Portfolio Selection With Higher Moments

Campbell R. Harvey*

Duke University, Durham, NC 27708 USA

John C. Liechty

Pennsylvania State University, University Park, PA 16803 USA

Merrill W. Liechty

Drexel University, Philadelphia, PA 19104 USA

Peter Müller

University of Texas M.D. Anderson Cancer Center, Houston, TX 77030 USA

Abstract

We propose a method for optimal portfolio selection using a Bayesian decision theoretic framework that addresses two major shortcomings of the Markowitz approach: the ability to handle higher moments and estimation error. We employ the skew normal distribution which has many attractive features for modeling multivariate returns. Our results suggest that it is important to incorporate higher order moments in portfolio selection. Further, our comparison to other methods where parameter uncertainty is either ignored or accommodated in an ad hoc way, shows that our approach leads to higher expected utility than the resampling methods that are common in the practice of finance.

KEYWORDS: Bayesian decision problem, multivariate skewness, parameter uncertainty, optimal portfolios, utility function maximization.

---

*Campbell R. Harvery is J. Paul Sticht Professor of International Business, Fuqua School of Business, Duke University, Durham NC 27708 and Research Associate, National Bureau of Economic Research, Cambridge, MA 02138 (E-mail: *cam.harvey@duke.edu*). John C. Liechty is Assistant Professor, Department of Marketing, Smeal College of Business, and Department of Statistics, Pennsylvania State University, University Park, PA 16803 (E-mail: *jcl12@psu.edu*). Merrill W. Liechty is Assistant Professor, Department of Decision Sciences, LeBow College of Business, Drexel University, Philadelphia, PA 19104 (E-mail: *merrill@drexel.edu*). Peter Müller is Professor, Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, TX 77030 (E-mail: *pm@odin.mdacc.tmc.edu*). We appreciate the comments of Philippe Jorion, Harry Markowitz, Haim Levy, Michael Brandt, Chris Adcock and Karl Shutes.

1. INTRODUCTION

Markowitz (1952a) 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. However, in the usual implementation of Markowitz's second stage, we are assumed to know with certainty the inputs from the first stage, i.e. the exact means, variances and covariances. This paper introduces a method for addressing both stages.

In a less well known part of Markowitz (1952a, p.91), he details a condition whereby mean-variance efficient portfolios will not be optimal – when an investor's utility is a function of mean, variance, *and* skewness. While Markowitz did not work out the optimal portfolio selection in the presence of skewness and other higher moments, we do. We develop a framework for optimal portfolio selection in the presence of higher order moments and parameter uncertainty.

Several authors have proposed advances to optimal portfolio selection methods. Some address the empirical evidence of higher moments; Athayde and Flôres (2003, 2004) and Adcock (2002) give methods for determining higher dimensional 'efficient frontiers', but they remain in the certainty equivalence framework (assuming exact knowledge of the inputs) for selecting an optimal portfolio. 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'. Other three moment optimization methods include using negative semi-variance in place of variance (see Markowitz 1959 and Markowitz, Todd, Xu, and Yamane 1993). A similar measure of downside risk is incorporated by Feiring, Wong, Poon, and Chan (1994), and Konno, Shirakawa, and Yamazaki (1993) who use an approximation to the lower semi-third moment in their Mean-Absolute Deviation-Skewness portfolio model. These methods rest on the assumption that an investor's expected utility is reasonably approximated by inserting estimates of the moments of an assumed sampling model.

A number of researchers have shown that mean-variance efficient portfolios, based on estimates and ignoring parameter uncertainty, are highly sensitive to perturbations of these estimates. Jobson, Korkie and Ratti (1979) and Jobson and Korkie (1980) detail these problems and suggest the use of shrinkage estimators. This 'estimation risk' comes from both choosing poor probability models and from ignoring parameter uncertainty, maintaining the assumption that expected utility can be evaluated by substituting point estimates of sampling moments in the utility function.

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. Jorion (1992) proposes a resampling method aimed at estimation error. Using a Bayesian approach, Britten-Jones (2002) 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 Bayesian perspective, Jorion (1986) use a shrinkage approach while Treynor and Black (1973) advocate the use of investors' views in combination with historical data. Kandel and Stambaugh (1996) examine predictability of stock returns when allocating between stocks and cash by a risk-averse Bayesian investor. (See also Johannes, Polson, and Stroud 2002 who examine the market timing relationship to performance of optimal portfolios using a model with correlation between volatility and returns in a Bayesian portfolio selection setting.) Zellner and Chetty (1965), Klein and Bawa (1976) and Brown (1978) emphasize 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). Pástor and Stambaugh (2000) study the implications of different pricing models on optimal portfolios, updating prior beliefs based on sample evidence. Pástor (2000) and Black and Litterman (1992) propose using asset pricing models to provide informative prior distributions for future returns. Pástor and Stambaugh (1999) show that the model used is less important than correctly accounting for parameter uncertainty in pricing assets.

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' (see Scherer 2002 for further discussion). 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. In addition, Jensen's inequality dictates that the resampled efficient frontier is in fact suboptimal. Polson and Tew (2000) argue for the use of posterior predictive moments instead of point estimates for mean and variance of an assumed sampling model. Their setup comes closest to the framework that we propose in this paper. Using posterior predictive moments, they accommodate parameter uncertainty. We follow their setup in our discussion in Section 3.1.1.

Our approach advances previous methods by addressing both higher moments and estimation risk in a coherent Bayesian framework. As part of our "stage one" approach (i.e., incorporating observation and experience), we specify a Bayesian probability model for the joint distribution of the asset returns, and discuss prior distributions. As for "stage two", the Bayesian methodology provides a straightforward framework 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. We empirically investigate the impact of simplifying the asset allocation task. For two illustrative data sets, we demonstrate the difference in expected utility that results from ignoring higher moments and using a sampling distribution, with point estimates substituted for the unknow parameter, instead of a predictive distribution. In addition, we demonstrate the loss in expected utility (explained by Jensen's inequality) from using the popular approach proposed by Michaud (1998). Markowitz and Usmen (2003) take a similar approach to us for comparing Bayesian methods to Michaud's (1998) approach, but they use diffuse priors.

Our paper is organized as follows. In the second section we discuss the importance of higher moments and provide the setting for portfolio selection and Bayesian statistics in finance. We discuss suitable probability models for portfolios and detail our proposed framework. In the third section, we show how to optimize portfolio selection based on utility functions in the face of parameter uncertainty using Bayesian methods. Section four empirically compares different methods and approaches to portfolio selection. Some concluding remarks are offered in the final section. The appendix contains some additional results and proofs.

2. HIGHER MOMENTS AND BAYESIAN MODELS

A prerequisite to the use of the Markowitz framework is either that the relevant distribution of asset returns be normally distributed or that utility is only a function of the first two moments. But it is well known that many financial returns 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 symmetric, see Kon (1984), Mills (1995), Peiro (1999) and Premaratne and Bera (2002). Also we argue that the relevant probability model is the posterior predictive distribution, which in general is not normal, not even under an assumed normal sampling model.

The approach proposed in our paper is closely related to the use of the Omega function introduced and discussed in Cascon, Keating and Shadwick (2003). They argue that point estimates of mean and variance of an assumed sampling distribution are insufficient summaries of the available information of future returns. Instead they advocate the use of a summary function, which they call "Omega", that represents all the relevant information contained within the observed data. We agree with the premise that a full probabilistic description of relevant uncertainties of future returns is needed. Instead of the Omega function, we base our approach on a traditional Bayesian decision theoretic framework which allows us to formally account for parameter uncertainty. Otherwise the rationale of the two methods is the same. The formalisms are different.

Our investigation of multiple assets builds on these empirical findings and indicates that the existence of 'coskewness', which can be interpreted as correlated extremes, is often hidden when assets are considered one at a time. To illustrate, Figure 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 2 contains a bivariate normal approximation of their joint returns. While the marginal summaries in Figure 1 suggest almost no deviation from the normality assumption, the joint summary appears to exhibit a degree of coskewness, suggesting that skewness may have a larger impact on the distribution of a portfolio than previously anticipated.

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 agree that *ceteris paribus* 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, Markowitz (1952b) and Arrow (1971) argue that desirable utility functions should exhibit decreasing absolute risk aversion, implying that investors should have preference for positively skewed asset returns. (Also see the discussion in Roy 1952.) 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 a strong preference for positive skewness in the study of 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.

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 reduce 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. 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 skewness will be important to a large class of investors and is evident in the historical returns of the underlying assets and portfolios, the question remains; how influential is skewness in terms of finding optimal portfolio weights? Cremers, Kritzman, and Page (2004) argue that only under certain utility is it worthwhile to consider skewness in portfolio selection. However, it is our experience that any utility function approximated by a third order Taylor's expansion can lead to more informatively selected portfolio weights if skewness is not ignored.

To illustrate, consider the impact of skewness on the empirical distribution of a collection of two-stock portfolios. For each portfolio, the mean is identical to the linear combination of the stock means and the variance is less than the combination of the stock variances, see Figure 3 for an illustration using three, two-stock portfolios. Unlike the variance, there is no guarantee that the portfolio skewness will be larger or smaller than the linear combination of the stock skewness, and in practice we observe a wide variety of behavior. This suggests that the mean-variance optimal criteria will lead to sub-optimal portfolios in the presence of skewness. To accommodate higher-order moments in the asset allocation task, we must introduce an appropriate probability model. After providing an overview of possible approaches, we formally state a model and discuss model choice tools.

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 for modeling a portfolio, there is considerable evidence that portfolio returns are non-normal. There are a number of alternative models that include higher moments. The multivariate Student 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 asset returns, 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 (2003) by using a multivariate cdf instead of univariate cdf, adding more flexibility, which often results in better fitting models. Because of the importance of coskewness in asset returns, we start with the multivariate skew normal probability model presented in Sahu et al. (2003) and offer a generalization of their model.

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

$$X = \mu + \Delta Z + \epsilon, \tag{1}$$

where μ and Δ are an unknown parameter vector and matrix respectively, ϵ is a normally distributed error vector with a zero mean 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 matrix and is restricted to be non-negative, or

$$p(Z) \propto (2/\pi)^{\ell/2}\, exp\left(-0.5Z'Z\right) \prod_{j=1}^{p} I_{\{Z_j>0\}}, \tag{2}$$

where $I_{\{\cdot\}}$ is the indicator function and Z_j is the j^{th} element of Z. In Sahu et al. (2003), Δ is restricted to being a diagonal matrix, which accommodates skewness, but does not allow for coskewness. We generalize the Sahu et al. (2003) model to allow Δ to be an unrestricted random matrix resulting in a modified density and moment generating function, see Appendix A.1 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. 2003 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.2 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. (2003)

are Lévy stable distributions. The skew normal can be generalized as a stable distribution (see Appendix $A.3$).

While the skew normal is similar in concept to a mixture of normal random variables, it is 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, and 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 coskewness 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 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 $vec()$ forms a vector from a matrix by stacking the columns of the matrix, and *a priori* Wishart for Σ^{-1}. The resulting full conditional posterior densities for μ and $vec\,(\Delta)$ are normal, the full conditional posterior density for Σ^{-1} is Wishart and the full conditional posterior density for the latent Z is a truncated normal. See Appendix $A.4$ for a complete specification of the prior densities and the full conditional posterior densities. Given these full conditional posterior densities, estimation is done using a Markov chain Monte Carlo (MCMC) algorithm based on the Gibbs sampler and the slice sampler, see Gilks, Richardson, and Spiegelhalter (1996) for a general discussion of the MCMC algorithm and the Gibbs sampler and see Appendix $A.5$ for a discussion of the slice sampler.

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 (see Berger 1985 and O'Hagan 1994 for further discussion of Bayes Factors).

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

$$BF = \text{posterior odds/prior odds} = p(x|M_1)/p(x|M_2).$$

We use a sampling based estimator proposed by Newton and Raftery (1994) to calculate the Bayes factor.

3. OPTIMIZATION

Markowitz defined the set of optimal portfolios as the portfolios that are on the efficient frontier, based on estimated moments of the sampling distribution. 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. When parameter uncertainty is explicitly considered, the efficient frontier, now written in terms of predictive moments, can still only identify optimal allocations for utility functions that are exactly linear in the moments of the portfolio. In all other cases, the utility function must be explicitly specified. 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.

3.1 Simplifications Made in Practice

3.1.1 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 needs to be a function of the future returns, not a function of the model parameters. This point is emphasized in Zellner and Chetty (1965) and Brown (1978). It is reasonable to assume that a 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 (DeGroot, 1970; Raiffa and Schlaifer, 1961). Following this paradigm, Polson and Tew (2000) propose the use of predictive moments for future returns to define mean-variance efficient portfolios which we also implement (see also Kandel and Stambaugh 1996 and Pástor and Stambaugh 2000). In the following discussion, we highlight the difference of this approach and the traditional approach.

Predictive returns are often ignored and utility is stated in terms of the posterior means of the model parameters because of computational complexity and the argument that the moments of the predictive distribution are approximated by the corresponding moments of the posterior distribution.

To illustrate, let x^o represent the history up to the current observation and let x represent future data. Let $\mathcal{X} = (x, V_x, S_x)$ be powers of future returns, where

$$m_p = \int x\, p(x|x^o) dx$$

is the predictive mean given x^o, $V_x = (x - m_p)(x - m_p)'$, and $S_x = V_x \otimes (x - m_p)'$. Assuming that utility is a third-order polynomial of future returns, predictive utility is given by

$$u_{pred}(\omega, \mathcal{X}) = \omega'\, x - \lambda\, [\omega'\,(x - m_p)]^2 + \gamma\, [\omega'\,(x - m_p)]^3 \tag{3}$$

where λ and γ determine the impact of predictive variance and skewness. Expected utility, becomes

$$EU_{pred}(\omega) = \omega'\, E[x|x^o] - \lambda\, \omega' E[V_x|x^o]\, \omega + \gamma\, \omega' E[S_x|x^o]\, \omega \otimes \omega = \omega'\, m_p - \lambda\, \omega'\, V_p\, \omega + \gamma\, \omega' S_p\, \omega \otimes \omega, \tag{4}$$

where $\theta_p = (m_p,\, V_p, S_p)$ are the predictive moments of x. We refer to utility function (3) as a *linear* utility function. Utility is linear in the sense that u_{pred} is linear in the predictive summaries $\mathcal{X}$, and thus EU_{pred} is linear in the predictive moments θ_p.

Often a function involving sampling moments corresponding to the predictive moments in (4) is used instead of actual future returns to define utility. Assuming an i.i.d. sampling $x_t \sim p_\theta(x_t)$ for returns at time t, let $\theta = (m, V, S)$ denote the moments of p_θ and define a utility function:

$$u_{param}(\omega, \theta) = \omega'\, m - \lambda\, \omega'\, V\, \omega + \gamma\, \omega'\, S\, \omega\, \otimes\, \omega. \tag{5}$$

See Appendix A.1 for the formulas under the skew normal model. The expected utility, becomes

$$EU_{param}(\omega) = \omega'\, \bar{m} - \lambda\, \omega'\, \bar{V}\, \omega + \gamma\, \omega'\, \bar{S}\, \omega\, \otimes\, \omega, \tag{6}$$

where $\bar{m}$, $\bar{V}$ and $\bar{S}$ are the posterior means of θ. Note that the expectation in (6) "plugs in" the expectation of the parameters, ignoring the contribution of parameter uncertainty to the expected utility function. (Kan and Zhou (2004) provide a thorough discussion of the difference of plug-in and Bayes estimator of the optimal decision under the parameter based utility (5). Their discussion highlights the difference between a proper Bayes rule, defined as the decision that maximizes expected utility, versus a rule that plugs in the Bayes estimate for the weights or the parameters in the sampling distribution.) The nature of this approximation is highlighted by considering the relationship with the predictive moments in (4). In fact, it is straightforward to show that the predictive mean equals the posterior mean and that the predictive variance and skewness equal the posterior means of V and S plus additional terms, or

$$m_p = \bar{m}$$

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

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

Polson and Tew (2000, proposition 1) highlight the implication of the difference between V_p and V for mean-variance efficient portfolios. Substituting this into (4) gives an alternative form that is composed of $EU_{param}(\omega)$ plus other terms.

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

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 expectation of m, V, and S, an assumption which often fails in practice. Formally, using $(\bar{m}, \bar{V}, \bar{S})$ in place of the predictive moments ignores the second and third line in the expression above. Our approach uses the predictive moments, capturing that extra information when maximizing the expected utility.

3.1.2 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 summaries, or assume

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

For linear utility functions this approximation is exact, as in (3) and (4). The Taylor's approximation removes any parameter uncertainty and leads directly to the certainty equivalence optimization framework, substituting predictive moments. It is easy to see that combining the Taylor's approximation and the much stronger 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

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

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. The maximum expected utility framework optimizes the expected utility of future returns; the certainty equivalence framework optimizes the utility of expected future returns, (*i.e.*, substituting posterior predictive moments in the utility function). 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 the approach could be viewed in terms of predictive returns, the sampling guidelines are arbitrary and could significantly impact the results. Given, that the main interest is to account 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 the algorithm are as follows. For a family of utility functions $(u_{param,1}, ..., u_{param,K})$, 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 $\bar{\omega}_k = 1/n \sum \omega_i$ define the optimal portfolio.

By Jensen's inequality, if $\omega_k^* \neq \bar{\omega}_k$, then for a large class of utility functions

$$E[u_{param,k}(\omega_k^*, \theta)|x^o] \neq E[u_{param,k}(\bar{\omega}_k, \theta)|x^o]. \tag{7}$$

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 $\omega^{**} \neq \omega^*$. From (7), clearly

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

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'.

3.1.3 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 sampling models and in the assumed utility functions. In order to illustrate the impact of ignoring skewness, Figure 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

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

where m_e, V_e, S_e are the empirical mean, variance and skewness, Figure 5 contrasts the optimal portfolios that result from assuming an investor only has an aversion to variance ($\lambda = 0.5, \gamma = 0$) and has both an aversion to variance 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.

3.2 Bayesian Optimization Methods

Bayesian methods offer a natural framework for both, the evaluation of expectations and the optimization of expected utilities for an arbitrary utility function, with respect to an arbitrarily complicated probability model. Given an appropriate Markov chain Monte Carlo (MCMC) estimation routine, it is straightforward and computationally trivial to generate draws from the posterior predictive density, or to draw by computer simulation

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

and then evaluate the predictive summaries $\mathcal{X}_i$. Given a set of n draws, the expected utility for an arbitrary utility function can be estimated as an ergodic average, or

$$EU(\omega) = E[u(\omega, \mathcal{X})|x^o] \approx \frac{1}{n}\sum u(\omega, \mathcal{X}_i).$$

The approximate expected utility can then be optimized numerically using a number of different approaches. One attractive algorithm is the Metropolis-Hastings (MH) algorithm. MH simulation is widely used for posterior simulation. But the same algorithm can be exploited for expected utility optimization. When used for formal optimization, it is known as simulated annealing. Stopping short of simulated annealing, we use the MH algorithm to explore the expected utility function, $f(\omega) = E[U(\omega)]$, as a function of the weights. Asymptotic properties of the MH chain lead to portfolio weights ω being generated with frequencies proportional to $EU(\omega)$. That is, promising portfolio weights with high expected utility are visited more often, as desired. See, for example, Gilks et al. (1996) and Meyn and Tweedie (1993) for a discussion of the MH algorithm. 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 non-negative and integrable, the MH can be used to numerically explore any 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, we need to ensure that our expected utility is non-negative and integrable. For the linear utility functions, integrability over the space of possible 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 non-negative function by subtracting the minimum expected utility, or the target function becomes

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

4. OPTIMAL PORTFOLIOS IN PRACTICE

In theory, simplifications of 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 (3) and (5), which have expected utilities given in terms of the predictive posterior and posterior moments respectively, see (4) and (6). 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.

4.1 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 technology 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.

4.2 Model Choice and Select Summary Statistics

To determine which skew normal model best fit 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 skew normal model proposed by Sahu et al. (2003) with both a diagonal and our modified full Δ matrix. The results for the technology stocks shows that the skew normal models with a diagonal Δ outperform the other models, with the Sahu et al. (2003) model fitting best. The skew normal model with the full Δ, however, performs better than the others in the case of the benchmark indices (see Table 1). The model with the full Δ accommodates coskewness, 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 coskewness than portfolios of highly diversified global summaries.

Insert Table 1 about here.

The posterior parameter estimates for μ, Σ, and Δ, for both the technology stocks and the global asset allocation benchmark indices are given in Tables 2 and 3. 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 elements tend to be negative.

Insert Tables 2 and 3 about here

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 skew normal (higher moment) probability model. For the two moment model, we consider two linear utility functions - see (4) and (6) - 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 (Table 4). 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 ($\lambda = 0.5$, $\gamma = 0.5$) (Table 5). The weights that resulted from each optimization were then used to calculate the expected predictive utility.

Insert Tables 4 and 5 about here.

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 technology sector 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 6. It could be argued that the weights from Michaud's 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 variance 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.

Insert Table 6 **about here.**

Not surprisingly, when there is only preference for skewness, the predictive optimization approach outperforms the parameter approach, *e.g.* $\lambda = 0, \gamma = 0.5$ in Table 4. 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 3.1. In the case of the technology 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 6.

Ignoring the weights from Michaud's approach, the biggest differences in expected predicted utilities comes from comparing optimizations using the normal model versus the skew normal model, see $\lambda = 0.5$ and $\gamma = 0.5$ in Table 4. For the benchmark indices, the optimal allocation changes markedly when skewness is estimated and included in the utility, see Table 7. 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.

Insert Table 7 about here.

5. CONCLUSION

Considering both higher moments and parameter uncertainty is important in portfolio selection. Up to now these issues have been treated separately. The multivariate normal distribution is an inappropriate probability model for portfolio returns primarily because it fails to allow for higher moments, in particular skewness and coskewness. We also demonstrate that the skew normal model of Sahu et al. (2003) is able to capture these higher moments. 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. We use Bayesian methods 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. In addition to the Jensen's inequality problem, we show that the resampling approach is outside the efficient utility maximization framework.

While we believe that we have made progress on two important issues in portfolio selection, there are at least three limitations to our approach. First, our information is restricted to past returns. That is, investors make decisions based on past returns and do not use other conditioning information such as economic variables that tell us about the state of the economy. Second, our exercise is an 'in-sample' portfolio selection. We have not applied our method to out-of sample portfolio allocation. Finally, the portfolio choice problem we examine is a static one. There is a growing literature that considers the more challenging dynamic asset allocation problem that allows for portfolio weights to change with investment horizon, labor income and other economic variables.

We believe that it is possible to make progress in future research on the first two limitations. In addition, we are interested in using revealed market preferences to determine whether 'the market' empirically exhibits preference for skewness. As a first step, we plan to use the observed market weights for a benchmark equity index and use the predictive utility function (3) to determine the implied market λ and γ. Finally, we intend to consider modifications to (3) that allow for asymmetric preferences over positive and negative skewness.

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.*

Arrow, K. J. (1971), "Essays in the Theory of Risk-Bearing," *Chicago: Markham Publishing Co.*

Athayde, G. M. de, and Flôres, R. G. Jr. (2003), "Advances in Portfolio Construction and Implementation", Satchell and Scowcroft, eds.

——— (2004), "Finding a maximum skewness portfolio–a general solution to three moments portfolio choice," *Journal of Economic Dynamics and Control,* vol. 28, issue 7, 1335–1352.

Azzalini, A., and Dalla Valle, A. (1996), "The Multivariate Skew-Normal Distribution," *Biometrika*, 83, 715–726.

Berger, J. (1985), "Statistical Decision Theory and Bayesian Analysis," *New York, NY: Springer Veralg.*

Black, F., and Litterman, R. (1992), "Global portfolio optimization," *Financial Analysts Journal,* September-October 1992.

Branco, M. D., and Dey, D. K. (2001), "A General Class of Multivariate Skew-Elliptical Distributions," *Journal of Multivariate Analysis*, 79, 99–113.

Britten-Jones, M. (2002),"Portfolio optimization and Bayesian regression," *Unpublished working paper,* London Business School.

Brown, S. J. (1978), "The portfolio choice problem: Comparison of certainty equivalent and optimal Bayes portfolios," *Communications in Statistics: Simulation and Computation* B7, 321–334.

Cascon, A., Keating, C., and Shadwick, W. F. (2003), "The Omega Function," *working paper.*

Cremers, J-H., Kritzman, M., and Page, S. (2004), "Portfolio Formation with Higher Moments and Plausible Utility," Working paper.

Damien, P. S., Wakefield, J., and Walker, G. (1999), "Gibbs Sampling For Bayesian Nonconjugate and Hierarchical Models Using Auxilliary Variables," *Journal of Royal Statistical Society,* Ser. B, 61, Part 2, 331-344.

DeGroot, M. H. (1970), "Optimal Statistical Decisions," *New York: McGraw-Hill.*

Feiring, B. R., Wong, W., Poon, M., and Chan, Y. C. (1994), "Portfolio Selection in Downside Risk Optimization Approach: Application to the Hong Kong Stock Market," *International Journal of Systems Science,* **25**, 1921–1929.

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.

——— (1988), "For better performance; Constrain portfolio weights," *Journal of Portfolio Management,* 14, 29–34

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996), "Markov Chain Monte Carlo in Practice," *Chapman and Hall, London.*

Harvey, C. R., and Siddique, A. (2000a), "Conditional Skewness in Asset Pricing Tests," *Journal of Finance,* 55, 1263–1295

——— (2000b), "Time-Varying Conditional Skewness and the Market Risk Premium," *Research in Banking and Finance,* 1, 27–60.

Jobson, J. D., Korkie, B., and Ratti, V. (1979), "Improved estimation for Markowitz portfolios using James-Stein type estimators," *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 279–284.

Jobson, J. D., and Korkie, B. (1980), "Estimation of Markowitz efficient portfolios," *Journal of the American Statistical Association* 75, 544–554.

Johannes, M., Polson, N., and Stroud, J. (2002), "Sequential Optimal Portfolio Performance: Market and Volatility Timing," *Unpublished working paper.*

Jorion, P. (1992), "Portfolio Optimization in Practice," *Financial Analysts Journal*, January/February

68–74.

—— (1986), "Bayes-Stein Estimation of Portfolio Analysis," *Journal of Financial and Quantitative Analysis*, 21, 279–292.

Kan, R., and Zhou, G. (2004), "Optimal Estimation for Economic Gains: Portfolio Choice with Parameter Uncertainty," *Working paper,* Washington University, St. Louis.

Kandel, S., and Stambaugh, R. F. (1996), "On the predictability of stock returns: An asset allocation perspective," *Journal of Finance*, 51, 2, 385–424.

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.

Konno, H., Shirakawa, H., and Yamazaki, H. (1993), "A Mean-Absolute Deviation-Skewness Portfolio Optimization Model," *Annals of Operations Research,* **45**, 205–220.

Levy, H., and Sarnat, M. (1984), "Portfolio and Investment Selection: Theory and Practice," *Prentice-Hall International.*

Markowitz, H., (1952a), "Portfolio Selection," *Journal of Finance*, **7**, 77–91.

——— (1952b), "The Utility of Wealth," *Journal of Political Economy* 152–158.

——— (1959), "Portfolio Selection: Efficient Diversification of Investments," Second Edition. *Malden: Blackwell.*

Markowitz, H., and Usmen, N. (2003), "Resampled Frontiers vs Diffuse Bayes: An Experiment," *Journal of Investment Management,* 1,**4**, 9–25

Markowitz, H., Todd, P., Xu, G., and Yamane, Y. (1993), "Computation of Mean-Semivariance Efficient Sets by the Critical Line Algorithm," *Annals of Operations Research,* **45**, 307–317.

Michaud, R. (1998), "Efficient Asset Management," *Cambridge: Harvard Business School Press.*

Mills, T. C. (1995), "Modelling Skewness and Kurtosis in the London Stock Exchange FT-SE Index Return Distributions," *The Statistician,* 44(3), 323–332.

Meyn, S. P., and Tweedie, R. L. (1993), "Markov Chains and Stochastic Stability," *Springer-Verlag.*

Newton, M. A., and Raftery, A. E. (1994), "Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion)," *Journal of the Royal Statistical Society, Series B 56*, 3-48.

Novshek, W. (1993), "Mathematics for Economists," *Academic Press New York*

O'Hagan, A. (1994), "Kendall's Advanced Theory of Statistics: Volume 2B Bayesian Inference," *New York, NY: John Wiley and Sons.*

Pástor, L. (2000), "Portfolio selection and asset pricing models," *Journal of Finance*, 55, 179–223.

Pástor, L., and Stambaugh, R. F. (2000), "Comparing asset pricing models: An investment perspective," *Journal of Financial Economics*, 56, 335–381.

——— (1999), "Costs of equity capital and model mispricing," *Journal of Finance*, 54, 67–121.

Peiro, A. (1999), "Skewness in Financial Returns," *Journal of Banking and Finance,* 23, 847–862.

Polson, N.G., and Tew, B.V. (2000), "Bayesian Portfolio Selection: An Empirical Analysis of the S&P 500 Index 1970-1996," *Journal of Business and Economic Statistics,* 18, 2, 164–73.

Premaratne, G., and Tay, A. S. (2002), "How Should We Interpret Evidence of Time Varying Conditional Skewness?," National University of Spingapore.

Raiffa, H., and Schlaifer, R. (1961), "Applied Statistical Decision Theory," *Cambridge: Harvard University Press.*

Roy, A. D. (1952), "Safety First and the Holding of Assets," *Econometrica* 20, 431–449.

Sahu, S. K., Branco, M. D., and Dey, D. K. (2003), "A New Class of Multivariate Skew Distributions with Applications to Bayesian Regression Models," *Canadian Journal of Statistics,* .

Scherer, B. (2002), "Portfolio Resampling: Review and Critique," *Financial Analysts Journal,* Vol.

58, No. 6.

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 Price, L. N. (1994), “Performance Measurement in a Downside Risk Framework,” *Journal of Investing,* Vol. 3, No. 3, 50–58.

Treynor, J. L., and Black, F. (1973), “How to use security analysis to improve portfolio selection,” *Journal of Business,* January, 66-88.

Zellner, A., and Chetty, V. K. (1965), “Prediction and decision problems in regression models from the Bayesian point of view,” *Journal of the American Statistical Association,* 60, 608-616.

Appendix: Skew normal probability model

A.1 Density and Moment Generating Function

The likelihood function and moment generating function given in Sahu et al. (2003) changes when we allow Δ to be a full matrix:

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

where ϕ_{ℓ} is the $\ell-$dimensional multivariate normal density function with mean zero and identity covariance, and Φ_{ℓ} is multivariate normal cumulative distribution also with mean zero and identity covariance.

The moment generating function becomes

$$M_{\mathbf{Y}}(\mathbf{t}) = 2^{\ell}e^{t'\mu+t'(\Sigma+\Delta\Delta')t/2}\Phi_{\ell}(\Delta t) \quad \text{(A-2)}$$

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

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

$$S = \Delta EZ\Delta'\otimes\Delta' + 3\mu'\otimes\{\Delta\Delta'(1-2/\pi)+2/\pi\Delta\mathbf{1}(\Delta\mathbf{1})'\} + \\ 3\{(2/\pi)^{1/2}(\Delta\mathbf{1})'\otimes[\Sigma+\mu\mu']\} + 3\mu'\otimes\Sigma \\ +\mu\mu'\otimes\mu' - 3m'\otimes V - m\,m'\otimes m', \quad \text{(A-3)}$$

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

$$\mathbf{EZ} = \begin{pmatrix} E[Z_1Z_1Z_1] & \dots & E[Z_1Z_1Z_{\ell}] & \dots & E[Z_{\ell}Z_1Z_1] & \dots & E[Z_{\ell}Z_1Z_{\ell}] \\ \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots \\ E[Z_1Z_{\ell}Z_1] & \dots & E[Z_1Z_1Z_{\ell}] & \dots & E[Z_{\ell}Z_{\ell}Z_1] & \dots & E[Z_{\ell}Z_{\ell}Z_{\ell}] \end{pmatrix}$$

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 a Linear Combination

Assume $X \sim SN(\mu, \Sigma, \Delta)$ and a set of constant portfolio weights $\omega = (\omega_1, \ldots, \omega_\ell)'$, the first three moments of $\omega' X$ are as follows

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

$$Var(\omega' X) = \omega' V \omega$$

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

where m, V and S are given above.

A.3 Lévy Stable Skew Normal

When there is a Z_i for each observation x_i, the moment generating function readily shows that the skew normal distributions of Sahu et al. (2003) and Azzalini and Dalla Valle (1996) are not Lévy stable. If, however, the latent variables Z are restricted to be time invariant, *i.e.* a single Z for all of the observations, then both models are Lévy stable. In addition, the Azzalini and Dalla Valle (1996) model maintains the property that the distribution of the portfolio is also skew normal.

A.4 Model Specification

0.4.1 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

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

where N_ℓ is a multivariate normal density,

$$Z_i \sim N_\ell(0, I_\ell)I\{Z_{ij} > 0\}, \text{ for all } j,$$

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

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

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

Full Conditionals.

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

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

where

$$A = I_\ell + \Delta'\Sigma^{-1}\Delta \text{ and } a = \sum_{i=1}^{n} \Delta'\Sigma^{-1}(x_i - \mu),$$

$$B = \sum_{i=1}^{n} y_i' \Sigma^{-1} y_i + \frac{1}{100} I_{\ell(\ell+1)} \text{ and } b = \sum_{i=1}^{n} y_i \Sigma^{-1} x_i,$$
$$C = \sum_{i=1}^{n} (x_i - (\mu + \Delta Z_i))(x_i - (\mu + \Delta Z_i))' + \ell I_\ell,$$
where $y_i = (I_\ell, Z_i' \otimes I_\ell)$.

A.5 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, Wakefield, and Walker (1999). To illustrate, assume that we want to sample from the following density,

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

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

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

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.

Table 1:

Evaluating the Distributional Representation of Four Equity Securities and global asset allocation benchmarks

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. And also 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. (2003) 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. Sahu et al. (2003) diagonal Δ model fits best overall.

a. Four equity securities

Moments	Distribution	Δ	Log(BF)	Max Log Likelihood
MV	normal	–	0.00	-2833.82
MVS	skew normal A	diagonal	2225.75	-2561.15
MVS	skew normal B	diagonal	2368.33	-2418.85
MVS	skew normal B	full	2192.83	-2595.56

b. Global asset allocation benchmarks

Moments	Distribution	Δ	Log(BF)	Max Log Likelihood
MV	normal	–	0.00	-2319.00
MVS	skew normal	diagonal	1946.11	-2096.26
MVS	skew normal	diagonal	2015.11	-2067.73
MVS	skew normal	full	2034.22	-2200.30

Table 2:

Parameter estimates for diagonal Δ skew normal on four securities

Parameter estimates for the diagonal Δ model of Sahu et al. (2003) 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 ($\mu+(2/\pi)^{1/2}\Delta\mathbf{1}$ and multiplied by 100), variance ($\Sigma + (1 - 2/\pi)\Delta\Delta'$), and skewness (see Appendix A.1 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

Table 3:
Parameter estimates for full Δ skew normal on global asset allocation benchmark
Parameter estimates for full Δ model of Sahu *et al.* (2003) 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*.1 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	-0.279	-0.567

Table 4:
Two and three moment optimization for four equity securities
This table contains predictive utilities for the weights that maximize utility as a linear function of the two and three 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 $\omega_{i,pred}$ where the i represents the number of moments in the model), the second is based on the posterior parameter estimates ($\omega_{i,param}$), and the third is the method proposed by Michaud ($\omega_{i,Michaud}$). 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, $\omega_{i,pred}$ gives the highest expected utility.

a. Two moments

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	0.123	0.123	0.113
0	0.5	0.098	0.098	0.011
0.5	0	-1.745	-1.758	-1.756
0.5	0.5	-1.733	-1.749	-1.75

b. Three moments

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	0.123	0.123	0.112
0	0.5	0.109	0.098	0.075
0.5	0	-1.745	-1.745	-1.802
0.5	0.5	-1.731	-1.732	-1.736

Table 5:

Two and three moment optimization for global asset allocation benchmark indices

Predictive utilities for the weights that maximize utility as a linear function of the two and three 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 $\omega_{i,pred}$ where the i represents the number of moments in the model), the second is based on the posterior parameter estimates ($\omega_{i,param}$), and the third is the method proposed by Michaud ($\omega_{i,Michaud}$). 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, $\omega_{i,pred}$ gives the highest expected utility.

a. Two moments

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	0.254	0.251	0.238
0	0.5	0.224	0.221	0.212
0.5	0	-0.647	-0.662	-0.674
0.5	0.5	-0.649	-0.666	-0.676

b. Three moments

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	0.254	0.252	0.24
0	0.5	0.223	0.223	0.211
0.5	0	-0.647	-0.647	-0.651
0.5	0.5	-0.649	-0.649	-0.656

Table 6:

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
$\omega_{3,pred}$	0.0009	0.0001	0.0005	0.9985
$\omega_{3,param}$	0.0017	0.0006	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	0
$\omega_{3,param}$	0.0004	0.0001	0.0006	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 7:

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	0	0.497	0.1	0.0984	0.183	0.0006
$\omega_{2,param}$	0.0015	0.0023	0.632	0.1	0.0777	0.111	0.075
$\omega_{2,Michaud}$	0.157	0.0097	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	0.0002	0.439	0.121	0.0874	0.193	0.0027
$\omega_{3,param}$	0.149	0.0002	0.461	0.0967	0.0905	0.18	0.0233
$\omega_{3,Michaud}$	0.225	0.0005	0.407	0.0885	0.0808	0.15	0.0479

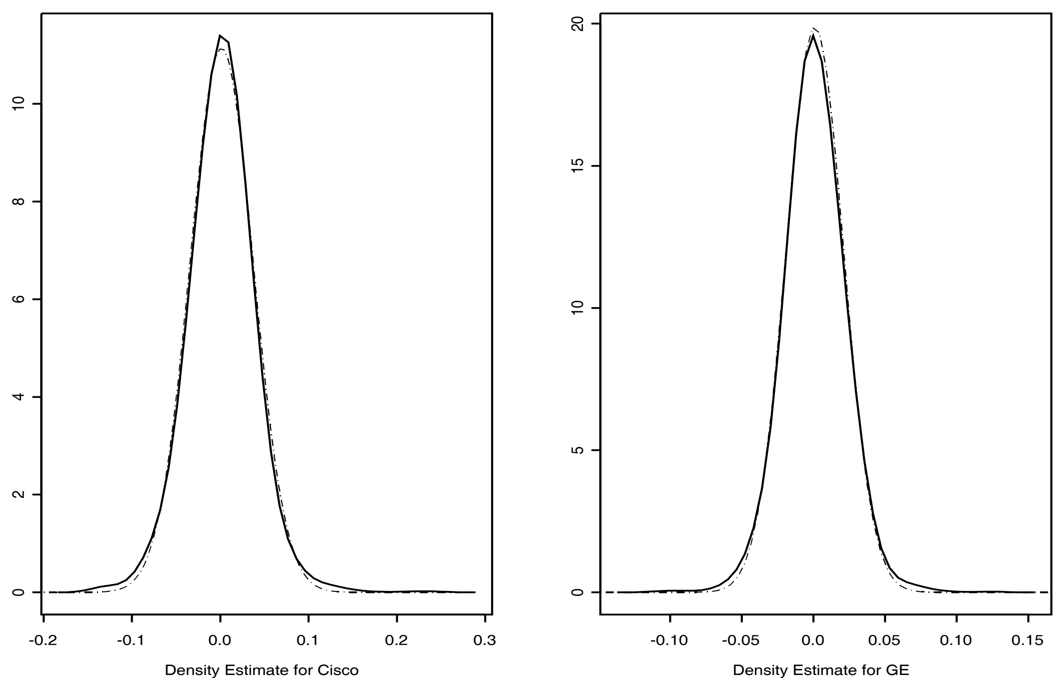

Figure 1: This figure contains univariate estimates for Cisco Systems and General Electric daily stock returns from April 1996 to March 2002. The solid lines represents the kernel density estimate, while the dotted lines are the normal density with sample mean and variance. In one dimension the normal distribution closely matches the returns for these two stocks.

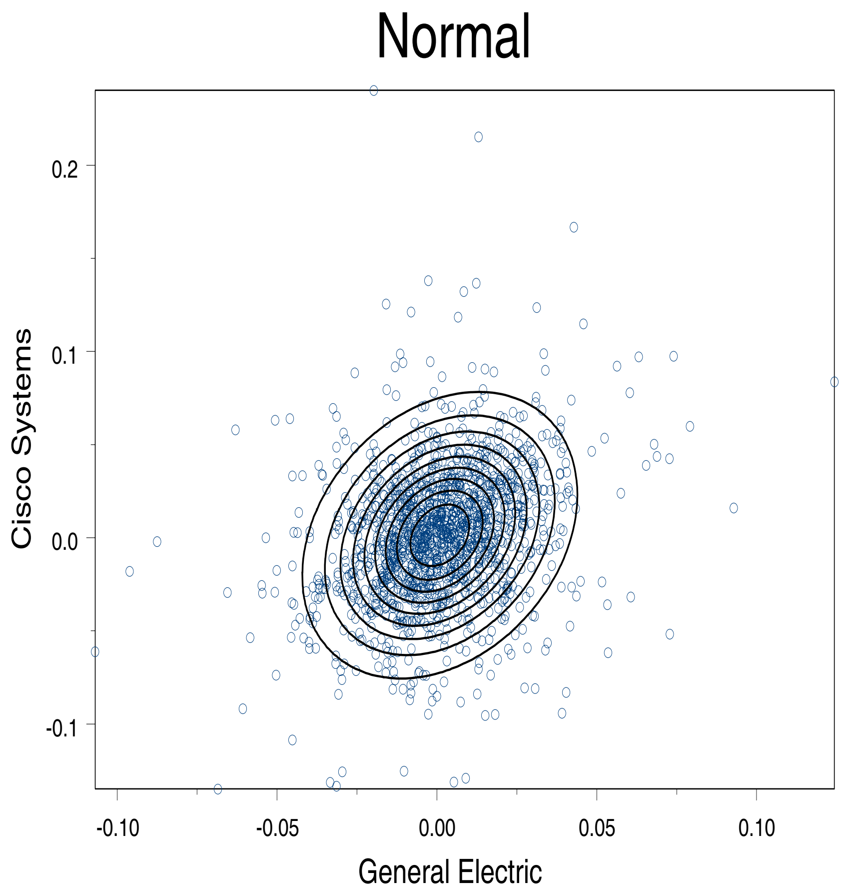

Figure 2: This figure contains a bivariate normal estimate for Cisco Systems and General Electric daily stock returns from April 1996 to March 2002. The plot is a bivariate normal with sample mean and covariance. The scatter points are the actual data. Unlike in one dimension, in two dimensions the normal distribution does not closely match these joint returns. The actual returns exhibit coskewness and much fatter tails than the normal approximation.

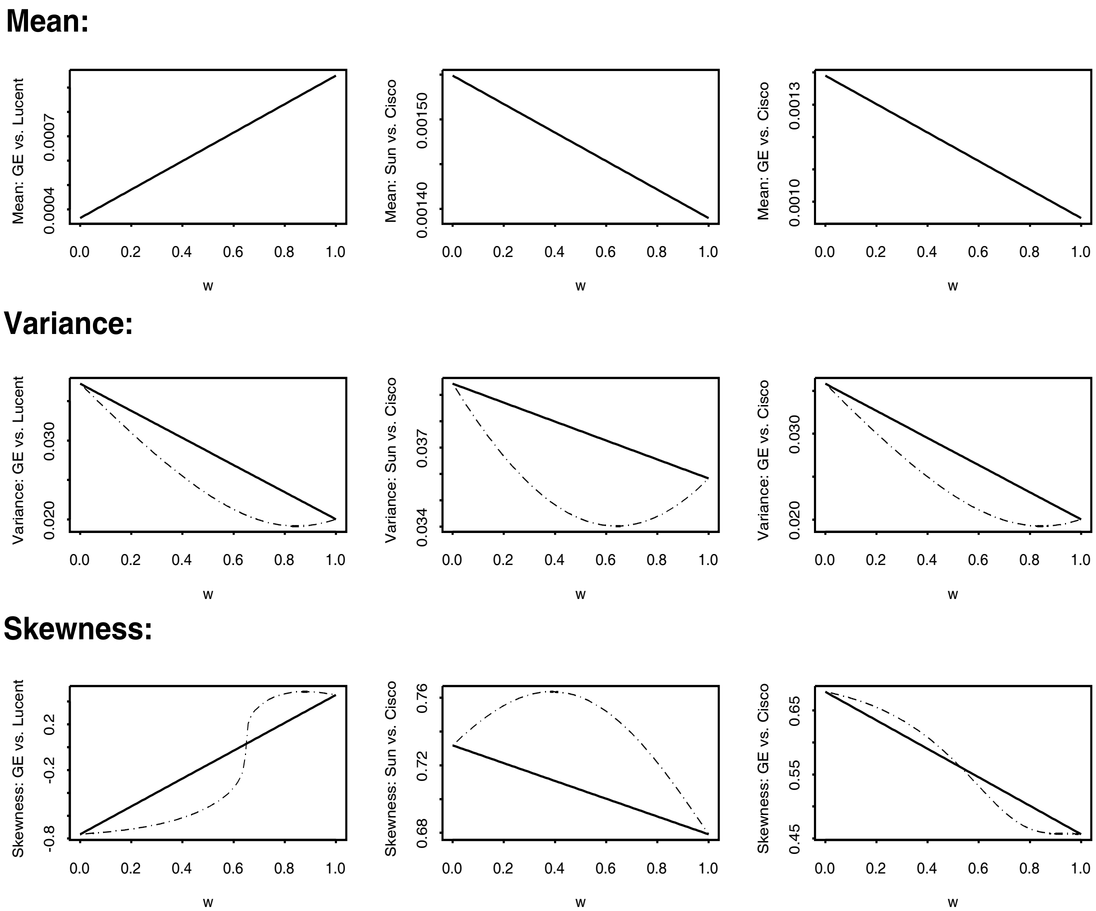

Figure 3: This figure contains plots of the mean, variance and skewness of portfolios consisting of two assets. Daily returns from April 1996 to March 2002 for General Electric and Lucent Technologies, Sun Microsystems and Cisco Systems, and General Electric and Cisco Systems are considered. The top 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 alway less or equal to the variance of the linear combination. 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. Selecting a portfolio based solely on minimum variance could lead to a portfolio with minimum skewness as well (see GE vs. Cisco).

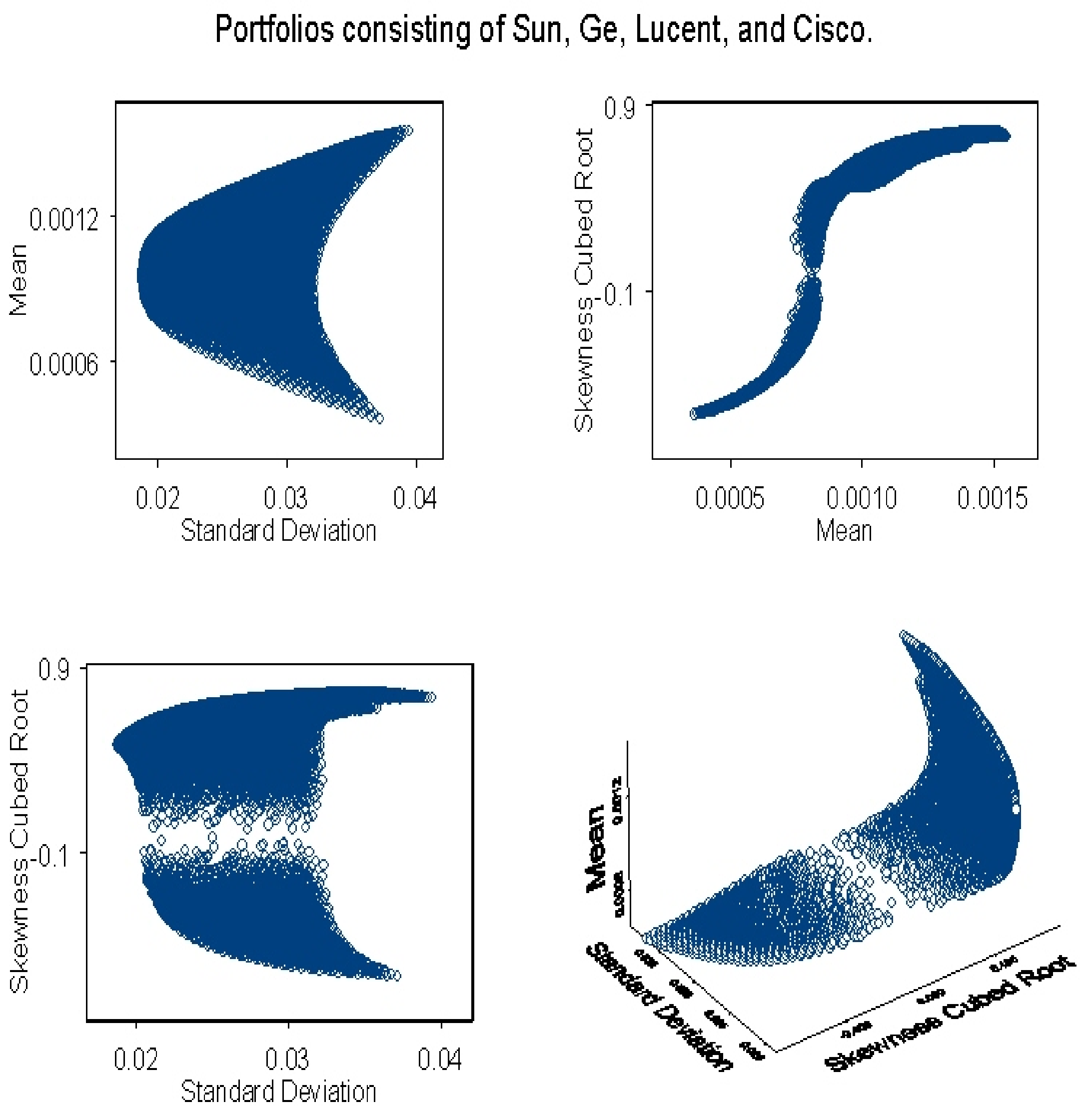

Figure 4: This figure shows the space of possible portfolios based on historical parameter estimates from the daily returns of General Electric, Lucent Technologies, Cisco Systems, and Sun Microsystems from April 1996 to March 2002. The top left plot is the mean-standard deviation space, the top right 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.

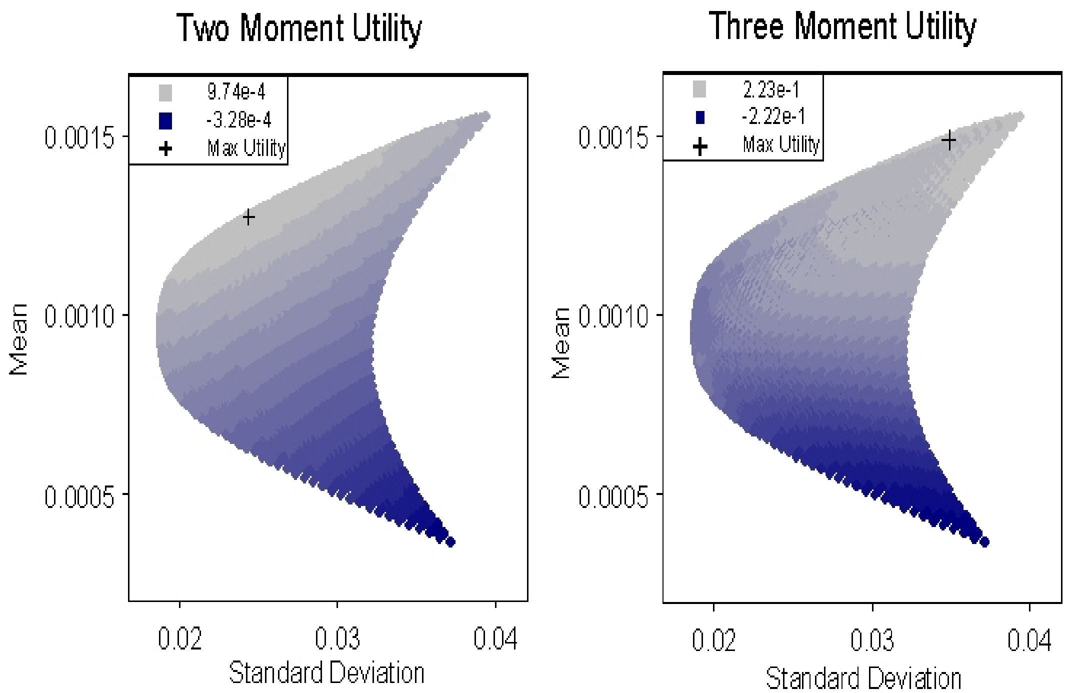

Figure 5: This figure shows the mean-variance space of possible portfolios based on historical parameter estimates from the daily returns of General Electric, Lucent Technologies, Cisco Systems, and Sun Microsystems from April 1996 to March 2002. The portfolios are shaded according to the utility associated 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 the 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 '+'.