Portfolio Selection with Higher Moments

Campbell R. Harvey

Duke University, Durham, NC 27708 USA

National Bureau of Economic Research, Cambridge, MA 02138 USA

John C. Liechty

Pennsylvania State University, University Park, PA 16803 USA

Merrill W. Liechty[*]

Duke University, Durham, NC 27708 USA

Peter Müller

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

July 17, 2003

Abstract

We build on the Markowitz portfolio selection process by incorporating higher order moments of the assets, as well as utility functions based on predictive asset returns. We propose the use of the skew normal distribution as a characterization of the asset returns. We show that this distribution has many attractive features when it comes to modeling multivariate returns. Preference over portfolios is framed in terms of expected utility maximization. We discuss estimation and optimal portfolio selection using Bayesian methods. These methods allow for a comparison to other optimization approaches where parameter uncertainty is either ignored or accommodated in a non-traditional manner. Our results suggest that it is important to incorporate higher order moments in portfolio selection. Further, we show that our approach leads to higher expected utility than the resampling methods common in the practice of finance.

KEYWORDS: Bayesian statistics, multivariate skewness, parameter uncertainty, portfolio selection, utility function maximization.

JEL Classification:

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 paper 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, we do.

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

---

[1]See Markowitz (1952, p.91)

proach, 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 become less obvious that an investor can easily interpret the geometry of the frontier and reasonably select a portfolio. Other three moment optimization methods include using negative semivariance inplace of variance (see Markowitz 1959, and Markowitz *et al.* 1993). A similar measure of downside risk is also incorporated by Feiring *et al.* (1994), and Konno *et al.* (1993) use an approximation to the lower semi-third moment in their Mean-Absolute Deviation-Skewness portfolio model. 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. Johannes *et al.* (2002) investigate market timing using a model with correlation between volatility and returns in a Bayesian portfolio selection setting.

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. Jobson, Korkie and Ratti (1979) and Jobson and Korkie (1980) point out these problems, and suggest the use of shrinkage estimators, also see Simbea and Mulvey (1998). This 'estimation risk' comes from both choosing poor probability models and from ignoring parameter uncertainty, 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. Jorion (1992) proposes a resampling method aimed at 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 Bayesian perspective, Treynor and Black (1973) propose the basic idea of using investors' view in combination with historical data. Jorion (1986) and Frost and Savarino (1986, 1988) use a traditional Bayesian shrinkage approach. Zellner

2

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). 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' (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.

Our approach advances previous methods by addressing both higher moments and estimation risk in a consistent 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 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 our estimation and optimization approach, 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, using a sampling distribution instead of a predictive distribution and assuming that the expected utility equals the utility of the expected values. In addition, we demonstrate the loss in expected utility (explained by Jensen's inequality) that comes from using the widely used approach proposed by Michaud (1998).

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

3

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 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 know 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 symmetric, see Kon (1984), Mills (1995), Peiro (1999) and Premaratne and Bera (2002).[2] Our 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 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 contour plots based on their joint returns.

Insert Figures 1 and 2 here

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.

---

[2]See 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)

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) do an empirical test on mutual funds and find a strong preference for positive skewness. 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. 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 portfo-

---

[3]Also 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), Chunhachinda *et al.* (1997), Fang and Lai (1997), Faff (1998), Adcock and Shutes (1999), Sornette *et al.* (2000a,b), Barone-Adesi *et al.* (2001), Bhattacharyya (2002), Dittmar (2002), Jurczenko and Maillet (2001), Malevergne and Sornette (2002), and Polimenis (2002)

lios), 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 portfolio 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 guaranteed 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.

Insert Figure 3 here

This variety 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 giving 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 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

6

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

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 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

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

where $I_{\{\cdot\}}$ is the indicator function and Z_j is the jth 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 an unrestricted 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]

However with additional restrictions on the truncated components, both models can be generalized as stable distributions (see Appendix A). The skew normal is similar in 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 $vec()$ 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

---

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

general discussion of the MCMC algorithm and the Gibbs sampler and see Appendix A 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. Several authors have discussed the advantages and appropriate uses of the Bayes factor; see Berger (1985) and O'Hagan (1994). In O'Hagan's (1994) discussion, he demonstrates the well-established relationship between the Bayesian Information Criteria (BIC), also known as the Schwartz Criteria, and the BF. As demonstrated, 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:

$$BF = \frac{\text{Posterior odds}}{\text{Prior odds}} = \frac{p(x|M_1)}{p(x|M_2)} \tag{3}$$

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

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.

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, or optimization criterion needs to target future returns and not be a function of model parameters, see Zellner and Chetty (1965), 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.

To illustrate, let x^o represent the current data up to the present and let x represent future data. Let $\mathcal{X} = (x, V_x, S_x)$ be predictive summaries, where $V_x = (x - \bar{m})(x - \bar{m})'$, $S_x = V_x \otimes (x - \bar{m})'$ and

$$\bar{m} = \int x \, p(x|x^o) dx$$

is the predictive mean given x^o. Assuming a linear utility function, the predictive utility is given by

$$u_{pred}(\omega, \mathcal{X}) = \omega' x - \lambda \left[\omega' (x - \bar{m})\right]^2 + \gamma \left[\omega' (x - \bar{m})\right]^3 \tag{4}$$

where λ and γ determine the impact of future variance and skewness. The expected utility, becomes

$$U_{pred}(\omega) = \omega' E[x|x^o] - \lambda \, \omega' E[V_x|x^o] \, \omega + \gamma \, \omega' E[S_x|x^o] \, \omega \otimes \omega = \omega' m_p - \lambda \, \omega' V_p \, \omega + \gamma \, \omega' S_p \, \omega \otimes \omega, \tag{5}$$

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), or

$$u_{param}(\omega, \theta) = \omega' m - \lambda \omega' V \omega + \gamma \omega' S \omega \otimes \omega \qquad (6)$$

where $\theta = (m, V, S)^5$ are parameters representing the first three moments of the probability model. The expected utility, becomes

$$U_{param}(\omega) = \omega' \bar{m} - \lambda \omega' \bar{V} \omega + \gamma \omega' \bar{S} \omega \otimes \omega, \qquad (7)$$

where $\bar{m}$, $\bar{V}$ and $\bar{S}$ are the posterior means of θ. 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

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

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

Substituting this into (5) gives an alternative form that is composed of $U_{param}(\omega)$ plus other terms.

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

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.

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,

---

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

1993) about the expected predictive summaries, or assume

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

For linear utility functions this approximation is exact. 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

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

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.[6] 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.[7] Given, that his main interest is with regards to

---

[6]The resampling technique that Michaud proposes is the same as the one proposed in Jorion (1992). However, Jorion does not claim that the resampled frontier is efficient. Jorion is able to use the resampling technique to gain important insight on the value of adding additional assets to a portfolio.

[7]Repeated estimates of predictive 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.

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 $(u_{param,1}, ..., u_{param,K})$, 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

$$E[u_{param,k}(\omega_k^*, \theta)|x^o] \neq E[u_{param,k}(\bar{\omega}_k, \theta)|x^o]. \tag{8}$$

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 $\omega^{**} \neq \omega^*$. From (8), 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 terms of the underlying probability model and 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 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.

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. 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

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

and then calculate draws of 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

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

The approximate expected utility can then be optimized numerically using a number of different approaches. One attractive algorithm, 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, we need to ensure that our expected utility is strictly positive 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 strictly positive function by subtracting the minimum expected utility, or the target function becomes

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

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 simplifica-

---

[8]Bielza *et al.* 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.

tions 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) and (6), which have expected utilities given in terms of the predictive posterior and posterior moments respectively, see (5) and (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.

4.1   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.

4.2   *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 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 shows that the skewed normal diagonal 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 1 and 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.

Insert Tables 1 **and** 2 **here**

The posterior parameter estimates for μ, Σ, and Δ, for both the high tech stocks and the global asset allocation benchmark indices are given in Tables 3 and 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.

Insert Tables 3 **and** 4 **here**

To illustrate the improvement that comes from using the best skewed normal model compared to a multivariate normal model, consider the bivariate summary of the actual returns, the predictive joint density (skewed normal) and the predictive joint density (normal) for Cisco Systems and General Electric, see Figure 6. The skewed normal density accommodates all of the features of the actual data including correlation, skewness, co- skewness and kurtosis (heavy tails); while the normal density only captures the correlation and significantly understates the tails of the distribution.

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 (5) and (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 5 and 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 7 and 8). The weights that resulted from each optimization were then used to calculate the expected predictive utility.

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 place on Sun Microsystems with the rest spread across the remaining stocks, see $\lambda = 0$ and $\gamma = 0$ from Table 9. 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 risk, then the appropriate portfolio would

18

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.

<div align="center">**Insert Table 9 here**</div>

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 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 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 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 5 and 6. For the benchmark indices, the optimal allocation changes markedly when skewness is estimated and included in the utility, see Tables 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.

<div align="center">**Insert Table 10 here**</div>

5   Conclusion

Considering both higher moments and parameter uncertainty is important in portfolio selection. Up to now these issues have been treated separately. We compare results with Michaud's application of the resampling technique. 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

<div align="center">19</div>

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 because of Jensen's inequality his averaging of weights produces sub-optimal portfolios, but also that his approach is outside the efficient utility maximization framework.

Using the framework we describe, we intend to investigate revealed market preferences and determine whether the 'market' empirically exhibits preference for skewness. As a first step we will use the observed market weights for a benchmark equity index and use the predictive utility function (4) to determine the implied market λ and γ. In addition, we intended to consider modifications to (4) 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.*

Adcock, C., and K. Shutes, 1999, Fat Tails and The Capital Asset Pricing Model. *Working Paper,* University of Bath, February 1999, 28 pages.

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., R.K.S. Rao and T. Hiraki, 1989, Skewness and Kurtosis in Japanese Equity Returns: Empirical Evidence, *Journal of Financial Research*, 12, 253–261

Arditti, F.D., 1971, Another Look at Mutual Fund Performance, *Journal of Financial and Quantitative Analysis*, 6 (3), 909–912.

Arditti, F.D., 1967, Risk and the Required Return on Equity, *Journal of Finance,* 22 (1), 19–36.

Arditti, F.D., and H. Levy, 1975, Portfolio Efficiency and Analysis in Three Moments: The Multiperiod Case, *Journal of Finance,* 30 (3), 797–809.

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

Athayde, G. M. de, and R.G. Flôres Jr., 2001, Finding a maximum skewness portfolio-a general solution to three moments portfolio choice, Fundacao Getulio Varas.

Avramov D., 2000, Stock return predictability and model uncertainty, *Unpublished working paper,* University of Maryland, College Park, MD.

Azzalini, A., and A. Capitanio, 1999, Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society,* B,61, 579–602.

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

Barone-Adesi, G., P. Gagliardini and G. Urga, 2001, Testing the Homogeneity of Asset Pricing Models, Universita' della Svizzera Italiana, Lugano, CH.

Barone-Adesi, G., 1985, Arbitrage Equilibrium with Skewed Asset Returns, *Journal of Financial and Quantitative Analysis,* 20 (3), 299–313.

Bates, D., 2001, The Market Price of Crash Risk, *working paper,* University of Iowa.

Bates, D., 2000, Post-'87 Crash fears in S&P 500 Futures Options, *Journal of Econometrics,* 94, 181–238.

Beedles, W., and M. Simkowitz, 1978, A Note on Skewness and Data Errors, *Journal of Finance,* 33 (1), 288–292.

Bera, A. and G. Premaratne, 2001, Modeling Asymmetry and Excess Kurtosis in Stock Return Data, *manuscript,* University of Illinois.

Bera, A., and G. Premaratne, 2002, A Test for Asymmetry with Leptokurtic Financial Data, *Unpublished working paper, University of Illinois, Champaign, IL.*

Berenyi, Z., 2002, Measuring Hedge Fund Risk with Multi-moment Risk Measures, *working paper,* University of Munich.

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

Bernardo, A.E. and O. Ledoit, 2000, Gain, Loss, and Asset Pricing, *Journal of Political Economy,* 108 (1), 144–172.

Bhattacharyya, N., 2002, From Mean Variance Space to Mean Skewness Space-Implications for Simultaneous Risk Seeking and Risk Averting Behavior, *Working Paper,* University of Manitoba.

Bielza C., P. Müller, D.R. Insua 1999 Decision analysis by augmented probability simulation, *Management Science.* 45, 995-1007

Blattberg, R., and N. Gonedes, 1974, A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices, *Journal of Business,* 47, 244–280.

Bond, S.A., 2001, Time Varying Skewness and Downside Risk Measurement: A Comparison of Models, University of Cambridge.

Bond, S.A., A review of asymmetric conditional density functions in autoregressive conditional heteroscedasticity models, *Return Distributions in Finance,* Ch. 2, 21–46.

Boyle, G.W. and R.K.S Rao, 1988, The Mean-Generalized Coefficient of Variation Selection Rule and Expected Utility Maximization, *Southern Economic Journal,* 55, 1, 1–8.

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

Brandt, M. W., A. Goyal, and P. Santa-Clara, 2001, Dynamic portfolio choice: A simulation approach, *Unpublished working paper, University of Pennsylvania, Philadelphia, PA.*

Brannas, K., and N. Nordman, 2001, Conditional Skewness Modeling for Stock Returns, Umea University.

Brennan, M. J. and Y. Xia, 2001, Assessing asset pricing anomalies, *Unpublished working paper, University of California at Los Angeles, Los Angeles, CA.*

Britten-Jones M., 1999, The sampling error in estimates of mean-variance efficient portfolio weights, *Journal of Finance,* 54, 655–671.

Britten-Jones M., 1998, Portfolio optimization and Bayesian regression, *Unpublished working paper,* London Business School.

Brockett, P., and Y. Kahane, 1992, Risk, Return, Skewness and Preference, *Management Science,* 38, 851–866.

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.

Brown, S.J., 1979 Optimal portfolio choice under uncertainty: a Bayesian approach *In Bawa, V.S., Brown, S.J., Klein, R.W. (Eds.), Estimation Risk and Optimal Portfolio Choice.* North-Holland, Amsterdam

Chung, Y.P., H. Johnson and M.J. Schill, 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., K. Danpani, S. Hamid and A. Prakash, 1997, Portfolio Selection and Skewness: Evidence from International Stocks Markets, *Journal of Banking and Finance,* 21, 143–167.

Corrado, C. And T. Su, 1996, Skewness and Kurtosis in S&P 500 Index Returns Implied by Option Prices, *Journal of Financial Research,* 19, 2, 175–192.

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

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 D. Pasternack, 2002, The Negative News Threshold- an Explanation for Negative Skewness in Stock Returns, *working paper,* Swedish School of Economics and Business Administration.

Faff, R., Y. Ho and L. Zhang, 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 T.-Y. Lai, 1997, Co-Kurtosis and Capital Asset Pricing, *Financial Review,* 32 (2), 293–307.

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

Friend, I., and R. Westerfield, 1980, Co-Skewness and Capital Asset Pricing, *Journal of Finance,* 35 (4), 897–913.

Frost, P.A., and J.E. Savarino, 1986, An Empirical Bayes Approach to Efficient Portfolio Selection, *Journal of Financial and Quantitative Analysis*, 21, No. 3, 293–305.

Frost, P.A., and J.E. Savarino, 1988, For better performance; Constrain portfolio weights, *Journal of Portfolio Management,* 14, 29–34

Goetzmann, W. N., J. E. Ingersoll Jr. and M. I. Spiegel, I. Welch, 2002, Sharpening Sharpe Ratios, *working paper,* Yale International Center For Finance.

Goldfarb, D. and G. Iyengar, 2001, Robust portfolio selection problems, *working paper.*

Greyserman, A., D. H. Jones, and W. E. Strawderman, 2002, Portfolio Selection Using Bayesian Analysis and Gibbs Sampling, *Unpublished.*

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

Harvey, C. R., and A. Siddique, 2000b, Time-Varying Conditional Skewness and the Market Risk Premium, *Research in Banking and Finance,* 1, 27–60.

Harvey, C. R., and A. Siddique, 1999, Autoregressive Conditional Skewness, *Journal of Financial and Quantitative Analysis,* 34, 4

Hueng, C.J., and R. Brooks, 2000, Forecasting Asymmetries in Stock Returns: Evidence from Higher Moments and Conditional Densities, University of Alabama.

Hwang, S., and S.E. Satchell, 1999, Modeling Emerging Market Risk Premia Using Higher Moments, *International Journal of Finance and Economics,* 4 (4), 271–296.

Ingersoll, J., 1987, Theory of Financial Decision Making, *Rowman & Littlefield, Totowa,* 474 pages.

Jean, W.H., 1973, More on multidimensional Portfolio Analysis, *Journal of Financial and Quantitative Analysis,* 8, 475–490.

Jean, W.H., 1971, The Extension of Portfolio Analysis to Three or More Parameters, *Journal of Financial and Quantitative Analysis,* 6 (1), 505–515.

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

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

Jondeau, E. and M. Rockinger, 2000, Conditional volatility, skewness, and kurtosis: existence and persistence, *manuscript,* HEC-School of Management.

Jones C., 1998, Bayesian estimation of continuous-time finance models, *Unpublished working paper,* University of Rochester, Rochester, NY.

Jorion P., 1992, Portfolio optimization in practice, *Financial Analysts Journal,* January/February 68–74.

Jorion P., 1991, Bayesian and CAPM estimators of the means: Implications for portfolio selection, *Journal of Banking and Finance,* 15, 717–727.

Jorion P., 1986, Bayes-Stein estimation of portfolio analysis, *Journal of Financial and Quantitative Analysis*, 21, 279–292.

Jorion P., 1985, International portfolio diversification with estimation risk, *Journal of Business*, 58, 259–278.

Jurczenko, E., and B. Maillet, 2001, The Three-moment CAPM: Theoretical Foundations and an Asset Pricing Model Comparison in a Unified Framework, *In: Dunis, Christian et. al. (Eds), Developments in Forecast Combination and Portfolio Choice,* John Wiley and sons, 2001.

Kahneman, D., and A. Tversky, 1979, Prospect Theory: An Analysis of Decision under Risk, *Econometrica,* 47, 263–291.

Kallberg J. G. and W. T. Ziemba 1983 Comparison of Alternative Functions in Portfolio Selection Problems, *Management Science*, 11, 1257–1267.

Kandel, S., R. McCulloch and R. F. Stambaugh, 1995, Bayesian inference and portfolio efficiency, *Review of Financial Studies*, 8, 1–53.

Kandel, S., and R. F. Stambaugh, 1995, Portfolio inefficiency and the cross-section of expected returns, *Journal of Finance*, 50, 157–184.

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

Kane, A., 1982, Skewness Preference and Portfolio Choice, *Journal of Financial and Quantitative Analysis,* 17 (1), 15–25.

Klein, R. W., and V. Bawa, 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., H. Shirakawa, and H. Yamazaki, 1993, A Mean-Absolute Deviation-Skewness Portfolio Optimization Model, *Annals of Operations Research,* **45**, 205–220.

Kozik, T.J., and A.M. Larson, 2001, The N-Moment Insurance CAPM, Allstate Insurance Company.

Kraus, A., and R. H. Litzenberger, 1976, Skewness preference and the valuation of risk assets, *Journal of Finance,* 31, 1085–1100.

Kroll, Y., H. Levy and H. Markowitz, 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 R.K.S. Rao, 1988, Mean-Lower Partial Moment Valuation with Lognormally Distributed Returns, *Management Science,* Vol. 34, No. 4, 446–453.

Levy, H., 1969, An Utility Function Depending on the First Three Moments, *Journal of Finance,* 24, 715–719.

Levy, H., and H. Markowitz, 1979, Approximating Expected Utility by a Function of Mean and Variance, *American Economic Review,* 69, 308–317.

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

Lewellen, J., and J. Shanken, 2002, Learning, Asset-Pricing Tests, and Market Efficiency, *Journal of Finance,* 57, 3, 1113–1146.

Lhabitant, F., 1997, On the (Ab)use of Expected Utility Approximations for Portfolio Selection and Portfolio Performance, *Working Paper,* University of Lausanne, 23 pages.

Lim, K.-G., 1989, A New Test of the Three-Moment Capital Asset Pricing Model, *Journal of Financial and Quantitative Analysis,* 205–216.

Liu, J., F. A. Longstaff, and J. Pan, 2002, Dynamic Asset Allocation With Event Risk, *Journal of Finance,* forthcoming.

Liu, J., J. Pan, and T. Wang, 2002, An Equilibrium Model of Rare Event Premia, *working paper,* Massachusetts Institute of Technology.

Machado-Santos, C., and A.C. Fernandes, 2002, Skewness In Financial Returns: Evidence From The Portuguese Stock Market, UTAD University.

Malevergne, Y., and D. Sornette, 2002, Multi-Moments Method for Portfolio Management: Generalized Capital Asset Pricing Model in Homogeneous and Heterogeneous markets, *Laboratoriere de Physique de la Matiere Condensee.*

Markowitz, H., 1952, Portfolio Selection, *Journal of Finance*, 7, 77–91.

Markowitz H., 1959, Portfolio Selection: Efficient Diversification of Investments. Second Edition. *Malden: Blackwell.*

Markowitz H., P. Todd, G. Xu, and Y. Yamane, 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 R.L. Tweedie, 1993, Markov Chains and Stochastic Stability, *Springer-Verlag.*

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

Nietert, B., 2001, Pricing Relevance of Stock Price Jumps: CAPM and Variance Decomposition, 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 Your, 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 R. F. Stambaugh, 2000, Comparing asset pricing models: An investment perspective, *Journal of Financial Economics*, 56, 335–381.

Pástor, L., and R. F. Stambaugh, 1999, Costs of equity capital and model mispricing, *Journal of Finance*, 54, 67–121.

Pederson, C., and S. Satchell, 2001, Asymmetric Equilibrium Risk Measures, *Journal of Empirical Finance,* 2001, forthcoming.

Pederson, C., and S. Satchell, 1998, An Extended Family of Financial Risk Measures, *Geneva Papers on Risk and Insurance Theory,* 23, 89–117.

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

Perez-Quiros, G. and A. Timmermann, 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, *working paper,* University of California, Riverside, CA.

Post, T., H. Levy, and P. van Vliet, 2003, Risk Aversion and Skewness Preference, *working paper*, Erasmus University Rotterdam.

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

Price, K., B. Price, and T. Nantell, 1982, Variance and Lower Partial Moments Measures of Systematic Risk: Some Analytical and Empirical Results, *Journal of Finance,* 37, 843–855.

Quintana, J.M., V. Lourdes, O. Aguilar, and J. Liu, 2002, Global Gambling, *Valencia 7.*

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

Rockinger, M. and E. Jondeau, 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., M. D. Branco, and D. K. Dey, 2002, A New Class of Multivariate Skew Distributions with Applications to Bayesian Regression Models, *Unpublished working paper.*

Sanchez-Torres, P., and E. Sentana, 1998, Mean-Variance-Skewness Analysis: An Application to Risk Premia in the Spanish Stock Market, *Investigaciones Economicas,* 22, 5–17.

Scherer, B., 2002, Portfolio Resampling: Review and Critique, *Financial Analysts Journal,* Vol. 58, No. 6.

Scott, R.C., and P.A. Horvath, 1980, On the Direction of Preference for Moments of Higher Order Than Variance, *Journal of Finance,* Vol. 40, No. 4, 1251–1253.

Sears, R.S., and K.C.J. Wei, 1988, The Structure of Skewness Preferences in Asset Pricing Models with Higher Moments: An Empirical Test, *Financial Review*, 23 (1), 25–38.

Sears, R.S., and K.C.J. Wei, 1985, Asset Pricing, Higher Moments, and the Market Risk Premium: A note, *Journal of Finance,* 40 (4), 1251–1253.

Sharpe, W.F., 1994, The Sharpe Ratio, *Journal of Portfolio Management,* fall 21 (1), 49–58.

Simkowitz, M.A., and W.L. Beedles, 1980, Asymmetric Stable Distributed Security Returns, *Journal of the American Statistical Association,* 75, 306–312.

Singleton, J.C., and J. Wingender, 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., J. V. Andersen and P. Simonetti, 2000a, Portfolio Theory for "Fat Tales", *International Journal of Theoretical and Applied Finance,* 3 (3), 523–535.

Sornette, D., P. Simonetti, and J. V. Andersen, 2000b, Field theory for portfolio optimization: "fat-tails" and non-linear correlations, *Physics Reports,* 335 (2), 19–92.

Sortino, F.A., and H.J. Forsey, 1996, On the Use and Misuse of Downside Risk, *Journal of Portfolio Management,* 22, 35–42.

Sortino, F.A., and L.N. Price, 1994, Performance Measurement in a Downside Risk Framework, *Journal of Investing,* Vol. 3, No. 3, 50–58.

Sortino, F.A., and R. van der Meer, 1991, Downside Risk, *Journal of Portfolio Investment,* 17, 27–31.

Spurgin, R.B., 2001, How to Game Your Sharpe Ratio, *The Journal of Alternative Investments,* (winter) 4 (3), 38–46.

Tamayo A., 2001, Stock return predictability, conditional asset pricing models and portfolio selection, *Unpublished working paper, 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 (2), 449–460.

Tsiang, S.C., 1972, The Rationale of the Mean-Standard Deviation Analysis, Skewness Preference, and the Demand for Money, *American Economic Review,* Vol. 62, No. 3, 354–371.

Xia Y., 2001, Learning about predictability: The effects of parameter uncertainty on dynamic asset allocation, *Journal of Finance*, 56, 205–246.

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

Appendix A: Skew normal probability model

0.1 *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 $(m, V, \text{and } S)$ 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 $m \times m^2$ 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.

0.2 *First three moments of linear combination*

Assume $X \sim SN(\mu, \Sigma, \Delta)$ and a set of constants $\omega = (\omega_1, \ldots, \omega_m)'$, 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.

0.3 *Model Specification*

0.3.1 *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 used conjugate prior densities, with hyper-parameters that reflect vague prior information, or *a priori* we assume

30

$$\beta \sim N_{m(m+1)}(0, 100I_{m(m+1)})$$

$$\Sigma \sim \text{Inverse-Wishart}(m, mI_m),$$

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

0.3.2 *Full conditionals*

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

$$Z_i|x, \mu, \Sigma, \Delta \sim N_m(A^{-1}a_i, A^{-1})I\{Z_{ij} > 0\}, \text{ for all } j,$$

$$\beta|x, \Sigma, Z \sim N_{m(m+1)}(B^{-1}b, B^{-1})$$

$$\Sigma|x, \mu, \Delta, Z \sim \text{Inverse-Wishart}(m + n, C),$$

where

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

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

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

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

0.3.3 *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.* (2002) and Azzalini *et al.* (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 *et*

al. (1996) model maintains the property that the distribution of the portfolio is also skew normal.

0.4 *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 Damian *et al.* (1999). To illustrate, assume that we want to sample from the following density,

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

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

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

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.

Table 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 *et al.* (2002)	2368.33	-2418.85
MVS-Normal F-Δ, Sahu *et al.* (2002)	2192.83	-2595.56

Table 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 *et al.* (2002)	2015.11	-2067.73
MVS-Normal F-Δ, Sahu *et al.* (2002)	2034.22	-2200.30

Table 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 ($\mu + (2/\pi)^{1/2}\Delta\mathbf{1}$), variance ($\Sigma + (1 - 2/\pi)\Delta\Delta'$), and skewness (see Appendix 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 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 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 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 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 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 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 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 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

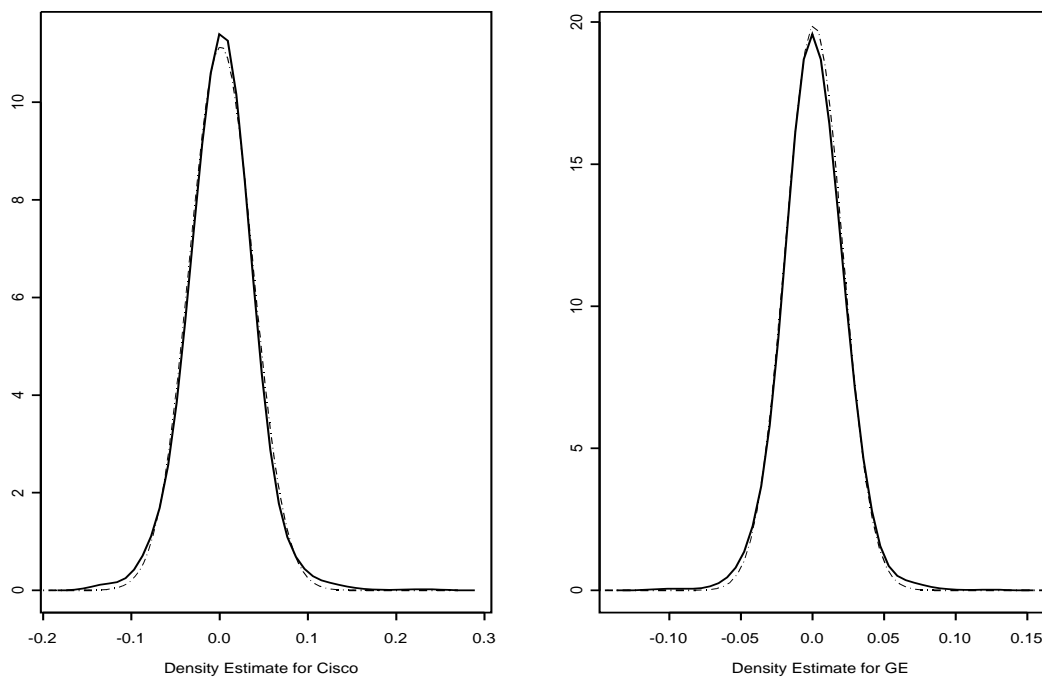| $\lambda = 0.5,\ \gamma = 0.5$	GB	CB	MBS	EAFE	EMF	R1000	R2000
$\omega_{3,pred}$	0.157	2.34e-4	0.439	0.121	0.0874	0.193	2.77e-3
$\omega_{3,param}$	0.149	2.11e-4	0.461	0.0967	0.0905	0.18	0.0233
$\omega_{3,Michaud}$	0.225	5.2e-4	0.407	0.0885	0.0808	0.15	0.0479

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.
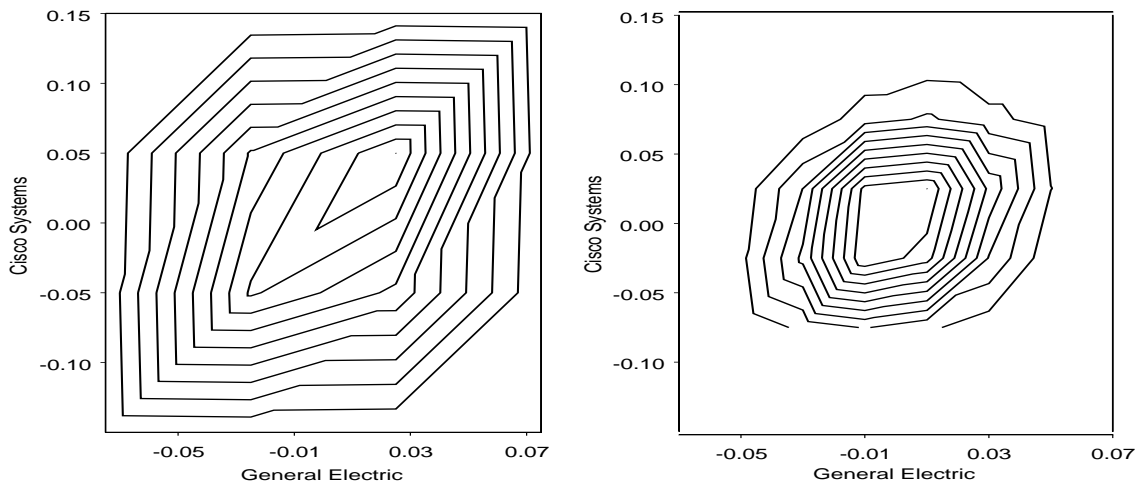
Figure 2: This figure contains bivariate estimates for Cisco Systems and General Electric daily stock returns from April 1996 to March 2002. The plot on the left is a contour plot of the actual data. The plot on the right is a bivariate normal with sample mean and covariance. In two dimensions the bivariate normal distribution is a poor model to use for these joint returns. The actual returns (left) exhibit co-skewness and much fatter tails than the normal approximation.

Mean:
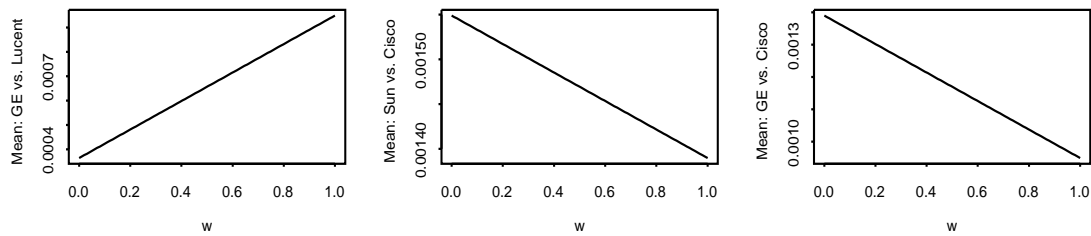
Variance:
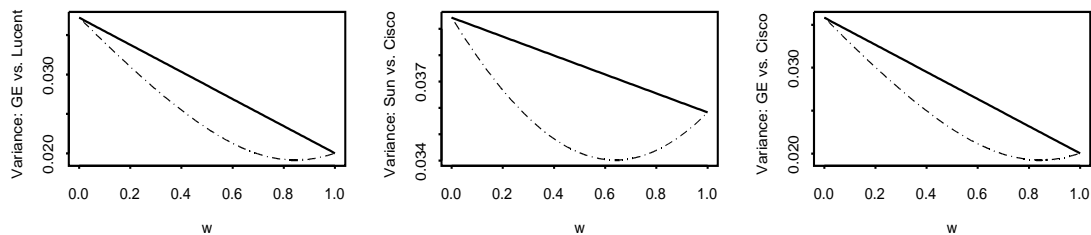
Skewness:
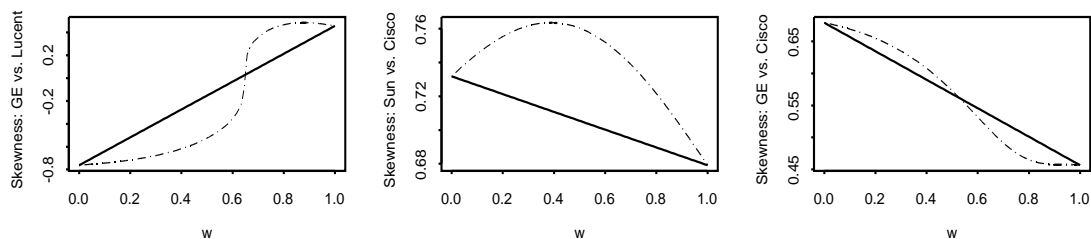
Figure 3: This figure contains plots of the mean, variance and skewness of portfolios consisting of two assets. Daily returns from April 1996 to March 2002 for General Electric and Lucent Technologies, Sun Microsystems and Cisco Systems, and General Electric and Cisco Systems are considered. The top 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. Picking a portfolio based solely on minimum variance could lead to a portfolio with minimum skewness as well (see GE vs. Cisco).
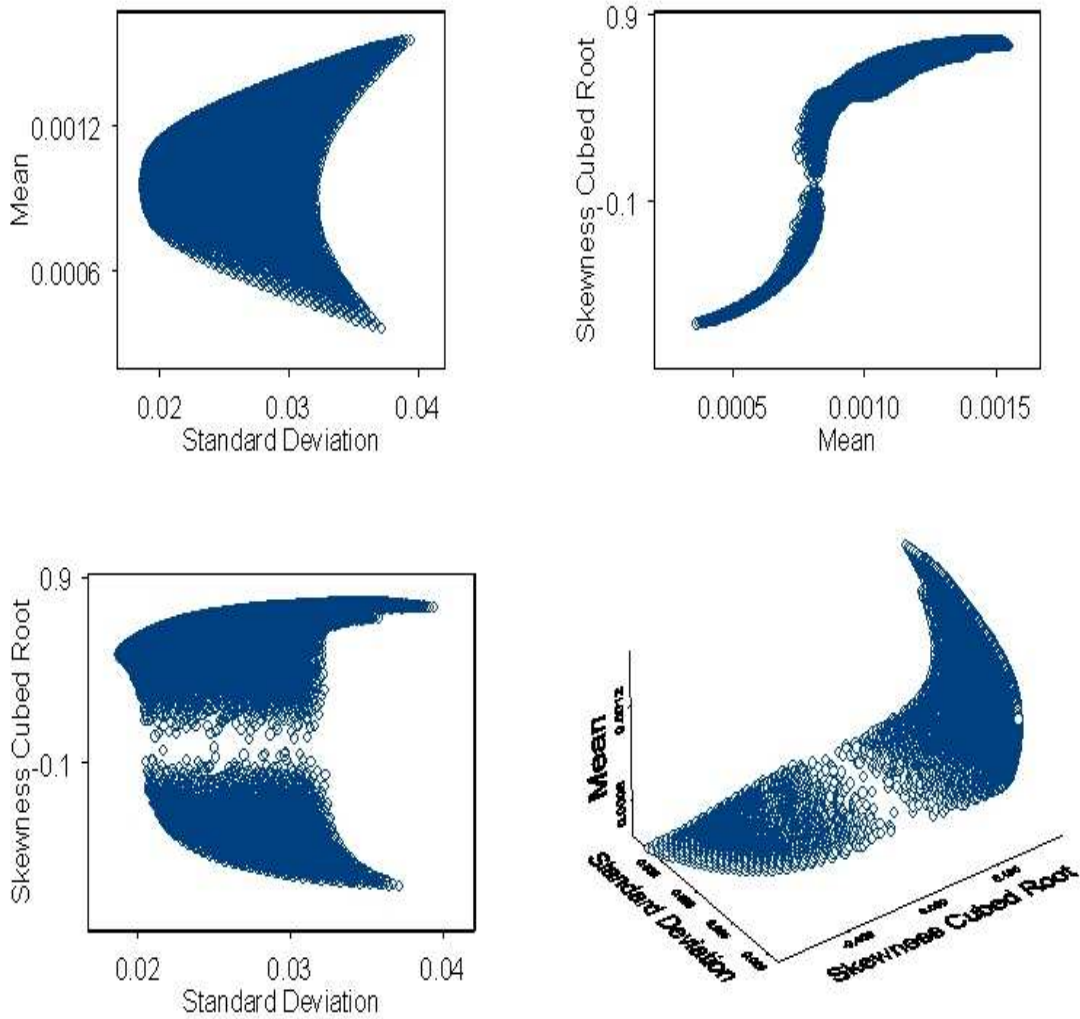
45

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.
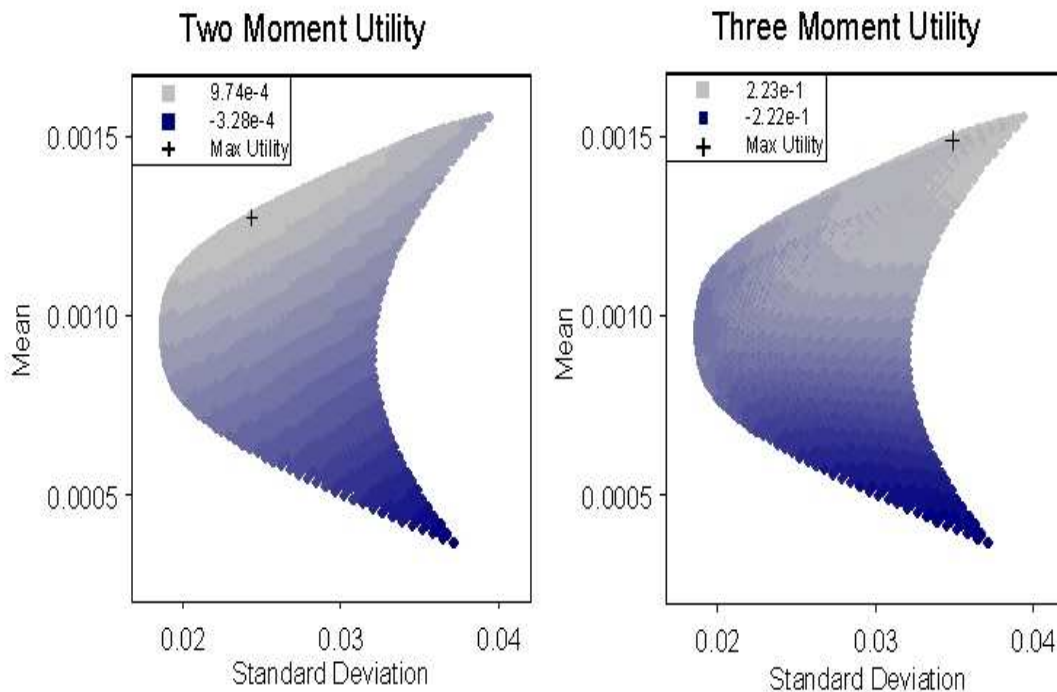
Figure 5: This figure shows the mean-variance space of possible portfolios based on historical parameter estimates from the daily returns of General Electric, Lucent Technologies, Cisco Systems, and Sun Microsystems from April 1996 to March 2002. The portfolios are shaded according to the utility associated 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 '+'.
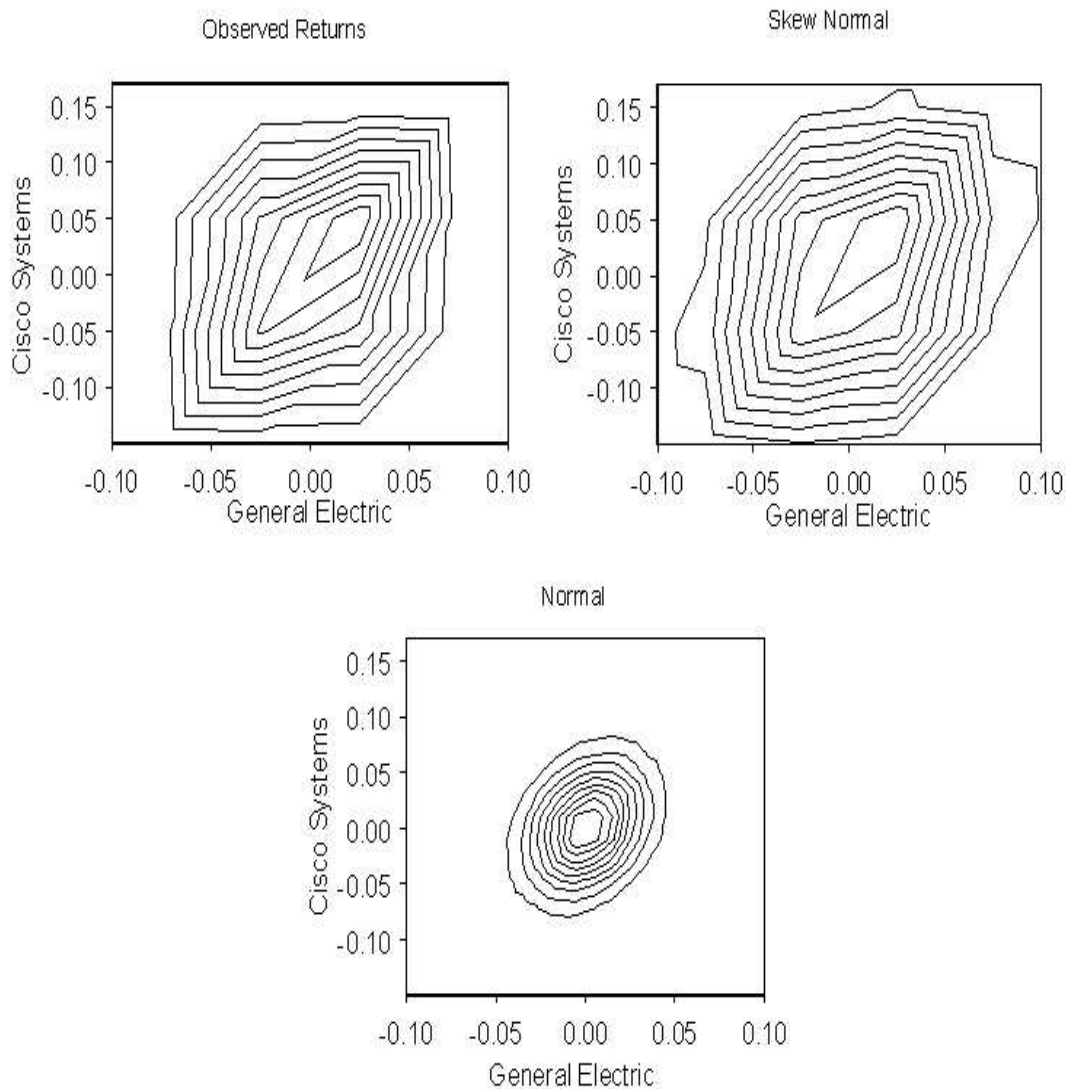
Figure 6: This figure contains bivariate estimates for Cisco Systems and General Electric daily stock returns from April 1996 to March 2002. The plot on the top left is a contour plot of the actual data. On the top right is a contour plot of predictive bivariate returns based on the skew normal model of Sahu *et al.* (2002) with a diagonal Δ. The plot on the bottom is a contour plot of predictive bivariate returns based on the normal distribution. The skew normal matches both skewness and kurtosis much better than the normal.