups of variables. This type of prior structure allows a researcher to incorporate substantive prior information related to grouping of variables and correlations. In many applications such prior information is available. For example, for financial time series it is often known that some returns are more closely related than others. Even in the absence of substantive prior information, the additional flexibility of the proposed models is important to mitigate dependence on strict parametric assumptions in standard models. The main goal of this study is to model correlation matrices. The resulting grouping of correlations or variables is an insightful byproduct. Alternative approaches to modeling correlation structure build on factor analysis. For example, recent work by West (2002) and Aguilar and West (2000) propose Bayesian factor models which are natural candidates for correlation estimation. While factor models effectively reduce

the dimensionality of the covariance matrix, it is difficult to interpret the factors and loadings, which restricts the ability for researchers to suggest informative prior distributions. These types of models may also overlook natural groupings or clusters of the underlying variables. Another alternative approach is explored in Karolyi (1992, 1993), who uses Bayesian methods to estimate the variance of individual stock returns based on stocks grouped *a priori* according to size, financial leverage, and trading volume.

The constraint to positive definiteness and the typically high dimensional nature of the parameter vector for the covariance matrix are important issues to consider when choosing a prior probability model for covariance matrices. Lack of conjugacy becomes a problem with departure from the inverse-Wishart parameterization. See, for example, Chib and Greenberg (1998) for a discussion. Another important consideration is the need to incorporate substantive prior information into the probability model. Finally, posterior simulation should be efficient and straightforward.

To focus the discussion we assume throughout a multivariate normal likelihood $y_i \sim N(0, \Sigma)$, for J -dimensional data y_i , $i = 1, \dots, n$. Extensions to normal regression models and hierarchical models with multivariate normal random effects distributions are straightforward (Daniels and Kass, 1999).

The most commonly used prior model is the conjugate inverse-Wishart (Bernardo and Smith, 1994). It allows closed form posterior inference, and efficient implementation of Gibbs sampling schemes for more complex models with additional parameters beyond the unknown covariance matrix. However, this prior model has the drawback that there is a single degree of freedom parameter ν , which is the only “tuning parameter” available to express uncertainty. Considering that with J variables, there are $J(J + 1)/2$ parameters in the matrix, one tuning parameter is very restrictive.

Several non-informative default priors have been proposed for covariance matrices. Jeffreys' prior is $p_J(\Sigma) = 1/|\Sigma|^{(J+1)/2}$. Alternatively, Yang and Berger (1994) propose a reference prior, $p_R(\Sigma) \propto 1/\{|\Sigma| \prod_{i < j} (d_i - d_j)\}$, where d_i are the eigenvalues of Σ . Like many other similarly parameterized models, the lack of intuition of the relationship between eigenvalues and correlations makes this model difficult to interpret. Daniels (1999) proposes a uniform shrinkage prior. The prior is based on considering the posterior mean as a linear combination of prior mean and sample average and assuming a uniform prior on the coefficient for the sample average (shrinkage parameter). See Christiansen and Morris (1997) and Everson and Morris (2000) for more discussion of shrinkage priors. The log matrix prior introduced by Leonard and Hsu (1992) uses a logarithmic transformation of the eigenvalue/eigenvector decomposition of Σ and allows for hierarchical shrinkage to be done with the eigenvalues. The dimension of the problem is reduced, but it is difficult to interpret the relationship of the log of the eigenvalues to the correlations and standard deviations.

Barnard *et al.* (2000) propose a separation strategy for modeling $\Sigma = SRS$ by assuming independent priors for the standard deviations S and the correlation matrix R . They propose two alternative prior models for R . One is the marginally uniform prior. The key property of this model is that the marginal prior for each r_{ij} in R is a modified beta distribution over $[-1, 1]$. With an appropriate choice of the beta parameters this becomes a uniform marginal prior distribution, thus the name. The other model for R is called the jointly uniform prior. Here the matrix R is assumed to be *a priori* uniformly distributed over all possible correlation matrices.

With improved computing facilities hierarchical priors for correlation matrices have become accessible. Daniels and Kass (1999) discuss three alternative hierarchical priors. The first is a hierarchical extension of the inverse-Wishart prior. The

extension is achieved by assuming priors on the degree of freedom parameter and on the unknown elements of a diagonal scale matrix. Alternatively they consider a separation strategy as in Barnard *et. al.* (2000) and assume a normal prior for a transformation of the correlation coefficients. The constraint to positive definiteness amounts to appropriate truncations of the normal prior. A third model uses an eigenvalue/eigenvector parameterization, with the orthogonal eigenvector matrix parameterized in terms of the Givens angles.

Wong *et. al.* (2002) propose a prior probability model on the precision matrix ($P = \Sigma^{-1}$) that is similar to my approach. Their application is geared towards graphical models and partial correlations, focusing on the sparseness of the precision matrix.

In this chapter we introduce additional hierarchical structure by allowing for correlations to group in natural ways. We consider three alternative probability models for covariance matrices. Throughout we assume a separation strategy, i.e., we model standard deviations S and the correlation matrix R separately. We focus on modeling R , as including S in the proposed posterior simulation schemes is straightforward. We start with a model that assumes a common normal prior for all correlations, with the additional restriction that the correlation matrix is positive definite. We call this model the “common correlation” model because all correlations r_{ij} are sampled from a common normal distribution, subject only to the positive definiteness constraint. This follows the frequentist work of Lin and Perlman (1985) that uses a version of the James-Stein estimator to model the off diagonal elements of the correlation matrix. The second model, the “grouped correlation” model, generalizes the common correlation model by allowing correlations to cluster into different groups, where each group has a different mean and variance. The third model, the “grouped variable” model, allows the observed variables, y_i , to

cluster into different groups. Correlations between variables in the same group have a common mean and variance and the correlations between variables in different groups have a mean and variance that depends on the group assignment for each variable.

Posterior simulation for these models presents several computational challenges. The main difficulties arise from the need to sample from truncated distributions that result from the positive definiteness constraint. The truncated distributions involve analytically intractable normalizing constants that are a function of the parameters. See Chen *et. al.* (2000, chap.6), for a general discussion of related problems. We investigate several strategies for effectively evaluating these normalizing constants. The three main strategies being: sidestepping the problem by assuming that ratios of these normalizing constants are approximately constant; using importance sampling strategies; and introducing an additional set of latent variables, denoted as ‘shadow’ priors, in the hierarchical structure. We discuss these three strategies in detail in Chapter 3.

The outline of Chapter 2 is as follows. Section 2.2 introduces the three proposed models. Implementation and posterior simulation are discussed in Section 2.3. Section 2.4 reviews possible areas of application for each model, with a discussion of feasible extensions. Section 2.5 gives examples and Section 2.6 concludes.

2.2 Models

I propose three alternative models for correlation matrices. The basic model assumes a common normal prior for all correlations r_{ij} , subject to positive definiteness. This is generalized into the grouped correlation model by allowing for groups of correlations. Finally, I introduce the grouped variable model where we group the correlations based on the variables. Throughout, we assume a multivariate normal

likelihood function and follow the separation strategy of Barnard *et. al.* (2000) writing $\Sigma = SRS$. Here S is a diagonal matrix of standard deviations and R is the $J \times J$ correlation matrix. Without loss of generality we assume $S = I$. Generalizations to include unknown variances are straightforward. We write \mathcal{R}^J for the space of all correlation matrices of dimension J . Before describing each of the models in detail, I consider a problem from the finance industry which motivates the grouped variables model.

2.2.1 Motivating example

In an effort to simplify the task of diversification, the finance community is interested in classifying companies into industries based on the type of products and services provided by a company. Typically this task of classification is done by individual's who have industrial expertise. While many of these classifications may be straightforward, there are a number of companies that engage in strategies where they expand and/or change the products and services that they offer, creating hybrid companies that may not fit into a specific industrial classification. As an example, before it's demise, Enron was traditionally an energy company that provided products and services related to the petroleum industry. However, during the 1990s they engaged in a strategy of transforming their basic business from an energy company to a finance company. As it is debatable whether an industry expert would be able to correctly classify Enron, it would be of interest to consider an analysis of Enron's stock performance and determine whether their stock behavior is correlated with energy companies or with finance companies. To illustrate, consider the monthly stock returns for nine companies which are either energy or finance companies from April, 1996 to May, 2002. The energy stocks include British Petroleum, Chevron, Exxon and Reliant; The financial stocks include Merrill Lynch, Bank of America,

Citi-Bank, and Lehman Brothers. The stock in the middle is Enron; see Table 2.1 for their empirical correlations.

Table 2.1: Empirical correlation matrix for monthly stock returns for nine equity securities from April, 1996 to May, 2002. The list includes energy companies: British Petroleum, Chevron, Exxon and Reliant; And financial services companies: Merrill Lynch, Bank of America, Citi-Bank, and Lehman Brothers. Enron could potentially be in either group based on different criterion.

Variable		1	2	3	4	5	6	7	8	9
Reliant	1	1	0.30	0.14	0.31	0.47	0.10	0.15	0.13	0.09
Chevron	2	0.30	1	0.75	0.60	0.54	0.34	0.22	0.13	0.41
BP	3	0.14	0.75	1	0.59	0.44	0.22	0.23	0.11	0.25
Exxon	4	0.31	0.60	0.59	1	0.32	0.38	0.15	0.21	0.26
Enron	5	0.47	0.54	0.44	0.32	1	0.08	0.11	-0.02	0.10
Citi-B.	6	0.10	0.34	0.22	0.38	0.08	1	0.69	0.71	0.68
Leh. Bros.	7	0.15	0.22	0.23	0.15	0.11	0.69	1	0.77	0.59
Merrill	8	0.13	0.13	0.11	0.21	-0.02	0.71	0.77	1	0.50
B.ofAm.	9	0.09	0.41	0.25	0.26	0.10	0.62	0.59	0.50	1

The grouped variable model classifies stocks into groups based on the correlation within and between each group, this offers a natural method for determining whether Enron successful made the transition from being an energy company to a financial company before they encountered recent troubles. This is explored in Section 2.5.

2.2.2 Common correlation

In the common correlation model we assume *a priori* that all correlations r_{ij} are sampled from a common normal distribution subject to $R \in \mathcal{R}^J$:

$$f(R|\mu, \sigma^2) = C(\mu, \sigma^2) \prod_{i < j} \exp \{-1/(2\sigma^2)(r_{ij} - \mu)^2\} I\{R \in \mathcal{R}^J\}, \quad (2.1)$$

where

$$C^{-1}(\mu, \sigma^2) = \int_{R \in \mathcal{R}^J} \prod_{i < j} \exp \{-1/(2\sigma^2)(r_{ij} - \mu)^2\} dr_{ij}, \quad (2.2)$$

and where $I\{\cdot\}$ represents an indicator function. We assume hyper-priors $\mu \sim N(0, \tau^2)$, and $\sigma^2 \sim IG(\alpha = \textit{shape}, \beta = \textit{scale})$, where τ^2 , α , and β are treated as known. The indicator function in (2.1) ensures that the correlation matrix is positive definite and introduces dependence across the r_{ij} 's. The implication of the constraint on the conditional prior is not the same for each coefficient. The full conditional posterior distribution $f(r_{ij}|-)$ will play a prominent role in the posterior simulation discussed later:

$$f(r_{ij}|-) \propto |R|^{-\frac{n}{2}} \exp \left\{ -1/2 \operatorname{tr}(R^{-1}B) \right\} \exp \left\{ -1/(2\sigma^2)(r_{ij} - \mu)^2 \right\} I\{R \in \mathcal{R}^J\}. \quad (2.3)$$

Where B is the empirical variance-covariance matrix. The full conditional densities for μ and σ^2 are similar to the conjugate densities with an additional factor due to the positive definiteness constraint on R :

$$f(\mu|-) \propto C(\mu, \sigma^2) \prod_{i < j} \exp \left\{ -1/(2\sigma^2)(r_{ij} - \mu)^2 \right\} \exp \left\{ -1/(2\tau^2)(\mu)^2 \right\} \quad (2.4)$$

and

$$f(\sigma^2|-) \propto C(\mu, \sigma^2) \prod_{i < j} \exp \left\{ -1/(2\sigma^2)(r_{ij} - \mu)^2 \right\} (1/\sigma^2)^{\alpha-1} \exp(-\beta/\sigma^2). \quad (2.5)$$

By symmetry the prior is centered at $R = I$ and thus implements shrinkage towards a diagonal correlation matrix. Alternative shrinkage to positive and negative correlations is possible by choosing different prior means for μ . However, model (2.1) constrains such subjective prior specification to be common to all correlation coefficients. The generalizations proposed below remove this constraint and allow clustering of correlation coefficients.

2.2.3 Grouped correlation

In many applications the common correlation model is too restrictive. For example, one might conjecture *a priori* that correlations cluster into groups of positive correlations and negative correlations. Substantive prior information about such clustering might arise from the interpretation of the corresponding pairs of variables.

Allowing *a priori* for groups of correlations can be achieved by generalizing the common correlation model to a mixture prior:

$$f(R|\mu, \sigma^2, \vartheta) = C(\mu, \sigma^2, \vartheta) \prod_{i < j} \left[\sum_{k=1}^K I\{\vartheta_{ij} = k\} \exp \left\{ -1/(2\sigma_k^2)(r_{ij} - \mu_k)^2 \right\} \right] I\{R \in \mathcal{R}^J\}, \quad (2.6)$$

with $\vartheta_{ij} \sim \text{multinomial}(p)$, and $C(\mu, \sigma^2, \vartheta)$ analogous to (2.2). The indicator $I\{\vartheta_{ij} = k\}$ selects one of the K clusters. To avoid trivial identifiability problems due to arbitrary permutation of indices, post processing may be necessary. See, for example, Celeux *et. al.* (1999) for a discussion of issues related to parameterizing mixture models. The full conditional distribution (2.3) remains almost unchanged:

$$f(r_{ij}|-) \propto |R|^{-\frac{n}{2}} \exp \left\{ -1/2 \text{tr}(R^{-1}B) \right\} \exp \left\{ -1/(2\sigma_{\vartheta_{ij}}^2)(r_{ij} - \mu_{\vartheta_{ij}})^2 \right\} I\{R \in \mathcal{R}^J\}. \quad (2.7)$$

As with the common correlation model, the full conditional densities for the μ 's and σ^2 's are not conjugate and are similar to (2.4) and (2.5). The full conditionals for ϑ_{ij} 's are multinomial distributions which will be dealt with in Section 2.3.

Besides accommodating substantive prior information about clustering of correlations, the mixture prior (2.6) is also motivated by concerns about the strict normality assumption in (2.1). In (2.1), outliers with high correlations could unduly influence final inference. Also, bi-modality arising from uncertainty about the direction of a correlation can not be represented with the single normal prior in

(2.1). For sufficiently large K the mixture model in the grouped correlation prior allows the model to approximate any random effects distribution, subject to some technical constraints only (Dalal and Hall, 1983).

Model (2.6) implements shrinkage of R towards a structure determined by clustering pairs (i, j) of variables. In many problems this is more appropriate than shrinkage towards a diagonal matrix. Additionally, the introduction of the mixture indicators ϑ_{ij} in the model allows a researcher to represent substantive prior information by choosing non-equal prior probabilities $Pr(\vartheta_{ij} = k)$. The posterior distribution under (2.6) includes inference about grouping r_{ij} into high and low correlations.

Model (2.6) includes as a special case model selection for different dependence structures. Specifically, by including as one term in the mixture a point mass δ_0 at zero, the model allows inference about the presence of (marginal) dependence of any two variables. Similar mixture models are commonly used for variable selection in regression models (e.g., Clyde and George, 2000):

$$f(R|\mu, \sigma^2, \vartheta) = C(\mu, \sigma^2, \vartheta) \prod_{i < j} \left[\sum_{k=1}^{K-1} I\{\vartheta_{ij} = k\} \exp\{-1/(2\sigma_k^2)(r_{ij} - \mu_k)^2\} \right. \\ \left. + I\{\vartheta_{ij} = K\} \delta_0(r_{ij}) \right] I\{R \in \mathcal{R}^J\}.$$

Alternatively, a point mass at zero could be replaced by a tight normal distribution as in George and McCulloch (1993).

2.2.4 Grouped variables

In many applications it is more natural to group the variables, rather than the correlations. For example, as discussed in the motivating example, it is natural to assume that there is a common correlation for percent changes in the stock price

(returns) of companies in similar industries and there is a different, smaller common correlation between companies in two different industries. One might expect a common correlation of returns between bank stocks and a different correlation between the returns of bank stocks and energy company stocks.

If we group the variables instead of the correlations, the prior (2.6) changes to:

$$f(R|\mu, \sigma^2, \vartheta) = C(\mu, \sigma^2, \vartheta) \prod_{i < j} \left[\sum_{k,h} I\{\vartheta_i = k\} I\{\vartheta_j = h\} \exp \left\{ -1/(2\sigma_{kh}^2)(r_{ij} - \mu_{kh})^2 \right\} \right] I\{R \in \mathcal{R}^J\}, \quad (2.8)$$

where again, $C(\mu, \sigma^2, \vartheta)$ is analogous to (2.2) and $\vartheta_i \sim \text{multinomial}(p)$. The full conditional posterior distribution for r_{ij} is as in (2.7), with μ_k replaced by μ_{kh} :

$$f(r_{ij}|\vartheta_i = k, \vartheta_j = h, \dots) \propto \quad (2.9)$$

$$|R|^{-\frac{n}{2}} \exp \left\{ -1/2 \text{tr}(R^{-1}B) \right\} \exp \left\{ -1/(2\sigma_{kh}^2)(r_{ij} - \mu_{kh})^2 \right\} I\{R \in \mathcal{R}^J\}.$$

Like (2.6), model (2.8) implements shrinkage of R towards a cluster structure, which now is determined by clustering variables, i.e., clustering is defined on indices i only. For each correlation r_{ij} the pair of indicators $(\vartheta_i, \vartheta_j)$ chooses the term in the prior mixture model. As with model (2.6), we can explore different dependence structures as a special case of model (2.8) by including a point mass at zero as a term in the model. This type of model could result in a block diagonal correlation matrix, potentially revealing independence between different groups of variables.

2.3 Implementation and posterior simulation

Posterior inference relies on Markov chain Monte Carlo posterior simulation. See, for example, Tierney (1994) for a discussion of Markov chain Monte Carlo schemes

referred to in the following discussion. Central to the Markov chain Monte Carlo scheme is sampling and evaluation of the complete conditional posterior distributions which result from the three models discussed in the previous section. Implementing such a Markov chain Monte Carlo scheme presents significant computational challenges. Most of these challenges arise from the need to evaluate the normalizing constant $C(\cdot)$ in (2.1), (2.6) and (2.8).

2.3.1 Sampling the full conditional of r_{ij}

Without loss of generality we consider only the full conditional (2.3) for r_{ij} in the common correlation model. The awkward manner in which r_{ij} is embedded in the likelihood complicates posterior simulation, leading us to use a Metropolis-Hastings algorithm to update one coefficient r_{ij} at a time (Barnard *et. al.* , 2000). See, for example, Chib and Greenberg (1995) for a review of the Metropolis-Hastings algorithm. The positive definiteness of R constrains $f(r_{ij}|-)$ to an interval (lb_{ij}, ub_{ij}) . Once this interval is found, there are several different proposal densities that could be used. The simplest proposal is a uniform density on that interval. Another possible proposal density is a Beta density that has been modified to fit the interval, with a mean equal to the current realization of r_{ij} and a variance that is a fraction of the interval length.

2.3.2 Sampling the full conditionals of μ and σ^2

We will focus the discussion on sampling from the full conditional density for μ based on the common correlation model, see (2.4). Extensions of these strategies to σ^2 for the common correlation model and to μ and σ^2 for the grouped models can be done in a natural way. Because μ is hopelessly entangled in the normalizing constant C , we again use a Metropolis-Hastings step to update μ . The proposal

density is the normal distribution that results if $I\{R \in \mathcal{R}^J\}$ is removed from (2.1). Let μ^* denote the generated proposal. The appropriate acceptance probability is

$$\alpha = \min \{1, C(\mu^*, \sigma^2)/C(\mu, \sigma^2)\}, \quad (2.10)$$

where C is given by (2.2). We propose several alternative strategies to evaluate α . The simplest approach is to assume $\alpha = 1$. This is the strategy that Daniels and Kass (1999) use when analyzing their model which uses a Fisher z transformation of the correlations. While this may be reasonable when μ is close to μ^* or σ^2 is small, these conditions may not hold in practice.

Since C is not data dependent, one may evaluate the value of $C(\mu, \sigma^2)$ once, up-front, for a range of values for μ and σ^2 and then use an interpolation strategy to evaluate α . While this approach has the advantage of only being dependent on the dimension of the problem, I chose to focus on strategies which estimate C as needed.

The normalizing constant C is proportional to the integral of a product of univariate normal densities restricted to a constrained space and can be estimated using an importance sampling scheme. See Chen *et. al.* (2000, chapter 5) for a general discussion of using importance sampling to estimate the ratio of multivariate integrals, as in (2.10). One importance sampling strategy is to sample from unconstrained normal densities, $r_{i,j}^m \sim N(\mu, \sigma^2)$, $i < j$, $m = 1, \dots, M$, define $R^m = [r_{ij}^m]$ and use

$$\hat{C}(\mu, \sigma^2) = 1/M \sum_{m=1}^M I\{R^m \in \mathcal{R}^J\}.$$

Another strategy is to consider a slight modification of the original model by inserting an additional layer of priors into the model hierarchy. We introduce latent variables δ_{ij} which enter the model hierarchy between r_{ij} and the prior moments

(μ, σ^2) by assuming

$$\delta_{ij} \sim N(\mu, \sigma^2) \tag{2.11}$$

and replace the prior (2.1) by

$$f(R|\delta) = C(\delta, \nu^2) \prod_{i < j} \exp \left\{ -1/(2\nu^2) (r_{ij} - \delta_{ij})^2 \right\} I\{R \in \mathcal{R}^J\},$$

where C^{-1} is similar to (2.2) with μ replaced by δ_{ij} and σ^2 replaced by ν^2 . We refer to (2.11) as a “shadow prior.” The resulting full conditional density for r_{ij} changes only slightly, with (δ_{ij}, ν^2) replacing (μ, σ^2) , but the full conditional density for μ and σ^2 are now conjugate. The nature of the full conditional density for δ_{ij} is similar to the full conditional (2.3) in the original model, requiring a Metropolis-Hastings step to update the δ_{ij} . However, there are important advantages to using the additional parameters and model structure. First, as the researcher has complete control over the value of ν^2 , it can be set to an arbitrary value. As ν^2 approaches zero, the ratio in (2.10) approaches one. In practice, it is often reasonable to set ν^2 to a small number and assume $C(\delta^*, \nu^2)/C(\delta, \nu^2) = 1$. Intuitively, by setting ν^2 small enough, the full conditional density of R essentially lies inside the constrained space \mathcal{R}^J , which allows the unconstrained normalizing constant to be a reasonably good approximation for the constrained normalizing constant $C(\delta, \nu^2)$. A second important advantage of introducing the shadow prior is the simplification of the computational burden associated with sampling the indicator variables in the two grouped models (see Section 2.3.3). Although the shadow prior offers a general way of dealing with constraints, the full extent of its effectiveness and limitations are not explored in this paper.

Perhaps the most important argument for using the shadow prior is a critical simplification in the full conditional for the indicators ϑ_i and ϑ_{ij} in the grouped variable and grouped correlation models, respectively. We discuss details in the

next section.

2.3.3 Sampling the full conditional of ϑ

Without the shadow priors, the full conditional distributions for ϑ for the grouped correlation model and grouped variables model are as follows:

$$f(\vartheta_{ij} = k | -) \propto C(\mu, \sigma^2, \vartheta_{ij} = k, \vartheta_{-ij}) \exp \left\{ -1/(2\sigma_k^2)(r_{ij} - \mu_k)^2 \right\} \quad (2.12)$$

and

$$f(\vartheta_i = k | -) \propto C(\mu, \sigma^2, \vartheta_i = k, \vartheta_{-i}) \prod_{j \neq i} \exp \left\{ -1/(2\sigma_{k, \vartheta_j}^2)(r_{ij} - \mu_{k, \vartheta_j})^2 \right\}. \quad (2.13)$$

Evaluating these full conditional densities requires that we calculate the normalizing constant C , which can be done using importance sampling strategies, as discussed previously. As the dimensionality of the correlation matrix increases, the computational task of evaluating (2.12) and (2.13) can make these models difficult if not impossible to analyze in practice.

By introducing the shadow prior we run into some good luck. Evaluating (2.12) and (2.13) significantly simplifies. With the shadow prior included in each model, the variables ϑ are no longer a part of the multivariate integral C , and as a result the full conditional densities (2.12) and (2.13) become

$$f(\vartheta_{ij} = k | -) \propto \exp \left\{ -1/(2\sigma_k^2)(\delta_{ij} - \mu_k)^2 \right\}$$

and

$$f(\vartheta_i = k | -) \propto \prod_{j \neq i} \exp \left\{ -1/(2\sigma_{k, \vartheta_j}^2)(\delta_{ij} - \mu_{k, \vartheta_j})^2 \right\}.$$

As a result, the full conditional densities for ϑ , using the shadow prior, are computationally easy to evaluate, even for high dimensional problems.

2.4 Extensions and applications

Probability models for covariance matrices are useful in many important hierarchical inference problems. We review some here.

2.4.1 Random effects distributions in hierarchical models

The proposed models were presented in the context of a multivariate normal sampling model for observed data $y_i \sim N(0, \Sigma)$. But the discussion and the proposed implementation of posterior simulation remain valid also for more complex probability models where the multivariate normal distribution defines one component or level of the probability model only. An important example are hierarchical models with multivariate normal random effects distributions $\theta_i \sim N(\mu, \Sigma)$ (Daniels and Kass, 1999). Here θ_i is a random effects vector specific to the i -th experimental unit in the hierarchical model. The proposed prior models would be used to define the hyper-prior for the covariance matrix Σ of the random effects. On top of this random effects model could be a sampling distribution for the observable data y_i . In many applications this takes the form of a nonlinear regression $f(y_i|\theta_i)$.

Example 1. Müller and Rosner (1997) describe a hematologic study. The data records white blood cell counts over time for each of n chemotherapy patients. Denote with y_{it} the measured response for patient i on day t . The profiles of white blood cell counts over time look similar for most patients. There is an initial base line, followed by a sudden decline when chemotherapy starts, and a slow, S-shaped recovery back to approximately base line after the end of treatment. Profiles can be reasonably well approximated with a piecewise linear-linear-logistic regression, using a 7-dimensional parameter vector (Müller and Rosner, 1997). But the nonlinear regression parameters differ significantly across patients. Thus we introduce

a patient specific random effects vector θ_i . Conditional on θ_i we assume a nonlinear regression using the piecewise linear-linear-logistic regression model

$$y_{it} = g_{\theta_i}(t) + \epsilon_{it}.$$

The model could be completed with a random effects model

$$\theta_i \sim N(m, \Sigma), \tag{2.14}$$

and a hyper-prior $h(m, \Sigma)$. Posterior predictive inference for future patients depends on the observed historical data only indirectly through learning about the random effects distribution (2.14). Conditional on knowing the parameters of the random effects model, observed and future data are independent. Thus a flexible hyper-prior for Σ is critical to effect the desired learning. The grouped correlation model provides a possible hyper-prior on Σ . Alternatively, Müller and Rosner (1997) use a flexible non-parametric model in place of the normal model (2.14).

2.4.2 ARCH models and time-varying correlations

ARCH (autoregressive conditional heteroskedasticity) models have achieved a considerable following in the econometrics and finance literature since their introduction by Engle (1982). An ARCH model is a discrete stochastic process with the characteristic feature that the variance at time t is related to previous squared values of the process in an autoregressive scheme:

$$\epsilon_t \sim N(0, h_t), \quad h_t = \alpha_0 + \sum_{j=1}^p \alpha_j \epsilon_{t-j}^2. \tag{2.15}$$

Bollerslev (1986) extended the ARCH formulation to include lagged values of the variance itself in the variance equation thereby providing a more flexible model

form. The GARCH(p,q) model for variance is

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \gamma_j h_{t-j}.$$

See, for example Bollerslev *et. al.* (1992) for a survey of related models.

A multivariate version of the GARCH(p,q) model was first studied by Bollerslev *et. al.* (1988). Let H_t denote the covariance matrix in a multivariate version of (2.15), $\epsilon_t \sim N(0, H_t)$. The high dimensional nature of H_t complicates modeling. In practice the dimensionality is greatly reduced by imposing additional structural assumptions. For example, Bollerslev (1990) has a time varying conditional covariance matrix, H_t , but assumes time invariant conditional correlations. The proposed separation strategy inherent in all three proposed models, (2.1), (2.6) and (2.8), provides a convenient approach to implement such models. It is natural to extend the framework presented in this chapter to model multivariate time-series data, by assuming a univariate (G)ARCH model for the marginal variances $h_{it} = (H_t)_{ii}$, and completing the model with a structured prior for the correlation matrix, using one of the models proposed in Section 2.2. Let S_t denote a diagonal matrix with the standard deviations on the diagonal. Assume $H_t = S_t R S_t$ with equations (2.1), (2.6), or (2.8) as prior models for a time invariant correlation matrix R .

Not only does this framework offer a natural way to model groupings of correlation among multivariate time-series data, it allows for a natural way to jointly model time-varying variances and time-varying correlation structures. The number of parameters in a correlation matrix can make it prohibitive to introduce time-varying structures for correlations. However, by shrinking correlations towards a set of common means, standard time-varying probabilistic models can be used to model the dynamic nature of these means, which offers a parsimonious way of modeling the time-varying correlation structure.

2.4.3 Probit models

The multivariate probit model implements regression of a set of binary response variables $y = (y_1, \dots, y_p)$ on covariates $x = (x_1, \dots, x_p)$. We introduce a p -dimensional normal latent variable vector y^* and assume

$$y_i = I\{y_i^* > 0\} \text{ and } y^* \sim N(\beta' x, \Sigma). \quad (2.16)$$

In words, conditional on a vector x of covariates the binary responses are assumed to have probabilities given by multivariate normal quantiles. Introducing the latent variables y^* reduces implementation to a standard normal linear regression problem. Albert and Chib (1995) discuss the corresponding univariate probit model. In the multivariate model (2.16), the covariance matrix Σ defines the covariate effects. For identifiability Σ needs to be suitably constrained, for example, as a correlation matrix. Models (2.1), (2.6) and (2.8) provide flexible prior models. Alternative Bayesian models for the identified parameters of a multivariate probit are discussed by Chib and Greenberg (1998) and McCulloch *et. al.* (2000). Müller *et. al.* (1999) discuss a parameterization which allows conjugate Gibbs sampling.

Example 2. Liechty *et. al.* (2001) discuss how the multivariate probit model can be used to estimate an empirical demand function for an information service. Potential commercial customers of an Internet Yellow Page service (IYPS) were shown a collection of scenarios where they could purchase enhancements/services to a basic free listing in an IYPS. Each enhancement had a monthly maintenance fee and some had an additional one time set-up fee. The levels of the fees along with several group discounting schemes were varied over the different scenarios. For each scenario, a potential customer indicated whether they would choose each of the possible enhancements, with the restriction that for some of the enhancements only one out of a sub-set of the enhancements could be selected. These choice

tasks resulted in a set of vectors of binary responses, which were modeled using a multivariate probit model. The price of the enhancement and the group discounting schemes were used as covariates. As revealed through the estimated correlation matrix, the products had relationships net of price and group discount effects. One way of interpreting these correlations is that the products with positive correlations were complementary products and the products with negative correlations were substitute products. A two group, grouped correlation model, would be a natural candidate for identifying complements and substitutes for this type of data set.

In addition, the choose many from many choice task is regularly encountered by consumers as they shop at supermarkets and the multivariate probit model is often used to model the content of analyzing a household’s shopping basket over time; see Manchanda *et. al.* (1999). For this class of problems, the resulting correlation matrix can become very large, but have a fairly simple underlying structure. For example it would be natural to consider a correlation structure which groups correlations between products based on product categories using a grouped variables model. It is also natural to look for zeros in the correlation matrix in much the same way as Wong *et. al.* (2002) look for zeros in the precision matrix. They interpret zeros in the partial correlation matrix in the graphical models context, but zeros in the correlation matrix, particularly as defined by the groups of variables, could be used to identify block diagonal matrices which could be used to define how consumers partition products in to different types of markets.

2.5 Examples

2.5.1 Simulated data

I generated an (8×8) correlation/covariance matrix with two different groups of correlations, using (2.6) with $\mu_1 = -0.54$, $\sigma_1^2 = 0.05$, $\mu_2 = 0.37$, and $\sigma_2^2 = 0.08$. This

model has $(8 \times 7)/2 = 28$ correlations; 14 correlations are in each group, and can be thought of as being in a negative (group 1) or positive (group 2) group. Based on this matrix I generated 500 observations. The true group of each correlation and the posterior probabilities are shown in Table 2.2. The posterior estimates recover the true structure of the correlation model in the simulation model.

Table 2.2: Simulation study with (8×8) correlation matrix consisting of two groups of correlations and 500 observations. The $(8 \times 7)/2 = 28$ correlations are evenly divided between two groups that have parameters values $\mu_1 = -0.54$, $\sigma_1^2 = 0.05$, $\mu_2 = 0.37$, and $\sigma_2^2 = 0.08$. Let ϑ_0 denote the true simulation model. Let $\hat{\vartheta}$ denote the posterior mean estimates. The matrix on the left shows the group that each of the correlations is in (ϑ_0), and the matrix on the right shows the posterior probability that the correlations are from group 1 ($\hat{\vartheta}$).

	ϑ_0									$\hat{\vartheta}$							
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8
1	0	1	1	1	2	2	2	2	1	0	.99	.99	.01	.01	.01	.00	.01
2	0	0	2	1	2	1	2	1	2	0	0	.00	.99	.00	.98	.00	.99
3	0	0	0	2	2	2	1	2	3	0	0	0	.00	.01	.00	.99	.00
4	0	0	0	0	1	2	1	2	4	0	0	0	0	.99	.00	.98	.00
5	0	0	0	0	0	1	1	1	5	0	0	0	0	0	.98	.02	.99
6	0	0	0	0	0	0	1	2	6	0	0	0	0	0	0	.97	.00
7	0	0	0	0	0	0	0	1	7	0	0	0	0	0	0	0	.99
8	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0

Additional studies with group means closer together and various values for group variances yielded similar results. When the residual variance σ^2 are large, the posterior probabilities are less concentrated and more spread out across groups, as would be expected for any mixture with overlapping distributions. This suggests that inference on K should be interpreted with caution, unless the clusters are well separated. I found similar results when simulating from the grouped variables model. To investigate the performance of the proposed model when there are no real groupings in the data, I considered the special case with $K = 1$ in the simulation

model. Posterior inference with $K > 1$ finds means close enough to conclude that there is only one group.

2.5.2 Structure in the stock market

As discussed in the motivating example, Enron was attempting to change from being an energy company to a finance company. As it may be difficult to classify Enron based on the range of products and services that they offer, it is possible to use the grouped variable model to determine whether to group Enron with energy stocks or finance stocks, based on the correlations between each group of stocks. Using the stocks introduced in the motivating example, we consider three different grouped variable models with different numbers of groups K , and use a reversible jump Markov chain Monte Carlo algorithm to infer the value of K . When $K = 1$, the grouped variable model reduces to the common correlation model (2.1). In addition to the common correlation model, we consider the grouped variables model (2.8) with two and three groups where $K = 2$, and $K = 3$ respectively. We also include a uniform prior as in Barnard *et. al.* (2000) for comparison. With a uniform prior on the model space, the reversible jump Markov chain Monte Carlo reduces to comparing the likelihoods of the current model and the proposed model at each step. Doing this shows that there is a 50% posterior probability that $K = 2$, see Table 2.3.

For the model with $K = 2$, the posterior parameter estimates are shown in Tables 2.4, 2.5, and 2.6. From Table 2.4 there is a distinct separation between the three group means. Table 2.5 shows that Enron is clearly grouped with the energy companies, and was unsuccessful, in terms of stock performance, in making the transition from being an energy company to a finance company. The posterior estimate of the correlation matrix for these variables is shown in Table 2.6.

Table 2.3: Result of reversible jump Markov chain Monte Carlo exercise to determine which model best fits monthly stock returns for nine equity securities from April, 1996 to May, 2002. The list includes energy companies: British Petroleum, Chevron, Exxon and Reliant; And financial services companies: Merrill Lynch, Bank of America, Citi-Bank, and Lehman Brothers. Enron could potentially be in either group based on different criterion. The reversible jump Markov chain Monte Carlo results in 50% posterior probability on the grouped variables model with two groups.

$K = 1$	$K = 2$	$K = 3$	Uniform Prior
0.06	0.5	0.25	0.19

Table 2.4: Posterior estimates of μ and σ^2 in grouped variables model with two groups for monthly stock returns for nine equity securities from April, 1996 to May, 2002. The list includes energy companies: British Petroleum, Chevron, Exxon and Reliant; And financial services companies: Merrill Lynch, Bank of America, Citi-Bank, and Lehman Brothers. Enron could potentially be in either group based on different criterion.

μ_{12}	μ_1	μ_2	σ_{12}^2	σ_1^2	σ_2^2
0.15	0.41	0.61	0.02	0.03	0.02

How does this classification of variables compare to classical factor analysis? Classical factor models with both the MLE and principal components method indicated five factors for this data. Loadings for both methods are very similar. But they offer no concise ‘groupings’ of variables as is found by the grouped variables model.

2.5.3 Population genetics

Murren *et. al.* (2002) consider a data set recording measurement of $J = 11$ traits over a population of $n = 40$ plants (*brassica*, i.e., broccoli). One of the questions of interest is how the different traits can be clustered into groups on the basis of correlation structure. Variables are *a priori* expected to cluster into groups of traits related to some common underlying themes. For example, possible groups might be variables related to “life history,” “plant size,” etc. The grouped variable model

Table 2.5: Posterior estimates of probability that variable i is in a particular group, in grouped variables model with two groups for monthly stock returns for nine equity securities from April, 1996 to May, 2002. The list includes energy companies: British Petroleum, Chevron, Exxon and Reliant; And financial services companies: Merrill Lynch, Bank of America, Citi-Bank, and Lehman Brothers. According to this model, Enron is classified as an energy stock.

Variable	$p(\vartheta = 1 y)$	$p(\vartheta = 2 y)$
Reliant Energy	0.99	0.01
Chevron	1	0
British Petroleum	1	0
Exxon	1	0
Enron	0.99	0.01
Citi-Bank	0.01	0.99
Lehman Brothers	0	1
Merrill Lynch	0	1
Bank of America	0.02	0.98

allows to formally investigate such inference.

I modeled the correlation matrix using a two-group, grouped variables model. The posterior probability that each variable is assigned to group one is summarized in Table 2.7. The three variables associated with the size of the leaves are classified into group 2. The remaining variables, placed in group 1, have to do with the size of the other parts of the plant and with the number, but not the size, of leaves. The average correlation for the leaf size group, group 2, is 0.91, while the average correlation for the plant size group, group 1, is 0.25. Interestingly, the average correlation between the two groups is negative, -0.31 . This negative correlation between leaf size and plant size variables seems to indicate that the plant either emphasizes leave growth or plant growth, but not both.

Table 2.6: Posterior estimates of R , the correlation matrix, in grouped variables model with two groups for monthly stock returns for nine equity securities from April, 1996 to May, 2002. The list includes energy companies: British Petroleum, Chevron, Exxon and Reliant; And financial services companies: Merrill Lynch, Bank of America, Citi-Bank, and Lehman Brothers. According to this model, Enron is classified as an energy stock.

Variable	1	2	3	4	5	6	7	8	9	
Reliant	1	0.32	0.18	0.33	0.43	0.13	0.16	0.16	0.11	
Chev.	2	0.32	1	0.66	0.5	0.49	0.26	0.17	0.1	0.32
BP	3	0.18	0.66	1	0.51	0.4	0.13	0.2	0.08	0.16
Exxon	4	0.33	0.5	0.51	1	0.29	0.3	0.09	0.17	0.16
Enron	5	0.43	0.49	0.4	0.29	1	0.11	0.15	0.04	0.11
C-B	6	0.13	0.26	0.13	0.3	0.11	1	0.63	0.66	0.62
L.Bros.	7	0.16	0.17	0.2	0.09	0.15	0.63	1	0.72	0.56
ML	8	0.16	0.1	0.08	0.17	0.04	0.66	0.72	1	0.46
BofA	9	0.11	0.32	0.16	0.16	0.11	0.62	0.56	0.46	1

2.6 Conclusion

I have proposed new mixture model priors for correlation matrices. The main arguments for the proposed models are the increased flexibility in representing substantive prior knowledge about the dependence structure, and less dependence on strict parametric assumptions implied in some of the standard models reviewed in the introduction. The main limitation of the proposed models is the computation intensive implementation. We have partially mitigated this problem by proposing appropriate Markov chain Monte Carlo strategies, including the shadow prior mechanism, which shows promise as a way of addressing computational challenges that arise from placing constraints on parameters. Using simulation studies and two empirical examples, we have demonstrated the usefulness of these models in both making inference about correlation structures and about gaining insight into the underlying physical phenomena.

The proposed models provide a framework to represent and learn about de-

Table 2.7: Posterior estimates of the probability that variable i is a particular group, in grouped variables model with two groups of 11 variables on 40 broccoli (*brassica*) plants. The variables include: days to leaf, days to flower, days to harvest, leaf length, fruit number, leaf width, petiole length, leaf number, height, leaf biomass, and stem biomass.

Variable	$P(\vartheta_i = 1)$	$P(\vartheta_i = 2)$
days to leaf	1	0
days to flower	1	0
days to harvest	1	0
leaf length	0	1
fruit number	1	0
leaf width	0	1
petiole length	0	1
leaf number	1	0
height	1	0
leaf biomass	1	0
stem biomass	1	0

pendence structure. An alternative approach is traditional factor analysis. Factor analysis explains dependence among a set of variables as arising from a few latent factors. The relative size of the factor loadings determines the strength of the correlations. In contrast, the grouped variables and grouped correlations models proceed by assuming a partition of the set of correlations and variables, respectively. The extent of the relationship between these two modeling approaches is an interesting topic for future research.

The discussion was focused on inference on covariance and correlation matrices. Inference in the proposed models is based on posterior simulation, as described in Section 2.3. However, since inference in the proposed models is based on posterior simulation, inference on any other function of the model parameters is possible. Of particular interest is inference about the precision matrix $P = R^{-1}$. This is of interest, for example, in graphical models. Small off-diagonal entries in the precision matrix correspond to small conditional correlation of the respective variables. This

allows inference about the presence and absence of connecting edges in a graphical representation of a multivariate distribution.

I have introduced the models in the context of independent multivariate normal sampling. But the models and proposed posterior simulation schemes are of more general interest and applicability. The proposed models lead to interesting generalizations of standard models in any modeling context which involves (hyper-) prior probability models on variance-covariance or precision matrices. Examples are repeated measurement models, hierarchical models, multivariate probit models, graphical models, or multivariate stochastic volatility models.

Chapter 3

Shadow Prior

3.1 Introduction

Use of Bayesian statistical methods involves specification of sampling distributions, priors and hyper-priors as well known distributions with known normalizing constants. When these distributions are conjugate, the normalizing constant is trivially found by recognizing the kernel of the density function. When they are not conjugate, the techniques used for posterior simulation generally consider the ratio of the density functions evaluated for different arguments, canceling out the normalizing constants.

While the above scenario is very common, sometimes the distributions of interest are truncated over a region that depends on (hyper-) parameters in the next level of the model. This can have severe ramifications in regard to the ability of Markov chain Monte Carlo methods to sample efficiently and effectively from these distributions. In some cases the dimension over which the normalizing constant is defined may be small enough that the researcher can do numerical integration at each step in the MCMC to estimate the constant of interest. Even when the dimension is small these numerical methods can be very computationally expensive.

To illustrate, consider a generic Bayesian model with truncation in the sampling distribution. Assume the likelihood $p(y|\theta)$ includes a constraint to some set \mathcal{A} such that

$$p(y|\theta) \propto q(y|\theta)I\{y \in \mathcal{A}\} = q(y|\theta)/q(\mathcal{A}|\theta)I\{y \in \mathcal{A}\}.$$

The probability $q(\mathcal{A}|\theta)$ that appears as a normalizing constant in the denominator gives rise to computational complications. Depending on the form of \mathcal{A} , $q(\mathcal{A}|\theta)$ is typically analytically intractable. This will complicate posterior updating for θ . The model is completed with a prior $p(\theta)$, typically chosen to be conjugate to the (unconstrained) sampling distribution $q(y|\theta)$. Formally, the problem of posterior inference in the presence of truncated sample spaces rests in the fact that the truncation destroys the conjugate setup. The additional factor $q(\mathcal{A}|\theta)$ in the denominator of the sampling distribution also appears in the posterior distribution. The constraint to the region \mathcal{A} complicates inference on θ because the evaluation of the normalizing constant $q(\mathcal{A}|\theta)$ is often an analytically intractable quantile of $p(y|\theta)$.

The same problem arises when the truncation appears in the prior of a hierarchical model. In particular, assume a hierarchical model with an unconstrained likelihood $p(y|\theta)$, but a prior restricted to a set \mathcal{B} , namely

$$p(\theta|\eta) \propto q(\theta|\eta)I\{\theta \in B\} = q(\theta|\eta)/q(B|\eta)I\{\theta \in B\},$$

with hyper-prior $p(\eta)$. We mitigate computational problems arising from truncated sample and parameter spaces by introducing shadow priors. The idea is to introduce an additional layer in the hierarchical model. We make θ depend on an intermediate, additional parameter δ . Conditional on δ , θ and η are independent. The advantage is that the conditional posterior for θ remains free of the analytically intractable normalization constant. The shadow prior can be introduced in such a way that the truncation does affect the conditional posterior of δ in only minor ways. The idea

is of particular interest when η has a complicated structure. Including the shadow prior by inserting δ between θ and η changes the prior to

$$p(\theta|\delta) \propto q(\theta|\delta)I\{\theta \in B\} = q(\theta|\delta)/q(B|\delta)I\{\theta \in B\}$$

which no longer depends on η . The hyper-priors are $p(\delta|\eta)$ and $p(\eta)$. Now the problem arises in updating η .

The goal of the shadow prior approach is to slightly change the model in such a way that the computational burden is reduced, while at the same time the effect of the change on the parameters is negligible. To illustrate see Figure 3.3, where the parameter estimates for μ and σ^2 in the common correlation model are plotted both with and without the shadow prior.

The model as described above is exactly what appears in Section 2.3 with $\theta = R$, $\eta = (\mu, \sigma^2, \vartheta)$, and $\mathcal{B} = \mathcal{R}^J$. The models used are for correlation matrices where the hyper-parameters include complicated indicators for clustering. The shadow parameter δ_{ij} 's placed between r_{ij} 's and μ did not inherit the same complicated structure as the r_{ij} 's, and consequently removed those constraints from $(\mu, \sigma^2, \vartheta)$. The full conditional distribution for the components of the correlation matrix are given by (2.3), (2.7), and (2.8). These are all restricted to the region of positive definiteness of R (\mathcal{R}^J) which is a very convoluted subspace of the J -dimensional hypercube $[-1, 1]^J$ (see Molenberghs and Rousseeuw (1994) for a general discussion of the shape of this positive definite region). Within the framework of this model the constant depends directly on μ and σ^2 and indirectly on ϑ . An analytic form for this region in high dimensions is sufficiently complex to be branded as intractable. This is a problem for sampling from the full conditionals for the r_{ij} 's, μ , σ^2 , but more so for sampling from the full conditional for ϑ . To efficiently deal with this normalizing constant we continue the discussion of the “shadow prior” introduced in Section 2.3.

3.2 Example: Correlation matrices

To focus the discussion consider again the full conditional density for μ based on the common correlation model,

$$f(\mu|-) \propto C(\mu, \sigma^2) \prod_{i < j} \exp \{ -1/(2\sigma^2)(r_{ij} - \mu)^2 \} \exp \{ -1/(2\tau^2)(\mu)^2 \}. \quad (3.1)$$

As pointed out previously, because μ is hopelessly entangled in the normalizing constant C , without the shadow prior we use a Metropolis-Hastings step to update μ . The proposal density is the normal distribution that results if $I\{R \in \mathcal{R}^J\}$ is removed from

$$f(R|\mu, \sigma^2, \vartheta) = C(\mu, \sigma^2) \prod_{i < j} \exp \{ -1/(2\sigma^2)(r_{ij} - \mu)^2 \} I\{R \in \mathcal{R}^J\}, \quad (3.2)$$

With μ^* denoting the generated proposal. The appropriate acceptance probability is

$$\alpha = \min \{ 1, C(\mu^*, \sigma^2)/C(\mu, \sigma^2) \}, \quad (3.3)$$

where C is given by

$$C^{-1}(\mu, \sigma^2) = \int_{R \in \mathcal{R}^J} \prod_{i < j} \exp \{ -1/(2\sigma^2)(r_{ij} - \mu)^2 \} dr_{ij}, \quad (3.4)$$

To deal with the normalizing constant C we inserted a shadow parameter into the model hierarchy. We introduced latent variables δ_{ij} between r_{ij} and the prior moments (μ, σ^2) by assuming

$$\delta_{ij} \sim N(\mu, \sigma^2) \quad (3.5)$$

and replaced the prior (3.2) by

$$f(R) = C(\delta, \nu^2) \prod_{i < j} \exp \{ -1/(2\nu^2)(r_{ij} - \delta_{ij})^2 \} I\{R \in \mathcal{R}^J\},$$

where C^{-1} is similar to (3.4) with μ replaced by δ_{ij} and σ^2 replaced by ν^2 . The resulting full conditional density for r_{ij} changes only slightly, with (δ_{ij}, ν^2) replacing (μ, σ^2) , but the full conditional density for μ and σ^2 are now conjugate. The nature of the full conditional density for δ_{ij} is similar to the full conditional (2.3) in the original model, requiring a Metropolis-Hastings step to update the δ_{ij} . There are admittedly a few disadvantages to using the shadow prior. Most importantly the size of ν^2 can have a strong impact on the mixing properties of the Markov chain. In this case since the r_{ij} 's and δ_{ij} 's are so closely related the Markov chain may mix very slowly. However, there are several important advantages to using the additional parameters and model structure. First, as the researcher has complete control over the value of ν^2 , it can be set to an arbitrary value. Second, using the shadow prior simplifies the computational burden associated with sampling the indicator variables in the two grouped models. Third, the dependence of the normalizing constant on μ is dispersed between the many δ_{ij} 's, making the impact of the normalizing constant less than it would be without the shadow prior. Further, as ν^2 approaches zero, the ratio in (3.3) approaches one. In practice, it is often reasonable to set ν^2 to a small number and assume $C(\delta^*, \nu^2)/C(\delta, \nu^2) = 1$. Intuitively, by setting ν^2 small enough, the full conditional density of R essentially lies inside the constrained space \mathcal{R}^J , which allows the constrained normalizing constant to be a reasonably good approximation of the unconstrained normalizing constant.

3.3 Simulation study

One concern with setting ν^2 to a small number is that it may effect the mixing properties of the MCMC algorithm with respect to μ and σ . In practice, I have found that setting ν^2 to a small value does not significantly impact the mixing properties of these variables as noted by the autocorrelation and the marginal posterior density

of the parameters.

This is illustrated by the following simulation study. Assuming a common correlation model, I generated an 8×8 correlation matrix and used that correlation matrix to generate 500 multivariate normal observations. I then analyzed this data using the shadow prior version of the common correlation model for a range of different values of ν^2 . For each analysis I calculated the mean of the estimated ratio of normalizing constants. As expected, as ν^2 becomes small the average ratio becomes one and the variance of this ratio approaches zero; see Figure 3.1. This suggests that for small values of ν^2 it is reasonable to assume that $C(\delta^*, \nu^2)/C(\delta, \nu^2) = 1$.

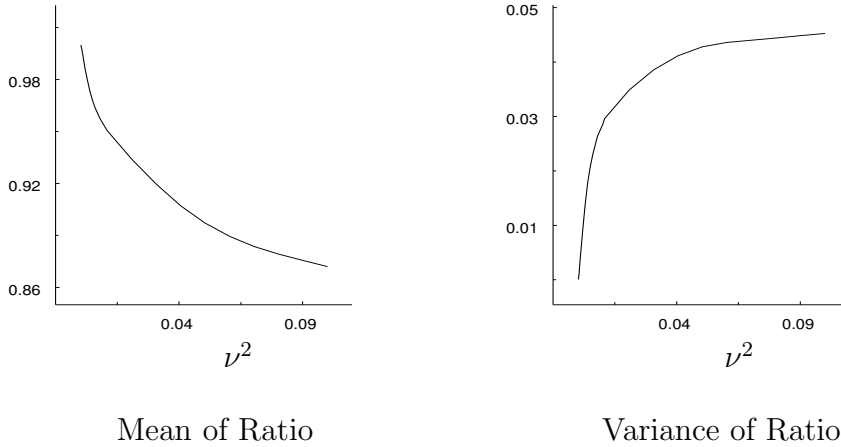


Figure 3.1: Using a common correlation model with 8 variables and varying values for ν^2 , estimates of the ratio of the normalizing constant $C(\cdot, \nu^2)$ are found. As ν^2 becomes small the average ratio becomes one and the variance of this ratio approaches zero. This suggests, that for small values of ν^2 it is reasonable to assume that $C(\delta^*, \nu^2)/C(\delta, \nu^2) = 1$.

With regards to the performance of the MCMC analysis for different values of ν^2 , I found that the mixing properties as summarized by a the auto correlation of μ and σ , and that the posterior inference as summarized by the marginal posterior density estimates of μ and σ , were almost identical across the range of ν^2 considered. To illustrate, consider the auto correlation functions and posterior density estimates

from the analysis when $\nu^2 = 0.1$; see Figures 3.2 and 3.3. In practice this type of analysis could be used as a guide for choosing ν^2 such that it is reasonable to assume that the ratio of normalizing constants equals one and where performance of the MCMC algorithm is not significantly impacted.

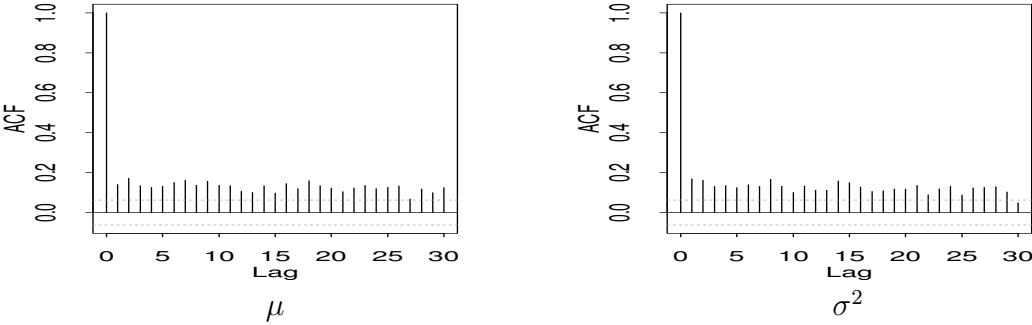


Figure 3.2: Autocorrelation plots for posterior draws of μ and σ^2 using the shadow prior.

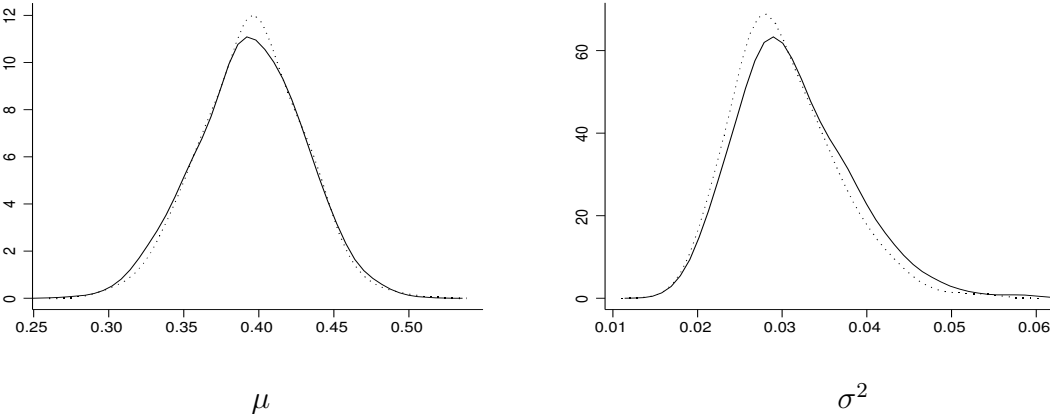


Figure 3.3: Kernel density estimates for posterior samples of μ and σ^2 using a common correlation model with 8 variables both with (\cdots) and without (—) the shadow prior. The resulting posterior distributions are similar enough to conclude that the modification of the model that comes from introducing the shadow prior is negligible.

3.4 Conclusion

Because of awkward constrained parameter spaces that often arise, normalizing constants can be difficult to estimate. We have given a general framework and a specific example for a method to deal with the implications of the parameter constraints. The example is that of the analytically intractable region over which a correlation matrix is positive definite. Integration over that region is essential to calculating the constant, but with the introduction of additional hierarchical parameters this constraint no longer has any bearing on the full conditional distributions of hyperparameters one level removed from the correlations. The shadow prior significantly reduces the computational burden of fitting this particular correlation model.

Chapter 4

Portfolio Selection with Higher Moments

4.1 Introduction

Markowitz (1952) provides the foundation for the current theory of asset allocation. He describes the task of asset allocation as having two stages. The first stage “... starts with observation and experience and ends with beliefs about the future performances of available securities.” The second stage “... starts with the relevant beliefs ... and ends with the selection of a portfolio.” Although Markowitz only deals with the second stage, he suggests that the first stage should be based on a “probabilistic” model. This chapter introduces methodology for addressing both stages.

In a less well known part of Markowitz (1952), he details a condition whereby mean-variance efficient portfolios will not be optimal – when an investor’s utility is a function of more than two moments, *e.g.* mean, variance, and skewness.¹ While Markowitz did not work out the optimal portfolio selection in the presence of skewness and other higher moments, I do.

The breadth of Markowitz’s approach is summarized in the resulting mean-

¹See Markowitz (1952, p.91)

variance efficient frontier. Assuming a ‘certainty equivalence framework’ where point estimates from stage one are given, an efficient frontier can be constructed by solving an appropriate set of quadratic programming problems. Under the assumption that investors have utility that is a function of the point estimates (*i.e.* their expected utility equals the utility of the expected value of the parameters), then the resulting frontier offers a guide for optimal portfolio selection for a large class of utility functions. This approach frees investors from having to explicitly state their utility and the second stage reduces to the task of choosing a point on the efficient frontier, see Kallberg and Ziemba (1983).

Several authors have proposed advances to this approach for selecting an optimal portfolio. Some address the empirical evidence of higher moments; Athayde and Flôres (2002) and Adcock (2002) give methods for determining higher dimensional ‘efficient frontiers’, but they remain in the certainty equivalence framework for selecting an optimal portfolio, addressing only the second stage of the asset allocation task. Like the standard efficient frontier approach, these approaches have the advantage that for a large class of utility functions, the task of selecting an optimal portfolio reduces to the task of selecting a point on the high dimensional ‘efficient frontiers’. As the dimensionality of the efficient frontier increases, it becomes less obvious that an investor can easily interpret the geometry of the frontier and reasonably select a portfolio. In addition, these methods rest on the assumption that an investor’s expected utility is reasonably approximated by the utility of the expected value of the parameters. In the mean-variance setting, a number of researchers have investigated this assumption and have shown that efficient frontier optimal portfolios, based on sample estimates, are highly sensitive to perturbations of these estimates; see Simbea and Mulvey (1998). This ‘estimation risk’ comes from both choosing poor probability models and from ignoring parameter uncer-

tainty, holding with the assumption that the expected utility equals the utility of the expected value.

Others have ignored higher moments, but address the issue of estimation risk. Frost and Savarino (1986, 1988) show that constraining portfolio weights, by restricting the action space during the optimization, reduces estimation error. Using a Bayesian approach, Britten-Jones (2002) builds on the idea of constraining weights and proposes placing informative prior densities directly on the portfolio weights. Others propose methods that address both stages of the allocation task and select a portfolio that optimizes an expected utility function given a probability model. From the classical perspective, Goldfarb and Iyengar (2002) present an optimization method where the optimal portfolio is the best portfolio given the worst case parameter scenario, where the worst case scenario is given by a probability model. From the Bayesian perspective, Jorion (1986) and Frost and Savarino (1986, 1988) use a traditional Bayesian shrinkage approach, Klein and Bawa (1976) emphasizes using a predictive probability model (highlighting that an investor’s utility should be given in terms of future returns and not parameters from a sampling distribution) and Greyserman *et. al.* (2002) consider a hierarchical Bayesian predictive probability model, estimate expected utilities using an ergodic average and select an optimal portfolio using a quadratic programming algorithm. Pástor (2000) and Black and Litterman (1992) propose using asset pricing models to provide informative prior distributions for future returns. In an attempt to maintain the decision simplicity associated with the efficient frontier and still accommodate parameter uncertainty, Michaud (1998) proposes a sampling based method for estimating a ‘resampled efficient frontier’. While this new frontier may offer some insight, using it to select an optimal portfolio implicitly assumes that the investor has abandoned the maximum expected utility framework.

My approach advances previous methods by addressing both higher moments and estimation risk in a consistent Bayesian framework. As part of the “stage one” approach (i.e., incorporating observation and experience), I specify a Bayesian probability model for the joint distribution of the assets, and discuss prior distributions. The Bayesian methodology provides a straight forward way to calculate and maximize expected utilities based on predicted returns. This leads to optimal portfolio weights in the second stage which overcome the problems associated with estimation risk. In addition to discussing the estimation and optimization approach, I empirically investigate the impact of simplifying the asset allocation task. For two illustrative data sets I demonstrate the difference in expected utility that results from ignoring higher moments, using a sampling distribution instead of a predictive distribution and assuming that the expected utility equals the utility of the expected values. In addition, I demonstrate the loss in expected utility that comes from using the widely used approach proposed by Michaud (1998).

Chapter 4 is organized as follows. In Section 4.2 I discuss the importance of higher moments and provide the setting for portfolio selection and Bayesian statistics in finance. I discuss suitable probability models for portfolios and detail the proposed framework. In Section 4.3, I show how to optimize portfolio selection based on utility functions in the face of parameter uncertainty using Bayesian methods. Section 4.4 empirically compares different methods and approaches to portfolio selection. Some concluding remarks are offered in Section 4.5. Appendix *A* contains some additional results and proofs.

4.2 Higher moments and Bayesian probability models

A prerequisite to the use of the Markowitz framework is either that asset returns are normally distributed or that utility is only a function of the first two moments.

The Markowitz framework implies normal distributions for asset returns, but it is well known that financial instruments are not normally distributed. Studying a single asset at a time, empirical evidence suggests that asset returns typically have heavier tails than implied by the normal assumption and are often not even symmetric; see Kon (1984), Mills (1995), Peiro (1999) and Premaratne and Bera (2002).² My investigation of multiple assets builds on these empirical findings and indicates that the existence of ‘co-skewness’, which could be viewed as correlated extremes, is often hidden when assets are considered one at a time. To illustrate, Figure 4.1 contains the kernel density estimate and normal distribution for the marginal daily returns of two stocks (Cisco Systems and General Electric from April 1996 to March 2002) and Figure 4.2 contains a bivariate normal approximation of their joint returns.

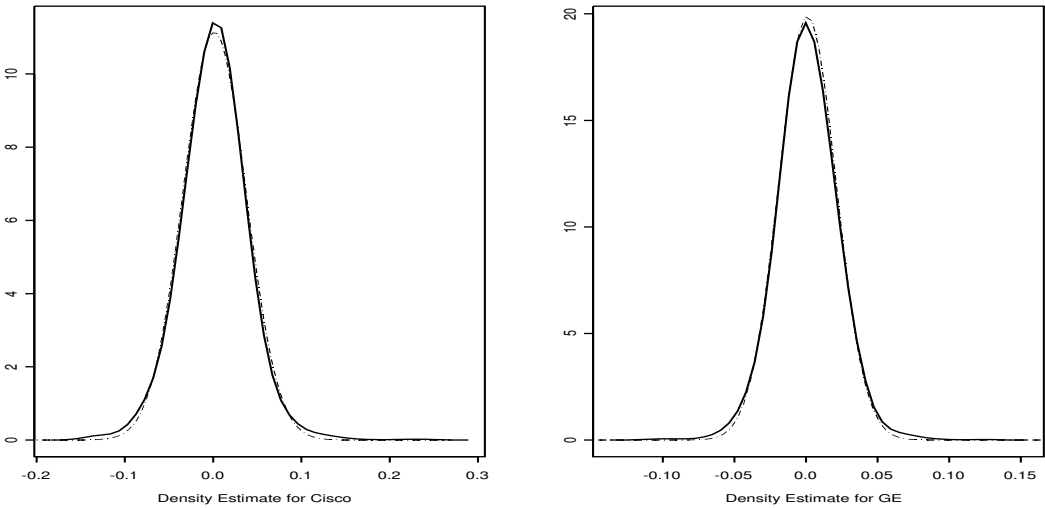


Figure 4.1: This figure contains univariate estimates for Cisco Systems and General Electric daily stock returns from April 1996 to March 2002. The solid lines represent the kernel density estimate, while the dotted lines are the normal density with sample mean and variance. In one dimension the normal distribution closely matches the returns for these two stocks.

²See also Fama (1965), Arditti (1971), Blattberg and Gonedes (1974), Aggarwal *et. al.* (1989), Aggarwal (1990), Sanchez-Torres and Sentana (1998), Hwang and Satchell (1999), Peiro (1999), Hueng and Brooks (2000), Machado-Santos and Fernandes (2002)

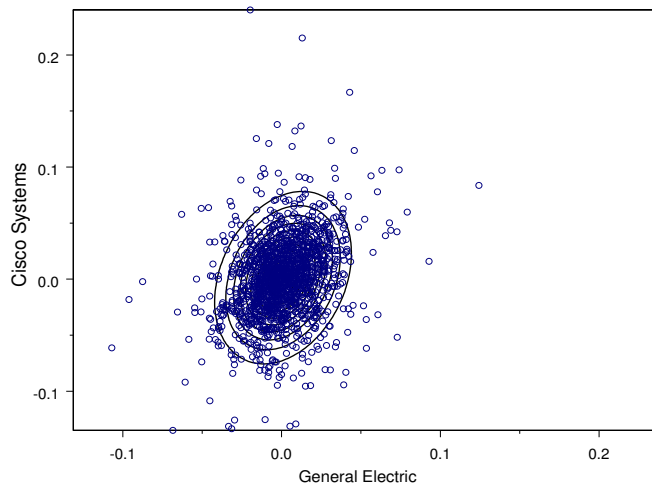


Figure 4.2: This figure contains a bivariate normal estimate for Cisco Systems and General Electric daily stock returns from April 1996 to March 2002. The plot is a bivariate normal with sample mean and covariance. The scatter points are the actual data. In two dimensions the bivariate normal distribution is a poor model to use for these joint returns. The actual returns exhibit co-skewness and much fatter tails than the normal approximation.

While the marginal summaries suggest almost no deviation from the normality assumption, the joint summary clearly exhibits a high degree of co-skewness, suggesting that skewness may have a larger impact on the distribution of a portfolio than previously anticipated.

4.2.1 Economic importance

Markowitz's intuition for maximizing the mean while minimizing the variance of a portfolio comes from the idea that the investor prefers higher expected returns and lower risk. Extending this concept further, most would agree that investors prefer a high probability of an extreme event in the positive direction over a high probability of an extreme event in the negative direction. From a theoretical perspective, Arrow (1971) argues that desirable utility functions should exhibit decreasing absolute

risk aversion, implying that investors should have preference for positively skewed asset returns. Experimental evidence of preference for positively skewed returns has been established by Sortino and Price (1994) and Sortino and Forsey (1996) for example. Levy and Sarnat (1984) find empirical evidence of strong preference for positive skewness in mutual funds. Harvey and Siddique (2000a,b) introduce an asset pricing model that incorporates conditional skewness, and show that an investor may be willing to accept a negative expected return in the presence of high positive skewness.³

An aversion towards negatively skewed returns summarizes the basic intuition that many investors are willing to trade some of their average return for a decreased chance that they will experience a large reduction in their wealth, which could significantly constrain their level of consumption. Some researchers have attempted to address aversion to negative returns in the asset allocation problem by abandoning variance as a measure of risk and defining a ‘downside’ risk that is based only on negative returns (see Markowitz, 1959 and Bawa, 1975). These attempts to separate “good” and “bad” variance can be formalized in a consistent framework by using utility functions and probability models that account for higher moments.

While it is clear that skewness will be important to a large class of investors and shows itself in the returns of the underlying assets (and as a result in the distribution of portfolios), the question remains; how influential is skewness in terms of finding optimal portfolio weights? To illustrate, consider the impact of skewness on the empirical distribution of a collection of two-stock portfolios. For each portfolio, the portfolio mean is identical to the linear combination of the stock means and the

³Also see Arditti (1967), Levy (1969), Jean (1971), Tsiang (1972), Jean (1973), Rubinstein (1973), Arditti and Levy (1975), Kraus and Litzenberger (1976), Friend and Westerfield (1980), Scott and Horvath (1980), Barone-Adesi (1985), Sears and Wei (1985, 1988), Ingersoll (1987), Lim (1989), Lai (1991), Tan (1991), Nummelin (1995), Chundhachinda *et. al.* (1997), Fang and Lai (1997), Faff (1998), Adcock and Shutes (1999), Sornette *et. al.* (2000a,b), Barone-Adesi *et. al.* (2001), Jurczenko and Maillet (2001), Bhattacharyya (2002), Dittmar (2002), Malevergne and Sornette (2002), and Polimenis (2002)

portfolio variance is less than the combination of the stock variances, see Figure 4.3 for an illustration using three, two-stock portfolios. Unlike the variance, there is no guaranteed that the portfolio skewness will be larger or smaller than the linear combination of the stock skewness, and in practice a wide variety of behavior is observed.

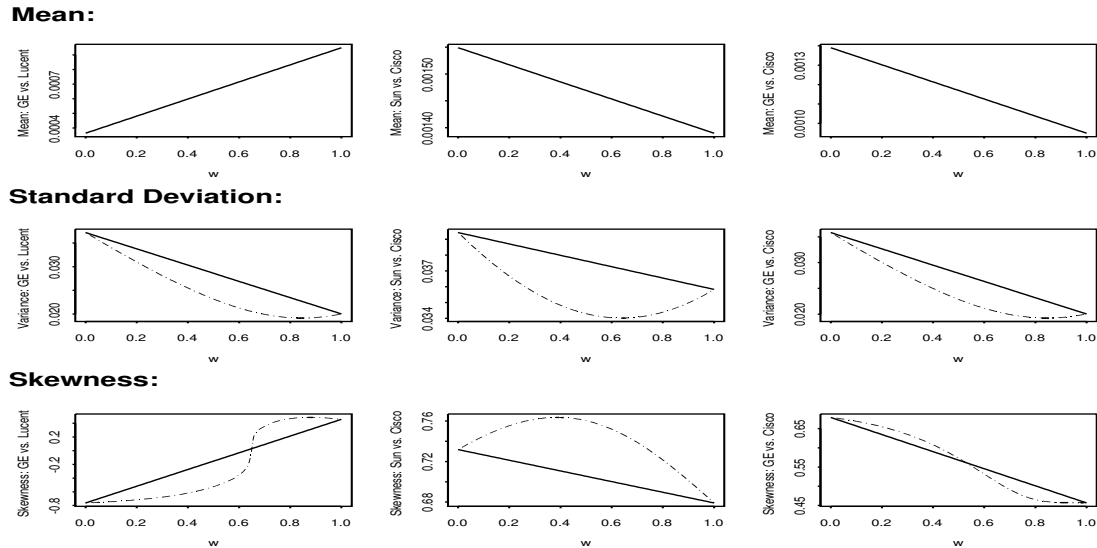


Figure 4.3: This figure contains plots of the mean, variance and skewness of portfolios consisting of two assets. Daily returns from April 1996 to March 2002 for General Electric and Lucent Technologies, Sun Microsystems and Cisco Systems, and General Electric and Cisco Systems are considered. The top row has the mean of the portfolio (equal to the linear combination of the asset means) as the weight of the first asset varies from 0 to 1. The solid line in the plots in the second row represents the linear combination of the variances of the assets, while the dotted line represents the variance of portfolios (variance of linear combination). The variance of the portfolio is always less than or equal to the linear combination of the variances of the underlying assets. The solid line in the third row of plots is the linear combination of the skewness of the two assets in the portfolio, and the dotted line is the skewness of the portfolio. The skewness of the portfolio does not dominate, nor is dominated by the linear combination of the skewness. Picking a portfolio based solely on minimum variance could lead to a portfolio with minimum skewness as well (see GE vs. Cisco).

This variety suggests that the mean-variance optimal criteria can lead to sub-optimal portfolios in the presence of skewness. To accommodate higher order mo-

ments in the asset allocation task, an appropriate probability model must be introduced. After giving an overview of possible approaches, I formally state a model and discuss model choice tools.

4.2.2 Probability models for higher moments

Though it is a simplification of reality, a model can be informative about complicated systems. While the multivariate normal distribution has several attractive properties when it comes to modeling a portfolio, there is considerable evidence that returns are non-normal. There are a number of candidate distributions that include higher moments. The multivariate t-distribution is good for fat tailed data, but does not allow for asymmetry. The non-central multivariate t-distribution also has fat tails and, in addition, is skewed. However the skewness is linked directly to the location parameter and is, therefore, somewhat inflexible.

The log-normal distribution has been used to model gross returns on assets, but its skewness is a function of the mean and variance, not a separate skewness parameter. Azzalini and Dalla Valle (1996) propose a multivariate skew normal distribution that is based on the product of a multivariate normal probability density function (pdf) and univariate normal cumulative distribution function (cdf). This is generalized into a class of multivariate skew elliptical distributions by Branco and Dey (2001), and improved upon by Sahu, Branco and Dey (2002) by using a multivariate cdf instead of univariate cdf, adding more flexibility, which often results in better fitting models. Because of the importance of co-skewness in asset returns, I propose advancing the multivariate skew normal probability model presented in Sahu *et. al.* (2002). An alternative to this model would be an advance of their skew multivariate Student t-model. As the modified skew normal allows for both asymmetry and heavy tails in a way that matches the data considered, I leave

investigation of the skew Student t-model as a point for future research.

The multivariate skew normal can be viewed as the sum of an unrestricted multivariate normal density and a truncated, latent multivariate normal density, or

$$X = \mu + \Delta Z + \epsilon, \tag{4.1}$$

where μ and Δ are an unknown parameter vector and matrix respectively, ϵ is a normally distributed error vector with mean 0 and covariance Σ and Z is a vector of latent random variables. Z comes from a multivariate normal with mean 0 and an identity covariance and is restricted to be strictly positive, or

$$Z = \left(\frac{2}{\pi}\right)^{p/2} \exp\left\{-\frac{1}{2}Z'Z\right\} I\{Z_j > 0\}, \text{ for all } j, \tag{4.2}$$

where $I\{\cdot\}$ is the indicator function and Z_j is the j^{th} element of Z . In Sahu *et. al.* (2002), Δ is restricted to being a diagonal matrix, which accommodates skewness, but does not allow for co-skewness. Removing this restriction and allowing Δ to be any non-singular random matrix results in a modified density and moment generating function, see Appendix A for details.

As with other versions of the skew normal model, this model has the desirable property that marginal distributions of subsets of skew normal variables are skew normal (see Sahu *et. al.* 2002 for a proof). Unlike the multivariate normal density, linear combinations of variables from a multivariate skewed normal density are not skew normal. This does not, however, restrict us from calculating moments of linear combinations with respect to the model parameters, see Appendix A for the formula for the first three moments.

Even though they can be written as the sum of a normal and a truncated normal random variable, neither the skew normal of Azzalini and Dalla Valle (1996) nor Sahu *et. al.* (2002) are Lévy stable distributions.⁴ The skew normal is similar in

⁴Lévy stable distributions are defined several ways. For example, a family of distributions X is

concept to a mixture of normal random variables, however they are fundamentally different. A mixture takes on the value of one of the underlying distributions with some probability and a mixture of Normal random variables results in a Lévy stable distribution. The skew normal is not a mixture of normal distributions, but it is the sum of two normal random variables, one of which is truncated, which results in a distribution that is not Lévy stable. Though it is not stable, the skew normal has several attractive properties. Not only does it accommodate co-skewness and heavy tails, but the marginal distribution of any subset of assets is also skew normal. This is important in the portfolio selection setting because it insures consistency in selecting optimal portfolio weights. For example, with short selling not allowed, if optimal portfolio weights for a set of assets are such that the weight is zero for one of the assets (*i.e.* the asset is not included because it is a poor performer in relation to the others), then removing that asset from the selection process and re-optimizing will not change the portfolio weights for the remaining assets.

Following the Bayesian approach, we assume conjugate prior densities for the unknown parameters, *i.e.* *a priori* normal for μ and $vec(\Delta)$, where $vec(\cdot)$ forms a vector from a matrix by stacking the columns of the matrix, and *a priori* Wishart for Σ^{-1} . The resulting full conditional densities for μ and $vec(\Delta)$ are normal, the full conditional density for Σ^{-1} is Wishart and the full conditional density for the latent Z is a truncated normal. See Appendix A for a complete specification of the prior densities and the full conditional densities. Given these full conditional densities, estimation is done using a Markov chain Monte Carlo (MCMC) algorithm based on the Gibbs sampler and the slice sampler, see Gilks *et. al.* (1998) for a general discussion of the MCMC algorithm and the Gibbs sampler and see Appendix A for

stable if for two independent copies of X , say X_1 and X_2 , the sum $X_1 + X_2$ is also a member of that family. The only stable distribution with finite variance is the normal distribution. It is easy to see that the sum of skew normals is not a skew normal by examination of the moment generating function.

a discussion of the slice sampler.

4.2.3 Model choice

The Bayes factor (BF) is a well developed and frequently used tool for model selection which naturally accounts for both the explanatory power and the complexity of competing models. Several authors have discussed the advantages and appropriate uses of the Bayes factor; see Berger (1985) and O’Hagan (1994). The Bayes Factor is closely related to the Bayesian Information Criteria (BIC), also known as the Schwartz Criteria (see, for example, O’Hagan, 1994). The BIC asymptotically approaches $-2\ln(\text{BF})$ under the assumption that each model is equally likely *a priori*, which is an assumption that we use.

For two competing models (M_1 and M_2), the Bayes factor is:

$$BF = \frac{\text{Posterior odds}}{\text{Prior odds}} = \frac{p(x|M_1)}{p(x|M_2)} \quad (4.3)$$

We use the fourth and final sampling based estimator proposed by Newton and Raftery (1994) to calculate the BF.

4.3 Optimization

Markowitz defined the set of optimal portfolios as the portfolios that are on the efficient frontier. Ignoring uncertainty inherent in the underlying probability model, the portfolio that maximizes expected utility for a large class of utility functions is in this set. This approach reduces the task of choosing an optimal portfolio to choosing the best portfolio from the efficient frontier. When parameter uncertainty is explicitly considered, then the efficient frontier can only be used as a guide for individuals with utility functions that are linear combinations of the moments of the underlying probability model. In all other cases the utility function must

be explicitly stated, and the expected utility must be calculated and maximized over the range of possible portfolio weights. For general probability models and arbitrary utility functions, calculating and optimizing the expected utility must be done numerically, a task that is straightforward to implement using the Bayesian framework.

In practice many of the challenges associated with solving the complete asset allocation problem are ignored leading to a number of simplifications.

4.3.1 Simplifications made in practice

Utility based on model parameters, not predictive returns

The relevant reward for an investor is the realized future return of their portfolio. Thus the utility function, or optimization criterion needs to target future returns and not be a function of model parameters, see Zellner and Chetty (1965), and Brown (1976). It can be argued (DeGroot, 1970; Raiffa and Schlaifer 1961) that a rational decision maker chooses an action by maximizing expected utility, the expectation being with respect to the posterior predictive distribution of the future returns, conditional on all currently available data. Because of computational issues and because the moments of the predictive distribution are approximated by the moments of the posterior distribution, predictive returns are often ignored and utility is frequently stated in terms of the model parameters.

Let ω denote the weights of the desired portfolio. Let x denote the future returns of the possible investments. The future return of portfolio ω is $\omega'x$. The utility function of a rational investor should be a function of $\omega'x$ only. Without loss of generality we write utility as a function of the first three moments. If necessary a third order Taylor series approximation expanded around $(\omega'\bar{m})$ is used:

$$u_{pred}(\omega, x) = \omega'x - \lambda [\omega'(x - \bar{m})]^2 + \gamma [\omega'(x - \bar{m})]^3 \quad (4.4)$$

The only unknown quantities in (4.4) are the future returns x . In particular, it does not include any function of the unknown parameters related to inference loss. We assume that an investor's decision is driven only by future returns. Considering expected utility, we integrate (4.4) with respect to the posterior predictive distribution $p(x|x^o)$. The expected utility becomes,

$$U_{pred}(\omega) = \omega' m_p - \lambda \omega' V_p \omega + \gamma \omega' S_p \omega \otimes \omega, \quad (4.5)$$

where m_p , V_p and S_p are the predictive moments of x . Often a parameter based utility function is used in place of (4.4), or

$$u_{param}(\omega, \theta) = \omega' m - \lambda \omega' V \omega + \gamma \omega' S \omega \otimes \omega \quad (4.6)$$

where $\theta = (m, V, S)^5$ are parameters representing the first three moments of the sampling distribution. Using (4.6) for optimal portfolio selection is equivalent to using (4.5) with posterior moments

$$\bar{m} = E(m|x^o), \bar{V} = E(V|x^o), \text{ and } \bar{S} = E(S|x^o)$$

instead of the posterior predictive moments. To see the error involved in this approximation consider the expected utility

$$U_{param}(\omega) = \omega' \bar{m} - \lambda \omega' \bar{V} \omega + \gamma \omega' \bar{S} \omega \otimes \omega. \quad (4.7)$$

It is straightforward to show that the predictive mean equals the posterior mean and that the predictive variance and skewness equal the posterior variance and skewness plus additional terms, or

$$m_p = \bar{m}, V_p = \bar{V} + Var(m|x^o) \text{ and}$$

$$S_p = \bar{S} + 3E(V \otimes m|x^o) - 3E(V|x^o) \otimes m_p - E[(m - m_p)'(m - m_p) \otimes (m - m_p)|x^o].$$

⁵see Appendix A for formulas for these under the skew normal model.

Substituting this into (4.5) gives an alternative form that is composed of $U_{param}(\omega)$ plus other terms.

$$\begin{aligned}
U_{pred}(\omega) &= \omega' \bar{m} - \lambda \omega' \bar{V} \omega + \gamma \omega' \bar{S} \omega \otimes \omega \\
&- \lambda \omega' Var(m|x^o) \omega + 3 \gamma \omega E(V \otimes m|x^o) \omega \otimes \omega \\
&- 3 \gamma \omega E(V|x^o) \otimes m_p \omega \otimes \omega \\
&- \gamma \omega E[(m - m_p)(m - m_p)' \otimes (m - m_p)'|x^o] \omega \otimes \omega.
\end{aligned}$$

The additional terms in rows 2 – 4 of the expression are missing when we use U_{param} as a proxy for U_{pred} . For linear utility functions, stating utility in terms of the probability model parameters implicitly assumes that the predictive variance and skewness are approximately equal to the posterior variance and skewness, an assumption which often fails in practice. This assumption becomes even more strained for arbitrary, non-linear utility functions.

Maximize something other than expected utility

Given that utility functions can be difficult to integrate, various approximations are often used. The simplest approximation is to use a first order Taylor’s approximation (see Novshek, 1993) about the expected predictive returns, or assume

$$U_{pred}(\omega) = E[u_{pred}(\omega, x)|x^o] \approx u_{pred}(\omega, E[x|x^o]).$$

For linear utility functions this approximation is exact. The Taylor’s approximation removes any parameter uncertainty and leads directly to the certainty equivalence optimization framework. It is easy to see that combining the Taylor’s approximation and the assumption that the posterior moments approximately equal predictive moments leads to a frequently used ‘two-times removed’ approximation of the expected utility of future returns, or

$$U_{pred}(\omega) = E[u_{pred}(\omega, x)|x^o] \approx u_{pred}(\omega, E[x|x^o]) \approx u_{param}(\omega, E[\theta|x^o]).$$

Moving away from the Taylor’s assumption requires either specially chosen probability models and utility functions, which result in analytic functions for expected utility, or require using numerical methods to estimate the expected utility function.

In an attempt to maintain the flexibility of the efficient frontier optimization framework but still accommodate parameter uncertainty, Michaud (1998) proposes an optimization approach that switches the order of integration (averaging) and optimization with respect to the portfolio weights ω . The maximum utility framework optimizes the expected utility; the certainty equivalence framework optimizes the utility of expected future returns (*e.g.* ergodic estimates of predictive moments based on draws from the predictive density). Michaud (1998) proposes creating a ‘resampled frontier’ by repeatedly maximizing the utility for a draw from a probability distribution and then averaging the optimal weights that result from each optimization. While his approach could be viewed in terms of predictive returns, his sampling guidelines are arbitrary and could significantly impact the results.⁶ Given that his main interest is with regards to accounting for parameter uncertainty, we consider a modified algorithm where parameter draws from a posterior density are used in place of the predictive moment summaries. To be explicit, assuming a utility of parameters, the essential steps of his algorithm are as follows. For a family of utility functions $(u_{param,1}, \dots, u_{param,K})$, perform the following steps.

1. For each utility function (*e.g.* $u_{param,k}$), generate n draws from a posterior density $\theta_{i,k} \sim p(\theta|x^o)$.
2. For each $\theta_{i,k}$ find weight $\omega_{i,k}$ that maximizes $u_{param,k}(\omega, \theta_{i,k})$.
3. For each utility function, let $\bar{\omega}_k = 1/n \sum \omega_{i,k}$ define the optimal portfolio.

⁶Repeated estimates of posterior moments are used to mimic parameter uncertainty. Each estimate is based on a number of draws from a multivariate normal density with an empirical mean and covariance. The number of draws for each estimate is equal to the number of data points used to form the empirical moments. Hence the variance of these estimates is arbitrarily determined by the size of the initial data set.

In general, if $\omega_k^* \neq \bar{\omega}_k$, then

$$E[u_{param,k}(\omega_k^*, \theta)|x^o] \neq E[u_{param,k}(\bar{\omega}_k, \theta)|x^o]. \quad (4.8)$$

Further if ω_k^* maximizes $E[u_{param,k}(\omega, \theta)|x^o]$, then

$$E[u_{param,k}(\omega_k^*, \theta)|x^o] \geq E[u_{param,k}(\omega^{**}, \theta)|x^o]$$

for all $\omega^{**} \neq \omega_k^*$. From (4.8), clearly

$$E[u_{param,k}(\omega_k^*, \theta)|x^o] > E[u_{param,k}(\bar{\omega}_k, \theta)|x^o],$$

or $\bar{\omega}_k$ results in a sub-optimal portfolio in terms of expected utility maximization. Stated in practical terms, on average, Michaud's approach 'leaves money on the table'.

Ignore Skewness

Although evidence of skewness and other higher moments in financial data are abundant, it is common for skewness to be ignored entirely in practice. Typically skewness is ignored both in the underlying probability model and in the assumed utility function. In order to illustrate the impact of ignoring skewness, Figure 4.4 shows the empirical summary of the distribution of possible portfolios for four equity securities (Cisco Systems, General Electric, Sun Microsystems, and Lucent Technologies).

The mean-variance summary immediately leads to Markowitz's initial insight, but the relationship between mean, variance and skewness demonstrates that Markowitz's two-moment approach offers no guidance for making effective trade offs between mean, variance and skewness. Using the certainty equivalence framework and a linear utility of the first three empirical moments, or

$$u_{empirical} = \omega' m_e - \lambda \omega' V_e \omega + \gamma \omega' S_e \omega \otimes \omega,$$

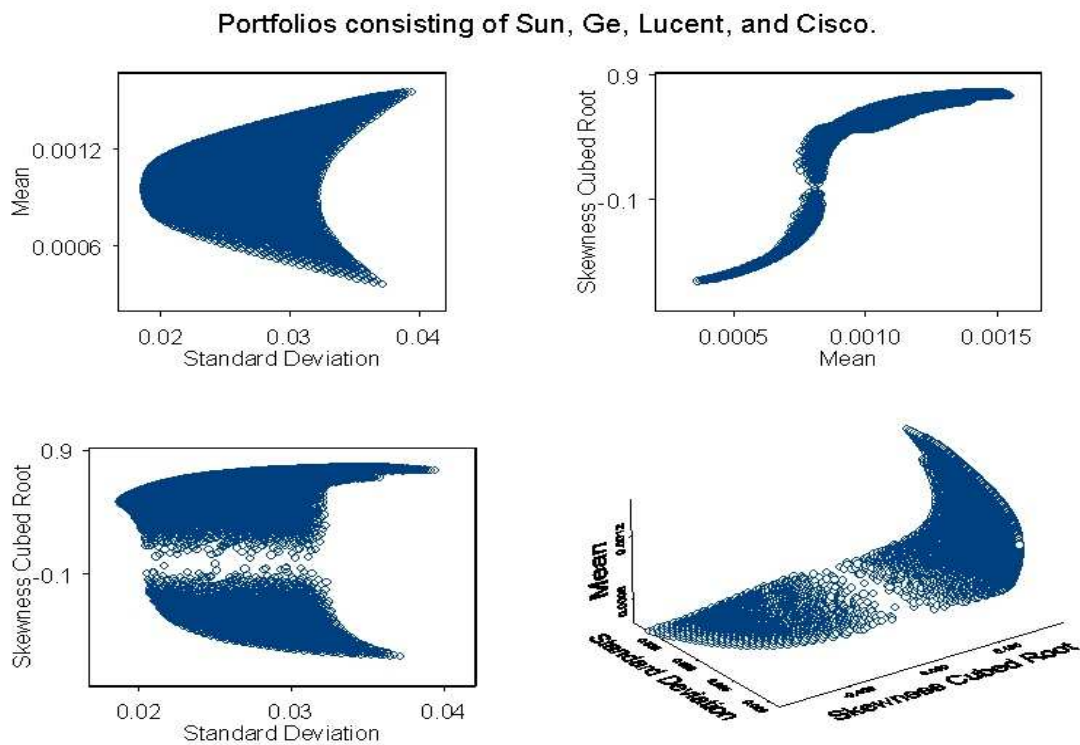


Figure 4.4: This figure shows the space of possible portfolios based on historical parameter estimates from the daily returns of General Electric, Lucent Technologies, Cisco Systems, and Sun Microsystems from April 1996 to March 2002. The top left plot is the mean-standard deviation space, the top right plot is the mean versus the cubed-root of skewness. The bottom left plot is the standard deviation versus the cubed-root of skewness, and the bottom right plot is a three dimensional plot of the mean, standard deviation and cubed-root of skewness. In all plots that contain the skewness there is a sparse region where the skewness is zero.

where m_e, V_e, S_e are the empirical mean, variance and skewness, Figure 4.5 contrasts the optimal portfolios that come from assuming an investor only has an aversion to ‘risk’ ($\lambda = 0.5, \gamma = 0$) and has both an aversion to ‘risk’ and a preference for positive skewness ($\lambda = 0.5, \gamma = 0.5$).

When skewness is considered, the optimal portfolio is pushed further up the efficient frontier signifying that for the same level of risk aversion, an investor can get a higher return if they include skewness in the decision process. In this case, the positive skewness of the portfolio effectively reduces the portfolio risk.

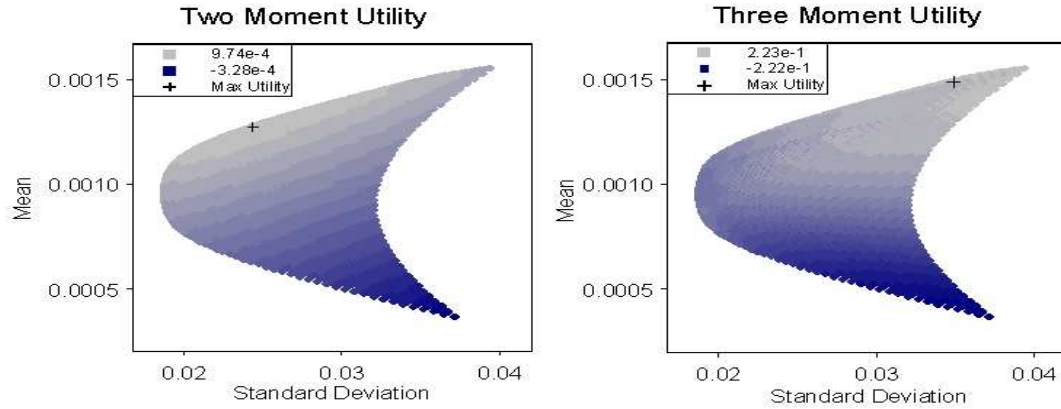


Figure 4.5: This figure shows the mean-variance space of possible portfolios based on historical parameter estimates from the daily returns of General Electric, Lucent Technologies, Cisco Systems, and Sun Microsystems from April 1996 to March 2002. The portfolios are shaded according to the utility associated with each. In the left plot the utility function is $E[u_{pred}(\omega)] = \omega' m_p - 0.5 \omega' V_p \omega$, which is a linear function of the first two moments. The maximum utility is obtained by a portfolio on the frontier and is marked by a '+'. The plot on the right is shaded according to the utility function $E[u_{pred}(\omega)] = \omega' m_p - 0.5 \omega' V_p \omega + 0.5 \omega' S_p \omega \otimes \omega$, which is a linear function of the first three moments. The maximum utility is obtained by a portfolio on the frontier and is marked by a '+'.

4.3.2 Bayesian optimization methods

Bayesian methods offer a natural framework for both estimating and optimizing an arbitrary utility function, given an appropriately complex probability model. In the context of Markov chain Monte Carlo (MCMC) posterior simulation, it is straightforward and computationally easy to generate draws from the posterior predictive density, *i.e.* to generate

$$x_i \sim p(x|x^o)$$

Appropriate Ergodic averages approximate the predictive moments (\bar{m}, V_p, S_p) . Given a set of n draws, the expected utility for an arbitrary utility function can be estimated as an ergodic average, or

$$U(\omega) = E[u(\omega, x)|x^o] \approx \frac{1}{n} \sum u(\omega, x_i).^7$$

The approximate expected utility can then be optimized numerically using a number of different approaches. One attractive algorithm, which is widely used in Bayesian statistics, is the Metropolis-Hastings (MH) algorithm.

Given reasonable regularity conditions for the target (objective) function, the MH algorithm forms an irreducible Markov chain on the parameter space (which would be the action space, or space of allowed portfolios) which converges, in distribution, to the target function, see Gilks *et. al.* (1996) and Meyn and Tweedie (1993) for a discussion of the MH algorithm.⁸ Intuitively, this Markov chain can be viewed as a type of ‘random walk’ with a drift in the direction of larger values of the target function. When the MH algorithm is used as a tool for performing statistical inference, the target density is typically a posterior probability density; however, this need not be the case. As long as the target function is strictly positive and integrable, the MH can be used to numerically explore the extreme values of the target function. Not only has the MH been shown to be very effective for searching high dimensional spaces, its irreducible property ensures that if a global maximum exists the MH algorithm will eventually escape from any local maximum and visit the global maximum.

In order to use the MH function, the expected utility must be strictly positive and integrable. For the linear utility functions, integrability over the space of possible

⁷Using Rao-Blackwellization may allow for more efficient evaluation of $U(\omega)$

⁸Bielza *et. al.* (1999) discuss an alternate optimization scheme where they skip the step of estimating the expected utility and give a one-step procedure, using the MH algorithm, to find optimal actions. We did not pursue this approach as we were interested both in making inference and determining an optimizing action.

portfolios where the portfolios are restricted to the unit simplex (*i.e.* we do not allow short selling) is easily established. We modify the utility function so that it is a strictly positive function by subtracting the minimum expected utility, or the target function becomes

$$\tilde{U}(\omega) = U(\omega) - \min_{\omega} U(\omega).$$

4.4 Optimal portfolios in practice

In theory, simplifications to the complete asset allocation task will result in a sub-optimal portfolio selection. In order to assess the impact that results from some of these simplifications in practice, we consider three different optimization approaches for two data sets using a family of linear utility functions. In particular, we consider the utility functions given in (4.4) and (4.6), which have expected utilities given in terms of the posterior predictive and posterior moments respectively; see (4.5) and (4.7). We consider a number of potential probability models and select the best model. Using results from both the multivariate normal model and the best higher moment model, we numerically determine the optimal portfolio based on the predictive returns, the parameter values and using Michaud’s (1998) non-utility maximization approach. We contrast the performance of each optimal portfolio in terms of expected predictive utility using the best model.

4.4.1 Data description

We consider two sets of returns. The first set comes from four equity securities. The second set comes from a broad-based portfolio of domestic and international equities and fixed income.

First we consider daily returns from April 1996 to March 2002 on four equity securities consisting of General Electric, Lucent Technologies, Cisco Systems, and

Sun Microsystems. These stocks are from the high tech sector, and are chosen to illustrate portfolio selection among closely related assets.

We also try to select securities that match the asset allocation choices facing individuals. To do so we consider the weekly returns from January 1989 to June 2002 on four equity portfolios: Russell 1000 (large capitalization stocks), Russell 2000 (smaller capitalization stocks), Morgan Stanley Capital International (MSCI) EAFE (non-U.S. developed markets), and MSCI EMF (emerging market equities that are available to international investors). We consider three fixed income portfolios: government bonds, corporate bonds, and mortgage backed bonds. Each of these fixed income return series are from Lehman Brothers and form the three major subcomponents of the popular Lehman aggregate index.

4.4.2 Model choice and select summary statistics

To determine which skew normal model best fits the respective data sets, we compute Bayes factors for the multivariate normal model, the skew normal model proposed by Azzalini and Dalla Valle (1996) with a diagonal Δ matrix, and the skewed normal model proposed by Sahu *et. al.* (2002) with both a diagonal and a full Δ matrix. The results for the high tech stocks show that the skewed normal models with a diagonal Δ out perform the other models, with the version of Sahu *et. al.* (2002) doing the best. The skewed normal model with the full Δ , however, performs better than the others in the case of the benchmark indices (see Tables 4.1 and 4.2). The model with the full Δ accommodates co-skewness, which could be viewed as correlated extremes, better than the model with the diagonal Δ . This suggests that portfolios of highly related stocks may have less co-skewness than portfolios of highly diversified global summaries.

The posterior parameter estimates for μ , Σ , and Δ , for both the high tech stocks

Table 4.1: Evaluating the distributional representation of four equity securities: Model choice results for analysis of the daily stock returns of General Electric, Lucent Technologies, Cisco Systems, and Sun Microsystems from April 1996 to March 2002. The four models that are used are the multivariate normal (MV-Normal), the multivariate skew-normal of Azzalini and Dalla Valle (1996) with a diagonal Δ matrix (MVS-Normal D- Δ), and the multivariate skew-normal of Sahu *et. al.* (2002) with both a diagonal and full Δ matrix (MVS-Normal F- Δ). Maximum log likelihood values are used to compute Bayes factors between the multivariate normal model and all of the other models and is reported on the log scale. The model with the highest Bayes factor best fits the data. The two models with diagonal Δ provide the best fit suggesting that there is little co-skewness. Sahu *et. al.* (2002) diagonal Δ model fits best overall.

Model	Log(BF)	Max Log Likelihood
MV-Normal	0.00	-2833.82
MVS-Normal D- Δ , Azzalini (1996)	2225.75	-2561.15
MVS-Normal D- Δ , Sahu <i>et. al.</i> (2002)	2368.33	-2418.85
MVS-Normal F- Δ , Sahu <i>et. al.</i> (2002)	2192.83	-2595.56

and the global asset allocation benchmark indices are given in Tables 4.3 and 4.4. The estimates for Δ for the four equity securities suggest that when considered jointly the skewness is significant, and all but Lucent exhibit positive skewness. For the global asset allocation benchmark indices, there are many positive and negative elements of Δ , though the largest δ 's tend to be negative.

To illustrate the improvement that comes from using the best skewed normal model compared to a multivariate normal model, consider the bivariate skewed normal model of Sahu *et. al.* (2002) with a diagonal Δ for Cisco Systems and General Electric; see Figure 4.6. The skewed normal density accommodates the features of the actual data including correlation and skewness; while the normal density only captures the correlation and significantly understates the tails of the distribution.

Table 4.2: Evaluating the distributional representation of global asset allocation benchmarks: Model choice results for weekly benchmark indices from January 1989 to June 2002 (Lehman Brothers government bonds, LB corporate bonds, and LB mortgage bonds, MSCI EAFE (non-U.S. developed market equity), MSCI EMF (emerging market free investments), Russell 1000 (large cap), and Russell 2000 (small cap)). The four models that are used are the multivariate normal (MV-Normal), the multivariate skew-normal of Azzalini and Dalla Valle (1996) with a diagonal Δ matrix (MVS-Normal D- Δ), and the multivariate skew-normal of Sahu *et. al.* (2002) with both a diagonal and full Δ matrix (MVS-Normal F- Δ). Maximum log likelihood values are used to compute Bayes factors between the multivariate normal model and all of the other models and is reported on the log scale. The model with the highest Bayes factor best fits the data. The Sahu *et. al.* (2002) full Δ model fits the data best.

Model	Log(BF)	Max Log Likelihood
MV-Normal	0.00	-2319.00
MVS-Normal D- Δ , Azzalini (1996)	1946.11	-2096.26
MVS-Normal D- Δ , Sahu <i>et. al.</i> (2002)	2015.11	-2067.73
MVS-Normal F- Δ , Sahu <i>et. al.</i> (2002)	2034.22	-2200.30

4.4.3 Expected utility for competing methods

Optimal weights are calculated for both data sets using the expected predicted utility, the expected parameter utility, and Michaud’s (1998) method. Each method assumes a normal (two moment) probability model and the best skewed normal (higher moment) probability model. For the two moment model we consider two linear utility functions - see (4.5) and (4.7) - one with no risk aversion ($\lambda = 0$) and no preference for skewness ($\gamma = 0$); another with a risk aversion of ($\lambda = 0.5$) and no preference for skewness (Tables 4.5 and 4.6). For the higher moment model, we considered linear utilities with no risk aversion and skewness preference ($\lambda = 0$, $\gamma = 0$), with risk aversion and no skewness preference ($\lambda = 0.5$, $\gamma = 0$), with no risk aversion and skewness preference ($\lambda = 0$, $\gamma = 0.5$) and with both risk aversion and skewness preference ($\lambda = 0.5$, $\gamma = 0.5$) (Tables 4.7 and 4.8). The weights that resulted from each optimization were then used to calculate the expected predictive utility.

Table 4.3: Parameter estimates for diagonal Δ skew normal on four securities: Parameter estimates for the diagonal Δ model of Sahu *et. al.* (2002) used to fit the daily stock returns of General Electric, Lucent Technologies, Cisco Systems, and Sun Microsystems from April 1996 to March 2002. These estimates are the result of a Bayesian Markov chain Monte Carlo iterative sampling routine. These parameters combine to give the mean ($\mu + \sqrt{2/\pi}\Delta\mathbf{1}$), variance ($\Sigma + (1 - 2/\pi)\Delta\Delta'$), and skewness (see Appendix A for formula).

μ	GE	Lucent	Cisco	Sun
	-0.203	1.088	-0.839	-1.214

Σ	GE	Lucent	Cisco	Sun
GE	4.13	1.478	2.331	2.304
Lucent	1.478	15.02	5.095	4.504
Cisco	2.331	5.095	13.868	10.711
Sun	2.304	4.504	10.711	17.485

Δ	GE	Lucent	Cisco	Sun
GE	0.331	0	0	0
Lucent	0	-1.192	0	0
Cisco	0	0	1.069	0
Sun	0	0	0	1.544

Our first observation is that Michaud’s (1998) optimization approach uniformly selects a sub-optimal portfolio even when there is no risk aversion or skewness preference. For example, in the global asset allocation benchmarks indices, the certainty equivalent loss is $(0.001227 - 0.001132) \times 52 = 0.00494$, or roughly 50 basis points per year. The implicit goal of no risk aversion or skewness preference is to select the securities with the best average return; for the high tech data set this is Sun Microsystems and for the benchmark indices this is the Russell 1000 index. For the expected utility approaches, essentially all of the weight is placed on these securities. In Michaud’s approach only 58% of the weight is placed on Sun Microsystems with the rest spread across the remaining stocks, see $\lambda = 0$ and $\gamma = 0$ from Table 4.9. It could be argued that the weights from Michaud’s

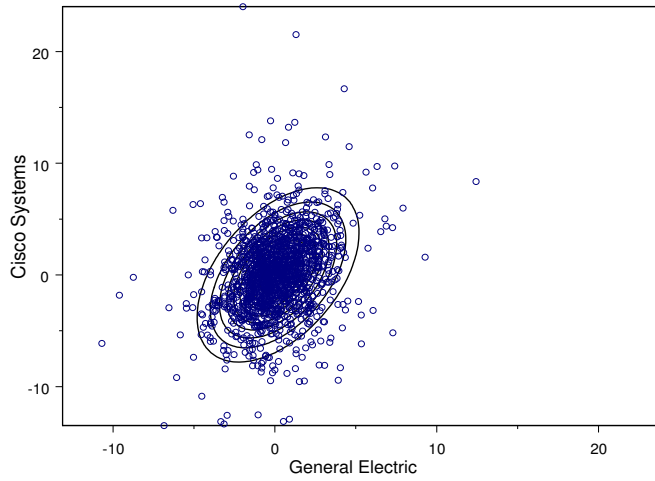


Figure 4.6: This figure contains a bivariate skew normal estimate for Cisco Systems and General Electric daily stock returns from April 1996 to March 2002. The plot is a bivariate skew normal distribution with parameters estimated using the described MCMC algorithm and the model of Sahu *et. al.* (2002) with a diagonal Δ . The scatter points are the actual data. In two dimensions the bivariate skew normal distribution is a better model to use for these joint returns than the normal.

approach should be preferred as they offer diversification and give some sort of protection against volatility. While this may be true, the stated utility function for this optimization ignores volatility (because $\lambda = 0$). Clearly if the investor has an aversion to risk, then the appropriate portfolio would be based on a utility function that explicitly accounts for this aversion. When evaluated in an maximum expected utility framework, Michaud's approach distorts the investor's preference by over diversifying.

Not surprisingly, when there is only preference for skewness, the predictive optimization approach out performs the parameter approach, *e.g.* $\lambda = 0$, $\gamma = 0.5$ in Table 4.6. This difference illustrates the fact that the predictive variance and skewness are only approximated by the estimates of the variance and skewness based on posterior parameter values, see Section 4.3.1. In the case of the high tech stocks,

with $\lambda = 0$ and $\gamma = 0.5$, the parameter optimization approach places almost all of the portfolio weight on Sun Microsystems, but in light of predictive skewness the predictive approach distributes almost all of the weight on the three other stocks, see Table 4.9.

Ignoring the weights from Michaud's approach, the biggest differences in expected predicted utilities comes from comparing optimizations using the normal model versus the skewed normal model; see $\lambda = 0.5$ and $\gamma = 0.5$ in Tables 4.5 and 4.6. For the benchmark indices, the optimal allocation changes markedly when skewness is estimated and included in the utility; see Table 4.10. The most noticeable change is that less weight is placed on the mortgage-backed securities and EMF and more weight is placed on the remaining securities, especially the US bonds and EAFE.

4.5 Conclusion

Considering both higher moments and parameter uncertainty is important in portfolio selection. Up to now these issues have been treated separately. We show that the multivariate normal distribution is an unrealistic probability model for portfolio returns primarily because it does not allow for higher moments, in particular skewness and coskewness. We also demonstrate that the skew normal model of Sahu *et al.* (2002) is a realistic model for portfolio returns. It is flexible enough to allow for skewness and coskewness, and at the same time, accommodates heavy tails. Additional features of the model include straightforward specification of conjugate prior distributions which allows for efficient simulation and posterior inference. Bayesian methods are also used to incorporate parameter uncertainty into the predictive distribution of returns, as well as to maximize the expected utility.

We show that predictive utility can be written in terms of posterior parameter

based utility plus additional terms. These additional terms can be very influential in an investor's utility. We compare results with Michaud's (1998) 'resampling' technique for portfolio selection. We show not only that his averaging of weights produces sub-optimal portfolios, but that his approach is outside the efficient utility maximization framework.

Using the framework described here, I will investigate revealed market preferences and determine whether the 'market' empirically exhibits preference for skewness. As a first step I will use the observed market weights for a benchmark equity index and use the predictive utility function (4.4) to determine the implied market λ and γ . In addition, I intended to consider modifications to (4.4) that allow for asymmetric preferences over positive and negative skewness.

To show that the market does have preference over skewness we will let λ_{param} and λ_{pred} be the coefficients of $U_{param}(\omega)$ and $U_{pred}(\omega)$ respectively (as described in Section 4.3), and let $\omega_{\lambda_{param}}(t)$ and $\omega_{\lambda_{pred}}(t)$ be the weights conditional on the data up to time t . Also let ξ_{param} and ξ_{pred} be the coefficients of $R_{param}(\omega)$ and $R_{pred}(\omega)$, where $R(\omega)$ is the two moment analog of $U(\omega)$ based on the normal distribution. Let $\omega^*(t)$ be the observed market weights at time t . The utility structure implied by the revealed preferences is described by the coefficients λ_{param}^* , λ_{pred}^* , ξ_{param}^* , and ξ_{pred}^* that best explains the market weights ω^* where each is found in a similar manner to

$$\lambda_{param}^* = \arg \min_{\lambda_{param}} \sum_t (\omega^*(t) - \omega_{\lambda_{param}}(t))^2.$$

and

$$\tilde{\lambda}^* = \arg \min_{\tilde{\lambda}} \sum_t (\omega^*(t) - \omega_{\tilde{\lambda}}(t))^2.$$

We use the sum of squared residuals $L_{U_{param}} = \sum_t (\omega^*(t) - \omega_{\lambda_{param}^*}(t))^2$, $L_{U_{pred}} = \sum_t (\omega^*(t) - \omega_{\lambda_{pred}^*}(t))^2$, $L_{R_{param}} = \sum_t (\omega^*(t) - \omega_{\xi_{param}^*}(t))^2$, and $L_{R_{pred}} = \sum_t (\omega^*(t) -$

$\omega_{\xi_{pred}^*}(t)^2$ as diagnostics of fit for the utility function. I hope to show that $L_{U_{pred}} < L_{R_{pred}}$, also noting how utilities based on parameter values do in comparison, *i.e.* $L_{U_{param}}$ and $L_{R_{pred}}$.

Table 4.4: Parameter estimates for full Δ skew normal on global asset allocation benchmark: Parameter estimates for full Δ model of Sahu *et. al.* (2002) used to fit the weekly benchmark indices Lehman Brothers government bonds, LB corporate bonds, and LB mortgage bonds, MSCI EAFE (non-U.S. developed market equity), MSCI EMF (emerging market free investments), Russell 1000 (large cap), and Russell 2000 (small cap) from January 1989 to June 2002. These estimates are the result of a Bayesian Markov chain Monte Carlo iterative sampling routine. These parameters combine to give the mean ($\mu + (2/\pi)^{1/2}\Delta\mathbf{1}$), variance ($\Sigma + (1 - 2/\pi)\Delta\Delta'$), and skewness (see Appendix A for formula).

μ	GB	CB	MBS	EAFE	EMF	R1000	R2000
	0.162	0.385	0.541	0.459	0.095	0.499	0.889

Σ	GB	CB	MBS	EAFE	EMF	R1000	R2000
GB	2.897	3.077	2.156	-0.012	-0.249	0.277	0.01
CB	3.077	3.734	2.405	0.25	0.076	0.549	0.383
MBS	2.156	2.404	2.163	0.269	0.021	0.459	0.403
EAFE	-0.012	0.25	0.269	5.16	2.89	2.574	2.694
EMF	-0.249	0.077	0.021	2.89	6.687	2.382	2.85
R1000	0.277	0.549	0.459	2.574	2.382	4.717	4.359
R2000	0.01	0.383	0.403	2.694	2.85	4.359	6.293

Δ	GB	CB	MBS	EAFE	EMF	R1000	R2000
GB	-0.049	0.227	-0.546	-0.119	-0.1	0.424	-0.1
CB	0.049	0.194	-0.583	-0.167	-0.119	0.484	-0.168
MBS	-0.138	0.161	-0.607	-0.135	-0.091	0.347	-0.045
EAFE	0.33	-0.238	-0.002	-0.11	-0.286	0.168	-0.34
EMF	0.041	0.336	0.111	0.142	-1.037	0.144	0.186
R1000	0.141	-0.015	0.108	0.122	-0.083	-0.451	-0.125
R2000	0.072	-0.083	-0.047	0.097	-0.021	-0.2790	-0.567

Table 4.5: Two moment optimization for four equity securities: This table contains predictive utilities for the weights that maximize utility as a linear function of the two moments of the multivariate normal model by three different methods for daily stock returns of General Electric, Lucent Technologies, Cisco Systems, and Sun Microsystems from April 1996 to March 2002. The first method is based on predictive or future values of the portfolio (results in $\omega_{2,pred}$ where the 2 represents the number of moments in the model), the second is based on the posterior parameter estimates ($\omega_{2,param}$), and the third is the method proposed by Michaud ($\omega_{2,Michaud}$). The weights that are found by each method are ranked by the three moment predictive utility they produce (*i.e* $E[u_{3,pred}(\omega)] = \omega' m_p - \lambda \omega' V_p \omega + \gamma \omega' S_p \omega \otimes \omega$, where the 3 signifies that the utility function is linear in the first three moments of the skew normal model, and m_p , V_p , and S_p are the predictive mean, variance and skewness) for varying values of λ and γ . The highest utility obtained signifies the method that is best for portfolio selection according to the investor's preferences. For each combination of λ and γ , $\omega_{2,pred}$ gives the highest expected utility.

λ	γ	$E[u_{3,pred}(\omega_{2,pred})]$	$E[u_{3,pred}(\omega_{2,param})]$	$E[u_{3,pred}(\omega_{2,Michaud})]$
0	0	0.122769	0.122747	0.113216
0	0.5	0.0978264	0.0974689	0.0104864
0.5	0	-1.74474	-1.75791	-1.75648
0.5	0.5	-1.73273	-1.74919	-1.74986

Table 4.6: Three moment optimization for four equity securities: Predictive utilities for the weights that maximize utility as a linear function of the three moments of the multivariate skew normal model by three different methods for daily stock returns of General Electric, Lucent Technologies, Cisco Systems, and Sun Microsystems from April 1996 to March 2002. The first method is based on predictive or future values of the portfolio (results in $\omega_{3,pred}$ where the 3 represents the number of moments in the model), the second is based on the posterior parameter estimates ($\omega_{3,param}$), and the third is the method proposed by Michaud ($\omega_{3,Michaud}$). The weights that are found by each method are ranked by the three moment predictive utility they produce (*i.e* $E[u_{3,pred}(\omega)] = \omega' m_p - \lambda \omega' V_p \omega + \gamma \omega' S_p \omega \otimes \omega$, where the 3 signifies that the utility function is linear in the first three moments of the skew normal model, and m_p , V_p , and S_p are the predictive mean, variance and skewness) for varying values of λ and γ . The highest utility obtained signifies the method that is best for portfolio selection according to the investor's preferences. For each combination of λ and γ , $\omega_{3,pred}$ gives the highest expected utility.

λ	γ	$E[u_{3,pred}(\omega_{3,pred})]$	$E[u_{3,pred}(\omega_{3,param})]$	$E[u_{3,pred}(\omega_{3,Michaud})]$
0	0	0.122735	0.122533	0.111573
0	0.5	0.109008	0.0976164	0.0747268
0.5	0	-1.74451	-1.74472	-1.80207
0.5	0.5	-1.73137	-1.73231	-1.73574

Table 4.7: Two moment optimization for global asset allocation benchmark indices: Predictive utilities for the weights that maximize utility as a linear function of the two moments of the multivariate normal model by three different methods for weekly benchmark indices Lehman Brothers government bonds, LB corporate bonds, and LB mortgage bonds, MSCI EAFE (non-U.S. developed market equity), MSCI EMF (emerging market free investments), Russell 1000 (large cap), and Russell 2000 (small cap) from January 1989 to June 2002. The first method is based on predictive or future values of the portfolio (results in $\omega_{2,pred}$ where the 2 represents the number of moments in the model), the second is based on the posterior parameter estimates ($\omega_{2,param}$), and the third is the method proposed by Michaud ($\omega_{2,Michaud}$). The weights that are found by each method are ranked by the three moment predictive utility they produce (*i.e.* $E[u_{3,pred}(\omega)] = \omega' m_p - \lambda \omega' V_p \omega + \gamma \omega' S_p \omega \otimes \omega$, where the 3 signifies that the utility function is linear in the first three moments of the skew normal, and m_p , V_p , and S_p are the predictive mean, variance and skewness) for varying values of λ and γ . The highest utility obtained signifies the method that is best for portfolio selection according to the investor's preferences. For each combination of λ and γ , $\omega_{3,pred}$ gives the highest expected utility.

λ	γ	$E[u_{3,pred}(\omega_{2,pred})]$	$E[u_{3,pred}(\omega_{2,param})]$	$E[u_{3,pred}(\omega_{2,Michaud})]$
0	0	0.254183	0.250803	0.2384
0	0.5	0.223508	0.22125	0.211609
0.5	0	-0.647172	-0.662493	-0.673768
0.5	0.5	-0.649377	-0.666131	-0.675885

Table 4.8: Three moment optimization for global asset allocation benchmark indices: Predictive utilities for the weights that maximize utility as a linear function of the three moments of the multivariate skew normal model by three different methods for weekly benchmark indices Lehman Brothers government bonds, LB corporate bonds, and LB mortgage bonds, MSCI EAFE (non-U.S. developed market equity), MSCI EMF (emerging market free investments), Russell 1000 (large cap), and Russell 2000 (small cap) from January 1989 to June 2002. The first method is based on predictive or future values of the portfolio (results in $\omega_{3,pred}$ where the 3 represents the number of moments in the model), the second is based on the posterior parameter estimates ($\omega_{3,param}$), and the third is the method proposed by Michaud ($\omega_{3,Michaud}$). The weights that are found by each method are ranked by the three moment predictive utility they produce (*i.e.* $E[u_{3,pred}(\omega)] = \omega' m_p - \lambda \omega' V_p \omega + \gamma \omega' S_p \omega \otimes \omega$, where the 3 signifies that the utility function is linear in the first three moments of the skew normal, and m_p , V_p , and S_p are the predictive mean, variance and skewness) for varying values of λ and γ . The highest utility obtained signifies the method that is best for portfolio selection according to the investor's preferences. For all combinations of λ and γ except one ($\lambda = 0$, $\gamma = 0.5$), $\omega_{2,pred}$ gives the highest expected utility.

λ	γ	$E[u_{3,pred}(\omega_{3,pred})]$	$E[u_{3,pred}(\omega_{3,param})]$	$E[u_{3,pred}(\omega_{3,Michaud})]$
0	0	0.253829	0.251662	0.239639
0	0.5	0.222629	0.223304	0.210897
0.5	0	-0.646989	-0.647157	-0.650585
0.5	0.5	-0.648895	-0.64923	-0.656476

Table 4.9: Portfolio weights: Four equity securities: Three moment (skew normal) utility based portfolio weights for daily stock returns of General Electric, Lucent Technologies, Cisco Systems, and Sun Microsystems from April 1996 to March 2002. The weights maximize the expected utility function $E[u_{3,pred}(\omega)] = \omega' m_p - \lambda \omega' V_p \omega + \gamma \omega' S_p \omega \otimes \omega$, (where the 3 signifies that the utility function is linear in the first three moments, and m_p , V_p , and S_p are the predictive mean, variance and skewness) for varying values of λ and γ . Three different methods of maximization are used. The first is based on predictive or future values of the portfolio (results in $\omega_{3,pred}$ where the 3 represents the number of moments in the model), the second is based on the posterior parameter estimates ($\omega_{3,param}$), and the third is the method proposed by Michaud ($\omega_{3,Michaud}$).

$\lambda = 0, \gamma = 0$	GE	Lucent	Cisco	Sun
$\omega_{3,pred}$	8.8e-4	1.21e-4	4.48e-4	0.9985
$\omega_{3,param}$	1.71e-3	6.35e-4	0.0112	0.986
$\omega_{3,Michaud}$	0.11	0.0764	0.225	0.588
$\lambda = 0, \gamma = 0.5$	GE	Lucent	Cisco	Sun
$\omega_{3,pred}$	0.405	0.409	0.186	2.15e-5
$\omega_{3,param}$	3.52e-4	1.0e-4	6.03e-4	0.9989
$\omega_{3,Michaud}$	0.0285	0.00297	0.0453	0.923
$\lambda = 0.5, \gamma = 0$	GE	Lucent	Cisco	Sun
$\omega_{3,pred}$	0.784	0.125	0.0535	0.0368
$\omega_{3,param}$	0.785	0.129	0.0557	0.0304
$\omega_{3,Michaud}$	0.677	0.15	0.0901	0.0807
$\lambda = 0.5, \gamma = 0.5$	GE	Lucent	Cisco	Sun
$\omega_{3,pred}$	0.785	0.129	0.0646	0.0214
$\omega_{3,param}$	0.787	0.125	0.0532	0.0356
$\omega_{3,Michaud}$	0.773	0.155	0.0565	0.0165

Table 4.10: Portfolio weights: Global asset allocation benchmark indices: Two moment (normal) utility based portfolio weights for weekly benchmark indices Lehman Brothers government bonds, LB corporate bonds, and LB mortgage bonds, MSCI EAFE (non-U.S. developed market equity), MSCI EMF (emerging market free investments), Russell 1000 (large cap), and Russell 2000 (small cap), from January 1989 to June 2002. The weights maximize the expected utility function $E[u_{2,pred}(\omega)] = \omega' m_p - \lambda \omega' V_p \omega$, (where the 2 signifies that the utility function is linear in the two moments of the normal model, and m_p and V_p are the predictive mean and variance) for varying values of λ and γ . Three different methods of maximization are used. The first is based on predictive or future values of the portfolio (results in $\omega_{2,pred}$ where the 2 represents the number of moments in the model), the second is based on the posterior parameter estimates ($\omega_{2,param}$), and the third is the method proposed by Michaud ($\omega_{2,Michaud}$).

$\lambda = 0.5$	GB	CB	MBS	EAFE	EMF	R1000	R2000
$\omega_{2,pred}$	0.12	2.4e-5	0.497	0.1	0.0984	0.183	6.34e-4
$\omega_{2,param}$	1.48e-3	2.32e-3	0.632	0.1	0.0777	0.111	0.075
$\omega_{2,Michaud}$	0.157	9.72e-3	0.461	0.116	0.0421	0.0812	0.134
$\lambda = 0.5, \gamma = 0.5$	GB	CB	MBS	EAFE	EMF	R1000	R2000
$\omega_{3,pred}$	0.157	2.34e-4	0.439	0.121	0.0874	0.193	2.77e-3
$\omega_{3,param}$	0.149	2.11e-4	0.461	0.0967	0.0905	0.18	0.0233
$\omega_{3,Michaud}$	0.225	5.2e-4	0.407	0.0885	0.0808	0.15	0.0479

Chapter 5

Conclusions

I explored applications and estimation of second and third moments. I have proposed new mixture model priors for correlation matrices. The main arguments for the proposed models are the increased flexibility in representing substantive prior knowledge about the dependence structure, and less dependence on strict parametric assumptions implied in some of the standard models reviewed in the introduction. The main limitation of the proposed models is the computation intensive implementation. We have partially mitigated this problem by proposing appropriate Markov chain Monte Carlo strategies, including the shadow prior mechanism.

The proposed models provide a framework to represent and learn about dependence structures. An alternative approach is traditional factor analysis. Factor analysis explains dependence among a set of variables as arising from a few latent factors. The relative size of the factor loadings determines the strength of the correlations. In contrast, the grouped variables and grouped correlations models proceed by assuming a partition of the set of correlations and variables, respectively. The extent of the relationship between these two modeling approaches is an interesting topic for future research.

The discussion was focused on inference on correlation matrices with inference

in the proposed models based on posterior simulation, as described in Section 2.3. Since inference in the proposed models is based on posterior simulation, inference on any other function of the model parameters is possible. Of particular interest is inference about the precision matrix $P = R^{-1}$. This is of interest, for example, in graphical models. Small off-diagonal entries in the precision matrix correspond to small conditional correlation of the respective variables. This allows inference about the presence and absence of connecting edges in a graphical representation of a multivariate distribution.

I have introduced the models in the context of independent multivariate normal sampling. But the models and proposed posterior simulation schemes are of more general interest and applicability. The proposed models lead to interesting generalizations of standard models in any modeling context which involves (hyper-) prior probability models on variance-covariance or precision matrices. Examples are repeated measurement models, hierarchical models, multivariate probit models, graphical models, or multivariate stochastic volatility models.

Computational issues for modeling correlation matrices are non-trivial. Because of the positive definite constraint, normalizing constants for any distribution of correlation matrices are analytically intractable. The easiest way to deal with this normalizing constant is to ignore the positive definite constraint. This may be unreasonable in some cases, however we propose a computational super trick that allows us to do just that. The shadow prior introduces an additional layer of hierarchy that effectively and efficiently side steps the issue.

Considering both higher moments and parameter uncertainty is important in portfolio selection. Up to now these issues have been treated separately. I show that the multivariate normal distribution is an unrealistic probability model for portfolio returns primarily because it does not allow for higher moments, in partic-

ular skewness and coskewness. I also demonstrate that the skew normal model of Sahu *et. al.* (2002) is a realistic model for portfolio returns. It is flexible enough to allow for skewness and coskewness, and at the same time, accommodates heavy tails. Additional features of the model include straightforward specification of conjugate prior distributions which allows for efficient simulation and posterior inference. Bayesian methods are also used to incorporate parameter uncertainty into the predictive distribution of returns, as well as to maximize the expected utility.

I show that predictive utility can be written in terms of posterior parameter based utility plus additional terms. These additional terms can be very influential in an investor's utility. I compare results with Michaud's (1998) 'resampling' technique for portfolio selection. I show not only that his averaging of weights produces sub-optimal portfolios, but that his approach is outside the efficient utility maximization framework.

Using the framework we describe, I intend to investigate revealed market preferences and determine whether the 'market' empirically exhibits preference for skewness. As a first step I will use the observed market weights for a benchmark equity index and use the predictive utility function (4.4) to determine the implied market λ and γ . In addition, I intended to consider modifications to (4.4) that allow for asymmetric preferences over positive and negative skewness.

Appendix A

Skew Normal Probability Model

A.1 Density and moment generating function

When Δ is a full matrix, the likelihood function and moment generating function given in Sahu *et. al.* changes. Their equation (9) becomes

$$f(y|\mu, \Sigma, \Delta) = 2^m |\Sigma + \Delta\Delta'|^{-\frac{1}{2}} \phi_m \left[(\Sigma + \Delta\Delta')^{-\frac{1}{2}} (y - \mu) \right] \times \\ \Phi_m \left[(I - \Delta'(\Sigma + \Delta\Delta')^{-1}\Delta)^{-\frac{1}{2}} \Delta'(\Sigma + \Delta\Delta')^{-1} (y - \mu) \right]. \quad (\text{A.1})$$

Where ϕ_m is the m -dimensional multivariate normal density function with mean zero and identity covariance, and Φ_m is multivariate normal cumulative distribution also with mean zero and identity covariance. The moment generating function becomes

$$M_{\mathbf{Y}}(\mathbf{t}) = 2^m e^{t'\mu + t'(\Sigma + \Delta\Delta')t/2} \Phi_m(\Delta t). \quad (\text{A.2})$$

The first three moments of the distribution (m , V , and S) can be written in terms of μ , Σ and Δ as follows,

$$m = \mu + (2/\pi)^{1/2} \Delta \mathbf{1}, \quad V = \Sigma + (1 - 2/\pi) \Delta \Delta', \quad \text{and}$$

$$\begin{aligned}
S = & \Delta E Z \Delta' \otimes \Delta' + 3\mu' \otimes \{\Delta \Delta'(1 - 2/\pi) + 2/\pi \Delta \mathbf{1}(\Delta \mathbf{1})'\} + \\
& 3\{(2/\pi)^{1/2}(\Delta \mathbf{1})' \otimes [\Sigma + \mu\mu']\} + 3\mu' \otimes \Sigma \\
& + \mu\mu' \otimes \mu' - 3m' \otimes V - m m' \otimes m', \quad (\text{A.3})
\end{aligned}$$

where $\mathbf{1}$ is a column vector of ones, and EZ is the $m \times m^2$ super matrix made up of the moments of a truncated standard normal distribution.

$$\mathbf{EZ} = \begin{pmatrix} E[Z_1 Z_1 Z_1] & \dots & E[Z_1 Z_1 Z_m] & \dots & E[Z_m Z_1 Z_1] & \dots & E[Z_m Z_1 Z_m] \\ \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots \\ E[Z_1 Z_m Z_1] & \dots & E[Z_1 Z_1 Z_m] & \dots & E[Z_m Z_m Z_1] & \dots & E[Z_m Z_m Z_m] \end{pmatrix}$$

Where $E[Z_i] = \sqrt{2/\pi}$, $E[Z_i^2] = 1$, and $E[Z_i^3] = \sqrt{8/\pi}$. Since the Z_i 's are independent, $E[Z_i^2 Z_j] = E[Z_i^2]E[Z_j] = \sqrt{2/\pi}$, and $E[Z_i Z_j Z_k] = E[Z_i]E[Z_j]E[Z_k] = (2/\pi)^{3/2}$ for any i, j , and k .

A.2 First three moments of linear combination

Assume $X \sim SN(\mu, \Sigma, \Delta)$ and a set of constants $\omega = (\omega_1, \dots, \omega_m)'$, the first three moments of $\omega' X$ are as follows

$$E(\omega' X) = \omega' m$$

$$\text{Var}(\omega' X) = \omega' V \omega$$

$$\text{Skew}(\omega' X) = \omega' S \omega \otimes \omega,$$

where m , V and S are given above.

A.3 Model specification

A.3.1 Likelihood and priors

The skew normal density is defined in terms of a latent (unobserved) random variable z , which comes from a truncated standard normal density. The likelihood is given by

$$X_i|Z_i, \mu, \Sigma, \Delta \sim N_m(\mu + \Delta Z_i, \Sigma),$$

where N_m is a multivariate normal density,

$$Z_i \sim N_m(0, I_m)I\{Z_{ij} > 0\}, \text{ for all } j,$$

and I_m is an m dimensional identity matrix. In all cases we use conjugate prior densities, with hyper-parameters that reflect vague prior information, or *a priori* assume

$$\begin{aligned}\beta &\sim N_{m(m+1)}(0, 100I_{m(m+1)}) \\ \Sigma &\sim \text{Inverse-Wishart}(m, mI_m),\end{aligned}$$

where $\beta' = (\mu', \text{vec}(\Delta)')$ and $\text{vec}(\cdot)$ forms a vector by stacking the columns of a matrix.

A.3.2 Full conditionals

Assuming n independent skew normal observations, the full conditional distributions are as follows:

$$\begin{aligned}Z_i|- &\sim N_m(A^{-1}a_i, A^{-1})I\{Z_{ij} > 0\}, \text{ for all } j, \\ \beta|- &\sim N_{m(m+1)}(B^{-1}b, B^{-1}) \\ \Sigma|- &\sim \text{Inverse-Wishart}(m + n, C),\end{aligned}$$

where

$$A = I_m + \Delta' \Sigma^{-1} \Delta \text{ and } a_i = \Delta' \Sigma^{-1} (x_i - \mu),$$

$$B = \sum_{i=1}^n y_i \Sigma^{-1} x_i + \frac{1}{100} I_{m(m+1)} \text{ and } b = \sum_{i=1}^n y_i \Sigma^{-1} (x_i),$$

$$C = \sum_{i=1}^n (x_i - (\mu - \Delta Z_i))(x_i - (\mu - \Delta Z_i))' + m I_m,$$

where $y_i = (I_m, Z_i' \otimes I_m)$.

A.4 Estimation using the slice sampler

The slice sampler introduces an auxiliary variable, which we will call u , in such a way that the draws from both the desired variable and the auxiliary variable can be obtained by drawing from appropriate uniform densities, for more details see Damien *et. al.* (1999). To illustrate, assume that we want to sample from the following density,

$$f(x) \propto \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\} I \{x \geq 0\}, \quad (\text{A.4})$$

where $I\{\cdot\}$ is an indicator function. We proceed by introducing an auxiliary variable u and form the following joint density,

$$f(x, u) \propto I \left\{ u \leq \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\} I \{x \geq 0\} \right\}. \quad (\text{A.5})$$

It is easy to see that based on (A.5), the marginal density of x is given by (A.4) and that the conditional density of u given x is a uniform density, or

$$f(u|x) \propto I \left\{ u \leq \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\} \right\}.$$

With a little more work, it is straightforward to see that the conditional density of x given u is also uniform, or

$$f(x|u) \propto I \left\{ \max \left(0, \mu - \sqrt{-2\sigma^2 \log(u)} \right) \leq x \leq \mu + \sqrt{-2\sigma^2 \log(u)} \right\}.$$

Samples from x can then be easily obtained by iteratively sampling from u conditional on x and then from x conditional on u .

Appendix B

Symbols

y_i - observations

Σ - variance-covariance matrix

$N(\cdot, \cdot)$ - normal random variable

n - the number of observations

J - the number of variables

$p_J(\Sigma) = 1/|\Sigma|^{(J+1)/2}$ - Jeffrey's prior for covariance matrix

$p_R(\Sigma) \propto 1/\{|\Sigma| \prod_{i < j} (d_i - d_j)\}$ - reference prior for covariance matrix

d_i - eigenvalues of Σ

$\Sigma = SRS$ - separation strategy

S - standard deviation matrix

R - correlation matrix

r_{ij} - correlation between variables i and j

$[-1, 1]$ - closed interval between -1 and 1

$P = \Sigma^{-1}$ - precision matrix

\mathcal{R}^J - space of all positive semi-definite correlation matrices

μ - mean of correlations

σ^2 - variance of correlations

$C(\mu, \sigma^2)$ - normalizing constant

$\prod_{i < j}$ - product for all unique pairs

$I\{\cdot\}$ - indicator function

τ^2 - prior variance for μ

$IG(\alpha, \beta)$ - inverse-gamma random variable

α - prior shape parameter for σ^2

β - prior scale parameter for σ^2

$f(r_{ij} | -)$ - full conditional distribution for r_{ij}

$tr(\cdot)$ - trace of a matrix

R^{-1} - inverse of correlation matrix

B - empirical variance-covariance matrix

I - identity matrix

$C(\mu, \sigma^2, \vartheta)$ - normalizing constant

K - number of groups of correlations or variables

$\sum_{k=1}^K$ - sum indexed from 1 to K

ϑ_{ij} - latent indicator variable for group that r_{ij} belongs to

μ_k - mean for correlations in group k

σ_k^2 - variance for correlations in group k

$Pr(\vartheta_{ij} = k)$ - probability density function for latent variable ϑ_{ij}

δ_0 - point mass

$\sum_{k,h}$ - sum over all pairs

ϑ_i - latent indicator variable for which group variable i belongs to

μ_{kh} - mean of correlations for variables in groups k and h

σ_{kh}^2 - variance of correlations for variables in groups k and h

(lb_{ij}, ub_{ij}) - interval over which r_{ij} maintains positive semi-definiteness of R

μ^* - proposed μ

$\alpha = \min \{1, C(\mu^*, \sigma^2)/C(\mu, \sigma^2)\}$ - acceptance probability

$r_{i,j}^m$ - importance sampling component for correlation

$R^m = [r_{ij}^m]$ - importance sampling component for correlation matrix

$\hat{C}(\mu, \sigma^2) = 1/M \sum_{m=1}^M I\{R^m \in \mathcal{R}^J\}$ - importance sampling estimate of normalizing constant

δ_{ij} - shadow prior latent mean

ν^2 - shadow prior latent variance
 ϑ_{-ij} - vector of latent indicator variables minus the ij^{th} variable
 y_{it} - observation for patient i on day t
 $g_{\theta_i}(t)$ - nonlinear regression function
 θ_i - random effects vector
 ϵ_{it} - regression error vector
 $h(m, \Sigma)$ - hyper-prior for random effects
 ϵ_t - time varying errors in ARCH model
 h_t - variance of time varying errors in ARCH model
 α_0 - intercept term in ARCH model
 α_j - coefficient for j^{th} lagged error ϵ_{t-j}^2 in ARCH model
 ϵ_{t-j} - j^{th} lagged error in ARCH model
 γ_j - coefficient for j^{th} lagged variance h_{t-j} in GARCH model
 h_{t-j} - j^{th} lagged variance in GARCH model
 H_t - multivariate variance of time varying errors in (G)ARCH model
 $y = (y_1, \dots, y_p)$ - multivariate binary response variables in probit model
 $x = (x_1, \dots, x_p)$ - covariates in probit model
 y - p -dimensional normal latent variable vector
 β' - coefficient parameters for covariates x

X - multivariate asset returns

μ - vector of “mean” parameters in skew normal distribution

Σ - covariance matrix

Δ - non-singular matrix of skewness and co-skewness parameters

Z - vector of independent latent standard normals truncated to be strictly positive

ϵ - normally distributed error vector with mean 0 and variance Σ

π - ratio of the area to the square of the radius of a circle (3.14159...)

Z_j - the j^{th} element of Z

X_i - a subset of assets in X

$\text{vec}(\cdot)$ - function that forms a vector from a matrix by stacking the columns of the matrix

Σ^{-1} - inverse of covariance matrix, or precision matrix

BF - Bayes Factor

BIC - Bayesian Information Criteria, or Schwartz Criteria

M_i - model i

x^o - current data up to the present

x - future data

\bar{m} - $\int x p(x|x^o)dx$, also posterior mean

\otimes - Kronecker matrix multiplication

$u_{pred}(\cdot, \cdot)$ - utility function that is a function predictive moments and portfolio weights

λ - risk aversion parameter

γ - skewness preference parameter

ω - portfolio weights

$U_{pred}(\omega)$ - expected utility function that is function of predictive moments

m_p - predictive mean of x

V_p - predictive variance/covariance of x

S_p - predictive skewness of x

$u_{param}(\cdot, \cdot)$ - utility function that is a function of parameters of the probability model, and portfolio weights

θ - parameters of probability model

m - mean of probability model

V - variance/covariance of probability model

S - skewness of probability model

$U_{param}(\cdot, \cdot)$ - expected utility function that is a function of parameters of the probability model, and portfolio weights

\bar{m} - posterior mean of m

\bar{V} - posterior variance/covariance of V

\bar{S} - posterior skewness of S

$u_{param,1}, \dots, u_{param,K}$ - a family of utility functions that are functions of parameters of a probability model

$\theta_{i,k} \sim p(\theta|x^o)$ - posterior parameter draws

$\omega_{i,k}$ - weights that optimizes utility function that is a function of posterior parameter draws

$\bar{\omega}_k = 1/n \sum \omega_{i,k}$ - average of weights

$u_{empirical}$ - utility function that is function of empirically observed moments

λ_{param} - coefficients of expected parameter based utility $U_{param}(\omega)$

λ_{pred} - coefficients of expected predictive based utility $U_{pred}(\omega)$

$\omega_{\lambda_{param}}(t)$ - parameter based utility weights conditional on the data up to time t

$\omega_{\lambda_{pred}}(t)$ - predictive based utility weights conditional on the data up to time t

$R.(\omega)$ - two moment analog of $U.(\omega)$

$\omega^*(t)$ - observed market weights at time t

L_U - sum of squared residuals

References

- Adcock, C. J. (2002) Asset pricing and portfolio selection based on the multivariate skew-student distribution. Paper presented at the Non-linear Asset Pricing Workshop, April, Paris.
- Adcock, C. J. and Shutes, K. (1999) Fat tails and the capital asset pricing model. University of Bath.
- Aggarwal, R. (1990) Distribution of spot and forward exchange rates: Empirical evidence and investor valuation of skewness and kurtosis. *Decision Science*, 588–595.
- Aggarwal, R., Rao, R. K. S. and Hiraki, T. (1989) Skewness and kurtosis in Japanese equity returns: Empirical evidence. *Journal of Financial Research*, **12**, 253–261.
- Aguilar, O. and West, M. (2000) Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics*, **18**, 338–357.
- Albert, J. and Chib, S. (1995) Bayesian residual analysis for binary response regression models. *Biometrika*, **82**, no. 4, 747–759.
- Arditti, F. D. (1967) Risk and the required return on equity. *Journal of Finance*, **22**, no. 1, 19–36.
- Arditti, F. D. (1971) Another look at mutual fund performance. *Journal of Financial and Quantitative Analysis*, **6**, no. 3, 909–912.
- Arditti, F. D. and Levy, H. (1975) Portfolio efficiency and analysis in three moments: The multiperiod case. *Journal of Finance*, **30**, no. 3, 797–809.
- Arrow, K.J. (1971) *Essays in the theory of risk-bearing*. Chicago: Markham Publishing Co.
- Athayde, G. M. de and Flôres Jr., R.G. (2001) Finding a maximum skewness portfolio—a general solution to three moments portfolio choice. Fundacao Getulio Vargas.
- Avramov, D. (2000) Stock return predictability and model uncertainty. University of Maryland, College Park, MD.

- Azzalini, A. and Capitanio, A. (1999) Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society*, no. B 61, 579–602.
- Azzalini, A. and Dalla Valle, A. (1996) The multivariate skew-normal distribution. *Biometrika*, **83**, 715–726.
- Barnard, J., McCulloch, R. E. and Meng, X. (2000) Modeling covariance matrices in terms of standard deviations and correlations, with applications to shrinkage. *Statistica Sinica*, **10**, 1281–1311.
- Barone-Adesi, G. (1985) Arbitrage equilibrium with skewed asset returns. *Journal of Financial and Quantitative Analysis*, **20**, no. 3, 299–313.
- Barone-Adesi, G., Gagliardini, P. and Urga, G. (2001) Testing the homogeneity of asset pricing models. Universita' della Svizzera Italiana, Lugano, CH.
- Bates, D. (2000) Post-'87 crash fears in s&p 500 futures options. *Journal of Econometrics*, **94**, 181–238.
- Bates, D. (2001) The market price of crash risk. University of Iowa.
- Bawa, V. S. (1975) Optimal rules for ordering uncertain prospects. *Journal of Financial Economics*, **2**, 95–121.
- Beedles, W. and Simkowitz, M. (1978) A note on skewness and data errors. *Journal of Finance*, **33**, no. 1, 288–292.
- Bera, A. and Premaratne, G. (2001) Modeling asymmetry and excess kurtosis in stock return data. University of Illinois.
- Bera, A. and Premaratne, G. (2002) A test for asymmetry with leptokurtic financial data. University of Illinois, Champaign, IL.
- Berenyi, Z. (2002) Measuring hedge fund risk with multi-moment risk measures. University of Munich.
- Berger, J. (1985) *Statistical Decision Theory and Bayesian Analysis*. New York, NY: Springer Verlag.
- Bernardo, A. E. and Ledoit, O. (2000) Gain, loss, and asset pricing. *Journal of Political Economy*, **108**, no. 1, 144–172.

- Bernardo, J. M. and Smith, A. F. M. (1994) *Bayesian Theory*. John Wiley & Sons, Inc.
- Bhattacharyya, N. (2002) From mean variance space to mean skewness space—implications for simultaneous risk seeking and risk averting behavior. University of Manitoba.
- Bielza, C., Müller, P. and Insua, D. R. (1999) Decision analysis by augmented probability simulation. *Management Science*, **45**, 995–1007.
- Blattberg, R. and Gonedes, N. (1974) A comparison of the stable and student distributions as statistical models for stock prices. *Journal of Business*, **47**, 244–280.
- Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, no. 31, 307–327.
- Bollerslev, T. (1990) Modeling the coherence in short-run nominal exchange rates: A multivariate generalized arch approach. *Review of Economics and Statistics*, **72**, 498–505.
- Bollerslev, T., Chou, R. and Kroner, K. F. (1992) Arch modeling in finance. *Journal of Econometrics*, **52**, 5–59.
- Bollerslev, T., Engle, R. F. and Wooldridge, M. (1988) A capital asset pricing model with time varying covariances. *Political Economy*, **96**, 116–131.
- Bond, S. A. *A review of asymmetric conditional density functions in autoregressive conditional heteroscedasticity models*. Ch. 2.
- Bond, S. A. (2001) Time varying skewness and downside risk measurement: A comparison of models. University of Cambridge.
- Boyle, G. W. and Rao, R. K. S (1988) The mean-generalized coefficient of variation selection rule and expected utility maximization. *Southern Economic Journal*, **55**, no. 1, 1–8.
- Branco, M. D. and Dey, D. K. (2001) A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, **79**, 99–113.
- Brandt, M. W., Goyal, A. and Santa-Clara, P. (2001) Dynamic portfolio choice: A simulation approach. University of Pennsylvania, Philadelphia, PA.

- Brannas, K. and Nordman, N. (2001) Conditional skewness modeling for stock returns. Umea University.
- Brennan, M. J. and Xia, Y. (2001) Assessing asset pricing anomalies. University of California at Los Angeles, Los Angeles, CA.
- Britten-Jones, M. (1998) Portfolio optimization and bayesian regression. London Business School.
- Britten-Jones, M. (1999) The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance*, **54**, 655–671.
- Brockett, P. and Kahane, Y. (1992) Risk, return, skewness and preference. *Management Science*, **38**, 851–866.
- Brown, S. J. (1979) In Bawa, V. S. Brown, S. J. Klein, R. W. (Eds.), *Estimation Risk and Optimal Portfolio Choice*, ch. Optimal portfolio choice under uncertainty: A Bayesian approach. North-Holland, Amsterdam.
- Celeux, G., Hurn, M. and Robert, C. P. (1999) Computational and inferential difficulties with mixture posterior distributions. Technical Report. Laboratoire de Statistique, CREST-ENSAE.
- Chen, M., Ibrahim, J. G. and Shao, Q. (2000) *Monte Carlo methods in Bayesian computation*. New York: Springer. Ch. 5.
- Chib, S. and Greenberg, E. (1995) Understanding the metropolis-hastings algorithm. *American Statistician*, **49**, 327–335.
- Chib, S. and Greenberg, E. (1998) Analysis of multivariate probit models. *Biometrika*, **85**, 347–361.
- Christiansen, C. and Morris, C. (1997) Hierarchical poisson regression modeling. *Journal of the American Statistical Association*, **92**, no. 438, 618.
- Chung, Y. P., Johnson, H. and Schill, M. J. (2001) Asset pricing when returns are nonnormal: Fama-french factors vs. higher-order systematic co-moments. A. Gray Anderson Graduate School of Management.
- Chunhachinda, P., Danpani, K., Hamid, S. and Prakash, A. (1997) Portfolio selection and skewness: Evidence from international stocks markets. *Journal of Banking and Finance*, **21**, 143–167.

- Clyde, M. and George, E. I. (2000) Flexible empirical bayes estimation for wavelets. *Journal of the Royal Statistical Society B*, **62**, 681–698.
- Corrado, C. and Su, T. (1996) Skewness and kurtosis in s&p 500 index returns implied by option prices. *Journal of Financial Research*, **19**, no. 2, 175–192.
- Dalal, S. and Hall, W. (1983) Approximating priors by mixtures of natural conjugate priors. *Journal of the Royal Statistical Society B*, **45**, 278–286.
- Damien, P. S., Wakefield, J. and Walker, G. (1999) Gibbs sampling for bayesian nonconjugate and hierarchical models using auxiliary variables. *Journal of Royal Statistical Society B*, **61**, 331–344; Part 2.
- Daniels, M. J. (1999) A prior for the variance in hierarchical models. *Canadian Journal of Statistics*, **27**, no. 3, 567–578.
- Daniels, M. J. and Kass, R. E. (1999) Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association*, **94**, 1254–1263.
- DeGroot, M. H. (1970) *Optimal Statistical Decisions*. New York: McGraw-Hill.
- Dittmar, R. F. (2002) Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *Journal of Finance*, **57**, 369–403.
- Ekholm, A. and Pasternack, D. (2002) The negative news threshold - an explanation for negative skewness in stock returns. Swedish School of Economics and Business Administration.
- Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, **50**, 987–1008.
- Everson, P. J. and Morris, C. N. (2000) Inference for multivariate normal hierarchical models. *Journal of the Royal Statistical Society B*, **62**, no. 2, 399; 14.
- Faff, R., Ho, Y. and Zhang, L. (1998) A gmm test of the three-moment capm in the australian equity market. *Asia Pacific Journal of Finance*, **1**, 45–60.
- Fama, E. F. (1965) Portfolio analysis in a stable paretian market. *Management Science*, **11**, 404–419.
- Fang, H. and Lai, T.-Y. (1997) Co-kurtosis and capital asset pricing. *Financial Review*, **32**, no. 2, 293–307.

- Friend, I. and Westerfield, R. (1980) Co-skewness and capital asset pricing. *Journal of Finance*, **35**, no. 4, 897–913.
- Frost, P. A. and Savarino, J. E. (1986) An empirical bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis*, **21**, no. 3, 293–305.
- Frost, P. A. and Savarino, J. E. (1988) For better performance; constrain portfolio weights. *Journal of Portfolio Management*, **14**, 29–34.
- George, E. I. and McCulloch, R. E. (1993) Variable selection via gibbs sampling. *Journal of the American Statistical Association*, **88**, 881–89.
- Goetzmann, W. N., Ingersoll Jr., J. E., Spiegel, M. I. and Welch, I. (2002) Sharpening sharpe ratios. Yale International Center For Finance.
- Goldfarb, D. and Iyengar, G. (2001) Robust portfolio selection problems. Columbia University.
- Greyserman, A., Jones, D. H. and Strawderman, W. E. (2002) Portfolio selection using bayesian analysis and gibbs sampling.
- Harvey, C. R. and Siddique, A. (1999) Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis*, **34**, 465–487.
- Harvey, C. R. and Siddique, A. (2000a) Conditional skewness in asset pricing tests. *Journal of Finance*, **55**, 1263–1295.
- Harvey, C. R. and Siddique, A. (2000b) Time-varying conditional skewness and the market risk premium. *Research in Banking and Finance*, **1**, 27–60.
- Hueng, C. J. and Brooks, R. (2000) Forecasting asymmetries in stock returns: Evidence from higher moments and conditional densities. University of Alabama.
- Hwang, S. and Satchell, S. E. (1999) Modeling emerging market risk premia using higher moments. *International Journal of Finance and Economics*, **4**, no. 4, 271–296.
- Ingersoll, J. (1987) *Theory of financial decision making*. Rowman & Littlefield, Totow.
- Jean, W. H. (1971) The extension of portfolio analysis to three or more parameters. *Journal of Financial and Quantitative Analysis*, **6**, no. 1, 505–515.

- Jean, W. H. (1973) More on multidimensional portfolio analysis. *Journal of Financial and Quantitative Analysis*, **8**, 475–490.
- Johannes, M., Polson, N. and J., Stroud (2002) Sequential optimal portfolio performance: Market and volatility timing.
- Jondeau, E. and Rockinger, M. (2000) Conditional volatility, skewness, and kurtosis: Existence and persistence. HEC-School of Management.
- Jones, C. (1998) Bayesian estimation of continuous-time finance models. University of Rochester, Rochester, NY.
- Jorion, P. (1985) International portfolio diversification with estimation risk. *Journal of Business*, **58**, 259–278.
- Jorion, P. (1986) Bayes-stein estimation of portfolio analysis. *Journal of Financial and Quantitative Analysis*, **21**, 279–292.
- Jorion, P. (1991) Bayesian and capm estimators of the means: Implications for portfolio selection. *Journal of Banking and Finance*, **15**, 717–727.
- Jurczenko, E. and Maillet, B. (2001) *The three-moment CAPM: Theoretical foundations and an asset pricing model comparison in a unified framework*. John Wiley and Sons. In: Dunis, Christian et al. (Eds), *Developments in Forecast Combination and Portfolio Choice*.
- Kahneman, D. and Tversky, A. (1979) Prospect theory: An analysis of decision under risk. *Econometrica*, **47**, 263–291.
- Kallberg, J. G. and Ziemba, W. T. (1983) Comparison of alternative functions in portfolio selection problems. *Management Science*, **11**, 1257–1267.
- Kandel, S., McCulloch, R. E. and Stambaugh, R. F. (1995) Bayesian inference and portfolio efficiency. *Review of Financial Studies*, **8**, 1–53.
- Kandel, S. and Stambaugh, R. F. (1995) Portfolio inefficiency and the cross-section of expected returns. *Journal of Finance*, **50**, 157–184.
- Kandel, S. and Stambaugh, R. F. (1996) On the predictability of stock returns: An asset allocation perspective. *Journal of Finance*, **51**, no. 2, 385–424.
- Kane, A. (1982) Skewness preference and portfolio choice. *Journal of Financial and Quantitative Analysis*, **17**, no. 1, 15–25.

- Karolyi, G. A. (1992) Predicting risk: Some new generalizations. *Management Science*, **1**, 57–74.
- Karolyi, G. A. (1993) A bayesian approach to modeling stock return volatility for option valuation. *Journal of Financial and Quantitative Analysis*, **28**, 579–594.
- Klein, R. W. and Bawa, V. (1976) The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, **3**, 215–231.
- Kon, S. (1984) Models of stock returns - a comparison. *Journal of Finance*, **39**, 147–165.
- Kozik, T. J. and Larson, A. M. (2001) The n-moment insurance capm. Technical Report. Allstate Insurance Company.
- Kraus, A. and Litzenberger, R. H. (1976) Skewness preference and the valuation of risk assets. *Journal of Finance*, **31**, 1085–1100.
- Kroll, Y., Levy, H. and Markowitz, H. (1984) Mean variance versus direct utility maximization. *Journal of Finance*, **39**, 47–61.
- Lai, T. (1991) Portfolio selection with skewness: A multiplicative-objective approach. *Review of Quantitative Finance and Accounting*, **1**, 293–305.
- Lee, W. V. and Rao, R. K. S. (1988) Mean-lower partial moment valuation with log normally distributed returns. *Management Science*, **34**, no. 4, 446–453.
- Leonard, T and Hsu, J. S. (1992) Bayesian inference for a covariance matrix. *Annals of Statistics*, **20**, no. 4, 1669–1696.
- Levy, H. (1969) An utility function depending on the first three moments. *Journal of Finance*, **24**, 715–719.
- Levy, H. and Markowitz, H. (1979) Approximating expected utility by a function of mean and variance. *American Economic Review*, **69**, 308–317.
- Levy, H. and Sarnat, M. (1984) *Portfolio and investment selection: Theory and practice*. Prentice-Hall International.
- Lewellen, J. and Shanken, J. (2002) Learning, asset-pricing tests, and market efficiency. *Journal of Finance*, **57**, no. 3, 1113–1146.

- Lhabitant, F. (1997) On the (ab)use of expected utility approximations for portfolio selection and portfolio performance. University of Lausanne.
- Liechty, J. C., Ramaswamy, V. and Cohen, S. H. (2001) Choice menus for mass customization: An experimental approach for analyzing customer demand with an application to a web-based information service. *Journal of Marketing Research*, **38**, 183–196.
- Lim, K.-G. (1989) A new test of the three-moment capital asset pricing model. *Journal of Financial and Quantitative Analysis*, 205–216.
- Lin, S. P. and Perlman, M. D. (1985) *An improved procedure for the estimation of a correlation matrix*, pp. 369–379. Elsevier Science Publishers B.V. In: Statistical Theory and Data Analysis.
- Liu, J., Longstaff, F. A. and Pan, J. (forthcoming) Dynamic asset allocation with event risk. *Journal of Finance*.
- Liu, J., Pan, J. and Wang, T. (2002) An equilibrium model of rare event premia. Massachusetts Institute of Technology.
- Machado-Santos, C. and Fernandes, A. C. (2002) Skewness in financial returns: Evidence from the portuguese stock market. UTAD University.
- Malevergne, Y. and Sornette, D. (2002) Multi-moments method for portfolio management: Generalized capital asset pricing model in homogeneous and heterogeneous markets. Technical Report. Laboratoire de Physique de la Matière Condensée.
- Manchanda, P., Ansari, A. and Gupta, S. (1999) The ‘shopping basket’: A model for multi-category purchase incidence decisions. *Marketing Science*, **18**, no. 2, 95–114.
- Markowitz, H. (1952) Portfolio selection. *Journal of Finance*, **7**, 77–91.
- Markowitz, H. (1959) *Portfolio selection: Efficient diversification of investments*, Second Edition edn. Malden: Blackwell.
- McCulloch, R. E., Polson, N. G. and Rossi, P. (2000) A bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, **99**, 173–193.

- Meyn, S. P. and Tweedie, R. L. (1993) *Markov chains and stochastic stability*. Springer-Verlag.
- Michaud, R. (1998) *Efficient asset management*. Cambridge: Harvard Business School Press.
- Mills, T. C. (1995) Modeling skewness and kurtosis in the london stock exchange ft-se index return distributions. *The Statistician*, **44**, no. 3, 323–332.
- Molenberghs, G. and Rousseeuw, P. J. (1994) The shape of correlation matrices. *The American Statistician*, **48**, 4.
- Müller, P., Parmigiani, G., Schildkraut, J. and Tardella, L. (1999) A bayesian hierarchical approach for combining case-control and prospective studies. *Biometrics*, **55**, 258–266.
- Müller, P. and Rosner, G. (1997) A bayesian population model with hierarchical mixture priors applied to blood count data. *Journal of the American Statistical Association*, **92**, 1279–1292.
- Murren, C. J., Pendelton, N. and Pigliucci, M. (2002) Evolution of phenotypic integration in brassica (brassicaceae). *American Journal of Botany*, **89**, 655–663.
- Newton, M. A. and Raftery, A. E. (1994) Approximate bayesian inference by the weighted likelihood bootstrap (with discussion). *Journal of the Royal Statistical Society B*, **56**, 3–48.
- Nietert, B. (2001) Pricing relevance of stock price jumps: Capm and variance decomposition. Technical Report. Passau University.
- Novshek, W. (1993) *Mathematics for economists*. Academic Press New York.
- Nummelin, K. (1997) Global coskewness and finish stocks: Empirical tests. *Journal of International Financial Markets, Institutions and Money*, **7**, 137–155.
- O’Hagan, A. (1994) *Kendall’s Advanced Theory of Statistics: Volume 2B Bayesian Inference*. New York, NY: John Wiley and Sons.
- Pan, J. (2002) The jump-risk premia implicit in options on the market portfolio with discontinuous returns. *Journal of Financial Economics*, **63**, 3–50.
- Pástor, L. (2000) Portfolio selection and asset pricing models. *Journal of Finance*, **55**, 179–223.

- Pástor, L. and Stambaugh, R. F. (1999) Costs of equity capital and model mispricing. *Journal of Finance*, **54**, 67–121.
- Pástor, L. and Stambaugh, R. F. (2000) Comparing asset pricing models: An investment perspective. *Journal of Financial Economics*, **56**, 335–381.
- Pederson, C. and Satchell, S. (1998) An extended family of financial risk measures. *Geneva Papers on Risk and Insurance Theory*, **23**, 89–117.
- Pederson, C. and Satchell, S. (forthcoming) Asymmetric equilibrium risk measures. *Journal of Empirical Finance*.
- Peiro, A. (1999) Skewness in financial returns. *Journal of Banking and Finance*, **23**, 847–862.
- Perez-Quiros, G. and Timmermann, A. (2001) Business cycle asymmetries in stock returns: Evidence from higher order moments and conditional densities. *Journal of Econometrics*, **103**, 259–306.
- Polimenis, V. (2002) The distributional capm: Connecting risk premia to return distributions. University of California, Riverside, CA.
- Premaratne, G. and Tay, A. S. (2002) How should we interpret evidence of time varying conditional skewness? National University of Singapore.
- Price, K., Price, B. and Nantell, T. (1982) Variance and lower partial moments measures of systematic risk: Some analytical and empirical results. *Journal of Finance*, **37**, 843–855.
- Quintana, J. M., Lourdes, V., Aguilar, O. and Liu, J. (2002) Global gambling. In *Valencia 7*.
- Raiffa, H. and Schlaifer, R. (1961) *Applied statistical decision theory*. Cambridge: Harvard University Press.
- Rockinger, M. and Jondeau, E. (2002) Entropy densities with an application to autoregressive conditional skewness and kurtosis. *Journal of Econometrics*, **106**, 119–142.
- Rubinstein, M. (1973) The fundamental theorem of parameter preference security valuation. *Journal of Financial and Quantitative Analysis*, **8**, 61–69.

- Sahu, S. K., Branco, M. D. and Dey, D. K. (2002) A new class of multivariate skew distributions with applications to bayesian regression models.
- Sanchez-Torres, P. and Sentana, E. (1998) Mean-variance-skewness analysis: An application to risk premia in the spanish stock market. *Investigaciones Economicas*, **22**, 5–17.
- Scott, R. C. and Horvath, P. A. (1980) On the direction of preference for moments of higher order than variance. *Journal of Finance*, **35**, no. 4, 915–919.
- Sears, R. S. and Wei, K. C. J. (1980) Asset pricing, higher moments, and the market risk premium: A note. *Journal of Finance*, **40**, no. 4, 1251–1253.
- Sears, R. S. and Wei, K. C. J. (1988) The structure of skewness preferences in asset pricing models with higher moments: An empirical test. *Financial Review*, **23**, no. 1, 25–38.
- Sharpe, W. F. (1994) The sharpe ratio. *Journal of Portfolio Management*, **21**, no. 1, 49–58.
- Simkowitz, M. A. and Beedles, W. L. (1980) Asymmetric stable distributed security returns. *Journal of the American Statistical Association*, **75**, 306–312.
- Singleton, J. C. and Wingender, J. (1986) Skewness persistence in common stock returns. *Journal of Financial and Quantitative Analysis*, **21**, 335–341.
- Sornette, D. (1998) Large deviations and portfolio optimization. *Physica A*, **256**, 251–283.
- Sornette, D., Andersen, J. V. and Simonetti, P. (2000) Portfolio theory for “fat tails”. *International Journal of Theoretical and Applied Finance*, **3**, no. 3, 523–535.
- Sornette, D., Simonetti, P. and Andersen, J. V. (2000) ϕ^q -field theory for portfolio optimization: “fat tails” and non-linear correlations. *Physics Reports*, **335**, no. 2, 19–92.
- Sortino, F. A. and Forsey, H. J. (1996) On the use and misuse of downside risk. *Journal of Portfolio Management*, **22**, 35–42.
- Sortino, F. A. and van der Meer, R. (1991) Downside risk. *Journal of Portfolio Investment*, **17**, 27–31.

- Sortino, F. A. and Price, L. N. (1994) Performance measurement in a downside risk framework. *Journal of Investing*, **3**, no. 3, 50–58.
- Spurgin, R. B. (2001) How to game your sharpe ratio. *The Journal of Alternative Investments*, **4**, no. 3, 38–46.
- Tamayo, A. (2001) Stock return predictability, conditional asset pricing models and portfolio selection. University of Rochester, Rochester, NY.
- Tan, K.-J. (1991) Risk return and the three-moment capital asset pricing model : Another look. *Journal of Banking and Finance*, **15**, no. 2, 449–460.
- Tierney, L. (1994) Markov chains for exploring posterior distributions. *Annals of Statistics*, **22**, 1701–1728.
- Tsiang, S. C. (1972) The rationale of the mean-standard deviation analysis, skewness preference, and the demand for money. *American Economic Review*, **62**, no. 3, 354–371.
- West, M. (2002) Bayesian factor regression models in the “large p, small n” paradigm. In *Bayesian Statistics 7*.
- Wong, F., Carter, C. K. and Kohn, R. (2002) Efficient estimation of covariance selection models. In *Valencia 7*, pp. 182–217.
- Y., Xia (2001) Learning about predictability: The effects of parameter uncertainty on dynamic asset allocation. *Journal of Finance*, **56**, 205–246.
- Yang, R. and Berger, J. O. (1994) Estimation of a covariance matrix using the reference prior. *Annals of Statistics*, **22**, 1195–1211.

Biography

I was born 27 August, 1974 in Provo, Utah. I earned a Bachelor of Science in Mathematics from Brigham Young University in Provo, Utah in 1998. In 2001, I earned a Masters of Science in Statistics from Duke University. I have co-authored the following:

1. Liechty, J. C., M. W. Liechty, and P. Müller, (2002), Bayesian correlation estimation, *Biometrika*, *forthcoming*