Portfolio Selection with Skew Normal Asset Returns

Quan Gan*

Draft: July 1, 2013

Abstract: This paper examines the portfolio selection problem with skew normal asset returns. By exploring an alternative parameterization of Azzalini & Dalla Valle (1996)'s multivariate skew normal distribution I show that the multivariate skew normal distribution is a special case of Simaan (1993)'s three-parameter model. All Simaan (1993)'s results are applicable to the skew normal asset returns. The three-parameter efficient frontier is spanned by three funds which include two funds from the mean-variance portfolio selection. Combining the skew normal asset returns with the CARA utility, I obtain the closed-form certainty equivalent and skewness premium. I show that the skewness premium is positive (negative) when asset returns have negative (positive) skewness. The magnitude of the skewness premium is increasing in market risk aversion. I use the skew normal certainty equivalent to evaluate the economic value of incorporating higher moments in portfolio selection. I find that when investors face broad investment opportunities, the economic value of considering higher moments is negligible under realistic margin requirements.

(JEL. G11, G12, D81)

Keywords: Portfolio selection, Skew normal, Certainty equivalent, Skewness premium

*Discipline of Finance, University of Sydney, Sydney 2006, NSW, Australia
quan.gan@sydney.edu.au, 61-2-91140831

1 Introduction

Research of portfolio selection with higher moments faces two obstacles. First, the theory of higher moments portfolio selection is underdeveloped comparing to the mean-variance portfolio selection theory. In the mean-variance portfolio selection theory, the expected return and risk characteristics of portfolios can be intuitively compared in a "mean-standard deviation" diagram. The mean-variance efficient frontier is spanned by two funds and the two-fund separation simplifies the calculation of efficient portfolios. The portfolio selection with higher moments however have several frameworks (Simaan 1993, Athayde & Flôres 2004, Briec et al. 2007, Mencía & Sentana 2009, and Low et al. 2012). Although each of these frameworks has its own merits, the lack of a unified framework is certainly a drawback. Second, the economic value of higher moments in portfolio selection is hard to capture through the comparison of expected utilities. Empirical studies (Patton 2004, Jondeau & Rockinger 2006, and Harvey et al. 2010) usually use numerical integrations and approximations which involve tedious computation.

This paper aims to advance both theoretical and empirical understanding of higher moments portfolio selection. From the perspective of theoretical development, I explore Simaan (1993)'s three-parameter model by combining it with Azzalini & Dalla Valle (1996)'s multivariate skew normal distribution. I examine an alternative parameterization of the multivariate skew normal distribution and show that under this parameterization the multivariate skew normal distribution is a special case of the three-parameter model. All theoretical results in Simaan (1993) are applicable to skew normal asset returns. Among all available higher moments portfolio selection frameworks, Simaan (1993)'s three-parameter model has many desirable features. The three-parameter efficient frontier is spanned by a three-fund separation which nests the classic two-fund separation and has nice geometric properties. The skew normal distribution provides flexibility of modeling higher moments behavior in asset returns and this distribution is well-known in the statistics literature. The skew normal distribution provides a way to operationalize the three-parameter model.

1

Combining the skew normal asset returns and the CARA (constant absolute risk aversion) utility, I obtain the closed-form certainty equivalent and skewness premium. Using the skewness premium, I show that when investors face positive skewness, they are willing to sacrifice some required return to chase positive skewness. On the other hand, when investors face negative skewness, they require higher return to compensate. Market risk aversion plays an important role of determining the magnitude of skewness premium. When market risk aversion is low (high), the magnitude of skewness premium is also low (high).

From the perspective of empirical study, the skew normal certainty equivalent allows me to empirically evaluate the economic value of higher moments in portfolio selection. It is well-known in the finance literature that combining the CARA utility and multivariate normal asset returns leads to the mean-variance certainty equivalent. Previous studies use the mean-variance certainty equivalent to analyze theoretical and empirical problems in portfolio selection (Pástor & Stambaugh 2000, Tu & Zhou 2004, and Kan & Zhou 2007, among others). Tu & Zhou (2004) compare the economic loss of a mean-variance investor when the multivariate normal data-generating process is replaced by a set of multivariate t distributions. Part of their study can be interpreted as a study on the economic value of higher moments because multivariate t has heavier tails than multivariate normal distribution. Although Tu & Zhou (2004) find nontrivial changes in optimal portfolio weights, they do not find any significant change in economic value when heavier tails are considered. The certainty equivalent of the CARA utility with student t asset returns however has no closed-form because the cumulant-generating function of multivariate t does not exist. Thus the economic value comparison conducted by Tu & Zhou (2004) might be better used for investors facing multivariate normal asset returns but mistakenly regard them as multivariate t returns. In my study, the certainty equivalent is obtained on skew normal asset returns, so I can explicitly measure the economic value difference when investors face multivariate skew normal asset returns but mistakenly regard them as multivariate normal returns.

Margin requirements can eliminate the benefits of higher moments by constraining

investors' ability of taking extreme long/short positions. Unconstrained portfolio selection results may provide biased value of higher moments because investors can take extreme positions. Patton (2004) considers short-sales constraints and finds that under short-sales constraints, the economic gain of considering higher moments is limited. Short-sales constraints however are different from margin requirement constraints. In my empirical study, I explicitly incorporate margin requirements. When investors face broad investment opportunities, I find that the economic gain of considering higher moments is negligible when realistic margin requirements exist.

The rest of the paper is organized as follows. Section 2 provides a review on the results of Simaan (1993) and relates them to other current developments in higher moments portfolio selection. Section 3 discusses the multivariate skew normal distribution and relate it to Simaan (1993)'s three-parameter model. Section 4 analyzes the portfolio selection with the CARA family of utilities by obtaining the closed-form certainty equivalent. It also characterizes the skewness premium, and shows a stochastic dominance result. Section 5 discusses the certainty equivalent approach of evaluating the economic value of higher moments, applies the approach to the data and reports the empirical results. Section 6 concludes.

2 Simaan's Three-Parameter Model

Simaan (1993)'s three-parameter model is generated from a perturbation of multivariate elliptical distribution. In his model, asset returns can be characterized by three parameters: location, scale and skewness. By solving a quadratic programming problem, Simaan (1993) shows a three-fund separation result which characterizes the efficient frontier for all risk-averse investors. The efficient frontier takes into account all moments of asset return distribution. Simaan (1993)'s three-parameter framework produces an elegant portfolio selection solution which extends intuitions obtained in Markowitz (1952)'s mean-variance framework.

2.1 Three-Parameter Model and Portfolio Selection

Simaan (1993) approaches the higher-moments portfolio selection problem by specifying skewed asset return distribution as a perturbation of elliptical distributions[1]. Let $y_t \in \Re^k$ be a vector of asset returns observed at time t, on which observations $(t = 1, ..., T)$ are available. The data-generating process of y_t is written as a multivariate regression:

$$y_t = \xi + Cs_t + e_t. \tag{1}$$

where s_t is a scalar perturbation with any non-elliptical distribution. The perturbation s_t can be centered to have zero mean so that the mean vector of asset returns $E(y_t) = \xi$. Conditioning on s_t, the random vector e_t follows a joint-elliptical distribution El_k with zero mean, positive-definite scale matrix Σ and characteristic function ϕ, i.e. $e_t|s_t \sim El_k(0, \Sigma; \phi)$ or equivalently $y_t|s_t \sim El_k(\xi + Cs_t, \Sigma; \phi)$. When elements of e_t are uncorrelated with each other, the regression (1) is a one-factor model with factor loading matrix C. In Section 3, I will show that Simaan (1993)'s construction is related to Engle et al. (1983)'s weak exogeneity concept.

In portfolio selection problems, the investor wants to maximize her expected utility with or without constraints by choosing the asset weight vector $\rho = (\rho_1, ..., \rho_k)'$. When asset returns follow the data-generating process (1), Simaan (1993) shows that the portfolio return $\rho'y_t$ is fully characterized by three parameters: 1) location parameter $\rho'\xi$, 2) scale parameter $\rho'\Sigma\rho$, and 3) skewness parameter $\rho'C$. And the optimal solutions of portfolio selection with or without a risk-free asset satisfy quadratic programming problems described in Proposition 1.

Proposition 1. For any concave utility there exists a pair (μ_0, s_0) such that the portfolio selection problem without a risk-free asset is equivalent to the following

---

[1]Simaan (1993) does not distinguish between spherical distributions and elliptical distributions. In standard statistics text (e.g. Fang et al. 1990), spherical and elliptical distributions are different. Simaan (1993) actually discusses elliptical distributions.

quadratic programming problem:

$$\min_{\rho \in \mathfrak{R}^k} \quad \frac{1}{2}\rho'\Sigma\rho$$

$$\text{s.t.} \quad \rho'\xi = \mu_0,$$

$$\rho'C = s_0,$$

$$\rho'\tau = 1. \tag{2}$$

When there exists a risk-free asset with return R_f, the portfolio selection problem is equivalent to the following quadratic programming problem:

$$\min_{\rho \in \mathfrak{R}^k} \quad \frac{1}{2}\rho'\Sigma\rho$$

$$\text{s.t.} \quad \rho'\xi + \rho_0 R_f = \mu_0,$$

$$\rho'C = s_0,$$

$$\rho'\tau + \rho_0 = 1. \tag{3}$$

where τ is the vector of ones and ρ_0 denotes the investment in the risk-free asset.

The implication of Proposition 1 is that solving the quadratic programming problem (2) or (3) by varying μ_0 and s_0 will provide a menu that includes efficient portfolios for all risk-averse investors. The return-risk trade-off on the efficient frontier is captured by a three-parameter vector (ξ, Σ, C). Simaan (1993)'s approach is similar to the Markowitz (1952)'s mean-variance portfolio selection approach. Both are fully compatible with expected utility maximization and both provide a menu of efficient portfolios to all risk averse investors. Proposition 2 shows a three-fund separation result which extends the two-fund separation in the mean-variance framework.

Proposition 2. (Three-Fund Separation) When asset returns follow the data-

generating process (1), the efficient frontier is spanned by the following three funds:

$$a_1 = \frac{V^{-1}\xi}{\tau'V^{-1}\xi}, a_2 = \frac{V^{-1}\tau}{\tau'V^{-1}\tau}, \text{ and } a_3 = \frac{V^{-1}C}{\tau'V^{-1}C}$$

where $V = Var(y_t)$. If there exists a risk-free asset with a rate R_f, then the efficient frontier is spanned by the risk-free asset, fund

$$a = \frac{V^{-1}(\xi - R_f\tau)}{\tau'V^{-1}(\xi - R_f\tau)}, \tag{4}$$

and a_3.

The fund a_1 (a) is the portfolio that maximizes (excess) portfolio expected return $\rho'\xi$ $(\rho'(\xi - R_f\tau))$ for a given portfolio variance. The fund a_2 is the global minimum variance portfolio. The fund a_3 is the portfolio that maximizes portfolio skewness for a given portfolio variance. Proposition 2 provides an elegant solution to the higher-moments portfolio selection problem. When perturbation in asset returns disappears, the three-fund separation naturally reduces to the classic two-fund separation of funds a_1 (a) and a_2.

2.2 Geometry of Three-Parameter Portfolio Selection

The efficient frontier constructed from Proposition 2 has nice geometric properties. Simaan (1993) shows that in the "(excess) mean - variance - skewness parameter C" space, the efficient set (with) without the risk-free asset) is the surface of an elliptical paraboloid which includes Markowitz (1952)'s minimum-variance set. The implications of this result are: 1) iso-variance contours are ellipses; 2) iso-mean and iso-C contours are parabolas; 3) the efficient set is the upper surface of an elliptical cone when variance is replaced by standard deviation. Figure 1 illustrates the geometric shapes of efficient set (with the risk-free asset) in two panels.

Panel (a) shows the efficient set in the "(excess) mean - variance - skewness parameter C" space and Panel (b) shows the efficient set in the "(excess) mean - standard deviation - skewness parameter C" space. I draw a relationship between the three-parameter model's geometric properties and properties of mean-variance-skewness efficient set shown by Athayde and Flôres (2004). Athayde and Flôres (2004) consider the geometry of optimal portfolios with first three sample moments in a "(excess) mean - standard deviation - cubic root of skewness" space. With the risk-free asset, they show an important linearity property of the efficient set. For a given k, suppose the efficient (minimum variance) portfolio is $\bar{\rho}$ when excess return $\mu = 1$ and cubic root of skewness is k. Portfolio $\bar{\rho}$'s standard deviation is $\bar{\sigma}$. Along the same direction k (in the "excess return - cubic root of skewness" space) with varying μ, the portfolios $\mu\bar{\rho}$ are also efficient with excess return μ and cubic root of skewness is $k\mu$. Their standard deviations are $\mu\bar{\sigma}$. Along each of these directions, the optimal portfolios are combinations of two funds: $\bar{\rho}$ and the risk-free asset. The linearity property comes from the mathematical fact that sample mean, variance and skewness are homogeneous functions of degree 1, 2 and 3 with respect to asset weights.

Panel (b) of Figure 1 intuitively shows that the linearity property holds with the three-parameter model when the cubic root of skewness is replaced by the skewness parameter C. Mathematically, the three-parameter efficient frontier is the solution of a quadratic programming in Proposition 1 and a portfolio's skewness parameter C is a homogeneous function of degree 1 with respect to asset weights. These two mathematical facts ensure the linearity property.

The iso-variance contours of the three-parameter efficient frontier are ellipses. The ellipse represents "a sort of ideal situation" (Athayde and Flôres 2004, Fig. 2(a)) with only one maximum, either for excess return or C, for a given variance. For positive skewness lover (negative skewness averter), the efficient set is the part of iso-variance curve from the maximum mean return point to the maximum skewness parameter point.

2.3 Simaan's Framework vs. Other Higher Moments Portfolio Selection Frameworks

There are several theoretical frameworks on higher moments portfolio selection. I review these theoretical frameworks including Simaan (1993) by summarizing them in Table 1.

[Insert Table 1 here.]

Table 1 is a two-by-two table which categorizes theoretical frameworks into four groups. "Parametric" ("Nonparametric") theories have (no) distributional assumption on asset returns. Theories consider all moments are identified as "all moments' theories. And theories only consider the first three moments are identified as "mean-variance-skewness" theories.

Mencía and Sentana (2009) consider the mean-variance-skewness portfolio selection with a location-scale mixture distribution of normals. The mixture distribution is flexible and its first three moments are fully characterized by three parameters. Mencía and Sentana (2009) show that the mean-variance-skewness frontier is spanned by three funds in the "mean-variance-skewness" space. Although their work has similarity to Simaan (1993). The mixture normal distribution however cannot be nested in Simaan (1993)'s three-parameter model.

Low et al. (2012) propose a skewness-aware measure and show that with this measure the efficient frontier in the presence of a risk-free asset has similar properties to the mean-variance efficient frontier. The skewness-aware measure has no distributional assumption and is based on the cumulant-generating function of a multivariate random variable which captures the information of all moments. However, for some distributions the cumulant-generating functions do not exist. Notable examples are multivariate t, stable and Cauchy distributions. Moreover, the skewness-aware framework is not fully compatible with the expected utility maximization.

Athayde and Flôres (2004) develop the mean-variance-skewness portfolio selection solution for general distributions based on their sample moments. They characterize several important geometric properties of the efficient frontier including the linearity property I have discussed in Section 2.2. They point out that the efficient frontier may not be well-behaved (e.g. may exist discontinuity) and emphasizes the importance of establishing a theoretical foundation to guide empirical studies. Briec et al. (2007), instead of solving the efficient frontier directly, propose a shortage function which guides the search of efficient portfolio in the mean-variance-skewness space. Their development is based on a dual approach and does not provide separation results or direct characterization of the efficient frontier.

Comparing to above frameworks, Simaan (1993)'s three-parameter model has several desirable properties: 1) it is fully compatible with the expected utility maximization; 2) it nests the mean-variance portfolio selection results; 3) it has nice geometric properties; 4) it considers information of all moments. Despite all these advantages, the three-parameter model is largely in oblivion. In next section, I will show that a well-known skew distribution in the statistics literature is a special case of the three-parameter model and thus can operationalize the model.

3 Skew Normal Distribution as A Three-Parameter Model

The key to operationalize the three-parameter model is to find a flexible skew distribution which satisfies the model (1). In this section, I show that the multivariate skew normal distribution proposed by Azzalini & Dalla Valle (1996) is a special case of the three-parameter model. The multivariate skew normal distribution can be generated by two equivalent methods: transformation method and conditioning method. I focus on the conditioning method because the multivariate skew normal distribution can be shown as a three-parameter model when conditioning method is used.

Let $y_t \in \Re^k$ and scalar s_t have a joint multivariate normal distribution with mean

vector μ and covariance matrix Γ, which is represented as

$$\begin{pmatrix} y_t \\ s_t \end{pmatrix} \sim N_{k+1} \left[\mu = \begin{pmatrix} \xi \\ 0 \end{pmatrix}, \Gamma = \begin{pmatrix} \Sigma + CC' & C \\ C' & 1 \end{pmatrix} \right] \qquad (5)$$

Here, both ξ and C are $(k \times 1)$ vector parameters and Σ is a $(k \times k)$ positive definite matrix. The multivariate skew normal distribution for y_t is defined as the conditional distribution of y_t upon $s_t > 0$, which I write as $(y_t|s_t > 0)$. The resulting multivariate skew normal density of $(y_t|s_t > 0)$ is

$$g_{sn}(y_t; \xi, \Omega, \alpha) = 2\phi_k(y_t; \xi, \Omega)\Phi\left(\alpha'\omega^{-1}(y_t - \xi)\right) \qquad (6)$$

where $\Omega = (\omega_{ij}) = \Sigma + CC'$ and $\omega = \text{diag}(\omega_{11}, ..., \omega_{kk})^{1/2}$. $\phi_k(.; \xi, \Omega)$ is a k-dimensional multivariate normal probability density function with mean ξ and covariance matrix Ω; $\Phi(.)$ denotes the standard normal cumulative distribution function.

α is a skewness parameter and $\alpha = \frac{1}{(1 - C'\Omega^{-1}C)^{1/2}}\Omega^{-1}C$. When $\alpha = 0_{k \times 1}$, the multivariate skew normal density in (6) reduces to the multivariate normal density $\phi_k(y_t; \xi, \Omega)$. To illustrate the role of the skewness parameter α, Figure 2 plots contours of bivariate skew normal distributions. The location parameter $\xi = (0, 0)'$ and the scale parameter Ω is a correlation matrix with off-diagonal element $\omega_{12} = 0.5$.

[Insert Figure 2 here.]

The varying skewness parameter α_1 (α_2) in each panel of Figure 2 affects both marginal skewness and dependence structure substantially. Positive (negative) α_1 (or α_2) accounts for positive (negative) skewness. The central panel corresponds to the contour plot of a bivariate normal distribution with $\alpha_1 = \alpha_2 = 0$. Figure 2 shows that dependence structures offered by the skew normal distribution are richer than those offered by the normal distribution.

By examining (5) and how α is calculated from C shows that C is an alternative

10

skewness parameter. When $C = 0_{k \times 1}$, y_t and s_t are independent and there is no skewness induced by conditioning $(y_t|s_t > 0)$. From this observation, there are at least three equivalent parameterizations of the multivariate skew normal distribution. The first parameterization (ξ, Σ, C) and the second parameterization (ξ, Ω, C) come directly from (5). The third parameterization (ξ, Ω, α) comes from the density formula (6). Both Azzalini & Dalla Valle (1996) and subsequent development in Azzalini & Capitanio (1999) focus on the third parameterization and another (the fourth) parameterization. They never discuss the first parameterization. When the third parameterization is used in describing linear transformation of the skew normal variables, the resulting expression of α is complex.

The first and the second parameterizations have the advantage of simplifying the result of linear transformation. This advantage is important in describing and analyzing portfolio selection problems. I focus on these two parameterizations and use them interchangeably. Proposition 3 is on linear transformation results using these two parameterizations.

Proposition 3.

Suppose A is a $k \times h$ matrix,

(a) under the first parameterization, if $y_t \sim SN_k(\xi, \Sigma, C)$, then $A'y_t \sim SN_h(A'\xi, A'\Sigma A, A'C)$;

(b) under the second parameterization, if $y_t \sim SN_k(\xi, \Omega, C)$, then $A'y_t \sim SN_h(A'\xi, A'\Omega A, A'C)$.

Proof. The proof is straightforward as multiplying y_t in (5) by A'.
Q.E.D.

With some algebraic work, it can be shown that α implied by Proposition 3 is the same as that obtained in Azzalini & Capitanio (1999).

From the conditioning method, the conditional distribution $(y_t|s_t) \sim N_k(\xi + Cs_t, \Sigma)$ and s_t is from a half-normal distribution $N(0, 1)I(s_t > 0)$. With this observation, (5) is readily written in a regression form $y_t = \xi + Cs_t + e_t$ which is the same as the three-

parameter model (1). The corresponding three parameters are ξ, Σ and C. s_t is non-elliptical with a half-normal distribution, it can be centered by deducting its mean value $\sqrt{2/\pi}$. $e_t | s_t \sim N_k(0, \Sigma)$ is a multivariate normal (joint-elliptical) random vector with the covariance matrix Σ. Noticing that the variance of a half-normal variate is $1 - 2/\pi$, the unconditional covariance of y_t can be decomposed to an elliptical variance component and a non-elliptical variance component as follows: $Var(y_t) = \Sigma + (1 - 2/\pi)CC'$. It is easy to check that this variance formula is equivalent to $Var(y_t) = \Omega - (2/\pi)CC'$ in Azzalini & Capitanio (1999). From the regression form (1), I obtain the following algorithm of simulating a k-dimensional skew normal random variable with parameterization (ξ, Σ, C).

Algorithm 1.

Step 1. Generate $s \sim N(0, 1)I(s > 0)$.

Step 2. Generate $y \sim N_k(\xi + Cs, \Sigma)$.

Since the multivariate skew normal distribution is a special case of the three-parameter model, all Simaan (1993)'s results are applicable to the multivariate skew normal distribution. I now examine whether other skew elliptical distributions generated by the same conditioning method are nested by the three-parameter model. To facilitate the examination, I draw a relationship between the three-parameter model and the weak exogeneity concept studied by Engle et al. (1983).

Consider the joint distribution of (y_t, s_t) as the product of a marginal distribution and a conditional distribution: $g(y_t, s_t) = g(y_t | s_t)g(s_t)$. The conditional distribution $g(y_t | s_t)$ embodies the regression equation in which y_t is the dependent variable and s_t is the regressor. If s_t is a weakly exogenous variable, the details of its marginal distribution can be safely ignored when making statistical inference about the parameters of the conditional distribution. In other words, the conditional information $y_t | s_t$ is sufficient to conduct statistical inference on these parameters.

The construction of the three-parameter model (1) guarantees that s_t is weakly

exogenous with respect to the parameters of the conditional distribution. As a typical example, the multivariate skew normal distribution's conditional Gaussian construction guarantees the weak exogeneity of s_t with respect to (ξ, Σ, C) (Engle et al. 1983, p. 287-288).

Branco & Dey (2001) (see also the discussion in Azzalini & Capitanio 2003) extends Azzalini and Dalla Valle (1996)'s multivariate skew normal distribution to a general class of skew elliptical distributions. They propose the following construction which is similar to (5): y_t and s_t have a joint-elliptical distribution $El_{k+1}(\mu, \Gamma; \phi)$ where μ is the location parameter, Γ is the scale matrix and ϕ is the characteristic function.

$$
\begin{pmatrix} y_t \\ s_t \end{pmatrix} \sim El_{k+1} \left[\mu = \begin{pmatrix} \xi \\ 0 \end{pmatrix}, \Gamma = \begin{pmatrix} \Sigma + CC' & C \\ C' & 1 \end{pmatrix}; \phi \right] \tag{7}
$$

Branco & Dey (2001) show that $(y_t|s_t > 0)$ has the density with the form similar to (6). However, in their construction s_t is not weakly exogenous to the parameters of $g(y_t|s_t)$ when the elliptical distribution is not multivariate normal. Specifically, $y_t|s_t$ indeed follows a joint-elliptical distribution but both its characteristic function and its scale matrix depend on the realization of s_t. Spanos (1994) shows that s_t is weakly exogenous with respect to the statistical parameterization of the conditional distribution $g(y_t|s_t)$ if and only if (y_t, s_t) follows a multivariate normal distribution. From Spanos (1994)'s result, the skew elliptical distributions generated from conditioning on the construction of (7) cannot be written as a three-parameter model unless the resulting skew distribution is multivariate skew normal.

There are other approaches to construct skew distributions. Sahu et al. (2003)'s skew elliptical distributions, for example, are not nested by the three-parameter model because the conditioning is on multiple non-elliptical variables rather than on one variable. In the finance literature, Ronn et al. (2009) use a distribution generated by conditioning on a two-sided truncation of a bivariate normal distribution. Ronn et al. (2009) study the correlation between the truncated variable and the non-truncated variable conditional on the truncation $|s_t| > a$. In their setting both y_t and s_t are

observable. Their construction can be easily extended into a multivariate one with latent scalar s_t. It is straightforward to show $(y_t||s_t| > a)$ can be written as a case of the three-parameter model. The resulting multivariate distribution is less studied and there is no obvious intuition on how to determine the truncation point for latent variable s_t. It remains to study whether there exist other well-known skew distributions that can be nested by the three-parameter model.

4   Certainty Equivalent: Some Theoretical Results

It is well known that a closed-form mean-variance certainty equivalent (objective function) can be obtained by combining the CARA utility and the multivariate normal distribution. Suppose the investor holding random wealth W_t at time t has the CARA utility function as follows:

$$U(W_t) = -exp(-\gamma W_t);\tag{8}$$

The Arrow-Pratt coefficient of absolute risk aversion is denoted by γ ($\gamma > 0$).

The certainty equivalent CE is defined as the certain return that makes the investor indifferent from obtaining the certain return and taking a risky investment. This definition is formally given by the equation: $CE(W_t) \equiv U^{-1}E(U(W_t))$. The indifference condition is: $U(CE(W_t)) = E(U(W_t))$. The certainty equivalent concept is particularly useful when an investor compares two financial choices A and B. If $CE_A > CE_B$, then the investor will prefer choice A to B and the economic gain of choosing A can be measured by $CE_A - CE_B$. The certainty equivalent is also useful in numerical computing because an investor who maximizes the certainty equivalent also maximizes the corresponding expected utility. If a closed-form certainty equivalent exists, numerical integration of computing expected utility can be avoided. For the CARA utility function, the certainty equivalent is closely related to the cumulant-generating function of (random) wealth. In particular, the certainty equivalent has the form

$$CE(W_t) = -\frac{1}{\gamma}ln(E(exp(-\gamma W_t))).\tag{9}$$

Thus the certainty equivalent is $-\frac{1}{\gamma}$ multiplied by the cumulant-generating function of W_t evaluated at $-\gamma$.

Assuming that the risk-free asset exists, at time $t-1$, the investor faces the following portfolio selection problem:

$$\rho^* = \arg\max_{\rho} E(-exp(-\gamma\rho'(y_t - R_{f,t}\tau))). \tag{10}$$

where y_t is a k-vector of asset returns, $R_{f,t}$ is the risk-free rate and τ is a vector of ones. When excess asset returns $(y_t - R_{f,t}\tau) \sim N_k(\mu, V)$ with $\mu = E(y_t) - R_f\tau$ and $V = Var(y_t)$, the excess portfolio return $\rho'(y_t - R_{f,t}\tau)$ follows a univariate normal distribution $N(\rho'\mu, \rho'V\rho)$. The cumulant-generating function of multivariate normal distribution evaluated at $-\gamma$ is $-\gamma\rho'\mu + \frac{\gamma^2}{2}\rho'V\rho$. The mean-variance certainty equivalent excess return is as follows:

$$\rho'\mu - \frac{\gamma}{2}\rho'V\rho. \tag{11}$$

The optimization problem (10) is equivalent to the following problem:

$$\rho^* = \arg\max_{\rho} \rho'\mu - \frac{\gamma}{2}\rho'V\rho. \tag{12}$$

It is well known that the solution to (12) is given by

$$\rho = \frac{1}{\gamma}V^{-1}\mu. \tag{13}$$

Both Pástor & Stambaugh (2000) and Tu & Zhou (2004) use the certainty equivalent (11) to investigate the economic value difference between portfolios. Kan and Zhou (2007) use it to examine the impacts of parameter uncertainty.

Now I will show that when asset returns are multivariate skew normal, the certainty equivalent can be obtained in its closed-form. To simplify notations, I use the second parameterization (ξ, Ω, C) and I assume that s_t is not centered. By Proposition 3, when excess asset returns $(y_t - R_{f,t}\tau) \sim SN_k(\xi, \Omega, C)$, the excess portfolio return follows a univariate skew normal distribution $SN(\rho'\xi, \rho'\Omega\rho, \rho'C)$. The cumulant-generating

function of a skew normal variate (Azzalini & Capitanio 1999) evaluated at point $-\gamma$ is $-\gamma\rho'\xi + \frac{\gamma^2}{2}\rho'\Omega\rho + log(2\Phi(-\gamma\rho'C))$. Therefore, the certainty equivalent excess return has the following closed form:

$$\rho'\xi - \frac{\gamma}{2}\rho'\Omega\rho - log(2\Phi(-\gamma\rho'C))/\gamma. \tag{14}$$

The certainty equivalent in (14) has three terms and each term involves only one parameter in the parameterization $(\rho'\xi, \rho'\Omega\rho, \rho'C)$. The investor facing problem (10) solves the following equivalent portfolio selection problem:

$$\rho^* = \arg\max_{\rho} \rho'\xi - \frac{\gamma}{2}\rho'\Omega\rho - log(2\Phi(-\gamma\rho'C))/\gamma. \tag{15}$$

The multivariate skew normal portfolio selection in (15) reduces to the mean-variance portfolio selection when skewness parameter C is a vector of zeros. When skewness exists, the mean-variance portfolio choice is not generally optimal for investors. Since $\Omega = \Sigma + CC'$, it is easy to check that the investor prefers higher $\rho'\Sigma\rho$ when $\rho'\xi$ and $\rho'C$ are fixed. This confirms that the optimal portfolio obtained in (15) satisfies the quadratic programming problem (3). I will use the closed-form certainty equivalent (14) to investigate the economic value of higher moments in Section 5. Now I explore some related theoretical results.

The first-order conditions of (15) are given by:

$$\xi - \gamma\Omega\rho + \frac{\phi(-\gamma\rho'C)}{\Phi(-\gamma\rho'C)}C = 0. \tag{16}$$

The second-order condition for maximizing the expected utility is satisfied because the CARA utility function is strictly concave. This guarantees that the optimal portfolio ρ^* solved from first-order conditions (16) is unique and is indeed the global maximum point. Although there is no closed-form solution from first-order conditions (16), numerical maximizer can be used to solve the optimal portfolio from (15). The programming problem (15) can be modified to handle nonnegative constraints or other

linear constraints and the Kuhn-Tucker conditions are both necessary and sufficient for this type of programming problems.

There are three implications from (15) and (16). First, suppose the maximum certainty equivalent obtained from (15) is CE^*, the envelop theorem gives $\frac{\partial CE^*}{\partial C_i} = \frac{\phi(-\gamma\rho^{*\prime}C)}{\Phi(-\gamma\rho^{*\prime}C)}\rho_i^*$. When $\rho'C \neq 0$, $\frac{\phi(-\gamma\rho^{*\prime}C)}{\Phi(-\gamma\rho^{*\prime}C)} > 0$. If in the optimal portfolio the investor takes a long (short) position in asset i, increasing asset i's skewness results in higher (lower) expected utility. The CARA investor has positive skewness preference.

Second, Cass and Stiglitz (1970)'s general result on two-fund separation holds. Cass and Stiglitz (1970) show that, with homogeneous probability beliefs and the existence of the risk-free asset, the composition of each CARA investor's optimal risky asset portfolio is the same. The two-fund separation holds regardless of asset returns' distribution. For markets to clear with two-fund separation, the optimal proportions of risky assets for each investor must be those of the market portfolio. In (16) suppose ρ^* is the optimal portfolio for the investor with risk aversion coefficient 1, the first order conditions (16) imply that $\frac{1}{\gamma}\rho^*$ is the optimal portfolio for the investor with risk aversion γ. Investor with higher (lower) risk aversion will invest less (more) in the risky fund.

Third, the closed-form risk premium in asset price can be obtained. Particularly, the skewness premium can be expressed in its exact form. Following Cochrane (2006), I assume the market (aggregate) risk aversion coefficient is γ_M. The investor who has risk aversion γ_M will put all her money (without borrowing or lending) into the market portfolio. Multiplying the portfolio weights on both sides of (16) and using the equations $E(y_t) = \xi + \sqrt{2/\pi}C + R_{f,t}$ and $Var(y_t) = \Omega - (2/\pi)CC'$ for the multivariate skew normal distribution, I obtain the expected excess return of individual asset i as follows:

$$E(y_{i,t}) - R_{f,t} = \gamma_M Cov(y_{i,t}, y_{M,t}) + g_M C_i. \tag{17}$$

where $g_M = (2/\pi)\gamma_M C_M + \sqrt{2/\pi} - \frac{\phi(-\gamma_M C_M)}{\Phi(-\gamma_M C_M)}$. Since g_M only depends on market risk aversion γ_M and the market portfolio's skewness parameter C_M, g_M can be interpreted as the systematic skewness return. γ_M and C_M play symmetric roles in the systematic skewness return. C_i, the skewness parameter of asset i, can be interpreted as the

17

sensitivity to the systematic skewness return g_M.

Substituting C_i by C_M in (17), the market price of risk can be written as:

$$E(y_{M,t}) - R_{f,t} = \gamma_M Var(y_M) + g_M C_M. \tag{18}$$

Comparing to the mean-variance results (Cochrane 2006, p. 154), the term $g_M C_i$ ($g_M C_M$) is the skewness premium in asset return (market price of risk). Simaan (1993) explains that the sign of the skewness premium is ambiguous because the skewness parameter C in the three-parameter model also captures higher moments' information. This ambiguity can not be eliminated by convergence of Taylor expansion on the utility function and skewness preference ($U''' > 0$). In the CARA investors' world with skew normal asset returns, the skewness premium is characterized in its exact form and the sign of the skewness premium is not ambiguous. Figure 3 illustrates the skewness premium in the market price of risk as a function of the market portfolio's skewness parameter C_M. The market risk aversion $\gamma_M = \{0.1, 0.5, 1\}$.

[**Insert Figure 3 here.**]

Two observations can be made from Figure 3. First, positive (negative) skewness in market portfolio returns induces negative (positive) skewness premium in the market price of risk. Investors sacrifice required market return to chase positive market skewness. On the other hand, they require return compensation for accepting negative market skewness. This result is consistent to the previous results which are based on approximations (Rubinstein 1973, Kraus & Litzenberger 1976, Harvey & Siddique 2000 among others). Second, the magnitude of the skewness premium also depends on the market risk aversion. When the market risk aversion is low (high), the absolute value of skewness premium is also low (high). Similar observations can be found when the asset i's skewness premium is expressed as a function of C_i. Proposition 5 formally presents these observations.

Proposition 4.

(a) $g_M \leq 0$. Strict inequality holds when $C_M \neq 0$.

(b) The skewness premium in asset i's price (market price of risk) is decreasing in the skewness parameter C_i (C_M).

(c) The magnitude of skewness premium in asset price is increasing in the market risk aversion γ_M.

Proof. See the Appendix.

An empirical implication of Proposition 4 is that testing (conditional) skewness preference is better implemented during high return skewness and (or) high market risk aversion periods. To conclude this section, I present a first-order stochastic dominance result as follows.

Proposition 5. Suppose $y_1 \sim SN(\xi, \Omega, C_1)$ and $y_2 \sim SN(\xi, \Omega, C_2)$ and $C_1 > C_2$, then y_1 dominates y_2 by first-order stochastic dominance.

Proof. See the Appendix.

The stochastic dominance results for normal, truncated normal and log-normal distributions are well studied (Levy 2006). Proposition 5 states a new result for the skew normal distribution under the second parameterization. All investors with increasing utility on wealth, *ceteris paribus*, prefer a skew normal distribution with larger skewness parameter. A direct consequence of Proposition 5 is that the certainty equivalent in (15) is larger (smaller) than the mean-variance (form of) certainty equivalent $\rho'\xi - \frac{\gamma}{2}\rho'\Omega\rho$ when $\rho'C > 0$ ($\rho'C < 0$).

5 Economic Value of Higher Moments: A Certainty Equivalent Approach

In this section I evaluate the economic value of higher moments using empirical data. I borrow the certainty equivalent framework used by Pástor & Stambaugh (2000) and Tu & Zhou (2004) and replace the mean-variance certainty equivalent by the certainty equivalent I have obtained in (14) to incorporate higher moments information.

Both Pástor & Stambaugh (2000) and Tu & Zhou (2004) analyze uncertainty of model mispricing. When asset returns y_t and model variables are jointly multivariate normal or multivariate t, $E(y_t|x_t)$ is a linear function of x_t. However, when y_t and x_t are jointly multivariate skew normal, model misspecification and model mispricing are entangled because $E(y_t|x_t)$ is nonlinear function of x_t (Azzalini and Capitanio 1999). The focal point of the current paper is not model mispricing. Instead, I focus on evaluating the economic value of higher moments when investors face broad investment opportunities.

5.1 Data Description

I consider two datasets. The first dataset contains 12 risky positions[2]. Three of the risky positions are the Fama and French (1993) benchmark positions, SMB, HML, and MKT. The other nine non-benchmark positions are constructed in Pástor and Stambaugh (2000). They are formed as spreads between portfolios selected from a larger universe of equity portfolios created by a three-way sorting on size, book-to-market, and HML beta. This three-way sorting creates 27 value-weighted portfolios that are identified by a combination of three letters which designate increasing values of size (S, M, B), book-to-market (L, M, H), and HML beta (l, m, h). Holding size and book-to-market constant, nine spread positions are long stocks with low HML betas and short stocks with high HML betas. The data are available as monthly returns from July 1963

---

[2]I am grateful to Ľuboš Pástor for providing me the dataset.

through December 1997. The second dataset is the 30 industry portfolios[3] constructed by sorting each NYSE, AMEX, and NASDAQ stock to an industry portfolio at the end of June of each year based on its four-digit SIC code at that time. Industry portfolios also well proxy the investment opportunity set. The original data are available back to year 1928. In my study I use daily returns from January 1, 1980 through December 31, 2012. I also construct weekly and monthly returns from daily data. I use one-month Treasury bill return to proxy the risk-free rate of return.

5.2 Margin Requirements

Margin requirements can reduce investors' ability of taking extreme positions in assets and dramatically reduce or eliminate the benefit of higher moments. Following Pástor and Stambaugh (2000) and Tu and Zhou (2004), I consider spread position i; constructed at the end of period $t-1$; as a purchase of one asset coupled with an offsetting short sale of another. The two assets are denoted as L_i and S_i; and their rates of return in period t are denoted as $R_{L_i;t}$ and $R_{S_i;t}$. Thus, a spread position of size X_i has a dollar payoff $X_i(R_{t_i;t} - R_{S_i;t})$. Since regulation T requires the use of margins for risky investments, a constant $c > 0$ is used to characterize the degree of margin requirements. The spread position involves at least one risky asset which, without loss of generality, is designated as asset L_i. If the other asset of position i, S_i of size X_i, is risky as well, then $(2/c)|X_i|$ dollars of capital are required. Otherwise, $(1/c)|X_i|$ dollars of capital are required. For example, $c = 2$ implies a 50% margin imposed by Regulation T. In addition to a 50% margin requirement, I also consider margins of only 20% ($c = 5$), 10% ($c = 10$) and the case of no margin ($c = \infty$).

The total capital required to establish the spread positions must be less than or equal to the investors wealth, W_{t-1}. That is

$$\sum_{i \in \Lambda}(2/c)|X_i| + \sum_{i \notin \Lambda}(1/c)|X_i| \leq W_{t-1}$$

---

[3]I am grateful to Ken French for making the data available. The data are downloaded from Ken French's homepage at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

where Λ denotes the set of positions in which S_i is risky, or alternatively,

$$\sum_{i \in \Lambda} 2|\rho_i| + \sum_{i \notin \Lambda} |\rho_i| \leq c \tag{19}$$

where $\rho_i = X_i/W_{t-1}$. For 12 risky positions, $R_{S_i,t} = R_{f,t}$ only for MKT and Λ has 11 elements. For 30 industry portfolios, $R_{S_i,t} = R_{f,t}$ for all i and Λ is a null set.

The total wealth in excess of the margin capital required to establish k spread positions is invested in the risk-free asset, earning the risk-free rate $R_{f,t}$ and the margin capital also earns this rate. The rate of return on the total portfolio is then given by

$$R_{p,t} = \frac{\sum_{i=1}^{k} X_i(R_{L_i,t} - R_{S_i,t}) + W_{t-1} R_{f,t}}{W_{t-1}}$$

so the excess portfolio return is simply weighted average of spread positions' returns.

$$R_{p,t} - R_{f,t} = \sum_{i=1}^{k} \rho_i(R_{L_i,t} - R_{S_i,t}). \tag{20}$$

Denote the spread returns $y_t = R_{L,t} - R_{S,t}$ and suppose $y_t \sim SN_k(\xi, \Omega, C)$, the certainty equivalent obtained in Section 4 is applicable to describing the portfolio selection problem. Combining the problem (15) and margin requirements (19), the optimal portfolio choice of the skew-normal-optimizing investor is the solution to:

$$\rho_{sn} = \arg\max_{\rho} \rho'\xi - \frac{\gamma}{2}\rho'\Omega\rho - log(2\Phi(-\gamma\rho'C))/\gamma$$

$$st. \sum_{i \in \Lambda} 2|\rho_i| + \sum_{i \notin \Lambda} |\rho_i| \leq c \tag{21}$$

Alternatively combining the optimization problem (12) and margin requirements (19), the optimal portfolio choice of a mean-variance-optimizing investor is the solution to:

$$\rho_{mv} = \arg\max_{\rho} \rho'\mu - \frac{\gamma}{2}\rho'V\rho$$

$$st. \sum_{i \in \Lambda} 2|\rho_i| + \sum_{i \notin \Lambda} |\rho_i| \leq c \qquad (22)$$

When $c = \infty$, (21) reduces to (15) and (22) reduces to (12).

5.3 Model Estimation and Portfolio Performance Evaluation

Following Pástor & Stambaugh (2000), Tu & Zhou (2004), and Kan & Zhou (2007), I assume independently and identically distributed (i.i.d.) asset returns. While the i.i.d. assumption rules out time-varying parameters and predictability, I maintain maintain this assumption for the similar reasons stated by Tu and Zhou (2004). I use the maximum likelihood estimation to estimate candidate models. fter obtaining the maximum likelihood estimates $\hat{\xi}$, $\hat{\Omega}$, $\hat{\alpha}$, the estimate of skewness parameter is computed as $\hat{C} = \frac{1}{(1 + \hat{\alpha}' \hat{\Omega} \hat{\alpha})^{1/2}} \hat{\Omega} \hat{\alpha}$ (the invariance property of the maximum likelihood estimator).

Then I plug the maximum likelihood estimates into (21) and (22) to compute optimal asset allocations ρ_{sn} and ρ_{mv} for a given margin requirement c and risk aversion level γ. After obtaining ρ_{sn} and ρ_{mv}, I can compare the certainty equivalents of investors when the actual data-generating process is multivariate skew normal. I compute the certainty equivalent for the skew-normal-optimizing investor as follows:

$$CE_{sn} = \rho'_{sn} \hat{\xi} - \frac{\gamma}{2} \rho'_{sn} \hat{\Omega} \rho_{sn} - log(2\Phi(-\gamma \rho'_{sn} \hat{C}))/\gamma \qquad (23)$$

Then I compute the certainty equivalent for the mean-variance-optimizing investor as follows:

$$CE_{mv} = \rho'_{mv} \hat{\xi} - \frac{\gamma}{2} \rho'_{mv} \hat{\Omega} \rho_{mv} - log(2\Phi(-\gamma \rho'_{mv} \hat{C}))/\gamma \qquad (24)$$

The economic gain of the skew-normal-optimizing investor is computed by:

$$CE_{sn} - CE_{mv}. \qquad (25)$$

If the gain computed by (25) is significant, then the skew-normal-optimizing investor has substantial increase in her expected utility and considering higher moments does

provide economic value. On the contrary, if the gain computed by (25) is negligible, then considering higher moments does not provide substantial economic value to investors.

5.4 Choice of Risk Aversion Parameter

The choice of risk aversion parameter is crucial for conducting economic value comparison between portfolios. In the limit of infinite risk aversion the investor would put all her wealth in the risk-free asset. In this limit case, the economic value of higher moments is zero.

Both Pástor & Stambaugh (2000) and Tu & Zhou (2004) interpret γ in (11) as the coefficient of relative risk aversion (RRA). Both studies set $\gamma = 2.83$, which is the value that results in an unconstrained allocation of all wealth to MKT when that is the only risky position available, i.e., the investor chooses neither to borrow. Specifically, utilizing (13) the number 2.83 is obtained by $\gamma = V_{MKT}^{-1}\mu_{MKT}$. If a skew-normal-optimizing investor put all her wealth to MKT when that is the only risky position available, then the corresponding risk aversion can be obtained by solving (16). The resulting risk aversion is not far from 2.83.

Although 2.83 might be a realistic number in the finance literature, the economics literature however seems to agree on much lower risk aversion. Holt and Laury (2002) show that the RRA is centered around range of $[0.3, 0.5]$. They also list many field studies which obtain similar estimates. Moreover, from the derivation of certainty equivalent in Section 4, γ is actually the coefficient of absolute risk aversion (ARA). The CRRA (constant relative risk aversion) utilities typically have no closed-form certainty equivalents. Maximizing those expected utilities has to be conducted through numerical integration.

As $RRA = -\frac{U''(W)W}{U'(W)} = ARA * W$, the magnitude of RRA should be much larger than ARA evaluated at a realistic wealth level. Patton (2004) use numerical integration to maximize the CRRA utilities. He use the range $[1, 20]$ for RRA. Even using $RRA = 20$ (which is about 40 times of 0.3-0.5 normal range in the economics literature), the implied ARA should still be far smaller than 2.83.

Although estimated values of *ARA* differ widely in the economics literature, none of the previous literature show γ higher than 0.538 (See a detailed literature review in Babcock et al. 1993). For a realistic level of $10000 loss, Babcock et al. (1993) estimate the range of absolute risk aversion as $[0.000002, 0.000462]$. In a more recent study, Rabin (2000, Table 3) use the absolute risk aversion in the range $[0.000009, 0.003]$ to obtain stock market investment range $[\$449, \$163,899]$. For about $30,000 stock market investment, the corresponding absolute risk aversion is around 0.00005.

The debate of realistic level of risk aversion can be traced back to the equity premium puzzle (Mehra and Prescott 1985). Many subsequent studies are devoted to solve the puzzle and justify the high level of risk aversion that is needed to explain the premium. Loosely linking the equity premium puzzle to the skewness premium plotted in Figure 3, it is straightforward to see that the skewness premium disappears when ARA parameter γ is small.

I am not in a position to judge whether 2.83 or 0.00005 is more suitable for computing the economic value of higher moments. Instead, I use both values as representatives of two academic camps (finance and economics) to ensure the robustness of my results.

5.5 Statistically Significant but Economically Negligible: The Role of Higher Moments

When conducting empirical studies on higher moments portfolio selection, there are two related questions to answer. First, whether the return series statistically depart from multivariate normal. Second, whether the departure (if exists) generates economically significant gain to investors who are aware of it.

Tu & Zhou (2004) apply Mardia (1970)'s multivariate skewness and kurtosis tests to the 12 risky positions and industry return data. They find significant departure from normality. My industry return data are relatively new comparing to those studied by Tu & Zhou (2004). I check whether industry return series show departure from joint normality using Mardia's tests. The test results are reported in Table 2. Both multivariate skewness and kurtosis tests show significant departure from multivariate

normal. All p-values in the table are 0.00 (less than 0.00001). The deviations from normality increase substantially when return frequency increases. These results are consistent with those reported in Tu & Zhou (2004).

Table 3 reports the maximum likelihood estimates $\hat{C}$ for monthly returns on 12 risky positions. The percentile method (Cameron and Trivedi 2005, p. 364) is used to calculate the 90% confidence intervals of estimates from 1000 bootstrapping replications. As discussed in Section 2, elements in the skewness parameter can be treated as factor loadings on the skew factor s_t. Table 3 shows that SMB has the highest loading of 1.51 and SH(l-h) has lowest loading of -1.11. The 90% confidence intervals of these two loadings do not include zero.

Table 4 reports the maximum likelihood estimates $\hat{C}$ for weekly returns on 30 industry portfolios along with bootstrapping confidence intervals. Textiles industry (Txtls) has the highest skew factor loading of 6.03 and Personal and Business Services industry (Servs) has the lowest skew factor loading of -0.22. Telecommunication has 0.00 loading which suggests no exposure to the skew factor s_t. Fifteen out of thirty confidence intervals do not include zero. Overall, the estimated results also suggest both datasets' departure from joint normality.

After establishing the statistical importance of higher moments in data, I obtain optimal portfolio weights of both the skew-normal-optimizing investor and the mean-variance-optimizing investor and then examine the economic gain of the skew-normal-optimizing investor. Table 5 reports the unconstrained asset allocations when investment opportunities include 12 risky positions and cash. As shown in Section 4, all the CARA investors will hold the same risky portfolio when there is no restriction. The only difference among them is the asset allocation between the risky portfolio and the risk-free asset. The unconstrained allocations are normalized in Table 5 so that the sum of the risky asset weights is 1 (100%). Table 5 shows that optimal portfolios constructed by the skew-normal-optimizing investor and the mean-variance-optimizing investor are similar. I examine the economic value of incorporating higher moments in portfolio selection using the approach described in Section 5.3. When $\gamma = 2.83$, the annualized

economic gain of considering higher moments is 0.0023% which is negligible. When $\gamma = 0.00005$, the annualized economic gain of considering higher moments is 249.00%. The high economic gain however mainly comes from financial leverage of borrowing $1.22 \times 10^7\%$ (holding $-1.22 \times 10^7\%$ in Cash). I conduct further examination which shows that to obtain annualized economic gain of 1%, the skew-normal-optimizing investor should have $\gamma = 0.006$ and borrow $10^5\%$. In real trading, such a high financial leverage is hard to achieve.

The I examine the results when margin requirements exist. I consider six cases that combine three different levels of margin requirements and two different levels of investor risk aversion. The parameters are: $c = \{2, 5, 10\}$ and $\gamma = \{0.00005, 2.83\}$.

Table 6 reports the optimal allocations when $c = 10$ and $\gamma = 0.0005$. Table 6 shows that asset allocations are different between the skew-normal-optimizing investor and the mean-variance-optimizing investor. For example, the mean-variance-optimizing investor longs 1.1710 (171.10%) in SMB and shorts 0.2431 (24.31%) in HML. The skew-normal-optimizing investor instead longs 1.7485 (174.85%) in HML and 0.0941 (9.41%) in SMB. In this case, the annualized economic gain of skew-normal-optimizing investor is 0.69%. This economic gain is the highest gain among all the six cases.

An interesting finding across all levels of margin requirements is that investors put significant and similar long positions (e.g. 3.4838 and 3.4963 respectively in Table 6 when $c = 10$) in MKT (the value-weighted market index portfolio). MKT dominates other risky positions in asset allocations. It might be reasonable to argue that a value-weighted market index proxies aggregate investor choice (including mean-variance-optimizing investors, skew-normal-optimizing investors and many other investors) so portfolio selection on a risky asset pool including MKT will concentrate on MKT and will not generate any significant economic value for considering higher moments. However, on the other hand, Chung et al. (2006) show that SMB and HML well proxy the higher moments in asset returns.

Portfolio selection results on 30 industry portfolios will help mitigate the above concerns because the investment universe does not include the market index portfolio.

27

Table 7 reports the unconstrained asset allocations. The asset allocations are different between the skew-normal-optimizing investor and the mean-variance-optimizing investor. When $\gamma = 2.83$, the annualized economic gain of considering higher moments is negligible as 0.24%. When $\gamma = 0.00005$, the annualized economic gain of considering higher moments is $6.67 \times 10^3\%$. Again, the extreme gain comes from leverage of borrowing $9.11 \times 10^6\%$. To obtain annualized economic gain of 1%, the skew-normal-investor should have $\gamma = 0.63$ and borrow 6120%. This leverage level is still very high in real trading.

Table 8 reports the asset allocations of investors with $\gamma = 0.00005$ and margin requirement $c = 10$. Table 8 shows that the skew-normal-optimizing investor borrows 893.51%. The annualized economic gain of the skew-normal-optimizing investor is 0.95%. Again, the 95 basis points annual gain is obtained by allowing relatively high financial leverage. This economic gain is again the highest gain among all the six cases.

When daily and monthly returns on 30 industry portfolios are considered, the economic gain of considering higher moments is still negligible. I also conduct out-of-sample performance comparison which is not reported here. I find that the out-of-sample economic gain is also negligible when margin requirements are considered. The benefits of higher moments disappear when margin requirements exist because investors cannot take extreme long/short positions.

My results however do not rule out the possibility of exploring higher moments benefits if investors can use high financial leverage and (or) if investors face a small and special investment universe. Real estate investors can establish substantial leverage. Momentum or special condition investors only focus on a small group of stocks. These investors may find higher moments portfolio selection useful and valuable.

5.6 Parameter Uncertainty and Bayesian Estimation: A Discussion

The standard plug-in approach described in Section 5.3 does not take into account of parameter uncertainty (Kan & Zhou 2007). To incorporate parameter uncertainty,

Bayesian methods can be used. In the Bayesian framework, parameters are treated as random variables. The parameter uncertainty is captured by their posterior distributions in light of data and prior beliefs. I assume that y_t is drawn independently from a multivariate skew normal distribution with unknown parameters ξ, Σ and C. An investor has prior beliefs $p(\xi, \Sigma, C)$. The investor forms posterior beliefs $p(\xi, \Sigma, C|Y)$ based on the data $Y = (y_1, ..., y_T)'$. The posterior beliefs on parameters is proportional to the product of likelihood and the prior beliefs:

$$p(\xi, \Sigma, C|Y) \propto p(Y|\xi, \Sigma, C)p(\xi, \Sigma, C). \tag{26}$$

Then the investor forms the predictive distribution for y_{T+1},

$$p(y_{T+1}|Y) = \int_\xi \int_\Sigma \int_C p(y_{T+1}|\xi, \Sigma, C)p(\xi, \Sigma, C|Y)d\xi d\Sigma dC$$

Similarly, the predictive distribution for utility U_{T+1} is,

$$p(U_{T+1}|Y) = \int_\xi \int_\Sigma \int_C p(U_{T+1}|\xi, \Sigma, C)p(\xi, \Sigma, C|Y)d\xi d\Sigma dC \tag{27}$$

By the law of iterated expectations,

$$E(U_{T+1}|Y) = E(E(U_{T+1}|Y, \xi, \Sigma, C)|Y). \tag{28}$$

This says the predictive expected utility is identical to the posterior expected utility. For the CARA utility, when the Bayesian sampling scheme converges to produce iterates $(\xi^{[k]}, \Sigma^{[k]}, C^{[k]}) \sim p(\xi, \Sigma, C|Y)$ for $k = 1, ..., K$, the predictive expected utility can be approximated by:

$$E(U_{T+1}|Y) \approx \frac{1}{K} \sum_{k=1}^{K} -exp(-\gamma\xi^{[k]} + \gamma^2\Omega^{[k]}/2 + log(2\Phi(-\gamma C^{[k]}))). \tag{29}$$

where $\Omega^{[k]} = \Omega^{[k]} + C^{[k]}C'^{[k]}$. The portfolio selection problem that maximizing $E(U_{T+1}|Y)$ is tedious to solve because (16) has no closed-form solution. Kan and Zhou (2007)'s analysis based on solution (13) cannot be easily extended here. And unfortunately the closed-form certainty equivalent does not help much in simplifying the computation. Plugging the posterior means of ξ, Ω, C into (14) will actually bias the result because of (conditional) Jensen's inequality. After obtaining the asset allocations which maximize $E(U_{T+1}|Y)$, the predicative certainty equivalent can be calculated using (9) and then (23) - (25) can still be used to compute the economic gain. The procedure is tedious and the economic gain is still negligible so I omit reporting the results here.

Bayesian estimation methods for the multivariate normal and multivariate t models are described by Tiao & Zellner (1964) and Zellner (1976). Pástor & Stambaugh (2000) and Tu & Zhou (2004) also provide comprehensive discussions on these two models respectively. Bayesian estimation method for the multivariate skew normal model is an extension of above methods. In the Appendix, I provide details of Bayesian estimation algorithm for the multivariate skew normal model.

6 Conclusion

In this paper, I examine higher moments portfolio selection by establishing a relationship between Simaan (1993)'s three-parameter model and Azzalini & Dalla Valle (1996)'s skew normal distribution. I show that the skew normal distribution is a special case of the three-parameter model and all Simaan (1993)'s results are applicable to the skew normal asset returns. The closed-form certainty equivalent and skewness premium can be obtained when the CARA investor faces skew normal asset returns. The certainty equivalent allows me to examine the economic value of incorporating higher moments in portfolio decision. Although I find that asset returns statistically depart from joint normality, I do not find significant economic value of considering higher moments in portfolio selection. Under reasonable margin requirements, the economic value of higher moments are negligible. One implication of my empirical results is that the value of

higher moments might be better explored by investors with relaxed financial constraints or by those facing a small and special investment universe.

Appendix

Proof of Proposition 4.

(a) The proof is straightforward when $C_M = 0$.

When $C_M \neq 0$, let $x = \gamma_M C_M$ and $f(x) = \phi(-x)/\Phi(-x)$. Then $f(x) = \frac{\phi(x)}{1-\Phi(x)} = 1/M(x)$ where $M(x)$ is the famous Mill's ratio.

Now using the Largrange remainder theorem for Taylor expansion:

$$f(x) = f(0) + f'(0)(x - 0) + f''(x^*)(x - 0)^2$$

where $0 < x^* < x$. From Sampford (1953), the second derivative $f''(x) > 0$ for all finite x. Substituting $f(0) = \sqrt{2/\pi}$ and $f'(0) = 2/\pi$ into the Taylor expansion above, $f(x) = \sqrt{2/\pi} + (2/\pi)x + Rem$ where the remainder $Rem > 0$. I obtain $g_M = \sqrt{2/\pi} + (2/\pi)x - f(x) < 0$.

(b) is straightforward to obtain when (a) is true.

(c) Since $g_M(x) = \sqrt{2/\pi} + (2/\pi)x - f(x) < 0$ (with $x = \gamma_M C_M$ and $C_M \neq 0$), it is sufficient to prove that $\frac{\partial g_M}{\partial \gamma_M} < 0$ when $C_M > 0$.

$\frac{\partial g_M}{\partial \gamma_M} = \frac{\frac{\partial g_M}{\partial x}}{\frac{\partial x}{\partial \gamma_M}} = [(2/\pi) - f'(x)]/C_M = [f'(0) - f'(x)]/C_M$

Again using the Largrange remainder theorem for Taylor expansion:

$$f'(x) = f'(0) + f''(x^*)(x - 0)$$

where $0 < x^* < x$ and $f''(x^*) > 0$.

Thus when $C_M > 0$, $[f'(0) - f'(x)]/C_M < 0$.

Q.E.D.[4]

---

[4]The proof of Proposition 5 gives a new bound for Mill's ratio ($M(x) = \frac{1-\Phi(x)}{\phi(x)}$): for all x, $\frac{1}{M(x)} > \sqrt{2/\pi} + (2/\pi)x$. The well-known bounds in the statistical literature are: $(\sqrt{4+x^2} - x)/2 < M(x) < 1/x, \quad x > 0$. The new bound improves the well-known bounds in some regions of x.

32

Proof of Proposition 5.

From Azzalini (1985), the CDF of $y_i = y$ is $CDF_i(y) = \Phi((y - \xi)/\sqrt{\Omega}) - 2T((y - \xi)/\sqrt{\Omega}, C_i/\sqrt{\Omega - C_i^2})$, $i = 1, 2$. $T(h, a)$ is the Owen's T function and it is strictly increasing in a. Since $C_1 > C_2$ implies $CDF_1(y) < CDF_2(y)$ for all y, y_1 dominates y_2 by first-order stochastic dominance.

Q.E.D.

Bayesian estimation algorithm for the multivariate skew normal posterior evaluation

Define $Y = (y_1, ..., y_T)'$, a $T \times k$ matrix, $S = (s_1, ..., s_T)'$, a $T \times 1$ vector, $X = (\tau_T \ S)$, where τ_T denotes a $T \times 1$ vector of ones. Also define $\beta = (\xi, C)'$, a $2 \times k$ matrix and $b = vec(\beta)$. The regression model can be written as

$$Y = X\beta + \epsilon, \quad vec(\epsilon) \sim N(0, \Sigma \otimes I_T), \tag{30}$$

where $\epsilon = (e_1, ..., e_T)'$.

Define the statistics $\hat{\beta} = (X'X)^{-1}X'Y$, $\hat{b} = vec(\hat{\beta})$ and $\hat{\Sigma} = (Y - X\beta)'(Y - X\beta)$. The likelihood function of Y and S can be factored as

$$p(Y, S) = p(Y|\beta, \Sigma, S)p(S) \tag{31}$$

where

$$\begin{aligned}
p(Y|\beta, \Sigma, S) &\propto |\Sigma|^{-T/2}\exp\left\{-\tfrac{1}{2}tr((Y - X\beta)'(Y - X\beta))\Sigma^{-1}\right\} \\
&\propto |\Sigma|^{-T/2}\exp\left\{-\tfrac{1}{2}tr\hat{\Sigma}\Sigma^{-1} - \tfrac{1}{2}(b - \hat{b})'(\Sigma^{-1} \otimes X'X)(b - \hat{b})\right\}
\end{aligned} \tag{32}$$

and

$$p(S) = N(0, I_T)p(S > 0) \tag{33}$$

The joint prior distribution of all parameters is

$$p(b, \Sigma) = p(b)p(\Sigma) \tag{34}$$

where

$$p(b) \propto 1 \tag{35}$$

$$p(\Sigma) \propto |\Sigma|^{-\frac{1}{2}(k+1)} \tag{36}$$

The priors of b and Σ are diffuse. Investors do not have particular beliefs on parameters and 'let data talk'.

The posterior distributions of b, Σ and S are:

$$p(b|Y, X, \Sigma) \sim N(\hat{b}, \Sigma \otimes (X'X)^{-1}); \tag{37}$$

$$p(\Sigma|Y, X, b) \sim IW(T - 2, \hat{\Sigma}); \tag{38}$$

The posterior of Σ is inverted Wishart with degrees of freedom $T - 2$ and parameter $\hat{\Sigma}$. And

$$p(S|Y, \Sigma, b) \sim N(\mu_s, V_s)I(S > 0) \tag{39}$$

where $\mu'_s = C'(\Sigma + CC')^{-1}(Y - \tau_T \xi)$ and $V_s = (1 - C'(\Sigma + CC')^{-1}C)I_T$. The inverse transform method can be used to generate S. To sample s from $N(\mu, \sigma^2)I(s > 0)$, simulate variate u from the uniform distribution $U[0, 1]$ and then taking $s = \mu + \sigma \Phi^{-1} \left[\Phi \left(\frac{-\mu}{\sigma} \right) + u \left(\Phi \left(\frac{\mu}{\sigma} \right) \right) \right]$ as the output.

Reference

Athayde, G. M. de, & Flôres, R. G. (2004). Finding a maximum skewness portfolio - a general solution to three-moments portfolio choice. Journal of Economic Dynamics and Control, 28(7), 1335-1352.

Azzalini, A. (1985). A class of distributions which includes the normal ones. Scandinavian journal of statistics, 171-178.

Azzalini, A., & Dalla Valle, A. (1996). The multivariate skew-normal distribution. Biometrika, 83(4), 715-726.

Azzalini, A., & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61(3), 579-602.

Azzalini, A., & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65(2), 367-389.

Babcock, B. A., Choi, E. K., & Feinerman, E. (1993). Risk and probability premiums for CARA utility functions. Journal of Agricultural and Resource Economics, 17-24.

Branco, M. D., & Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. Journal of Multivariate Analysis, 79(1), 99-113.

Briec, W., Kerstens, K., & Jokung, O. (2007). Mean-variance-skewness portfolio performance gauging: A general shortage function and dual approach. Management Science, 53(1), 135-149.

Cameron, A. C., & Trivedi, P. K. (2005). Microeconometrics: methods and applications. Cambridge university press.

Cass, D., & Stiglitz, J. E. (1970). The structure of investor preferences and asset returns, and separability in portfolio allocation: A contribution to the pure theory of mutual funds. Journal of Economic Theory, 2(2), 122-160.

Chung, Y., Johnson, H., & Schill, M. (2004). Asset pricing when returns are nonnormal: Fama-French factors vs. higher-order systematic co-moments. Journal of Business, 79(2), 923-940.

Cochrane, J. H. (2006). Asset pricing (Revised edition). Princeton University Press.

Engle, R. F., Hendry, D. F., & Richard, J. F. (1983). Exogeneity. Econometrica: Journal of the Econometric Society, 277-304.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of financial economics, 33(1), 3-56.

Fang, K.-T., Kotz, S. and Ng, K. (1990) Symmetric Multivariate and Related Distributions. London: Chapman and Wiley.

Geweke, J., & Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory. Review of Financial Studies, 9(2), 557-587.

Harvey, C. R., Liechty, J. C., Liechty, M. W., & Müller, P. (2010). Portfolio selection with higher moments. Quantitative Finance, 10(5), 469-485.

Harvey, C. R., & Siddique, A. (2000). Conditional skewness in asset pricing tests. The Journal of Finance, 55(3), 1263-1295.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. The American Economic Review, 92(5), 1644-1655.

Jondeau, E., & Rockinger, M. (2006). Optimal portfolio allocation under higher moments. European Financial Management, 12(1), 29-55.

Kan, R., & Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. Journal of Financial and Quantitative Analysis, 42(3), 621-656.

Kraus, A., & Litzenberger, R. H. (1976). Skewness preference and the valuation of risk assets. The Journal of Finance, 31(4), 1085-1100.

Levy, H. (2006). Stochastic dominance: investment decision making under uncertainty (Vol. 12). Springer Science+ Business Media.

Low, C., Pachamanova, D., & Sim, M. (2012). Skewness-Aware Asset Allocation: A New Theoretical Framework and Empirical Evidence. Mathematical Finance, 22(2), 379-410.

Markowitz, H. (1952). Portfolio Selection. The Journal of Finance, 7(1), 77-91.

Mehra, R., & Prescott, E. C. (1985). The equity premium: A puzzle. Journal of monetary Economics, 15(2), 145-161.

Mencía, J., & Sentana, E. (2009). Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation. Journal of Econometrics, 153(2), 105-121.

Patton, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. Journal of Financial Econometrics, 2(1), 130-168.

Pástor, L., & Stambaugh, R. F. (2000). Comparing asset pricing models: an investment perspective. Journal of Financial Economics, 56(3), 335-381.

Post, T., Van Vliet, P., & Levy, H. (2008). Risk aversion and skewness preference. Journal of Banking & Finance, 32(7), 1178-1187.

Ronn, E. I., Sayrak, A., & Tompaidis, S. (2009). The Impact of Large Changes in Asset Prices on Intra-Market Correlations in the Domestic and International Markets. Financial Review, 44(3), 405-436.

Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. Econometrica, 68(5), 1281-1292.

Rubinstein, M. (1973). The fundamental theorem of parameter-preference security valuation. Journal of Financial and Quantitative Analysis, 8(1), 61-69.

Sahu, S. K., Dey, D. K., & Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. Canadian Journal of Statistics, 31(2), 129-150.

Sampford, M. R. (1953). Some inequalities on Mill's ratio and related functions. The Annals of Mathematical Statistics, 24(1), 130-132.

Simaan, Y. (1993). Portfolio Selection and Asset PricingThree-Parameter Framework. Management Science, 39(5), 568-577.

Spanos, A. (1994). On modeling heteroskedasticity: The student's t and elliptical linear regression models. Econometric Theory, 10, 286-286.

Tiao, G. C., & Zellner, A. (1964). On the Bayesian estimation of multivariate regression. Journal of the Royal Statistical Society. Series B (Methodological), 277-285.

Tu, J., & Zhou, G. (2004). Data-generating process uncertainty: What difference does it make in portfolio decisions?. Journal of Financial Economics, 72(2), 385-421.

Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error terms. Journal of the American Statistical Association, 71(354), 400-405.
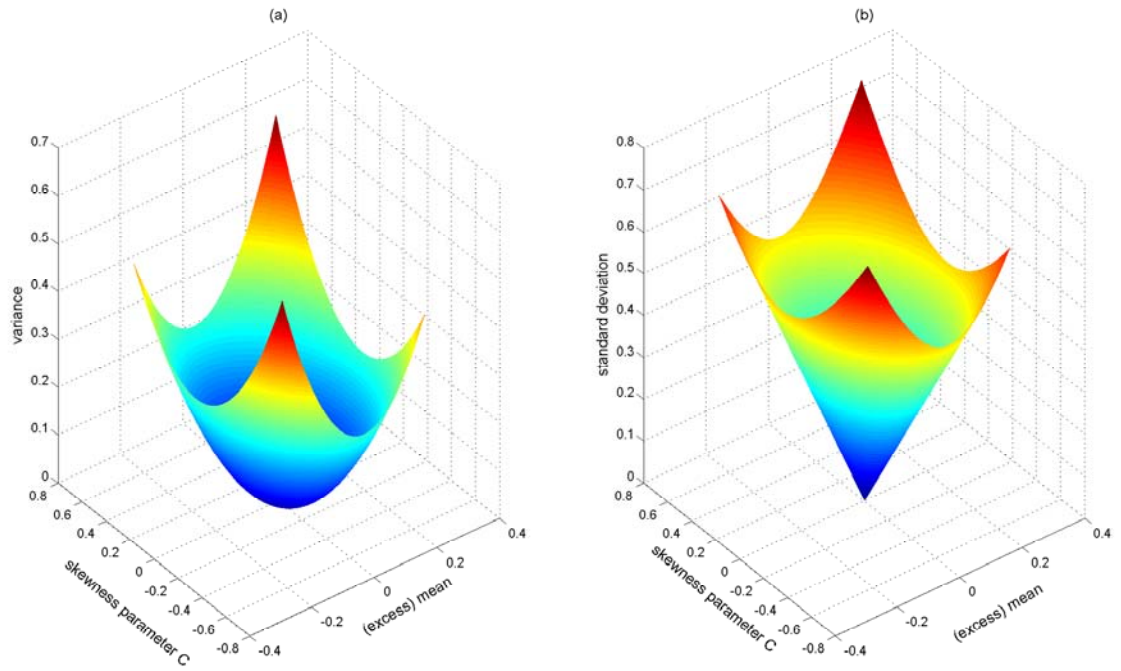
Figure 1: The efficient set of the three-parameter model with risk-free asset. Panel (a) shows the efficient set in the "(excess) mean - variance - skewness parameter C" space. Panel (b) shows the efficient set in the "(excess) mean - standard deviation - skewness parameter C" space.
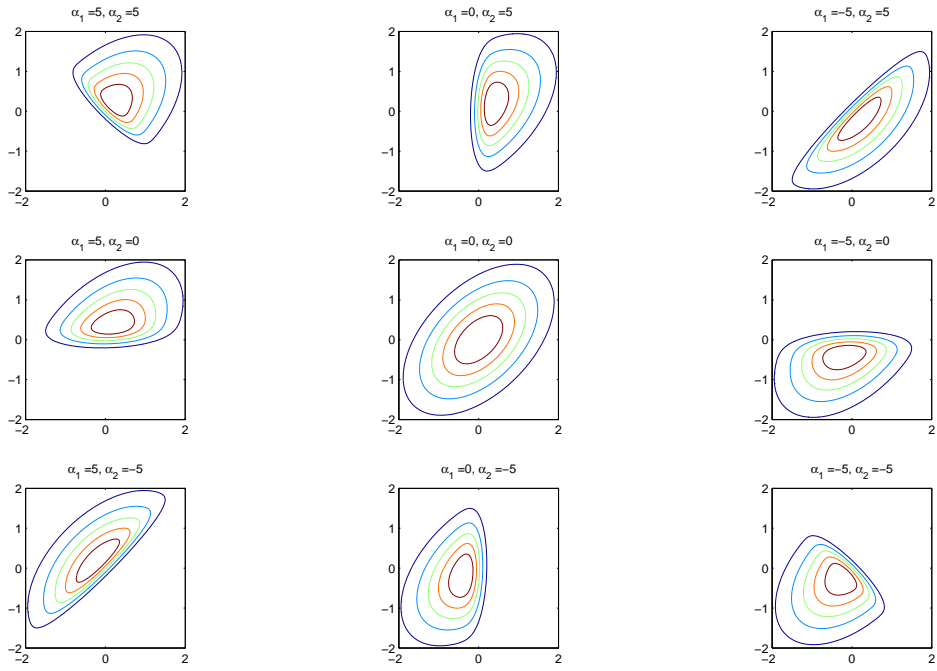
Figure 2: Contours of bivariate distributions with $\xi = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\Omega = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ and different values of α. The central panel corresponds to a normal distribution.
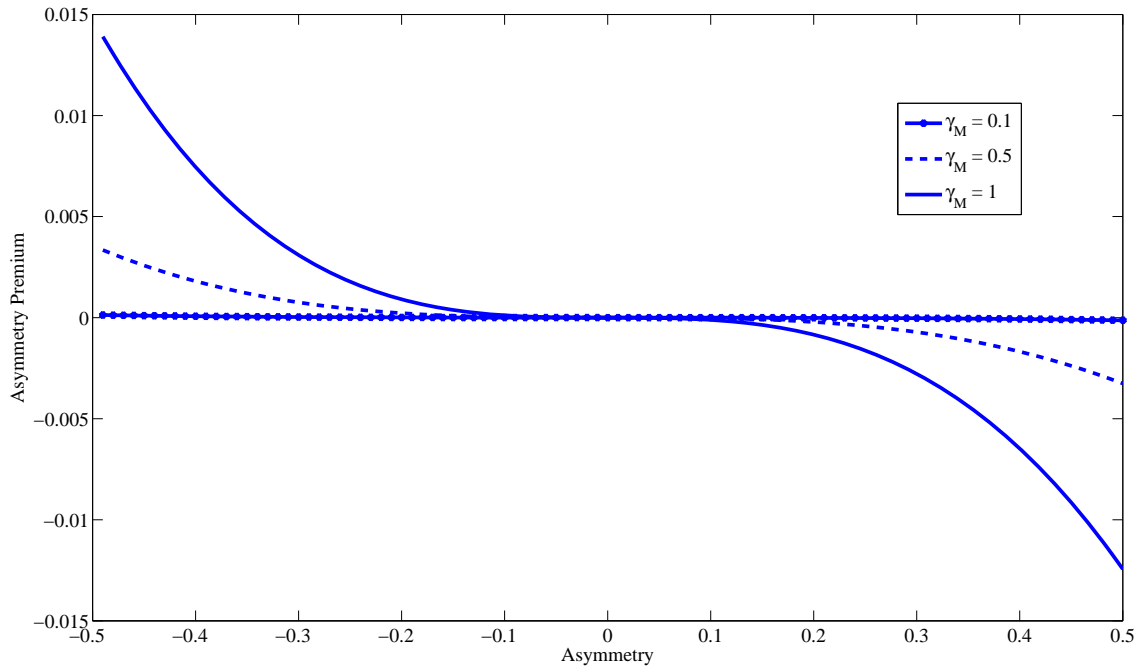
Figure 3: Skewness premium in the market price of risk with varying level of market portfolio skewness C_M. The market risk aversion coefficients $\gamma_M = \{0.1, 0.5, 1\}$

Tables

	parametric	nonparametric
all moments	Simaan (1993)	Low et al. (2012)
mean-variance-skewness	Mencía & Sentana (2009)	Athayde & Flôres (2004) Briec et al. (2007)

Table 1: Summary of theories on portfolio selection with higher moments. The theories are categorized into four groups. "Parametric" ("Nonparametric") theories have (no) distributional assumption. Theories consider all moments are identified as "all moments" theories. Theories only consider the first three moments are identified as "mean-variance-skewness" theories.

30 industry portfolios		
	Skewness	Kurtosis
Daily	5459670.65 (0.00)	6108.32 (0.00)
Weekly	274571.89 (0.00)	874.45 (0.00)
Monthly	30180.95 (0.00)	177.30 (0.00)

Table 2: Summary of Mardia's multivariate skewness and kurtosis test statistics along with their p-values in parentheses. The return series are 30 industry returns from January 1980 through December 2012 in daily, weekly and monthly frequencies.

	$\hat{C}$	90% Conf. Inter.
SL(l-h)	1.08	(-0.38, 2.28)
SM(l-h)	0.67	(-0.36, 1.72)
SH(l-h)	-1.11	(-2.08, -0.01)
ML(l-h)	1.05	(-0.13, 2.16)
MM(l-h)	-0.81	(-1.83, 0.30)
MH(l-h)	-0.19	(-2.29, 1.36)
BL(l-h)	0.14	(-1.01, 1.02)
BM(l-h)	-0.07	(-1.36, 1.38)
BH(l-h)	0.08	(-1.07, 1.16)
SMB	1.51	(0.41, 2.49)
HML	0.51	(-0.43, 1.35)
MKT	0.94	(-0.93, 2.46)

Table 3: The maximum likelihood estimates (in percentage points) of the skewness parameter C, along with 90% confidence intervals (in percentage points) in parentheses for the nine spread positions created by Pástor and Stambaugh (2000) and the three Fama-French factors based on monthly returns from July 1963 through December 1997. The confidence intervals are evaluated by the percentile method through 1000 bootstrapping replications.

	$\hat{C}$	90% Conf. Inter.
Food	0.89	(0.26, 1.49)
Beer	0.74	(0.20, 1.24)
Smoke	0.84	(0.13, 1.51)
Games	3.11	(1.60, 4.46)
Books	2.73	(1.70, 3.67)
Hshld	1.00	(0.31, 1.61)
Clths	3.60	(2.48, 4.59)
Hlth	1.10	(0.48, 1.65)
Chems	2.44	(1.28, 3.49)
Txtls	6.03	(4.50, 7.32)
Cnstr	3.64	(0.51, 5.30)
Steel	2.42	(-0.22, 4.23)
FabPr	1.91	(-0.07, 3.39)
ElcEq	2.18	(0.85, 3.29)
Autos	3.61	(0.00, 5.47)
Carry	1.49	(0.33, 2.51)
Mines	1.49	(-0.67, 3.19)
Coal	3.54	(1.06, 5.76)
Oil	1.13	(-0.55, 2.34)
Util	0.35	(-0.43, 1.08)
Telcm	0.00	(-1.04, 1.09)
Servs	-0.22	(-1.44, 1.02)
BusEq	-0.07	(-1.34, 1.14)
Paper	0.07	(-0.96, 1.16)
Trans	0.56	(-0.66, 1.78)
Whlsl	0.18	(-0.89, 1.22)
Rtail	0.71	(-0.36, 1.79)
Meals	0.38	(-0.63, 1.42)
Fin	2.67	(0.92, 4.12)
Other	0.83	(-0.32, 1.85)

Table 4: The maximum likelihood estimates (in percentage points) of the skewness parameter C, along with 90% confidence intervals (in percentage points) in parentheses for the 30 industry portfolios weekly returns from Jan 2006 through December 2012. The confidence intervals are evaluated by the percentile method through 1000 bootstrapping replications.

	mean-variance	skew-normal
	$c = \infty$	
SL(l-h)	0.0603	0.0604
SM(l-h)	0.0725	0.0732
SH(l-h)	0.1797	0.1774
ML(l-h)	-0.1707	-0.1680
MM(l-h)	0.0010	-0.0008
MH(l-h)	0.0380	0.0382
BL(l-h)	0.1436	0.1439
BM(l-h)	-0.0357	-0.0357
BH(l-h)	-0.0202	-0.0196
SMB	0.0957	0.0969
HML	0.4381	0.4379
MKT	0.1977	0.1964

Table 5: The unconstrained optimal allocations for the mean-variance-optimizing investor and the skew-normal-optimizing investor. The investment universe includes 12 risky positions (the nine spread positions created by Pástor and Stambaugh (2000) and the three Fama-French factors) and cash (risk-free asset). The risky asset weights are reported and they are normalized to sum to 1.

	mean-variance	skew-normal
	c = 10	
SL(l-h)	0.2043	0.0038
SM(l-h)	-0.1919	-0.0315
SH(l-h)	0.1783	-0.1296
ML(l-h)	-0.7466	-0.7639
MM(l-h)	0.0000	-0.0183
MH(l-h)	-0.1271	0.0080
BL(l-h)	-0.0049	-0.1742
BM(l-h)	-0.1318	-0.2789
BH(l-h)	0.1695	0.0000
SMB	1.1710	0.0941
HML	-0.2431	1.7485
MKT	3.4838	3.4963
Cash	-2.4838	-2.4963

Table 6: The optimal allocations with 10% margin requirement ($c = 10$) for the mean-variance-optimizing investor and the skew-normal-optimizing investor with risk aversion $\gamma = 0.00005$. The investment universe includes 12 risky positions (the nine spread positions created by Pástor and Stambaugh (2000) and the three Fama-French factors) and cash (risk-free asset).

	mean-variance	skew-normal
$c = \infty$		
Food	0.0491	0.6151
Beer	0.4444	0.5086
Smoke	1.2714	1.4613
Games	-0.0835	0.0407
Books	-1.1782	-1.4028
Hshld	-0.0818	-0.2545
Clths	0.4627	0.6196
Hlth	-0.1312	-0.0551
Chems	0.9334	1.1926
Txtls	0.6823	0.9302
Cnstr	-0.9090	-0.9218
Steel	-0.6358	-0.6376
FabPr	0.6340	0.8524
ElcEq	0.3895	0.4546
Autos	0.4517	0.6190
Carry	0.3832	0.4071
Mines	-0.0447	-0.0117
Coal	-0.0045	-0.0734
Oil	0.4563	0.3302
Util	-0.5274	-0.4182
Telcm	1.1085	1.1268
Servs	-0.4850	-0.6414
BusEq	-0.7084	-1.0247
Paper	-1.3055	-1.5362
Trans	-0.0458	-0.2691
Whlsl	0.1740	-0.2026
Rtail	-0.4337	-0.5218
Meals	1.2911	1.2262
Fin	-0.3191	-0.4312
Other	-0.8380	-0.9823

Table 7: The unconstrained optimal allocations for the mean-variance-optimizing investor and the skew-normal-optimizing investor. The investment universe includes 30 industry portfolios and cash (risk-free asset). The risky asset weights are reported and they are normalized to sum to 1.

	mean-variance	skew-normal
	c = 10	
Food	0.0083	-0.0036
Beer	0.0070	0.0830
Smoke	0.0001	0.0114
Games	-0.0045	0.0747
Books	-0.0050	0.0023
Hshld	0.0011	0.0010
Clths	0.2450	0.2319
Hlth	-0.0018	0.0122
Chems	2.7658	3.1348
Txtls	5.2462	6.0873
Cnstr	-0.0072	0.0004
Steel	0.0010	0.0221
FabPr	-0.0053	0.0087
ElcEq	0.0286	0.0076
Autos	0.0082	0.0013
Carry	0.0000	0.0063
Mines	0.5574	0.0107
Coal	1.0585	0.0251
Oil	0.0011	0.2012
Util	0.0045	-0.0058
Telcm	0.0011	-0.0070
Servs	-0.0031	0.0145
BusEq	-0.0011	0.0039
Paper	0.0034	-0.0053
Trans	-0.0034	0.0049
Whlsl	0.0018	0.0213
Rtail	0.0192	-0.0004
Meals	0.0089	-0.0041
Fin	0.0004	0.0009
Other	-0.0011	-0.0051
Cash	-8.9351	-8.9362

Table 8: The optimal allocations with 10% margin requirement ($c = 10$) for the mean-variance-optimizing investor and the skew-normal-optimizing investor with risk aversion $\gamma = 0.00005$. The investment universe includes 30 industry portfolios and cash (risk-free asset).