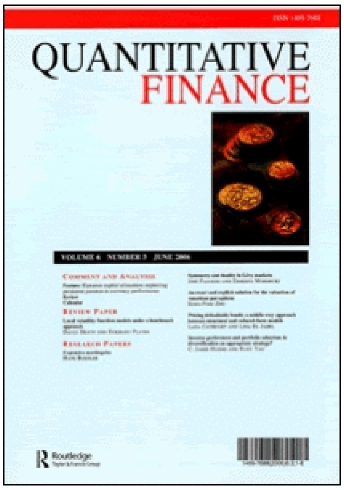

Quantitative Finance

Portfolio selection with higher moments

Campbell R. Harvey [ab]; John C. Liechty [c]; Merrill W. Liechty [d];Peter Müller [e]
[a] Duke University, Durham, NC 27708, USA [b] National Bureau of Economic Research, Cambridge, MA 02138, USA [c] Pennsylvania State University, University Park, PA 16803, USA [d] Drexel University, Philadelphia, PA 19104, USA [e] University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA

Online publication date: 29 April 2010

To cite this Article Harvey, Campbell R. , Liechty, John C. , Liechty, Merrill W. andMüller, Peter(2010) 'Portfolio selection with higher moments', Quantitative Finance, 10: 5, 469 — 485
To link to this Article: DOI: 10.1080/14697681003756877
URL: http://dx.doi.org/10.1080/14697681003756877

Quantitative Finance, Vol. 10, No. 5, May 2010, 469–485

Portfolio selection with higher moments

CAMPBELL R. HARVEY*†‡, JOHN C. LIECHTY§,
MERRILL W. LIECHTY¶ and PETER MÜLLER⊥

†Duke University, Durham, NC 27708, USA
‡National Bureau of Economic Research, Cambridge, MA 02138, USA
§Pennsylvania State University, University Park, PA 16803, USA
¶Drexel University, Philadelphia, PA 19104, USA
⊥University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA

(*Received 26 February 2010; in final form 6 March 2010*)

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

Keywords: Bayesian decision problem; Multivariate skewness; Parameter uncertainty; Optimal portfolios; Utility function maximization

1. Introduction

Markowitz (1952a) provides the foundation for the current theory of portfolio choice. He describes the task of asset allocation as having two stages. The first stage "starts with observation and experience and ends with beliefs about the future performances of available securities." The second stage "starts with the relevant beliefs ... and ends with the selection of a portfolio." Although Markowitz only deals with the second stage, he suggests that the first stage should be based on a "probabilistic" model. However, in the usual implementation of Markowitz's second stage, it is assumed that we know with certainty the inputs from the first stage, i.e. the exact means, variances and correlations. This paper introduces a method for addressing both stages.

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

Several authors have proposed advances to optimal portfolio selection methods. Some address the empirical evidence of higher moments; Athayde and Flôres (2001) and Adcock (2009) give methods for determining higher dimensional 'efficient frontiers', but they remain in the certainty equivalence framework (assuming exact knowledge of the inputs) for selecting an optimal portfolio. Like the standard efficient frontier approach, these approaches have the advantage that for a large class of utility functions, the task of selecting an optimal portfolio reduces to the task of selecting a point on the high dimensional 'efficient frontiers'. Other three moment optimization methods include using negative semi-variance in place of variance (Markowitz 1959, Markowitz *et al.* 1993). A similar measure of downside risk is incorporated by Feiring *et al.* (1994) and Konno *et al.* (1993) who use an approximation to the lower

*Corresponding author. Email: cam.harvey@duke.edu

Quantitative Finance
ISSN 1469–7688 print/ISSN 1469–7696 online © 2010 Taylor & Francis
http://www.informaworld.com
DOI: 10.1080/14697681003756877

semi-third moment in their Mean-Absolute Deviation-Skewness portfolio model.† These methods rest on the assumption that an investor's expected utility is reasonably approximated by inserting estimates of the moments of an assumed sampling model. Chiu (2008) examines the relationship between expected utility maximization and stochastic dominance.

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

Others have ignored higher moments, but address the issue of estimation risk. Frost and Savarino (1986, 1988) show that constraining portfolio weights, by restricting the action space during the optimization, reduces estimation error. Jorion (1992) proposes a resampling method aimed at estimation error. Using a Bayesian approach, Britten-Jones (2002) proposes placing informative prior densities directly on the portfolio weights. Others propose methods that address both stages of the allocation task and select a portfolio that optimizes an expected utility function given a probability model.‡

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, 2006 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 as Jensen's inequality dictates that the resampled efficient frontier is in fact suboptimal. Polson and Tew (2000) argue for the use of posterior predictive moments instead of point estimates for mean and variance of an assumed sampling model. Their setup comes closest to the framework that we propose in this paper. Using posterior predictive moments, they accommodate parameter uncertainty. We follow their setup in our discussion in section 3.1.1.

Our approach advances previous methods by addressing both higher moments and estimation risk in a coherent Bayesian framework. As part of our 'stage one' approach (i.e. incorporating observation and experience), we specify a Bayesian probability model for the joint distribution of the asset returns, and discuss prior distributions. As for 'stage two', the Bayesian methodology provides a straightforward framework to calculate and maximize expected utilities based on predicted returns. This leads to optimal portfolio weights in the second stage which overcome the problems associated with estimation risk. We empirically investigate the impact of simplifying the asset allocation task for a class of utility functions which can be approximated by a third-order Taylor series expansion (i.e. contributions to higher order moments are assumed to be negligible for this class of investors).§ For two illustrative data sets, we demonstrate the difference in expected utility that results from ignoring higher moments and using a sampling distribution, with point estimates substituted for the unknown parameter, instead of a predictive distribution. In addition, we demonstrate the loss in expected utility from using the popular resampling approach proposed by Michaud (1998).

Our paper is organized as follows. In the section 2 we discuss the importance of higher moments and provide the setting for portfolio selection and Bayesian statistics in finance. We discuss suitable probability models for portfolios and detail our proposed framework. In the section 3 we show how to optimize portfolio selection based on utility functions in the face of parameter uncertainty using Bayesian methods. Section 4 empirically compares different methods and approaches to portfolio selection. Some concluding remarks are offered in the

†Several other authors have investigated the importance of higher moments in financial applications, including Campbell *et al.* (2001), Chen *et al.* (2001), Dittmar (2002), Athayde and Flores (2004), Burger and Warnock (2004), Goetzman and Kumar (2004), Jondeau and Rockinger (2004), Levy and Levy (2004), Patton (2004), Adcock (2005), Brunnermeier and Parker (2005), Jurczenko *et al.* (2005), Liew and French (2005), Sfiridis (2005), Ang *et al.* (2006), Bakshi and Madan (2006), Barro (2006), Williams and Ioannidis (2006), Barberis and Huang (2007), Briec *et al.* (2007), Brunnermeier *et al.* (2007), Chiang and Li (2007), Guidolin and Timmermann (2007), Mitton and Vorkink (2007), Martellini and Ziemann (2007), Chabi-Yo (2008a, b), Cvitanić *et al.* (2008), DeMiguel *et al.* (2009, 2010), Post *et al.* (2008), Bacmann and Benedetti (2009), Da Silva *et al.* (2009), Hall *et al.* (2009), Knight and Satchell (2009), Mencia and Sentana (2009), Morton and Popova (2009), Wilcox and Fabozzi (2009), Zhou (2009), Bali and Cakici (2010), Blau and Pinegar (2009), Brandt *et al.* (2010), Conrad *et al.* (2010), Fabozzi *et al.* (2010), Martin (2010), Poti (2010), Jondeau and Rockinger (2006), Vorkink *et al.* (2009).

‡From the Bayesian perspective, Jorion (1986) uses a shrinkage approach while Treynor and Black (1973) advocate the use of investors' views are combined with historical data. Kandel and Stambaugh (1996) examine predictability of stock returns when allocating between stocks and cash by a risk-averse Bayesian investor. Johannes *et al.* (2002) examine the market timing relationship to performance of optimal portfolios using a model with correlation between volatility and returns in a Bayesian portfolio selection setting. Zellner and Chetty (1965), Klein and Bawa (1976) and Brown (1978) emphasize using a predictive probability model (highlighting that an investor's utility should be given in terms of future returns and not parameters from a sampling distribution). Pástor and Stambaugh (2000) study the implications of different pricing models on optimal portfolios, updating prior beliefs based on sample evidence. Pástor (2000) and Black and Litterman (1992) propose using asset pricing models to provide informative prior distributions for future returns.

§We restrict ourselves to third moment approximations and probability models in order to simplify the exposition; clearly these methods and models can be extended in a straightforward manner to accommodate fourth-order and higher moments.

final section. The appendix contains some additional results and proofs.

2. Higher moments and Bayesian models

A prerequisite to the use of the Markowitz framework is either that the relevant distribution of asset returns be normally distributed or that utility is only a function of the first two moments. But it is well known that many financial returns are not normally distributed. Studying a single asset at a time, empirical evidence suggests that asset returns typically have heavier tails than implied by the normal assumption and are often not symmetric (Kon 1984, Mills 1995, Markowitz and Usmen 1996, Peiro 1999, Premaratne and Tay 2002). Also, we argue that the relevant probability model is the posterior predictive distribution, which in general is not normal, even under an assumed normal sampling model.

Since we are integrating over the predictive distribution to get the predictive moments, the skew normal distribution will accommodate higher order moments (skewness, coskewness and kurtosis) in the predictive density. We use Bayes factors (see section 2.3) to show that the skew normal fits better than the normal distribution, but acknowledge that there are other competing densities which could be considered such as multivariate versions of the Pearson Type IV and skew-*t* distributions. As one of our main goals is demonstrating the value of incorporating distributions with higher moments into the portfolio optimization problem, we have left an exhaustive comparison of competing distributions for future research.

The approach proposed in our paper is somewhat related to Shadwick and Keating (2002) and Cascon *et al.* (2003) who argue that point estimates of the mean and variance of an assumed sampling distribution are insufficient summaries of the available information of future returns. Instead they advocate the use of a summary function, which they call 'Omega', that represents all the relevant information contained within the observed data. This 'Omega' function suggests a decision rule where investors select a portfolio that maximizes Omega for each possible level of average return. Given an average return level, this approach provides a complicated, nonlinear utility function which can accommodate higher order moments. While we strongly agree with the premise that investors have utilities which accommodate higher order moments of the predictive distribution, we are restricting our focus to utility functions which can be approximated by a third-order Taylor series. Future work could easily extend our approach to consider the Omega summary and decision rules.†

Our investigation of multiple assets builds on these empirical findings and indicates that the existence of 'coskewness', which can be interpreted as correlated extremes, is often hidden when assets are considered one at a time. To illustrate, figure 1 contains the kernel density estimate and normal distribution for the marginal daily returns of Carnival, Starwood Hotels and Resorts, L-3 Communications Holdings, and Raytheon from July 2001 to June 2006, and figure 2 contains bivariate normal approximations of select pairs of joint returns.

While the marginal summaries in figure 1 suggest only a slight deviation from the normality assumption, the joint summaries appear to exhibit a degree of coskewness, suggesting that skewness may have a larger impact on the distribution of a portfolio than previously thought.

2.1. Economic importance

Markowitz's intuition for maximizing the mean while minimizing the variance of a portfolio comes from the idea that the investor prefers higher expected returns and lower risk. Extending this concept further, most agree that, *ceteris paribus*, investors prefer a high probability of an extreme event in the positive direction over a high probability of an extreme event in the negative direction. From a theoretical perspective, Markowitz (1952b) and Arrow (1971) argue that desirable utility functions should exhibit decreasing absolute risk aversion, implying that investors should have preference for positively skewed asset returns. (Also see the discussion in Roy (1952).) Experimental evidence of preference for positively skewed returns has been established by, for example, Sortino and Price (1994) and Sortino and Forsey (1996). Levy and Sarnat (1984) find a strong preference for positive skewness in the study of mutual funds. Harvey and Siddique (2000a,b) introduce an asset pricing model that incorporates conditional skewness, and show that an investor may be willing to accept a negative expected return in the presence of high *ex ante* positive skewness.

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

While skewness will be important to a large class of investors and is evident in the historical returns of the underlying assets and portfolios, the question remains: how influential is skewness in terms of finding optimal portfolio weights? Cremers *et al.* (2004) argue that only

†While this approach accommodates higher order moments it does not account for parameter uncertainty. A proper extension of the Omega function would include the uncertainty about cumulative predictive densities. In addition to extending our proposed methods to accommodate utility functions, we could also extend our probabilistic models to build upon non-parametric Bayesian estimates of the predictive densities, inherently modeling the uncertainty in the 'cumulative CDF'.

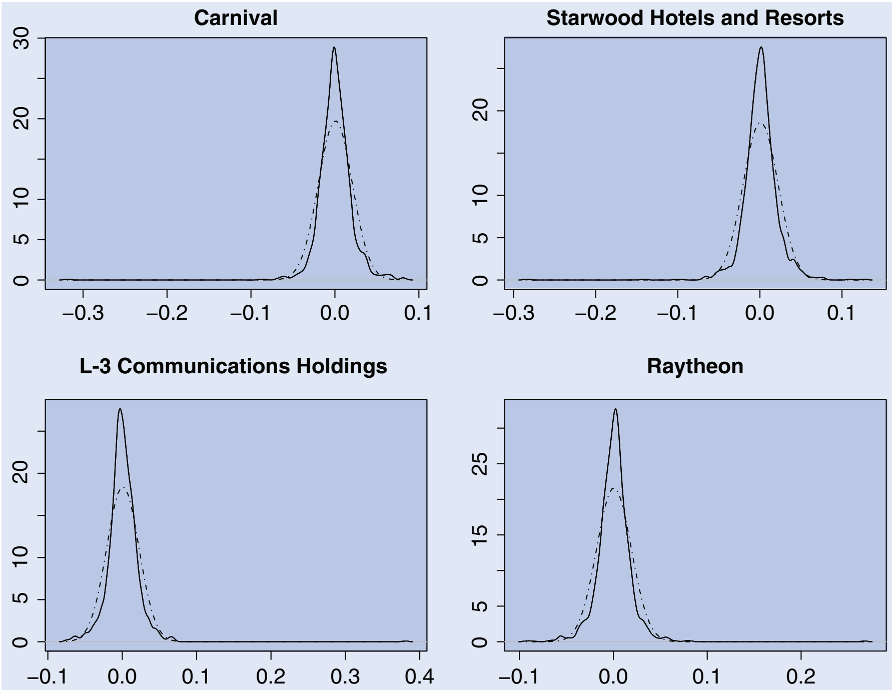

Figure 1. Univariate density estimates for Carnival, Starwood Hotels and Resorts, L-3 Communications Holdings, and Raytheon daily stock returns from July 2001 to June 2006. The solid lines represent the kernel density estimate, while the dotted lines are the normal density with sample mean and variance. In one dimension, the normal distribution is a close match of the returns for these four stocks.

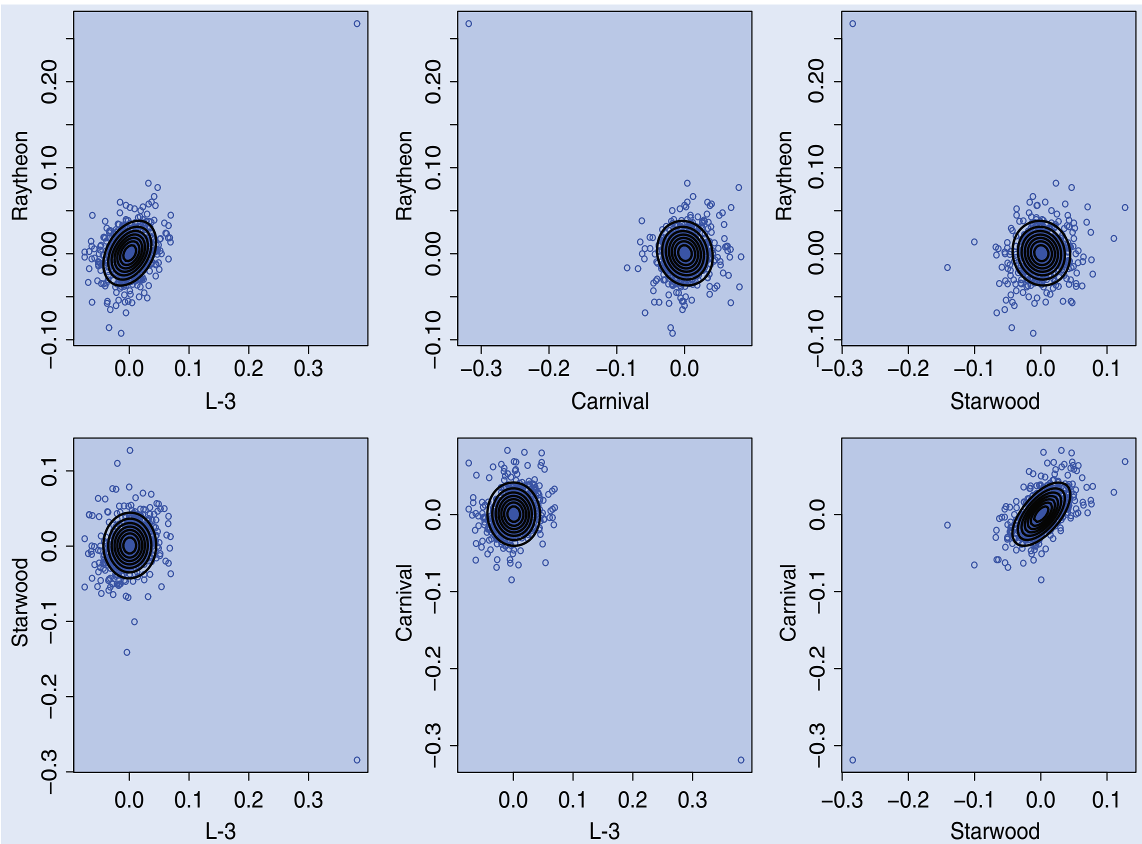

Figure 2. Bivariate normal estimates for pairs of Carnival, Starwood Hotels and Resorts, L-3 Communications Holdings, and Raytheon from July 2001 to June 2006. The plots are bivariate normal densities with sample mean and covariance. The scatter points are the actual data. Unlike in one dimension, in two dimensions the normal distribution does not closely match these joint returns. The actual returns exhibit coskewness and much fatter tails than the normal approximation.

under certain utility specifications is it worthwhile to consider skewness in portfolio selection. However, it is our experience that any utility function approximated by a third-order Taylor's expansion will lead to superior portfolio weights if skewness is included. To accommodate higher-order moments in the asset allocation task, we must introduce an appropriate probability model. After providing an overview of possible approaches, we formally state a model and discuss model choice tools.

2.2. Probability models for higher moments

Though it is a simplification of reality, a model can be informative about complicated systems. While the multivariate normal distribution has several attractive properties for modeling a portfolio, there is considerable evidence that portfolio returns are non-normal. There are a number of alternative ways to accommodate higher moments. The multivariate Student-t distribution is good for fat-tailed data, but does not allow for asymmetry. The non-central multivariate t distribution also has fat tails and, in addition, is skewed. However, the skewness is linked directly to the location parameter and is, therefore, somewhat inflexible. The log-normal distribution has been used to model asset returns, but its skewness is a function of the mean and variance, not a separate skewness parameter.

Azzalini and Dalla Valle (1996) propose a multivariate skew normal distribution that is based on the product of a multivariate normal probability density function (pdf) and univariate normal cumulative distribution function (cdf). This is generalized into a class of multivariate skew elliptical distributions by Branco and Dey (2001), and improved upon by Sahu *et al.* (2003) by using a multivariate cdf instead of univariate cdf, adding more flexibility, which often results in better-fitting models. Because of the importance of coskewness in asset returns, we start with the multivariate skew normal probability model presented by Sahu *et al.* (2003) and offer a generalization of their model.

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

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

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

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

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

As with other versions of the skew normal model, this model has the desirable property that marginal distributions of subsets of skew normal variables are skew normal (see Sahu *et al.* 2003 for a proof). Unlike the multivariate normal density, linear combinations of variables from a multivariate skewed normal density are not skew normal. This does not, however, restrict us from calculating moments of linear combinations with respect to the model parameters (see appendix A.2 for the formula for the first three moments).

While the skew normal is similar in concept to a mixture of normal random variables, it is fundamentally different. A mixture takes on the value of one of the underlying distributions with some probability and a mixture of normal random variables results in a Lévy stable distribution.† The skew normal is not a mixture of normal distributions, but it is the sum of two normal random variables, one of which is truncated, and results in a distribution that is not Lévy stable. Though it is not stable, the skew normal has several attractive properties. Because it is the sum of two distributions, it can accommodate heavy tails along with coskewness, but the marginal distribution of any subset of assets is also skew normal. This is important in the portfolio selection setting because it insures consistency in selecting optimal portfolio weights. For example, with short selling not allowed, if optimal portfolio weights for a set of assets are such that the weight is zero for one of the assets then removing that asset from the selection process and re-optimizing will not change the portfolio weights for the remaining assets.

Following the Bayesian approach, we assume conjugate prior densities for the unknown parameters, i.e. *a priori* normal for μ and $vec(\Delta)$, where $vec(\cdot)$ forms a vector from a matrix by stacking the columns of the matrix, and *a priori* Wishart for Σ^{-1}. The resulting full conditional posterior densities for μ and $vec(\Delta)$ are normal, the full conditional posterior density for Σ^{-1} is Wishart and the full conditional posterior density for the latent Z is a truncated normal. See appendix A.3 for a complete specification of the prior densities and the full conditional posterior densities. Given these full conditional posterior densities, estimation is done using a Markov Chain Monte Carlo (MCMC) algorithm based on the Gibbs sampler and the slice sampler (see Gilks *et al.* (1996) for a general discussion of the MCMC algorithm and the Gibbs sampler and see appendix A.4 for a discussion of the slice sampler).

†A family of distributions X is said to be Lévy stable if for two independent draws 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.

2.3. *Model choice*

The Bayes factor (BF) is a well developed and frequently used tool for model selection which naturally accounts for both the explanatory power and the complexity of competing models (see Berger 1985 and O'Hagan 1994 for further discussion of Bayes factors).

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

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

We use a sampling-based estimator proposed by Newton and Raftery (1994) to calculate the Bayes factor. In practice, we use the Bayes factor to determine the best probability model (the model that provides the best empirical fit for a set of returns data) by calculating the Bayes factor with respect to the Multivariate normal probability model, which is used as a baseline probability model.

3. Optimization

Markowitz defined the set of optimal portfolios as the portfolios that are on the efficient frontier, based on estimated moments of the sampling distribution. Ignoring uncertainty inherent in the underlying probability model, the portfolio that maximizes expected utility for a large class of utility functions is in this set. When parameter uncertainty is explicitly considered, the efficient frontier, now written in terms of predictive moments, can still only identify optimal allocations for utility functions that are exactly linear in the moments of the portfolio. In all other cases, the utility function must be explicitly specified. For general probability models and arbitrary utility functions, calculating and optimizing the expected utility must be done numerically, a task that is straightforward to implement using the Bayesian framework.

3.1. *Simplifications made in practice*

There are several simplifications that are made in practice with regards to selecting optimal weights in the face of parameter uncertainty: such as using parameter estimates instead of predictive returns, maximizing something other than expected utility or ignoring higher-order moments, such as skewness. As we will detail, each of these simplifications can, for different reasons, lead to poor decisions, in that the investor will select weights that 'leave expected utility on the table'. We complete this comparison by illustrating the extent to which these simplifications can impact portfolio decisions in the empirical study that is presented in section 4.

3.1.1. Simplification 1: Utility based on model parameters, not predictive returns. The relevant reward for an investor is the realized future return of their portfolio. Thus the utility function needs to be a function of the future returns, not a function of the model parameters. This point is emphasized by Zellner and Chetty (1965) and Brown (1978). It is reasonable to assume that a decision maker chooses an action by maximizing expected utility, the expectation being with respect to the posterior predictive distribution of the future returns, conditional on all currently available data (DeGroot 1970, Raiffa and Schlaifer 1961). Following this paradigm, Polson and Tew (2000) propose the use of predictive moments for future returns to define mean–variance efficient portfolios which we also implement (see also Kandel and Stambaugh 1996 and Pástor and Stambaugh 2000, 2002). In the following discussion, we highlight the difference of this approach and the traditional approach.

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

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

$$m_p = \int x\, p(x \mid x^o)\, \mathrm{d}x$$

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

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

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

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

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

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

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

where the moments m, V, and S are given in terms of μ, Σ, and Δ, assuming a skew normal model

†A third-order Taylor series approximation of utilities are dominated by the first three predictive moments.

(see appendix A.1 for details). The expected utility becomes

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

where $\bar{m}$, $\bar{V}$ and $\bar{S}$ are the posterior means of θ. Note that the expectation in (6) is in terms of posterior moments which is a distribution of the model parameters and are point estimates or 'plug in'. The predictive moments in (4) contain the posterior moments given in (6) and additional terms.† It is straightforward to show that the predictive mean equals the posterior mean and that the predictive variance and skewness equal the posterior means of V and S plus additional terms, or

$$\begin{aligned} m_p &= \bar{m}, \\ V_p &= \bar{V} + \text{Var}(m \mid x^o), \\ S_p &= \bar{S} + 3E(V \otimes m \mid x^o) - 3E(V \mid x^o) \otimes m_p \\ &\quad - E[(m - m_p)'(m - m_p) \otimes (m - m_p) \mid x^o]. \end{aligned}$$

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

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

When considering linear utility functions, it is clear that using the parameter utility will result in a very different utility when compared with the predictive utility. For example, the parameter utility ignores the uncertainty about the first moment, or the variance of the mean, and as a result the parameter variance is smaller than the predictive variance, which includes this uncertainty about the mean. Similar omissions are made for skewness and other higher-order terms. As a result the parameter utility function is only an approximation to the predictive utility function. Clearly when one considers other utility functions, particularly ones which are not linear, the difference in the parameter and predictive expected utilities will have marked differences which may make the parameter utility an even worse approximation to the predictive utility than in the case of linear utility functions.

3.1.2. Simplification 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 (Novshek 1993) about the expected predictive summaries, or assume

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

For linear utility functions this approximation is exact, as in (3) and (4). The Taylor's approximation removes any parameter uncertainty and leads directly to the certainty equivalence optimization framework, substituting predictive moments. It is easy to see that combining the Taylor's approximation and the much stronger assumption that the posterior moments approximately equal predictive moments leads to a frequently used 'two-times removed' approximation of the expected utility of future returns, or

$$\begin{aligned} EU_{\text{pred}}(\omega) = E[u_{\text{pred}}(\omega, \mathcal{X}) \mid x^o] &\approx u_{\text{pred}}(\omega, E[\mathcal{X} \mid x^o]) \\ &\approx u_{\text{param}}(\omega, E[\theta \mid x^o]). \end{aligned}$$

In an attempt to maintain the flexibility of the efficient frontier optimization framework but still accommodate parameter uncertainty, Michaud (1998) proposes an optimization approach that switches the order of integration (averaging) and optimization. The maximum expected utility framework optimizes the expected utility of future returns; the certainty equivalence framework optimizes the utility of expected future returns (i.e. substituting posterior predictive moments in the utility function). Michaud (1998) proposes creating a 'resampled frontier' by repeatedly maximizing the utility for a draw from a probability distribution and then averaging the optimal weights that result from each optimization. While the approach could be viewed in terms of predictive returns, the sampling guidelines are arbitrary and could significantly impact the results.

Given that the main interest is to account for parameter uncertainty, we consider a modified resampling algorithm where parameter draws from a posterior density are used in place of the predictive moment summaries.‡ To be explicit, assuming a utility of parameters, the essential steps of the algorithm are as follows. For a family of utility functions $(u_{\text{param},1}, \ldots, u_{\text{param},K})$, perform the following steps.

(1) For each utility function (e.g. $u_{\text{param},k}$), 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 $u_{\text{param},k}(\omega, \theta_{i,k})$.

†Kan and Zhou (2007) provide a thorough discussion of the difference of plug-in and Bayes estimator of the optimal decision under the parameter-based utility (5). Their discussion highlights the difference between a proper Bayes rule, defined as the decision that maximizes expected utility, versus a rule that plugs in the Bayes estimate for the weights or the parameters in the sampling distribution.

‡In a related study, Markowitz and Usmen (2003) take a similar approach to ours for comparing Bayesian methods to Michaud's (1998) approach. However, in their study, they only consider two moments and use an importance sampling approach for Bayesian inference, where the number of samples that they use to estimate the expected utility is too small compared to the dimension of the integral. In related papers, Harvey *et al.* (2008, 2011) correct this problem by using the standard MCMC method for Bayesian inference and using a much larger number of samples to estimate the expected utility. The results from the new study are sharply different from Markowitz and Usmen (2003).

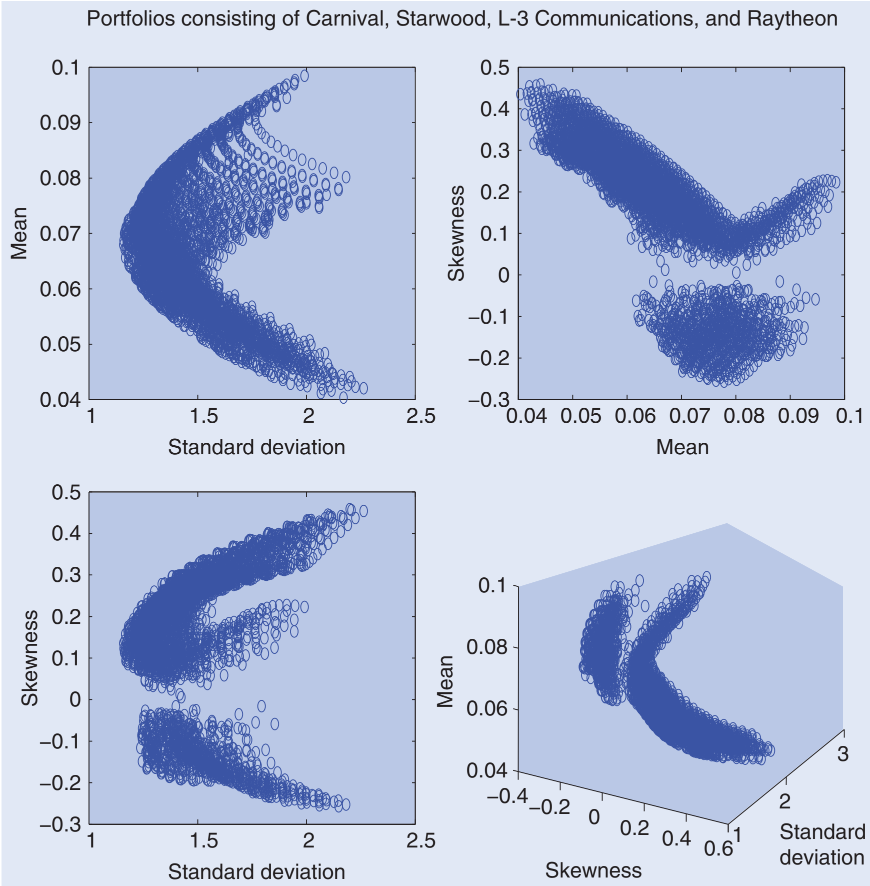

Figure 3. The space of possible portfolios based on historical parameter estimates from the daily returns of Carnival, Starwood Hotels and Resorts, L-3 Communications Holdings, and Raytheon from July 2001 to June 2006. The top left plot is the mean–standard deviation space, the top right plot is the mean versus the cubed-root of skewness. The bottom left plot is the standard deviation versus the cubed-root of skewness, and the bottom right plot is a three-dimensional plot of the mean, standard deviation and cubed-root of skewness.

(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

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

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

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

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

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

or $\bar{\omega}_k$ results in a sub-optimal portfolio in terms of expected utility maximization.

3.1.3. Simplification 3: Ignore skewness. Although evidence of skewness and other higher moments in financial data are abundant, it is common for skewness to be ignored entirely in practice. Typically skewness is ignored both in the sampling models and in the assumed utility functions. In order to illustrate the impact of ignoring skewness, figure 3 shows the empirical summary of the distribution of possible portfolios for four example securities (Carnival, Starwood Hotels and Resorts, L-3 Communications Holdings, and Raytheon). Here we have calculate the portfolio mean, standard deviation, and cubed-root of skewness on a grid of portfolio weights between zero and one.

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

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

where m_e, V_e, and S_e are the empirical mean, variance and skewness.

3.2. Bayesian optimization methods

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

$$x_i \sim p(x \mid x^o),$$

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

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

The approximate expected utility can then be optimized numerically using a number of different approaches. One attractive algorithm is the Metropolis–Hastings (MH) algorithm. While MH simulation is widely used for posterior simulation, the same algorithm can be exploited for expected utility optimization. We use the MH algorithm to explore the expected utility function, $f(\omega) = E[U(\omega)]$, as a function of the weights. Asymptotic properties of the MH chain lead to portfolio weights ω being generated with frequencies proportional to $EU(\omega)$. That is, promising portfolio weights with high expected utility are visited more often, as desired.† One advantage of using the MH algorithm for optimization is that it can easily accommodate constraints on the weights.

Intuitively, this Markov chain can be viewed as a type of 'random walk' with a drift in the direction of larger values of the target function. When the MH algorithm is used as a tool for performing statistical inference, the target density is typically a posterior probability density; however, this need not be the case. As long as the target function is non-negative and integrable, the MH can be used to numerically explore any target function. Not only has the MH been shown to be very effective for searching high-dimensional spaces, its irreducible property ensures that if a global maximum exists the MH algorithm will eventually escape from any local maximum and visit the global maximum.

In order to use the MH function, we need to ensure that our expected utility is non-negative and integrable. For the linear utility functions, integrability over the space of possible portfolios, where the portfolios are restricted to the unit simplex (i.e. we do not allow short selling), is easily established. We modify the utility function so that it is a non-negative function by subtracting the minimum expected utility which is found by tracking the minimum as the algorithm runs. The target function becomes

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

4. Optimal portfolios

In theory, simplifications of the complete asset allocation task will result in a sub-optimal portfolio selection. In order to assess the impact that results from some of these simplifications in practice, we consider three different optimization approaches for two data sets using a family of linear utility functions. In particular, we consider the utility functions given in (3) and (5), which have expected utilities given in terms of the predictive posterior and posterior moments, respectively (see equations (4) and (6)). We consider a number of potential probability models and select the best model. Using results from both the multivariate normal model and the best higher moment model, we numerically determine the optimal portfolio based on the predictive returns, the parameter values and using Michaud's (1998) non-utility maximization approach. We contrast the performance of each optimal portfolio in terms of expected predictive utility using the best model.

Table 1. Marginal summaries for four stocks. This table contains the maximum likelihood estimates for the first three moments of Carnival, Starwood Hotels and Resorts, L-3 Communications Holdings, and Raytheon daily from July 2001 to June 2006.

	Carnival	Starwood	L-3	Raytheon
Mean ($\times$1000)	0.519	0.848	0.773	0.653
Std. dev.	0.02	0.021	0.022	0.018
Skewness$^{1/3}$	−1.42	−1.22	1.62	1.31

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 mixed with some important commodities.

First, we consider daily returns from July 2001 to June 2006 on four equity securities which we considered earlier. In addition, we select assets that more closely match the portfolio decision that most investors face. The daily returns are from January 2002 to June 2006 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 also consider crude oil futures, gold, and the 10-year Treasury bond.

One reason that skewness and in particular coskewness is overlooked in portfolio choice is that simple summaries of asset returns, particularly marginal summaries, often show very little evidence of higher order-moments; see, for example, table 1 and figure 1 for marginal summaries for the four stocks in our first data set.

While the marginal summaries may not appear to support including higher-order moments, as we will demonstrate in the next subsection, there is overwhelming statistical evidence that coskewness exists in both sets of data. As a result, the impact of this coskewness on

†See, for example, Meyn and Tweedie (1993) and Gilks *et al.* (1996) for a discussion of the MH algorithm.

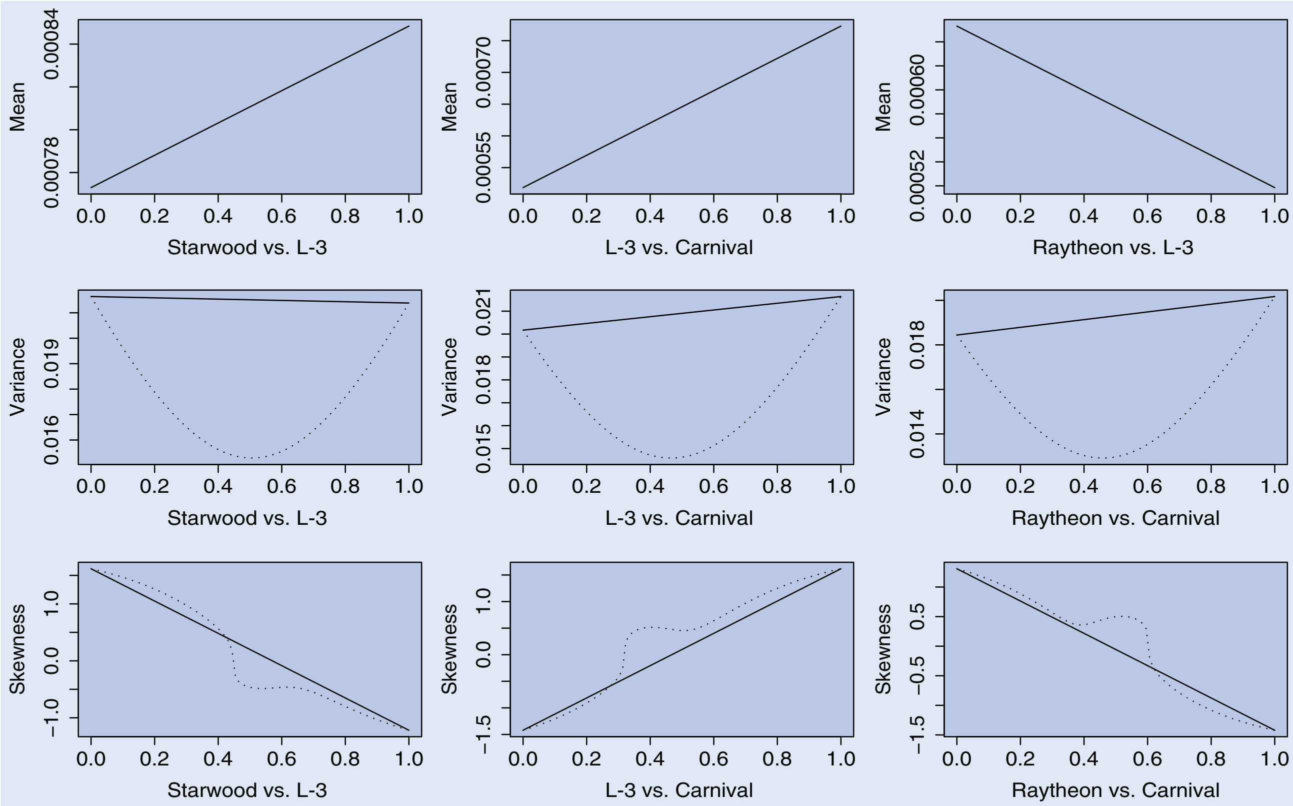

Figure 4. Mean, variance and skewness of portfolios consisting of two assets. Daily returns from July 2001 to June 2006 for Carnival, Starwood Hotels and Resorts, L-3 Communications Holdings, and Raytheon are considered. The top row has the mean of the portfolio (equal to the linear combination of the 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 the portfolios (variance of linear combination). The variance of the portfolio is always 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 it dominated by the linear combination of the skewness. Selecting a portfolio based solely on sub-optimal variance could lead to a portfolio with minimum skewness as well.

portfolio choice is important if an investor's utility function is sensitive to skewness. To illustrate, consider the mean, standard deviation and cubed root of skewness for three sub-portfolios, which are based on the maximum likelihood estimates for pairs of the four stocks in the initial data set (see figure 4). For each of these two asset portfolios the mean and standard deviation behave as we would expect, the linear combination of the mean and the mean of the linear combination are equal and the linear combination of the standard deviation is greater than the standard deviation of the linear combination. The cubed root of the skewness is a different matter, as the skewness of the linear combination can be above or below the linear combination of the skewness.

This suggests that an investor that is interested in skewness must consider an 'extended efficient frontier' which includes the additional dimension of skewness (see figure 3). Even though there is little evidence of higher-order moments from a simple empirical investigation of marginal properties of these four stocks, it is clear that the possible portfolios can vary dramatically with respect to skewness. Empirically there is strong evidence that skewness matters in portfolio selection.

4.2. Model choice and select parameter summaries

To formalize our empirical investigation, we calibrate several competing models to both sets of data and use the Bayes factors calculation to discriminate between these competing models. The models that we consider are the multivariate normal model, the skew normal model proposed by Azzalini and Dalla Valle (1996) with a diagonal Δ matrix, and the skew normal model proposed by Sahu *et al*. (2003) with both a diagonal and our modified full Δ matrix. The results for both the four stocks and the benchmark stocks show that the skew normal models with a diagonal Δ outperform the other models, with the Sahu *et al*. (2003) model fitting best (see table 2).

The posterior parameter estimates for μ, Σ, and Δ, for both the four equity securities and the global portfolio choice benchmark indices, are given in tables 3 and 4. The estimates for Δ for both sets of returns suggest that when considered jointly the skewness is significant.

4.3. Changes in expected utility

As we anticipated, based on earlier arguments, the more simplifications that are imposed, the more sub-optimal the

Table 2. Evaluating the distributional representation of four equity securities and global asset allocation benchmarks. Model choice results for analysis of the daily stock returns of Carnival, Starwood Hotels and Resorts, L-3 Communications Holdings, and Raytheon from July 2001 to June 2006, and also for daily benchmark indices from January 2002 to June 2006 (Russell 1000, Russell 2000, Oil, MSCI EAFE, MSCI EMF, Gold, and the 10 year T-bond). The four models that are used are the multivariate normal, the multivariate skew normal of Azzalini and Dalla Valle (1996) with a diagonal Δ matrix (skew normal A), and the multivariate skew normal of Sahu *et al.* (2003) with both a diagonal and full Δ matrix (skew normal B). Bayes factors are computed between the multivariate normal model and all of the other models and are reported on the log scale. The model with the highest Bayes factor best fits the data.

Distribution	Δ	Log(BF)
(a) Four stocks		
Normal		0.00
Skew normal A	Diagonal	2272.83
Skew normal B	Diagonal	2384.22
Skew normal B	Full	2359.38
(b) Global asset allocation benchmarks		
Normal		0.00
Skew normal A	Diagonal	3271.25
Skew normal B	Diagonal	3618.51
Skew normal B	Full	3458.34

Table 3. Parameter estimates for diagonal Δ skew normal on four securities. Parameter estimates for the diagonal Δ model of Sahu *et al.* (2003) used to fit the daily stock returns of Carnival, Starwood Hotels and Resorts, L-3 Communications Holdings, and Raytheon from July 2001 to June 2006. These estimates are the result of a Bayesian Markov Chain Monte Carlo iterative sampling routine. These parameters combine to give the mean $(\mu+(2/\pi)^{1/2}\Delta\mathbf{1})$, variance $(\Sigma+(1-2/\pi)\Delta\Delta')$, and skewness (see appendix A.1 for formula).

μ	Carnival	Starwood	L-3	Raytheon
	−0.295	−0.421	0.365	0.534
Σ	Carnival	Starwood	L-3	Raytheon
Carnival	3.0839	2.1323	−0.2218	−0.4981
Starwood	2.1323	3.6151	−0.1770	−0.5255
L-3	−0.2218	−0.1770	3.5001	1.2067
Raytheon	−0.4981	−0.5255	1.2067	2.7837
Δ	Carnival	Starwood	L-3	Raytheon
Carnival	0.409	0	0	0
Starwood	0	0.565	0	0
L-3	0	0	−0.391	0
Raytheon	0	0	0	−0.580

decisions. Our first observation, for both sets of data and for all of the utilities that we investigated, was that the Michaud's (1998) method is markedly sub-optimal when compared with the expected utility based on the predictive density for the corresponding best probability model (see table 5). Michaud's (1998) method gives expected utility that is between 0.63% and 17.2% lower than the expected utility found with the predictive density.

Interestingly, the parameter expected utility was very close and in some cases almost exactly equal to the predictive expected utility. This suggests that the

Table 4. Parameter estimates for full Δ skew normal on global asset allocation benchmark. Parameter estimates for diagonal Δ model of Sahu *et al.* (2003) used to fit the daily benchmark indices of Russell 1000, Russell 2000, Oil, MSCI EAFE, MSCI EMF, Gold, and the 10 year T-bond from January 2002 to June 2006. These estimates are the result of a Bayesian Markov Chain Monte Carlo iterative sampling routine. These parameters combine to give the mean $(\mu+(2/\pi)^{1/2}\Delta\mathbf{1})$, variance $(\Sigma+(1-2/\pi)\Delta\Delta')$, and skewness (see appendix A.1 for the formula).

μ	R1000	R2000	Oil	EAFE	EMF	Gold	10T
	−0.033	0.096	−0.175	0.163	0.328	0.441	−0.478
Σ	R1000	R2000	Oil	EAFE	EMF	Gold	10T
R1000	0.792	0.805	−0.077	0.318	0.199	−0.093	0.366
R2000	0.805	1.099	−0.085	0.362	0.318	−0.009	0.309
Oil	−0.077	−0.085	3.902	0.017	0.022	0.047	0.140
EAFE	0.318	0.362	0.017	0.679	0.484	0.166	0.120
EMF	0.199	0.318	0.022	0.484	0.794	0.279	−0.041
Gold	−0.093	−0.009	0.047	0.166	0.279	0.947	−0.342
10T	0.366	0.309	0.140	0.120	−0.041	−0.342	1.704
Δ	R1000	R2000	Oil	EAFE	EMF	Gold	10T
R1000	0.060	0	0	0	0	0	0
R2000	0	−0.077	0	0	0	0	0
Oil	0	0	0.321	0	0	0	0
EAFE	0	0	0	−0.162	0	0	0
EMF	0	0	0	0	−0.339	0	0
Gold	0	0	0	0	0	−0.506	0
10T	0	0	0	0	0	0	0.597

Table 5. Three moment optimization for four equity securities. This table contains predictive utilities for the weights that maximize utility as a linear function of the three moments of the multivariate normal model by three different methods for daily stock returns of Carnival Corp., Starwood Hotels & Resorts, L-3 Communications Holdings, and Raytheon Co. from July 2001 to June 2006. The first method is based on predictive or future values of the portfolio (results in $\omega_{3,\mathrm{pred}}$), the second is based on the posterior parameter estimates ($\omega_{3,\mathrm{param}}$), and the third is the method proposed by Michaud ($\omega_{3,\mathrm{Michaud}}$). The weights that are found by each method are ranked by the three moment predictive utility they produce (i.e. $E[u_{3,\mathrm{pred}}(\omega)]=\omega' m_p-\lambda\omega' V_p\omega+\gamma\omega' S_p\omega\otimes\omega$, 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, where λ represents risk aversion and γ represents skewness preference. 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, $\omega_{3,\mathrm{pred}}$ gives the highest expected utility.

λ	γ	$EU(\omega_{3,\mathrm{pred}})$	$EU(\omega_{3,\mathrm{param}})$	$EU(\omega_{3,\mathrm{Michaud}})$
0	0	0.099	0.099	0.094
0	0.1	0.100	0.097	0.087
0	0.2	0.101	0.093	0.090
0	0.5	0.104	0.094	0.086
0.1	0	−0.073	−0.073	−0.079
0.1	0.1	−0.073	−0.073	−0.076
0.1	0.2	−0.073	−0.073	−0.075
0.1	0.5	−0.072	−0.072	−0.075
0.25	0	−0.291	−0.291	−0.294
0.25	0.1	−0.291	−0.291	−0.294
0.25	0.2	−0.291	−0.291	−0.293
0.25	0.5	−0.290	−0.290	−0.296
0.5	0	−0.655	−0.655	−0.659
0.5	0.1	−0.655	−0.655	−0.659
0.5	0.2	−0.654	−0.654	−0.663
0.5	0.5	−0.654	−0.654	−0.660

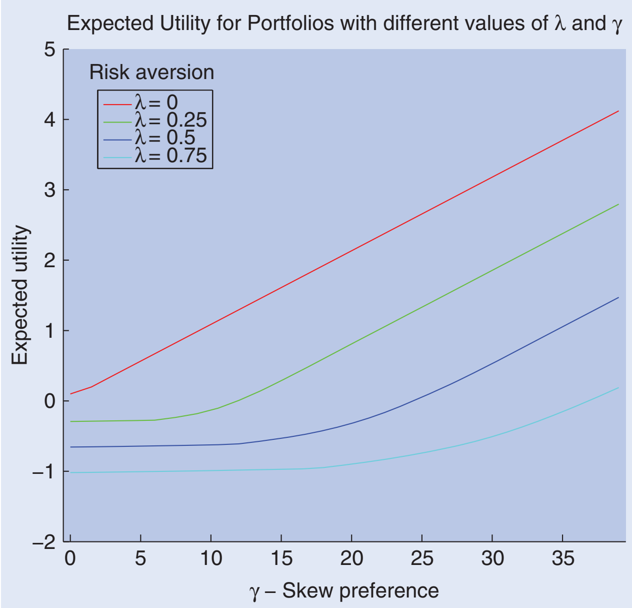

Figure 5. How the predictive expected utility for portfolios consisting of Carnival, Starwood Hotels and Resorts, L-3 Communications Holdings, and Raytheon change with the risk aversion λ and skewness preference γ.

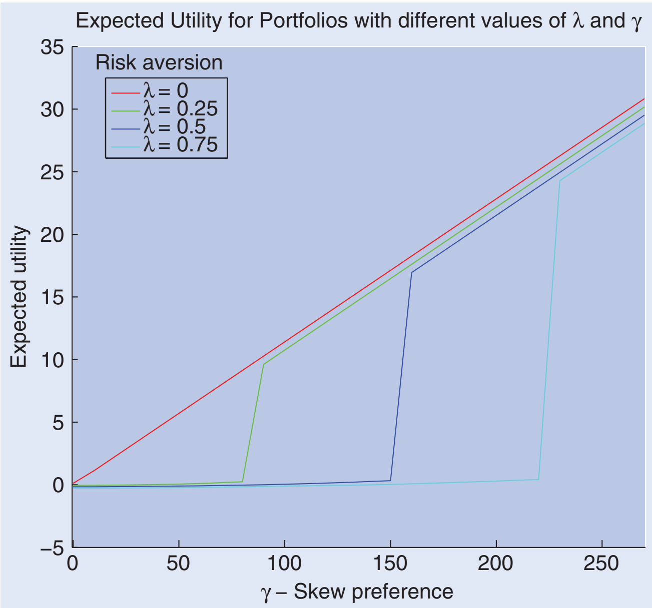

Figure 6. How the predictive expected utility for portfolios consisting of Russell 1000, Russell 2000, Oil, MSCI EAFE, MSCI EMF, Gold, and the 10 year T-bond change with the risk aversion λ and skewness preference γ.

difference in the posterior and predictive moments is relatively small for these data sets. In practice, however, if investors have a nonlinear utility function or larger risk aversion and skewness preference (as indicated by larger values of λ and γ), these differences should become more pronounced.

In order to explore the improvements that can come from including skewness, we plotted the predictive expected utility as a function of the skewness preference for a range of different risk preferences (see figures 5 and 6). For both sets of data, including the skewness clearly improves the expected utility (compare expected utility with no skewness preference to the remaining levels of skewness preference) and the increasing utility indicates that substantially better portfolios are available for both sets of data over a wide range of different skewness preferences.

It is interesting, however, to note that there are sudden changes in the expected utility as the skewness preference increases. These sudden changes in expected utility are due to large changes in the portfolio weights as a function of the skewness preferences (see figures 7 and 8). For both sets of data, there appears to be a change point, with respect to the skewness preference, given that the risk preference is held constant. Before and after these change points the portfolio allocations change gradually and at the change point the allocation switches dramatically, putting most of the weight on a single asset. The evolution of the maximum, predictive expected utility is graphically summarized in the extended efficient frontier graph for the four stock data set (see figure 9). In this graph the same large shift in the portfolio allocation is observed when the optimal portfolio suddenly jumps to another portion of the extended efficient frontier.

Interestingly this pattern of suddenly switching to a distinctly different set of portfolio weights repeats itself for larger and larger values of skewness preference as the risk preference increases, suggesting that there is an underlying impact of skewness which leads to a set of weights which dominate the impact of variance as skewness preference increases.

5. Conclusion

We provide a framework for portfolio selection that addresses three major issues: (1) asset returns are not normal, (2) investors prefer assets with large upside to those with large downside, and (3) the inputs for the portfolio optimization are uncertain. We also demonstrate that our generalization of the skew normal model of Sahu *et al*. (2003) is able to capture higher moments. It is flexible enough to allow for skewness and coskewness and at the same time accommodates heavy tails. Additional features of the model include straightforward specification of conjugate prior distributions which allows for efficient simulation and posterior inference. We use Bayesian methods to incorporate parameter uncertainty into the predictive distribution of returns, as well as to maximize the expected utility.

We show that predictive utility can be written in terms of posterior parameter based utility plus additional terms. These additional terms can be very influential in an investor's utility. We compare our results to Michaud's (1998) resampling technique for portfolio selection. In addition to the Jensen's inequality problem, we show how the resampling approach is outside the efficient utility maximization framework.

While we believe that we have made progress on important issues in portfolio selection, there are at least

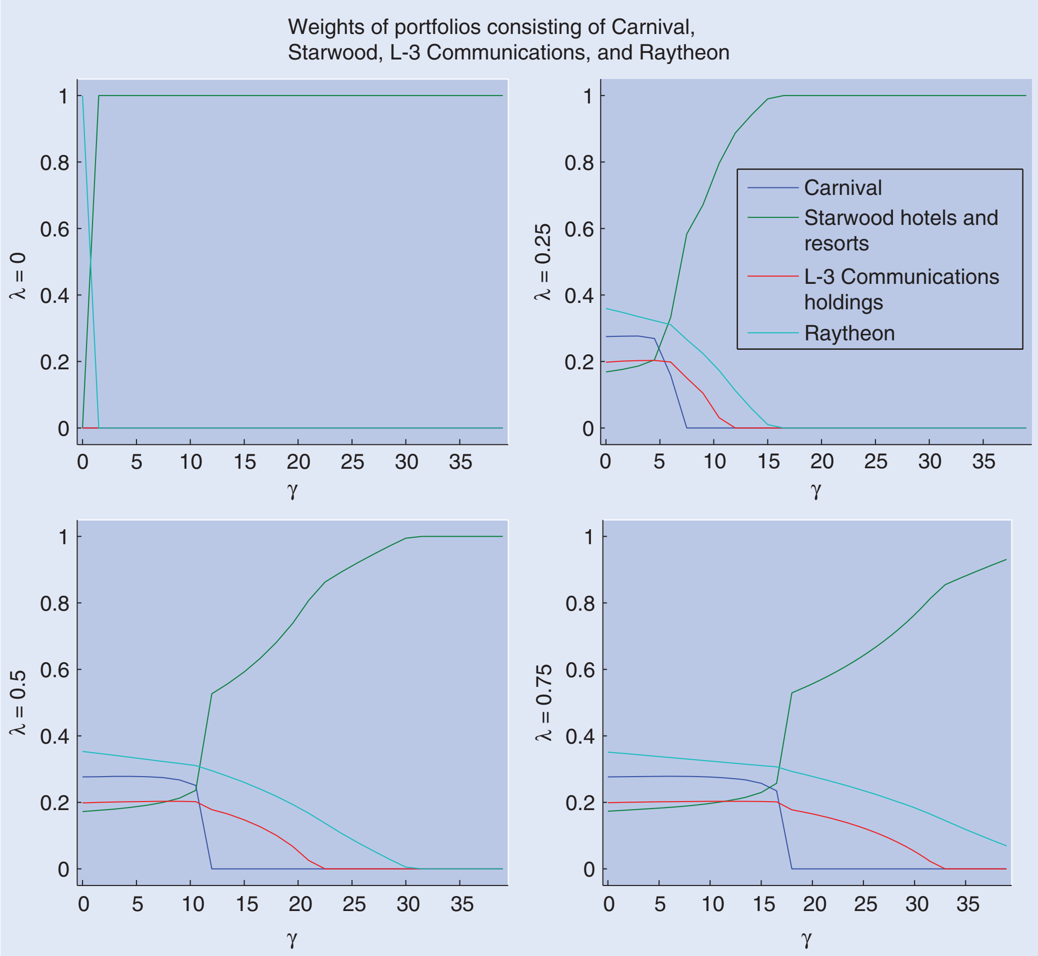

Figure 7. How the portfolio weights for portfolios consisting of Carnival, Starwood Hotels and Resorts, L-3 Communications Holdings, and Raytheon change with the risk aversion λ and skewness preference γ.

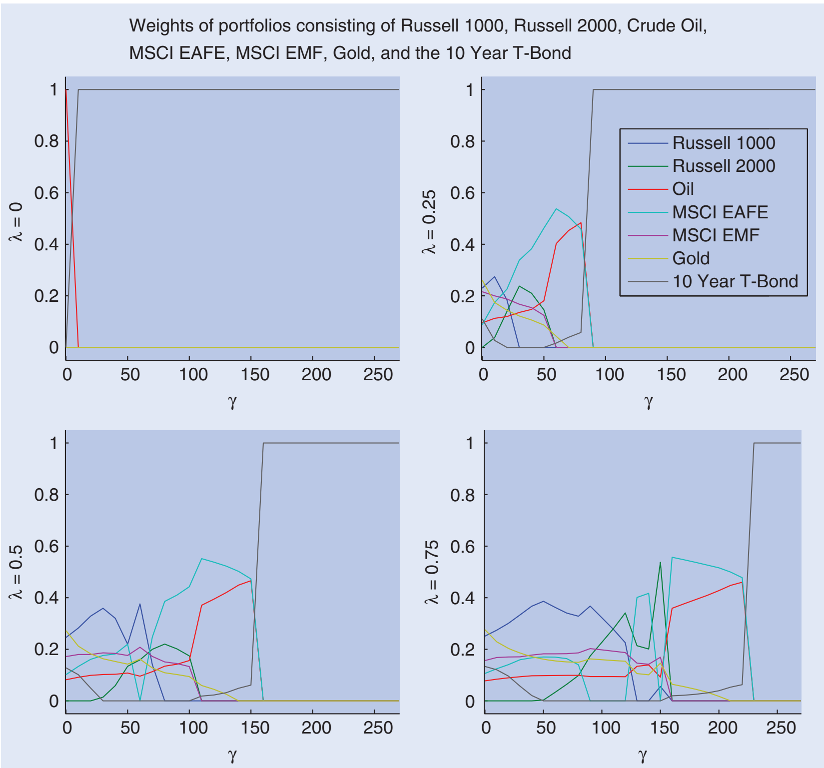

Figure 8. How the portfolio weights for portfolios consisting of Russell 1000, Russell 2000, Oil, MSCI EAFE, MSCI EMF, Gold, and the 10 year T-bond change with the risk aversion λ and skewness preference γ.

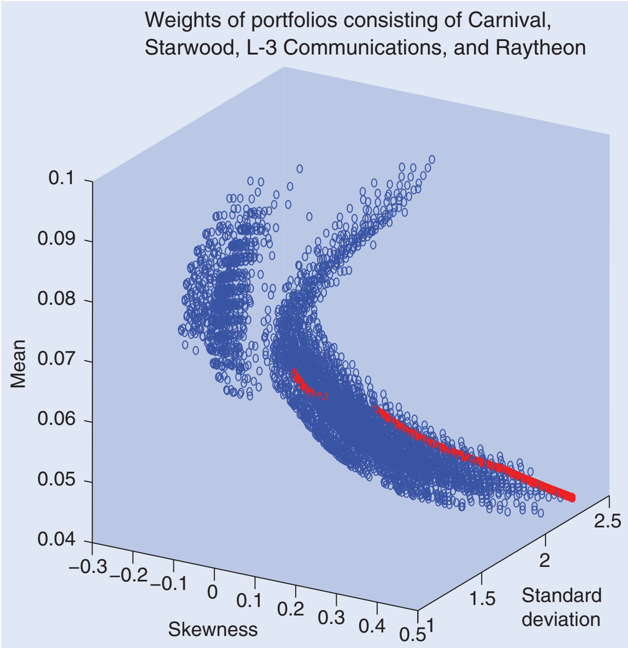

Figure 9. Mean–variance–skewness space of possible portfolios based on historical parameter estimates from the daily returns of Carnival, Starwood Hotels and Resorts, L-3 Communications Holdings, and Raytheon from July 2001 to June 2006. This is a three-dimensional plot of the mean, standard deviation and cubed-root of skewness.

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

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

Acknowledgements

We appreciate the comments of Philippe Jorion, Harry Markowitz, Haim Levy, Michael Brandt, Chris Adcock and Karl Shutes, as well as seminar participants at Boston College and participants of the Inquire-UK conference in Bournemouth.

References

Adcock, C.J., Exploiting skewness to build an optimal hedge fund with a currency overlay. *Eur. J. Finance*, 2005, **11**, 445–462.

Adcock, C.J., Asset pricing and portfolio selection based on the multivariate extended skew-Student-*t* distribution. *Ann. Oper. Res.*, 2009, **176**, 221–234.

Ang, A., Chen, J. and Xing, Y., Downside risk. *Rev. Financial Stud.*, 2006, **19**, 1191–1239.

Arrow, K.J., *Essays in the Theory of Risk-Bearing*, 1971 (Markham: Chigaco).

Athayde, G.M. de and Flôres, R.G., Finding a maximum skewness portfolio—a general solution to three moments portfolio choice. Working Paper no. 434, Graduate School of Economics, Getulio Vargas Foundation (Brazil), 2001.

Athayde, G.M. de and Flôres, R.G., Finding a maximum skewness portfolio—A general solution to three moments portfolio choice. *J. Econ. Dynam. Control*, 2004, **28**, 1335–1352.

Azzalini, A. and Dalla Valle, A., The multivariate skew-normal distribution. *Biometrika*, 1996, **83**, 715–726.

Bacmann, J.F. and Benedetti, S.M., Optimal Bayesian portfolios of hedge funds. *Int. J. Risk Assess. Mgmt*, 2009, **11**, 39–58.

Bakshi, G. and Madan, D., A theory of volatility spread. *Mgmt Sci.*, 2006, **52**, 1945–1956.

Bali, T.G. and Cakici, N., Market, industry, and firm level skewness and expected returns. Working Paper, Baruch College, 2010.

Barberis, N. and Huang, M.Stocks as lotteries: The implications of probability weighting for security prices. Working Paper, Yale University, 2007.

Barro, R.J., Rare disasters and asset markets in the twentieth century. *Q. J. Econ.*, 2006, **121**, 823–866.

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

Black, F. and Litterman, R., Global portfolio optimization. *Financial Anal. J.*, 1992, **48**, 28–43.

Blau, B.M. and Pinegar, J.M., Shorting Skewness, 2009. Available online at SSRN: http://ssrn.com/abstract=1468047.

Branco, M.D. and Dey, D.K., A general class of multivariate skew-elliptical distributions. *J. Multiv. Anal.*, 2001, **79**, 99–113.

Brandt, M.W., Brav, A., Graham, J.R. and Kumar, A., The idiosyncratic volatility puzzle: Time trend or speculative episodes? *Rev. Financial Stud.*, 2010, **23**, 863–899.

Briec, W., Kerstens, K. and Jokung, O., Mean–variance–skewness portfolio performance gauging: A general shortage function and dual approach. *Mgmt. Sci.*, 2007, **53**, 135–149.

Britten-Jones, M., Portfolio optimization and Bayesian regression. Unpublished Working Paper, London Business School, 2002.

Brown, S.J., The portfolio choice problem: Comparison of certainty equivalent and optimal Bayes portfolios. *Commun. Statist.: Simul. Comput.*, 1978, **B7**, 321–334.

Brunnermeier, M., Gollier, C. and Parker, J., Optimal beliefs, asset prices and the preference for skewed returns. *Am. Econ. Rev. Pap. Proc.*, 2007, **97**, 159–165.

Brunnermeier, M. and Parker, J., Optimal expectations. *Am. Econ. Rev.*, 2005, **95**, 1092–1118.

Burger, J.D. and Warnock, F.E., Foreign participation in local-currency bond markets. Board of Governors of the Federal Reserve System International Finance Discussion Papers Number 794, 2004.

Campbell, J.Y., Lettau, M., Malkiel, B.G. and Xu, Y., Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *J. Finance*, 2001, **56**, 1–43.

Cascon, A., Keating, C. and Shadwick, W.F., The Omega function. Working Paper, 2003.

Chabi-Yo, F., Conditioning information and variance bounds on pricing kernels with higher-order moments: Theory and evidence. *Rev. Financial Stud.*, 2008a, **21**, 181–231.

Chabi-Yo, F., Pricing kernels with coskewness and volatility risk. Charles A. Dice Center Working Paper No. 2008–25 and Fisher College of Business Working Paper No. 2008-03-023, 2008b. Available online at SSRN: http://ssrn.com/abstract=1317205.

Chen, J., Hong, H. and Stein, J., Forecasting crashes: Trading volume, past returns, and conditional skewness in stock prices. *J. Financial Econ.*, 2001, **61**, 345–381.

Chiang, T. and Li, J., Empirical analysis of asset returns with skewness, kurtosis, and outliers: Evidence from 30 Dow-Jones industrial stocks. Working Paper, Drexel University, 2007.

Chiu, W.H., Skewness preference, risk taking and expected utility maximization. Working Paper, University of Manchester, 2008.

Conrad, J., Dittmar, R.F. and Ghysels, E., Ex ante skewness and expected stock returns. Working Paper, University of North Carolina, 2010.

Cremers, J.-H., Kritzman, M. and Page, S.Portfolio formation with higher moments and plausible utility. Working Paper, 2004.

Cvitanić, J., Polimenis, V. and Zapatero, F., Optimal portfolio allocation with higher moments. *Ann. Finance*, 2008, **4**, 1–28.

Da Silva, A.S., Lee, W. and Pornrojnangkool, B., The Black–Litterman model for active portfolio management. *J. Portfol. Mgmt*, 2009, **35**, 61–70.

Damien, P.S., Wakefield, J. and Walker, G., Gibbs sampling For Bayesian nonconjugate and hierarchical models using auxiliary variables. *J. R. Statist. Soc. Ser. B*, 1999, **61**, 331–344.

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

DeMiguel, V., Garlappi, L. and Uppal, R., Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy? *Rev. Financial Stud.*, 2009, **22**, 1915–1953.

DeMiguel, V., Plyakha, Y., Uppal, R. and Vilkov, G., Improving portfolio selection using option-implied volatility and skewness, 2010. Available online at SSRN: http://ssrn.com/abstract=1474212

Dittmar, R.F., Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *J. Finance*, 2002, **57**, 369–403.

Fabozzi, F.J., Huang, D. and Zhou, G., Robust portfolios: Contributions from operations research and finance. *Ann. Oper. Res.*, 2010, in press.

Feiring, B.R., Wong, W., Poon, M. and Chan, Y.C., Portfolio selection in downside risk optimization approach: Application to the Hong Kong stock market. *Int. J. Syst. Sci.*, 1994, **25**, 1921–1929.

Frost, P.A. and Savarino, J.E., An empirical Bayes approach to efficient portfolio selection. *J. Financial Quant. Anal.*, 1986, **21**, 293–305.

Frost, P.A. and Savarino, J.E., For better performance: Constrain portfolio weights. *J. Portfol. Mgmt*, 1988, **14**, 29–34.

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

Goetzman, W.N. and Kumar, A., Why do individual investors hold under-diversified portfolios? Working Paper, Yale University, 2004.

Guidolin, M. and Timmermann, A., Optimal portfolio choice under regime switching, skew and kurtosis preferences. *J. Econ. Dynam. Control*, 2007, **31**, 3503–3544.

Hall, N.G., Long, Z., Qi, J. and Sim, M., A satisficing model for project selection. Working Paper, Ohio State University, 2009.

Harvey, C.R., Liechty, J.C. and Liechty, M.W., Bayes verses resampling: A rematch. *J. Invest. Mgmt*, 2008, **6**, 29–45.

Harvey, C.R., Liechty, J.C. and Liechty, M.W., Parameter uncertainty and asset allocation. In *Handbook of Quantitative Asset Management*, edited by B. Scherer and K. Winston, 2011 (Oxford University Press: Oxford), forthcoming.

Harvey, C.R. and Siddique, A., Conditional skewness in asset pricing tests. *J. Finance*, 2000a, **55**, 1263–1295.

Harvey, C.R. and Siddique, A., Time-varying conditional skewness and the market risk premium. *Res. Bank. Finance*, 2000b, **1**, 27–60.

Ingersoll, J.E., Multidimensional security pricing. *J. Financial Quant. Anal.*, 1975, **10**, 785–798.

Jobson, J.D., Korkie, B. and Ratti, V., Improved estimation for Markowitz portfolios using James–Stein type estimators. *Proc. Am. Statist. Assoc., Business Econ. Statist. Sect.*, 1979, 279–284.

Jobson, J.D. and Korkie, B., Estimation of Markowitz efficient portfolios. *J. Am. Statist. Assoc.*, 1980, **75**, 544–554.

Johannes, M., Polson, N. and Stroud, J.Sequential optimal portfolio performance: Market and volatility timing. Unpublished Working Paper, 2002.

Jondeau, E. and Rockinger, M., Optimal portfolio allocation under higher moments. *EFMA 2004 Basel Meetings Paper*, 2004. Available online at SSRN: http://ssrn.com/abstract=498322

Jondeau, E. and Rockinger, M.G., The economic value of distributional timing. Swiss Finance Institute Research Paper No. 06–35, 2006. Available online at SSRN: http://ssrn.com/abstract=957514.

Jorion, P., Bayes–Stein estimation of portfolio analysis. *J. Financial Quant. Anal.*, 1986, **21**, 279–292.

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

Jurczenko, E., Maillet, B.B. and Merlin, P., Hedge funds portfolio selection with higher-order moments: A non-parametric mean–variance–skewness–kurtosis efficient frontier, 2005. Available online at SSRN: http://ssrn.com/abstract=676904

Kan, R. and Zhou, G., Optimal portfolio choice with parameter uncertainty. *J. Financial Quant. Anal.*, 2007, **42**, 621–656.

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

Klein, R.W. and Bawa, V., The effect of estimation risk on optimal portfolio choice. *J. Financial Econ.*, 1976, **3**, 215–231.

Knight, J. and Satchell, S.E., Exact properties of measures of optimal investment for benchmarked portfolios. *Quant. Finance*, 2009, 1469–7696.

Kon, S., Models of stock returns – A comparison. *J. Finance*, 1984, **39**, 147–165.

Konno, H., Shirakawa, H. and Yamazaki, H., A mean–absolute deviation–skewness portfolio optimization model. *Ann. Oper. Res.*, 1993, **45**, 205–220.

Levy, H., The CAPM is alive and well: A review and synthesis. *Eur. Financial Mgmt*, 2010, **16**, 43–71.

Levy, H. and Levy, M., Prospect theory and mean–variance analysis. *Rev. Financial Stud.*, 2004, **17**, 1015–1041.

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

Liechty, M.W. and Lu, J., Multivariate normal slice sampling. *J. Comput. Graph. Stat.*, 2009, forthcoming.

Liew, J. and French, C., Quantitative topics in hedge fund investing. *J. Portfol. Mgmt*, 2005, **31**, 21–32.

Markowitz, H., Portfolio selection. *J. Finance*, 1952a, **7**, 77–91.

Markowitz, H., The utility of wealth. *J. Polit. Econ.*, 1952b, 152–158.

Markowitz, H., *Portfolio Selection: Efficient Diversification of Investments*, 2nd ed., 1959 (Blackwell: Malden).

Markowitz, H., Todd, P., Xu, G. and Yamane, Y., Computation of mean–semivariance efficient sets by the critical line algorithm. *Ann. Oper. Res.*, 1993, **45**, 307–317.

Markowitz, H. and Usmen, N., The likelihood of various stock market return distributions, Part 2: Empirical results. *J. Risk Uncert*., 1996, **13**, 221–247.

Markowitz, H. and Usmen, N., Resampled frontiers vs diffuse Bayes: An experiment. *J. Invest. Mgmt*, 2003, **1**, 9–25.

Martellini, L. and Ziemann, V., Extending Black–Litterman analysis beyond the mean–variance framework. *J. Portfol. Mgmt*, 2007, **33**, 33–44.

Martin, I.W.R., Consumption-based asset pricing with higher cumulants. Working Paper, Stanford University, 2010.

Mencia, J. and Sentana, E., Multivariate location–scale mixtures of normals and mean–variance–skewness portfolio allocation. *J. Econometr*., 2009, **153**, 105–121.

Meyn, S.P. and Tweedie, R.L., *Markov Chains and Stochastic Stability (Communications and Control Engineering)*, 1993 (Springer: New York).

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

Mills, T.C., Modelling skewness and kurtosis in the London Stock Exchange FT-SE index return distributions. *Statistician*, 1995, **44**, 323–332.

Mitton, T. and Vorkink, K., Equilibrium underdiversifacation and the preference for skewness. *Rev. Financial Stud*., 2007, **20**, 1255–1288.

Morton, D. and Popova, I., Asset allocation with subsistence: The role of returns' higher moments, 2009. Available online at SSRN: http://ssrn.com/abstract=1517929

Newton, M.A. and Raftery, A.E., Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion). *J. R. Statist. Soc. Ser. B*, 1994, **56**, 3–48.

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

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

Pástor, L., Portfolio selection and asset pricing models. *J. Finance*, 2000, **55**, 179–223.

Pástor, L. and Stambaugh, R.F., Comparing asset pricing models: An investment perspective. *J. Financial Econ*., 2000, **56**, 335–381.

Pástor, L. and Stambaugh, R.F., Investing in equity mutual funds. *J. Financial Econ*., 2002, **63**, 351–380.

Patton, A.J., On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *J. Financial Econometr*., 2004, **2**, 130–168.

Peiro, A., Skewness in financial returns. *J. Bank. Finance*, 1999, **23**, 847–862.

Polson, N.G. and Tew, B.V., Bayesian portfolio selection: An empirical analysis of the S&P 500 index 1970–1996. *J. Business Econ. Statist*., 2000, **18**, 164–73.

Post, T., Levy, H. and van Vliet, P., Risk aversion and skewness preference. *J. Bank. Finance*, 2008, **32**, 1178–1187.

Poti, V., Estimating factor models with coskewness: Stochastic discount factor volatility bounds and the problem of luck. Working Paper, Dublin City University, 2010.

Premaratne, G. and Tay, A.S., How should we interpret evidence of time varying conditional skewness? National University of Singapore, 2002.

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

Roy, A.D., Safety first and the holding of assets. *Econometrica*, 1952, **20**, 431–449.

Rubinstein, M., The fundamental theorem of parameter preference security valuation. *J. Financial Quant. Anal*., 1973, **8**, 61–69.

Sahu, S.K., Branco, M.D. and Dey, D.K., A new class of multivariate skew distributions with applications to Bayesian regression models. *Can. J. Statist*., 2003, **12**, 129–150.

Scherer, B., Portfolio resampling: Review and critique. *Financial Anal. J*., 2002, **58**, 98–109.

Scherer, B., A note on the out of sample performance of resampled efficiency. *J. Asset Mgmt*, 2006, **7**, 170–178, (9).

Sfiridis, J.M., Incorporating higher moments into financial data analysis, 2005. Available online at SSRN: http://ssrn.com/abstract=887967

Shadwick, W.F. and Keating, C., A universal performance measure. *J. Perform. Measure*., 2002, **6**, 59–84.

Sortino, F.A. and Forsey, H.J., On the use and misuse of downside risk. *J. Portfol. Mgmt*, 1996, **22**, 35–42.

Sortino, F.A. and Price, L.N., Performance measurement in a downside risk framework. *J. Invest*., 1994, **3**, 50–58.

Treynor, J.L. and Black, F., How to use security analysis to improve portfolio selection. *J. Business*, 1973, 66–88, January.

Vorkink, K., Mitton, T. and Boyer, B.H., Expected idiosyncratic skewness, 2009. Available online at SSRN: http://ssrn.com/abstract = 1089071.

Wilcox, J. and Fabozzi, F.J., A discretionary wealth approach to investment policy. Yale ICF Working Paper No. 09-03, 2009. Available online at SSRN: http://ssrn.com/abstract=1355922

Williams, J.M. and Ioannidis, C., Static and dynamic asset allocation with higher moments, 2006. Available online at SSRN: http://ssrn.com/abstract=954995

Zellner, A. and Chetty, V.K., Prediction and decision problems in regression models from the Bayesian point of view. *J. Am. Statist. Assoc*., 1965, **60**, 608–616.

Zhou, G., Portfolio management and estimation risk, 2009. Available online at SSRN: http://ssrn.com/abstract=1452985

Appendix A: Skew normal probability model

A.1. Density and moment generating function

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

$$\begin{aligned} f(y \mid \mu, \Sigma, \Delta) &= 2^{\ell} \, |\Sigma + \Delta\Delta'|^{-1/2} \phi_{\ell}[(\Sigma + \Delta\Delta')^{-1/2}(y - \mu)] \\ &\quad \times \Phi_{\ell}[(I - \Delta'(\Sigma + \Delta\Delta')^{-1}\Delta)^{-1/2}\Delta'(\Sigma + \Delta\Delta')^{-1}(y - \mu)], \end{aligned} \tag{A1}$$

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

The moment generating function becomes

$$M_{\mathbf{Y}}(\mathbf{t}) = 2^{\ell} e^{t'\mu + t'(\Sigma + \Delta\Delta')t/2} \Phi_{\ell}(\Delta t). \tag{A2}$$

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

$$\begin{aligned} m &= \mu + (2/\pi)^{1/2}\Delta\mathbf{1}, \quad V = \Sigma + (1 - 2/\pi)\Delta\Delta', \\ S &= \Delta EZ\Delta' \otimes \Delta' + 3\mu' \otimes \{\Delta\Delta'(1 - 2/\pi) + 2/\pi\Delta\mathbf{1}(\Delta\mathbf{1})'\} \\ &\quad + 3\{(2/\pi)^{1/2}(\Delta\mathbf{1})' \otimes [\Sigma + \mu\mu']\} + 3\mu' \otimes \Sigma \\ &\quad + \mu\mu' \otimes \mu' - 3m' \otimes V - mm' \otimes m', \end{aligned} \tag{A3}$$

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

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

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

A.2. First three moments of a linear combination

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

$$E(\omega' X) = \omega' m,$$
$$\mathrm{Var}(\omega' X) = \omega' V \omega,$$
$$\mathrm{Skew}(\omega' X) = \omega' S\omega \otimes \omega,$$

where m, V and S are given above.

A.3. Model specification

A.3.1. Likelihood and priors. The skew normal density is defined in terms of a latent (unobserved) random variable Z, which comes from a truncated standard normal density. The likelihood is given by

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

where N_ℓ is a multivariate normal density,

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

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

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

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

A.3.2. Full conditionals. Assuming n independent skew normal observations, the full conditional distributions are as follows:

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

where

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

$$B = \sum_{i=1}^{n} y_i'\Sigma^{-1}y_i + \frac{1}{100} I_{\ell(\ell+1)}, \quad b = \sum_{i=1}^{n} y_i \Sigma^{-1} x_i,$$

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

and $y_i = (I_\ell, Z_i' \otimes I_\ell)$.

A.4. Estimation using the slice sampler

The slice sampler introduces an auxiliary variable, which we will call u, in such a way that the draws from both the desired variable and the auxiliary variable can be obtained by drawing from appropriate uniform densities (for more details, see Damien *et al.* 1999. Also see Liechty and Lu 2009, who introduce a multivariate slice sampler that could be used to extend the complexity of the model by allowing for correlated latent Z variables). To illustrate, assume that we want to sample from the following density:

$$f(x) \propto \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} I\{x \geq 0\}, \qquad \text{(A4)}$$

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

$$f(x, u) \propto I\left\{u \leq \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} I\{x \geq 0\}\right\}. \qquad \text{(A5)}$$

It is easy to see that based on (A5), the marginal density of x is given by (A4) and that the conditional density of u given x is a uniform density, or

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

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

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

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