Institute of Insurance Economics
University of St.Gallen

SKEWED DISTRIBUTIONS IN FINANCE AND ACTUARIAL SCIENCE

CHRIS ADCOCK

MARTIN ELING

NICOLA LOPERFIDO

WORKING PAPERS ON RISK MANAGEMENT AND INSURANCE NO. 122

EDITED BY HATO SCHMEISER

CHAIR FOR RISK MANAGEMENT AND INSURANCE

JUNE 2012

Skewed Distributions in Finance and Actuarial Science: A Review

by

C. J. Adcock[1], M. Eling[2], and N. Loperfido[3]

[1] University of Sheffield, United Kingdom, [2] University of St Gallen, Switzerland, [3] University of Urbino, Italy

Abstract

That the returns on financial assets and insurance claims are not well described by the multivariate normal distribution is generally acknowledged in the literature. This paper presents a review of the use of the skew-normal distribution and its extensions in finance and actuarial science, highlighting known results as well as potential directions for future research. When skewness and kurtosis are present in asset returns, the skew-normal and skew-Student distributions are natural candidates in both theoretical and empirical work. Their parameterisation is parsimonious and they are mathematically tractable. In finance, the distributions are interpretable in terms of the efficient markets hypothesis. Furthermore, they lead to theoretical results that are useful for portfolio selection and asset pricing. In actuarial science, the presence of skewness and kurtosis in insurance claims data is the main motivation for using the skew-normal distribution and its extensions. The skew-normal has been used in studies on risk measurement and capital allocation, which are two important research fields in actuarial science. Empirical studies consider the skew-normal distribution because of its flexibility, interpretability, and tractability.

This paper is comprised of four main sections: an overview of skew-normal distributions; a review of skewness in finance, including asset pricing, portfolio selection, time series modeling, and a review of its applications in insurance, in which the use of alternative distribution functions is widespread. The final section summarizes some of the challenges associated with the use of skew-elliptical distributions and points out some directions for future research.

Key Words: Skew-Normal Distribution, Asset Pricing, Portfolio Selection, Risk Measurement, Capital Allocation

JEL Classification: C16, C46, G10, G22

Corresponding Author

C J Adcock
The University of Sheffield
Management School
9 Mappin Street
Sheffield, S1 4DT
United Kingdom
Tel: +44 (0)114 - 222 3402
Email: (c.j.adcock@sheffield.ac.uk)

1. Introduction

A great deal of finance theory is based on the assumption, either explicitly or implicitly, that the multivariate probability distribution of asset returns is normal. However, many published studies indicate that this assumption is not supported empirically; indeed, models based on this assumption fail to satisfactorily fit real-world data (see, for example, Fama, 1965; Fung and Hsieh, 2000; Eling, 2008). In addition to fat tails, evidence which has been widely reported, it is recognized that there is often asymmetry in asset returns. In practice, this is usually taken to mean skewness.

That skewness is a feature of the returns of some financial assets was recognized early on in modern finance theory, with earliest credit perhaps going to von Neumann and Morgenstern (1944) or Friedman and Savage (1948). The seminal work of Samuelson (1970), followed by those of Arditti and Levy (1975) and Kraus and Litzenberger (1976), marks the point at which the modern theory of finance started revealing evidence of skewness analogous to that provided by Markowitz, Sharpe, Lintner, Tobin, and others.

In the context of actuarial science it is important to recognize that insurance risks have skewed distributions (see, for example, Lane, 2000), which is why in many cases the classical normal distribution is an inadequate model for insurance risks or losses. Some insurance risks also exhibit heavy tails, especially those exposed to catastrophes (see Embrechts *et al*, 2002). The skew-normal distribution and its extensions thus might be promising since they preserve the advantages of the normal distribution with the additional benefit of flexibility with regard to skewness and kurtosis. The skew-normal distribution and its extensions have been used recently in studies on risk management, capital allocation, and goodness of fit (Vernic, 2006; Bolance *et al*; 2008; Eling, 2012).

This review paper is motivated by the belief that asymmetry in asset returns, and skewness in particular, is an important field of research. Such research has two clear themes. The first is the never-ending and entirely natural wish to develop better univariate models for asset returns. Such model developments are in turn driven by the hope that new models may lead to better financial decisions. The second theme involves the idea that effective portfolio selection, and hence asset pricing, requires as a starting point the specification of a coherent multivariate probability distribution for asset returns. The classes of distributions that are the direct descendants of that described in Azzalini (1985) and those implied in Simaan (1987, 1993) are coherent; they also lead directly to asset pricing models, are parsimonious in their parameterisation and capable of extension. They are not, of course, without challenges and difficulties, but an assessment of them is included in this review.

The structure of this paper is as follows. Section 2 presents an overview of skew-elliptical distributions. It covers the methods of deriving such distributions since these are potentially relevant to a description of financial market behavior. The emphasis of this section is on the skew-normal distribution. There is also a brief presentation of related distributions, such as the skew-Student t and a more complex class of distributions referred to in the statistics literature as closed skew-normal. The section also covers issues associated with parameter estimation and inference. Section 3 is concerned with finance. The section covers asset pricing and portfolio selection, time series modeling based on the skew-normal and option pricing. Section 4 deals with applications in insurance, in which the use of alternative distributions is widespread. The section covers applications in claims modeling, capital allocation, and risk management. As will be discussed in this section, a number of authors recently have used skew-elliptical distributions as flexible alternatives to other commonly used distributions. Section 5 discusses some of the many challenges associated with the use of skew-elliptical distributions and offers some directions for future research.

In this paper, an n-vector is denoted by $\mathbf{X}$. The subscript n, denoting its length, is omitted. The vector $\boldsymbol{\mu}$ and matrix $\boldsymbol{\Theta}$ denote the unconditional expected value and covariance matrix of $\mathbf{X}$, respectively. It is assumed that both exist and that $\boldsymbol{\Theta}$ is positive definite. The vector of excess returns over the risk-free rate r is denoted by $\tilde{\boldsymbol{\mu}}$. In the usual notation, the corresponding scalar quantities are denoted X_i, μ_i, θ_i^2, θ_{ij}, and $\tilde{\mu}_i$ with subscripts omitted when appropriate. An n-vector of zeros is denoted by $\boldsymbol{0}_n$. A standardized random vector, with zero mean vector and unit covariance matrix, is denoted by $\mathbf{Z}$. The symbols $\varphi(x)$ and $\Phi(x)$ denote, respectively, the density and distribution functions of the standard normal distribution evaluated at x. Various additional notations and abbreviations are defined in the paper. Other notation is that in common use.

2. Skew-Elliptical Distributions

Skew-elliptical distributions form a large class of multivariate probability distributions. Indeed, it could be argued that they constitute a number of overlapping classes. Some properties of these distributions are complicated and there are difficulties associated with parameter estimation and inference. Nonetheless, they are an attractive class of models for both theoretical and empirical finance and actuarial science applications, chiefly because the parameterisation is parsimonious. It is often possible to interpret the distribution in terms of the behavior of the application under study and there are coherent multivariate distributions.

The first reported work on what is now called the skew-normal distribution is credited to the Italian mathematician de Helguero (1908). Roberts (1966) wrote an important paper on the topic, but the true progenitor of this class of distributions is Adelchi Azzalini. His initial paper, Azzalini (1985), was the start of a remarkable research effort that continues largely unabated to this day, more than 25 years later. The papers cited below include numerous works by Azzalini himself and his co-authors, as well as other researchers in the field. Many contributions, although important, are not cited, as they are not directly related to the subject of this review.

Many multivariate distributions belong to the class of skew-elliptical distributions. This review focuses on those specific skew-elliptical distributions and their properties that are most likely to be of use in finance, for either theoretical or empirical purposes. The review covers the basic multivariate skew-normal (henceforth, MSN) distribution and its properties. This distribution is first reported in Azzalini and Dalla Valle (1996). This is followed by a subsection describing the multivariate extended skew-normal (MESN) distribution. The MESN is attributed to Arnold and Beaver (2000) but was independently reported in Adcock and Shutes (2001) and as a conditional distribution in Azzalini and Capitanio (1999). It is referred to in the review paper by Azzalini (2005). This distribution has one more parameter than the skew-normal and so is more flexible, particularly when modeling the moments and the tails of the distribution. There are different kinds and levels of non-normality, but the skew-normal distribution, extended or otherwise, models only some of them. In particular, skewness and excess kurtosis are both limited. A natural generalization is the multivariate (extended) skew-Student t (MST or MEST) distribution. Like its symmetric counterpart, this has a degree of freedom parameter that makes it possible to model both higher moments and the tails of the distribution.

As the following sections show, there are several ways of representing skew-normal distributions. Despite the additional complexity, different forms are reported either because they serve different purposes or they offer different insights into the mechanisms that generate the distributions. In the following sections, presentation of the skew-normal and related distributions uses a notation that varies from that usually employed in the statistics literature. This is done in the belief that the notation chosen is more appropriate for use in financial applications.

2.1 Multivariate Skew-Normal Distribution

In its simplest form, a random *n*-vector $\mathbf{X}$ has the multivariate skew-normal distribution if its probability density function is

$$f(\boldsymbol{x};\boldsymbol{\eta},\boldsymbol{\Omega},\varsigma)=2\varphi_n(\boldsymbol{x};\boldsymbol{\eta},\boldsymbol{\Omega})\Phi\{\varsigma^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\eta})\},$$

where $\varphi_n(\boldsymbol{x};\boldsymbol{\eta},\boldsymbol{\Omega})$ is the probability density function of an *n*-dimensional multivariate normal distribution with mean vector $\boldsymbol{\eta}$ and (positive definite) covariance matrix $\boldsymbol{\Omega}$ evaluated at $\boldsymbol{x}$. The distribution of $\mathbf{X}$ is said to be multivariate skew-normal with location parameter vector $\boldsymbol{\eta}$, scale matrix $\boldsymbol{\Omega}$, and shape (or skewness) parameter vector ς. The shorthand notation $\mathbf{X}\sim MSN_n(\boldsymbol{\eta},\boldsymbol{\Omega},\varsigma)$ is used. This distribution is symmetric if and only if $\varsigma=\boldsymbol{0}_n$. In this case, $\mathbf{X}$ has the multivariate normal distribution $N_n(\boldsymbol{\eta},\boldsymbol{\Omega})$. The multivariate skew-normal distribution retains some useful properties of the multivariate normal. For example, the density function is unimodal, its support is $\mathcal{R}^n$, and the marginal distribution of any sub-vector of $\mathbf{X}$ is also multivariate skew-normal. This parameterisation is generally known as the “directional” parameterisation (Loperfido, 2010).

The shape or skewness vector ς is sometimes referred to as the directional parameter, since it highlights the directions where departures from normality are more pronounced. Some properties of the MSN distribution are better expressed via the canonical parameter $\boldsymbol{\delta}$, which is defined as

$$\boldsymbol{\delta}=\frac{\boldsymbol{\Omega}\varsigma}{\sqrt{1+\varsigma^{T}\boldsymbol{\Omega}\varsigma}}\Leftrightarrow\varsigma=\frac{\boldsymbol{\Omega}^{-1}\boldsymbol{\delta}}{\sqrt{1-\boldsymbol{\delta}^{T}\boldsymbol{\Omega}^{-1}\boldsymbol{\delta}}}.$$

The probability density function of the univariate skew-normal distribution using this parameterisation is

$$f(x;\eta,\omega^2,\varsigma) = \frac{2}{\omega}\varphi\{(x-\eta)/\omega\}\Phi\{\varsigma(x-\eta)\},$$

$$\delta = \omega\varsigma \big/ \sqrt{1+\omega^2\varsigma^2} \Leftrightarrow \varsigma = \delta \big/ \sqrt{\omega^2 - \delta^2}.$$

This distribution is denoted $SN(\eta,\omega^2,\varsigma)$.

The Genesis of Skew-Normal Distributions

The parameters in the model are only meaningful if the theory purporting to explain the data and the model used to describe them are related. Thus it is necessary to understand the random mechanisms that lead to the model. This is sometimes referred to as the "model's genesis," an expression attributed to Johnson and Kotz (1970). The multivariate skew-normal distribution can be obtained from the multivariate normal in several ways, each of which has a separate interpretation. Three potentially relevant approaches to model genesis are convolution, conditioning, and maximization.

Convolution

This method of generating the multivariate skew-normal distribution is based on the convolution of a multivariate normal vector with a half-normal variable, which is distributed independently. It was first reported in Azzalini and Dalla Valle (1996). Let $\mathbf{U}$ have a multivariate normal distribution $N_n(\boldsymbol{\zeta},\boldsymbol{\Sigma})$ and let V be independently distributed as $N(0,1)$. The vector $\mathbf{X} = \mathbf{U} + \boldsymbol{\lambda}|V|$ has the multivariate skew-normal distribution

$$MSN_n(\boldsymbol{\zeta}, \boldsymbol{\Sigma} + \boldsymbol{\lambda}\boldsymbol{\lambda}^T, \varsigma), \varsigma = \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}\sqrt{1+\boldsymbol{\lambda}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}}.$$

The probability density function is

$$f(\boldsymbol{x};\boldsymbol{\zeta},\boldsymbol{\Sigma},\boldsymbol{\delta}) = 2\varphi_n(\boldsymbol{x};\boldsymbol{\zeta},\boldsymbol{\Sigma}+\boldsymbol{\lambda}\boldsymbol{\lambda}^T)\Phi\left\{\boldsymbol{\lambda}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\zeta}) \big/ \sqrt{1+\boldsymbol{\lambda}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}}\right\}.$$

As pointed out in Adcock (2010), this method of derivation is potentially attractive for finance applications because a possible interpretation of $|V|$ is that it represents a shock causing a departure from market efficiency as defined by Fama (1970). An individual asset's sensitivity to the shock is proportional to the corresponding element of $\boldsymbol{\lambda}$, which can be positive, negative, or zero. This result also provides a simple way of generating multivariate skew-normal vectors for simulation purposes. An equivalent approach is to define the vector $\mathbf{X} = \mathbf{U} + \boldsymbol{\lambda}V$ where V is an independently distributed $N(0,1)$ variable truncated from below at zero.

Conditioning

The n-dimensional MSN distribution can be obtained from an $n+1$ dimensional multivariate normal distribution by conditioning on the event that one component is greater than its expectation. This method of derivation of the MSN distribution was first reported by Azzalini and Dalla Valle (1996) and is defined as follows. Let $(\mathbf{X}^T, Y)$ have the full-rank multivariate normal distribution

$$N_{n+1}\left\{\begin{pmatrix}\boldsymbol{\eta}\\ 0\end{pmatrix},\begin{pmatrix}\boldsymbol{\Omega} & \boldsymbol{\gamma}^T\\ \boldsymbol{\gamma} & \sigma^2\end{pmatrix}\right\}.$$

The conditional distribution of $\mathbf{X}$ given that $Y > 0$ (or equivalently less than zero) is multivariate skew-normal $MSN_n(\boldsymbol{\eta}, \boldsymbol{\Omega}, \varsigma)$, $\varsigma = \boldsymbol{\Omega}^{-1}\boldsymbol{\gamma}\big/\sqrt{\sigma^2 - \boldsymbol{\gamma}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\gamma}}$. This result also provides a simple way of generating multivariate skew-normal vectors for simulation purposes. This method of derivation is easy to interpret in applications for which an observation on a random vector $\mathbf{X}$ is included in a sample if the value of another scalar variable Y is greater (less) than its mean. As the value of variable Y does not appear in the conditional distribution of $\mathbf{X}$, these models are sometimes referred to as "hidden truncation" models after Arnold and Beaver (2002). In finance in particular, if Y is taken to be the return on a given portfolio and $\mathbf{X}$ the returns on other assets, the implication of conditioning is that the conditional distribution of asset returns given that portfolio return is greater (less) than its mean is multivariate skew-normal even though the unconditional distribution of $\mathbf{X}$ is normal.

Maximization

The third method of derivation is known as maximization. Let $\mathbf{X}$ and $\mathbf{Y}$ be two random n-vectors each distributed as $N_n(\boldsymbol{\eta}, \boldsymbol{\Omega})$ and with cross-covariance matrix $\boldsymbol{\Omega} - \boldsymbol{\beta}\boldsymbol{\beta}^T$. Define $\mathbf{W}$ to be an n-vector with i^{th} element $W_i = max(X_i, Y_i)$, $i = 1, \cdots, n$. Loperfido (2008) shows that $\mathbf{W}$ has the multivariate skew-normal distribution

$$\mathrm{M}SN_n(\boldsymbol{\eta}, \boldsymbol{\Omega}, \varsigma),\ \varsigma = \boldsymbol{\Omega}^{-1}\boldsymbol{\beta}\big/\sqrt{2 - \boldsymbol{\beta}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\beta}}\ .$$

This property applies, with minor changes, to the vector formed by $\min(X_i, Y_i)$. If the corresponding elements of $\mathbf{X}$ and $\mathbf{Y}$ represent returns corresponding to two randomly selected price quotes for the same security, the result is somewhat similar to Siegel's lemma, (Siegel, 1993).

2.2 Extended Skew-Normal Distribution

The skew-normal distribution described above has potential limitations in financial applications. The convolution derivation implies that the shock to market efficiency has a half-normal distribution, which may be restrictive. The conditioning property above is defined only in relation to the mean return on the conditioning asset. Furthermore, as shown in Section 2.4, the relationship between the higher moments is fixed entirely by the shape parameter and both skewness and kurtosis are limited. The multivariate extended skew-normal distribution has an extra parameter and thus has the potential to overcome some of these limitations. This probability distribution may be derived using the convolution or conditioning methods, with appropriate changes of notation.

The multivariate extended skew-normal distribution was first described in Arnold and Beaver (2000). Using the notation of Section 2.1, the probability density function is

$$\varphi_n(\boldsymbol{x};\boldsymbol{\eta},\boldsymbol{\Omega})\Phi\left\{\tau\sqrt{1+\varsigma^T\boldsymbol{\Omega}\varsigma}+\varsigma^T(\boldsymbol{x}-\boldsymbol{\eta})\right\}/\Phi(\tau). \tag{1.}$$

For some purposes it is useful to write this probability density function as

$$\varphi_n(\boldsymbol{x};\boldsymbol{\eta},\boldsymbol{\Omega})\Phi\left\{\tau\sqrt{1+\boldsymbol{\psi}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\psi}}+\boldsymbol{\psi}^T\boldsymbol{\Omega}^{-1}(\boldsymbol{x}-\boldsymbol{\eta})\right\}/\Phi(\tau);\ \boldsymbol{\psi}=\boldsymbol{\Omega}\varsigma\,. \tag{2.}$$

The shorthand notation $\mathbf{X}\sim MESN_n(\boldsymbol{\eta},\boldsymbol{\Omega},\varsigma,\tau)$ or $MESN_n(\boldsymbol{\eta},\boldsymbol{\Omega},\boldsymbol{\Omega}^{-1}\boldsymbol{\psi},\tau)$ is used. Adcock and Shutes (2001) show that multivariate extended skew-normal distribution may also be derived using the convolution $\mathbf{X}=\mathbf{U}+\lambda V$, where $\mathbf{U}$ is defined as above, but V is now an independently distributed $N(\tau,1)$ variable truncated from below at zero. The corresponding probability density function is

$$\varphi_n\left(\boldsymbol{x};\boldsymbol{\zeta}+\boldsymbol{\lambda}\tau,\boldsymbol{\Sigma}+\boldsymbol{\lambda}\boldsymbol{\lambda}^T\right)\Phi\left[\tau\sqrt{1+\boldsymbol{\lambda}\ \boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}}+\left\{\boldsymbol{\lambda}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\zeta}-\boldsymbol{\lambda}\tau)\Big/\sqrt{1+\boldsymbol{\lambda}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}}\right\}\right]\Big/\Phi(\tau). \tag{3.}$$

Briefly, the roles played by these three representations are as follows. The MESN distribution defined in Equation (1) is considered the most suitable representation for parameter estimation. The representation in Equation (3) offers the simplest expressions for the moments and for the derivation of conditional distributions when the vector $\mathbf{X}$ is partitioned into two components. Equation (2) demonstrates the connection between Equations (1) and (3) and is more suitable for parameter estimation than Equation (3). To avoid confusion, the representation in Equation (3) is denoted

$$\mathbf{X}\sim MESN_n^C(\boldsymbol{\zeta},\boldsymbol{\Sigma},\boldsymbol{\lambda},\tau),$$

with the superscript C denoting convolution. Under this parameterisation, the variable V may still be regarded as a shock that causes a departure from market efficiency. There is now more flexibility in modeling the distribution of this shock. Similarly, if the unconditional distribution of returns on all assets is multivariate normal, the conditional distribution of asset returns, given that the return on a given portfolio is greater (or less) than a specified level, is MESN. The univariate form of the distribution is denoted $ESN^C(\zeta,\sigma^2,\lambda,\tau)$.

A slightly different parameterisation was used in Capitanio *et al* (2003) for inferential purposes. That paper also provides further motivation for the use of MESN distribution. Other parameterisations appear in Azzalini and Capitanio (1999), Arnold and Beaver (2000), and Loperfido and Guttorp (2008). Arnold and Beaver (2002, 2004) review theoretical properties of the MESN and univariate ESN distributions.

The extended skew-normal distribution appears in several areas of statistical theory, including Bayesian statistics (O'Hagan and Leonard, 1976), regression analysis (Copas and Li, 1997), and graphical models (Capitanio *et al*, 2003). It also appears in several areas of applied statistics: environmetrics, medical statistics, econometrics, and finance. Environmental applications of the ESN include modeling data from pollution monitoring stations (Loperfido and Guttorp, 2008) and uncertainty analysis related to the economics of climate change control (Sharples and Pezzey, 2007). In medical statistics, the ESN has been used as a predictive distribution for cardiopulmonar functionality (Crocetta and Loperfido, 2009) and for visual acuity (Loperfido, 2008). Financial applications mainly deal with portfolio selection (Adcock and Shutes, 2001; Adcock, 2007) and the market model, which relates asset returns to the return on the market portfolio (Adcock, 2004). In econometrics, the extended skew-normal distribution is known for its connection with bias modeling in Heckman's model (Heckman, 1979) and with stochastic frontier analysis (Kumbhakar *et al*, 1991).

The Fisher information matrix of both the MSN and MESN distributions is singular when $\boldsymbol{\psi} = \boldsymbol{\lambda} = \boldsymbol{0}$; that is, when the distribution is multivariate normal. This prevents straightforward application of standard likelihood-based methods to test the null hypothesis of normality. This is a familiar problem in the MSN case, but can be overcome via centered parameterisation. Centered parameterisation might also be useful for the MESN case, but satisfactory theoretical results for this distribution are difficult to obtain (Arellano-Valle and Azzalini, 2008). Problems with the information matrix of the MESN distribution are made worse by the truncation parameter τ, which indexes the distribution only when it is not normal. As a direct consequence of the above arguments, the rank of the information matrix is at least two less than its order, thus preventing application of results in Rotnitzky *et al* (2000).

2.3 Transformations of the Skew-Normal Distribution

The multivariate normal distribution has three familiar and useful properties. It is closed under affine transformations, convolution, and conditioning. An affine transformation of a normal random vector is another normal vector, the sum of two independent normal vectors with the same dimension is normally distributed. Similar properties hold for the extended skew-normal and hence the skew-normal distribution. Under the normal distribution, the outer product of two centered normal vectors has a Wishart distribution with one degree of freedom. The same property holds for the skew-normal but not for the extended skew-normal distribution. Finally, if an extended skew-normal vector is partitioned into two components, the conditional distribution of one component given the other is extended skew-normal.

Affine Transformation

If $\mathbf{X} \sim MESN_n(\boldsymbol{\eta}, \boldsymbol{\Omega}, \varsigma, \tau)$, $\mathbf{A}$ is a $m \times n$ real matrix and $\boldsymbol{b}$ a real *m-vector*, then $\mathbf{AX} + \mathbf{b} \sim MESN_n\left(\boldsymbol{A\eta} + \boldsymbol{b}, \boldsymbol{A\Omega A}^{\mathbf{T}}, \boldsymbol{A}\varsigma, \tau\right)$.

Convolution

Let $\mathbf{X} \sim MESN_n(\boldsymbol{\eta}_0, \boldsymbol{\Omega}_0, \varsigma, \tau)$ and $\mathbf{Y} \sim N_n(\boldsymbol{\eta}_1, \boldsymbol{\Omega}_1)$ independently, then $\mathbf{X} + \mathbf{Y} \sim MESN_n(\boldsymbol{\eta}_0 + \boldsymbol{\eta}_1, \boldsymbol{\Omega}_0 + \boldsymbol{\Omega}_1, \varsigma, \tau)$.

However, convolutions of (extended) skew-normal vectors are not (extended) skew-normal. *Ceteris paribus*, they are members of the closed skew-normal distribution, which is summarized in Section 2.5.

Wishartness

When $\mathbf{X} \sim MESN_n(\boldsymbol{\eta}, \boldsymbol{\Omega}, \varsigma, 0)$, the outer product $(\mathbf{X} - \boldsymbol{\eta})(\mathbf{X} - \boldsymbol{\eta})^T$ has the Wishart distribution $W(\boldsymbol{\Omega}, 1)$. This implies that the distribution of even functions of $(\mathbf{X} - \boldsymbol{\eta})$, including quadratic forms, do not depend on ς. These properties are of limited use in finance and actuarial science as they do not hold for nonzero values of τ.

These three properties may also be expressed using the $MESN_n^C$ parameterisation. Finally, if an extended skew-normal vector is partitioned into two components, the conditional distribution of one component given the other is extended skew-normal. As noted above, this result is better expressed using the $MESN_n^C$ parameterization.

Conditional Distributions

Let $\mathbf{X} \sim MESN_n^C(\boldsymbol{\zeta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$ be partitioned into two components $\mathbf{X}_{1,2}$ of lengths n_1 and $n_2 = n - n_1$, with corresponding partitions for $\boldsymbol{\zeta}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\lambda}$. The conditional distribution of $\mathbf{X}_2$ given $\mathbf{X}_1 = \boldsymbol{x}_1$ is $MESN_{n_1}^C(\boldsymbol{\zeta}_{2|1}, \boldsymbol{\Sigma}_{2|1} \boldsymbol{\lambda}_{2|1}, \tau_{2|1})$ with

$$\boldsymbol{\zeta}_{2|1} = \boldsymbol{\zeta}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\zeta}_1),\ \boldsymbol{\Sigma}_{2|1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12},$$

$$\boldsymbol{\lambda}_{2|1} = \frac{\boldsymbol{\lambda}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\lambda}_1}{\sqrt{1 + \boldsymbol{\lambda}_1^T \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\lambda}_1}},\ \tau_{2|1} = \frac{\tau + \boldsymbol{\lambda}_1^T \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{r}_1 - \boldsymbol{\zeta}_1)}{\sqrt{1 + \boldsymbol{\lambda}_1^T \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\lambda}_1}}.$$

This result facilitates the use of regression-type models in finance, particularly the market model. As the formulae above indicate, the models are nonlinear in the conditioning variables.

2.4 Moments of the Skew-Normal Distribution

Moments of the MSN and MESN distributions may be derived using any of the representations in the sections above. As the MSN distribution is a specific case of the corresponding MESN distribution with the parameter τ set equal to zero, the results in this section are for the extended version of the distribution. When $\mathbf{X} \sim MESN_n(\boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}^{-1}\boldsymbol{\Psi}, \tau)$, the cumulant generating function, ignoring constants, is

$$log\left\{E\left(e^{\boldsymbol{t}^T\mathbf{X}}\right)\right\} = \boldsymbol{\eta}^T\boldsymbol{t} + \boldsymbol{t}^T\boldsymbol{\Omega}\boldsymbol{t}/2 + \xi_0\left(\tau + \boldsymbol{t}^T\boldsymbol{\psi}\Big/\sqrt{1+\boldsymbol{\psi}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\psi}}\right),$$

where $\xi_0(x) = log\Phi(x)$. For the representation $\mathbf{X} \sim MESN_n^C(\boldsymbol{\zeta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$, it is

$$log\left\{E\left(e^{\boldsymbol{t}^T\mathbf{X}}\right)\right\} = (\boldsymbol{\zeta} + \boldsymbol{\lambda}\tau)^T\boldsymbol{t} + \boldsymbol{t}^T\left(\boldsymbol{\Sigma} + \boldsymbol{\lambda}\boldsymbol{\lambda}^{\mathrm{T}}\right)\boldsymbol{t}/2 + \xi_0\left(\tau + \boldsymbol{t}^T\boldsymbol{\lambda}\right).$$

Expressions for the vector of expected values, covariance matrix, co-skewness, and co-kurtosis are shown for both representations in Table 1. The function $\xi_k(x)$ is defined as follows: $\xi_k(x) = \partial^k\xi_0(x)/\partial x^k$.

Table 1 about here

The choice of representation is material for applications in finance and actuarial science. Under the representation defined in Equation (2.), the variance of returns on an asset or liability is

$$var(X_i) = \Omega_{ii} + \psi_i^2\xi_2(\tau)/\left(1 + \boldsymbol{\psi}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\psi}\right).$$

Since $\xi_2(\tau) \leq 0$, the variance decreases as the magnitude of skewness increases. Given the interpretation that the truncated variable represents a departure from market efficiency, it seems unlikely that this would be accompanied by a decrease in variance. Furthermore, an increase in skewness will increase the expected value. A fall in variance coupled with an increase in expectation contradicts the founding premise of modern finance that expected return and volatility are related along Markowitz’s efficient frontier. It is accepted that the estimation may be best performed using the central parameterization. Better interpretations, however, seem more likely under the convolution interpretation: variance increases with the magnitude of skewness.

There are several measures of skewness and kurtosis for multivariate distributions, but those proposed by Mardia (1970) are by far the most popular. When moments of appropriate order exist, Mardia’s skewness and kurtosis for a random vector $\mathbf{X}$ are

$$\varpi_1 = E\left[\left\{(\mathbf{X}-\boldsymbol{\mu})^T\boldsymbol{\Theta}^{-1}(\mathbf{Y}-\boldsymbol{\mu})\right\}^3\right], \quad \varpi_2 = E\left[\left\{(\mathbf{X}-\boldsymbol{\mu})^T\boldsymbol{\Theta}^{-1}(\mathbf{X}-\boldsymbol{\mu})\right\}^2\right],$$

respectively, where $\mathbf{X}$ and $\mathbf{Y}$ are independent and identically distributed random vectors with mean vector $\boldsymbol{\mu}$ and nonsingular covariance matrix $\boldsymbol{\Theta}$. Mardia (1975) also proposed $\tilde{\varpi}_2 = \varpi_2 - n(n+2)$ as a measure of excess kurtosis, this being the difference between the kurtosis of a random n-vector $\mathbf{X}$ and a normally distributed vector of the same dimension.

Mardia's skewness and kurtosis for the multivariate skew-normal distribution were originally presented in Azzalini and Capitanio (1999). When $\mathbf{X} \sim MESN_n\left(\boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega}^{-1}\boldsymbol{\psi}, \tau\right)$, it is straightforward to show that

$$\varpi_1 = \xi_3(\tau)^2\left(\boldsymbol{\psi}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\psi}\right)^3 \Big/ \left[1 + \boldsymbol{\psi}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\psi}\{1 + \xi_2(\tau)\}\right]^3,$$

$$\tilde{\varpi}_2 = \xi_4(\tau)\left(\boldsymbol{\psi}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\psi}\right)^2 \Big/ \left[1 + \boldsymbol{\psi}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\psi}\{1 + \xi_2(\tau)\}\right]^2.$$

Under the $MESN_n^C(\boldsymbol{\zeta}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau)$ representation, the results are the same since $\boldsymbol{\psi}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\psi} = \boldsymbol{\lambda}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\lambda}$. Upper bounds for these measures are

$$\varpi_1 \le \xi_3(\tau)^2 \big/ \{1 + \xi_2(\tau)\}^3, \tilde{\varpi}_2 \le \xi_4(\tau) \big/ \{1 + \xi_2(\tau)\}^2.$$

When $\tau = 0$, these values are, respectively, 0.99 and 3.07. For positive values of τ, the properties of the function $\xi_0(\ )$ and its derivatives imply that the measures of skewness and kurtosis rapidly tend to zero as τ increases. For negative values of τ, numerical investigations indicate the as $\tau \to -\infty$ the limiting values of the Mardia measures are 2 and 6 respectively. This result is confirmed by considering the expressions for $\xi_0(\ )$ and its derivatives that result from using the standard series representation for $\Phi(\ )$ (see for example Abramowitz and Stegun, 1965, page 932, section 26.2.12). These results provide a further demonstration which shows the additional flexibility of the extended version of the Skew-normal distribution.

2.5 Multivariate Skew-Student and Extended Skew-Student t Distributions

The multivariate skew-Student t (or MSN) distribution is a generalization of the skew-normal under which skewness and excess kurtosis may take larger values. It was introduced by Branco and Dey (2001), independently in the extended form by Adcock (2002), and, in a slightly different notation, by Azzalini and Capitanio (2003). Multivariate skew-Student t distributions may be derived by conditioning or by convolution. Using the notation of Section 2.2, the probability density function of the multivariate extended skew-Student t (or MEST) distribution derived using convolution is

$$f(\boldsymbol{x}) = t_n\left(\boldsymbol{x}; \boldsymbol{\zeta} + \lambda\tau, \boldsymbol{\Sigma} + \lambda\lambda^T, \upsilon\right) \frac{T_{\upsilon+n}\left\{\sqrt{\frac{\upsilon+n}{\upsilon}} \frac{\tau + (\boldsymbol{x} - \boldsymbol{\zeta})^T \boldsymbol{\Sigma}^{-1}\lambda}{\sqrt{\{1 + Q(\boldsymbol{x}, \zeta, \lambda, \tau, \boldsymbol{\Sigma})/\upsilon\}\left(1 + \lambda^T\boldsymbol{\Sigma}^{-1}\lambda\right)}}\right\}}{T_\upsilon(\tau)}.$$

The quadratic form $Q(\boldsymbol{x}, \boldsymbol{\zeta}, \boldsymbol{\lambda}, \tau, \boldsymbol{\Sigma})$ is defined as

$$Q(\boldsymbol{x}, \boldsymbol{\zeta}, \boldsymbol{\lambda}, \tau, \boldsymbol{\Sigma}) = (\boldsymbol{x} - \boldsymbol{\zeta} - \boldsymbol{\lambda}\tau)^T \left(\boldsymbol{\Sigma} + \boldsymbol{\lambda}\boldsymbol{\lambda}^T\right)^{-1} (\boldsymbol{x} - \boldsymbol{\zeta} - \boldsymbol{\lambda}\tau).$$

The function $t_n(\mathbf{x};\varsigma,\mathbf{\Omega},\upsilon)$ denotes the probability density function of an *n*-variate Student distribution with υ degrees of freedom, location parameter vector ς, and dispersion matrix $\mathbf{\Omega}$ evaluated at $\mathbf{x}$, and $T_\upsilon(x)$ denotes the distribution function for the Student-t distribution with υ degrees of freedom evaluated at x. The notation $\mathbf{X} \sim MEST_n^C(\boldsymbol{\zeta},\mathbf{\Sigma},\boldsymbol{\delta},\tau,\upsilon)$ is used. The multivariate skew-Student t (or MSN) distribution arises when $\tau = 0$. There is a similar result that corresponds to the centered parameterization. As the degrees of freedom $\upsilon \to \infty$, the MEST and MST densities tend to the corresponding multivariate skew-normal distributions. When the shape parameter $\boldsymbol{\psi} = \boldsymbol{\delta} = \boldsymbol{0}$ and $\tau = 0$, the distribution is the multivariate Student t. For the extended case with nonzero values of τ, the resulting symmetric multivariate distribution is described in Arellano-Valle and Genton (2010) and Adcock (2010). Univariate forms of the distribution are described by the shorthand EST^C.

Applications of the skew-t distribution include income distribution (Azzalini *et al*, 2003) and coastal flooding (Thompson and Shen, 2004). Finance applications are described in Section 3. Interest in skew-t distributions may be motivated by the properties of likelihood-based inference (Azzalini and Capitanio, 2003; Azzalini and Genton, 2008). It is possible that maximum likelihood estimates of the shape vector the degrees of freedom might be on the boundary of the parameter space; that is, take infinite values. The problem has received some attention for the (limiting case of the) skew-normal distribution for boundary estimates of the shape parameter (Pewsey, 2000), but it seems that this remains an issue for the skew-Student t distribution. In general, deriving the properties of estimators of the parameters of the MEST and MST distributions remains an outstanding research task.

2.6 Generalizations of Skew-Elliptical Distributions

Two other classes of skew-elliptical distributions are briefly summarized below.

Generalized Skew-Normal Distribution

Azzalini and Capitanio (2003) and Genton and Loperfido (2005) independently introduced the generalized skew-normal distribution (GSN), whose pdf is

$$f(\boldsymbol{x};\boldsymbol{\xi},\mathbf{\Omega},\boldsymbol{\alpha}) = 2\varphi_n(\boldsymbol{x};\boldsymbol{\xi},\mathbf{\Omega})\pi(\boldsymbol{x}-\boldsymbol{\xi}),$$

where $\pi(\boldsymbol{x})$ is a scalar function satisfying $0 \le \pi(x) = 1 - \pi(x) \le 1$ and is known as the skewing function. The distribution is denoted $GSN_n(\boldsymbol{\xi},\mathbf{\Omega},\pi)$. The invariance properties of skew-normal distributions still hold in the GSN class; that is, even functions of $(\boldsymbol{X}-\boldsymbol{\xi})$, $\mathbf{X} \sim GSN_n(\boldsymbol{\xi},\mathbf{\Omega},\pi)$ have a distribution that does not depend on the skewing function. Moreover, the class is flexible enough to include multimodal as well as platikurtic distributions. Other properties of GSN distributions, including those related to convolutions, linear transformations, and moment generating functions, are discussed in Loperfido (2004), Ma *et al* (2005). Ma and Tsiatis (2006) and Azzalini *et al* (2010) propose point estimates for the parameters of GSN distributions. To date, no one has addressed hypotheses testing problems for GSN distributions when no prior knowledge about their parameters is assumed. To the best of our knowledge, there are no reported applications of this distribution in either finance or actuarial science.

Closed Skew-Normal and Related Distributions

Gonzàlez-Farías *et al* (2004) introduced what is known as the closed skew-normal distribution (or CSN). As the name indicates, this multivariate distribution is closed with respect to conditioning, linear transformations, and convolutions. An extended version of the CSN distribution is described in Adcock (2007). Closed skew-Student t (or CST) distributions, extended or otherwise, are described in Arellano-Valle and Genton (2010) and Adcock (2009). The expressions for the probability density functions and moment generating function (for the CSN cases) are complicated and thus omitted here.

CSN and CST distributions can be generated by considering the event that several components of a normal vector are greater than given constants. Specifically, if two vectors $\mathbf{U}_n$ and $\mathbf{V}_k$ have a full rank multivariate normal distribution in $n+k$ dimensions, the distribution of $\mathbf{U}_n$ given that $\mathbf{V}_k > \boldsymbol{0}$ (or $\mathbf{V}_k < \boldsymbol{0}$) is CSN. CSN and CST distributions may also be derived from the convolution $\mathbf{U}_n + \boldsymbol{\Lambda}\mathbf{V}_k$, where each element of $\mathbf{V}_k$ is truncated from below at zero and $\boldsymbol{\Lambda}$ is a real $n \times k$ matrix of shape parameters. The CSN class includes as special cases several other generalizations of the skew-normal distribution, for example, the MSN distribution (Azzalini and Dalla Valle, 1996), the MESN distribution (Arnold and Beaver, 2000; Adcock and Shutes, 2001), and the hierarchical skew-normal distribution (Liseo and Loperfido, 2003). CSN distributions have proven useful in modeling environmental data (Loperfido and Guttorp, 2008), distributions of order statistics (Loperfido *et al*, 2007), and linear combinations of order statistics (Crocetta and Loperfido, 2005). They have also been used for other applications in finance (Roch and Valdez, 2011) and econometrics (Domínguez-Molina *et al*, 2004).

Although mainly of theoretical interest, in finance the convolution case may correspond to situations in which departures from symmetry in the distribution of returns are caused by more than one shock, which could occur, for example, in an international portfolio where different markets are affected by different factors.

3. Finance

The past 35 years have witnessed the publication of many papers concerned with skewness in asset returns, which, for the purpose of this review, can be sorted into four related, but nonetheless distinct, groups: (i) empirical studies, (ii) portfolio selection, (iii) asset pricing, and (iv) time series models. Any paper, of course, may contain a contribution in more than one of these areas. Although asset pricing and portfolio selection require the specification of multivariate probability distributions, there are many applications in finance for which it is sufficient to use univariate models. Ever since publication of Engle’s (1982) initial paper on ARCH models, which presented a model to deal with the observed empirical feature of heterogeneity of variance, there has been a substantial stream of research devoted to time series models. Development of these models in the finance, econometrics, and statistics literature over the past 30 years owes much to developments in computer facilities and to basic research into new probability distributions. Many of these models are also motivated by empirical studies into skewness in asset returns.

This section proceeds as follows. We begin with a short section that summarizes some of the main conclusions from a number of key empirical studies. This part of the review is not intended to be a complete survey of empirical studies into skewness; its purpose is to summarize some of the main themes in both theoretical and empirical research. Section 3.2 is concerned with portfolio selection, an area of research that has received considerable attention and has provided numerous interesting and useful results. Section 3.3 is concerned with asset pricing models and Section 3.4 with the use of skew-normal-based time series models and other applications.

3.1 Empirical Studies of Skewness

The seminal paper by Kraus and Litzenberger (1976) (henceforth, K&L) and those by Beedles (1979) and Friend and Westerfield (1980) are considered the starting point of empirical research into skewness. K&L's main purpose is to develop an asset pricing model that incorporates skewness and extends the CAPM. The empirical part of their paper is based on the returns from stocks that were listed continuously on the New York Stock Exchange from January 1926 until December 1935. Friend and Westerfield (1980) reevaluate the results reported by K&L. They conclude that the "*attempt to develop and substantiate a modified form of the Sharpe-Lintner CAPM is not successful.*" However, they also report that there is some evidence that investors may pay a premium for skewness but that such evidence is not conclusive. Today, it seems reasonable to believe that conclusions based on a period that could be regarded as unusual using data that were already around 50 years old at the time of their use are unlikely *a priori* to be relevant to the present day. Nonetheless, these papers, which found either in favor of or against skewness, set the tone for subsequent empirical work.

An indication of the equivocal nature of research into skewness may be discerned in the papers by Arditti and Levy (1975), Fogler and Radcliffe (1974), and Lau and Wingender (1989). Arditti and Levy consider multi-period returns. They report that multi-period skewness may not be zero even if single period skewness is. By contrast Fogler and Radcliffe (1974) and Lau and Wingender (1989) report the opposite: skewness is sensitive to sample size, the time period studied, and the frequency of the returns. In particular, skewness will tend toward zero as the frequency decreases. This difference in results may be due to the fact that the findings reported in Arditti and Levy are based on returns computed as percentages, whereas the other authors employ logarithms.

Another early study into asymmetry in stock returns is that of Kon (1984). His study used eighteen and a half years of daily returns data for 30 U.S. stocks, the S&P500 index, and the CRSP equally weighted and value weighted indices. Although the time period for this data set is not reported, according to Table 1 of Kon's paper, 30 of these securities and indices exhibit skewness in returns that are significant at the 1% level. Using a finite mixture of three or four normal distributions, he found a better fit to returns than that given by Student's t distribution. His conclusions, however, do not explain how many of the fitted mixture distributions exhibit skewness. Kon makes the point that both skewness and kurtosis in the entire data set may be explained by changes in mean and volatility. This is similar in spirit to the studies of skewness persistence described below.

Badrinath and Chatterjee (1988) study both skewness and kurtosis in daily returns on the CRSP value weighted market index. They report that this index exhibits a heavier right tail. Mills (1995) reports a study of daily returns on the FTSE100 and two related indices for the period 1986 to 1992. For the sample period as a whole, the returns on the indices are negatively skewed. If 1986 and 1987 are excluded, skewness is positive but smaller in magnitude and there is a decline in kurtosis. For returns after 1987, Mills's findings are consistent with those of Badrinath and Chatterjee and those of Akgiray and Booth (1988), as well as with other studies of U.S. stock returns cited in Singleton and Wingender (1986). Lim (1989) uses monthly stock return data from the CRSP database for the period January 1933 to December 1982 to test the K&L three-moment CAPM. He reports some evidence that systematic skewness is priced.

In an earlier paper, Beedles (1979) considers the impact of the definition of return and the sample period on the skewness measures for monthly data over the period 1927–1976. He finds that the estimates of skewness are persistently significant. However, he also reports that they are not necessarily stationary; that is, they are not constant in time. The papers by Muralidhar (1993), DeFusco *et al* (1996), and Sun and Yan (2003) are also concerned with the evidence that skewness in asset returns may be a transient feature. In a notable paper, Singleton and Wingender (1986) consider the persistence of skewness using monthly returns data for the period 1961–1980. They find that positively skewed assets are as likely to exhibit negative skewness in the next period as positive, and vice versa. Using Spearman's rank-order correlation test, the relationship between current and future skewness is found to be weak and statistically insignificant. Furthermore, they find that when assets are combined in portfolios, skewness is generally reduced. The transience of skewness is studied in the context of emerging markets by Bekaert *et al* (1998), who suggest that the phenomenon is part of the emergence process. The study by Singleton and Wingender (1986) also is concerned with skewness persistence. They use monthly returns for 551 U.S. stocks that were continuously listed in the CRSP database from 1961 to 1980. As well as monthly returns, they also compound the data to give semi-annual and annual returns. They report that the incidence of skewness is "*relatively stable*" over the time period of the study, but that the skewness of individual stocks does not persist. In keeping with this theme, Lau *et al* (1989) argue that the estimates used are random variables and that it is possible for the sample values to show evidence of change even when the population values remain constant.

An important paper by Harvey and Siddique (2000) proposes a model for asset returns that gives new insight into asset pricing and offers conclusions about the skewness of U.S. stocks. The paper deals with conditional skewness and, in particular, with the co-skewness of asset returns with returns on the market portfolio in addition to covariance. This model is discussed further below in the subsection on asset pricing. Harvey and Siddique use monthly returns on U.S. stocks from the CRSP database, mainly from the period July 1963 to December 1993. They report that conditional skewness helps explain the cross-sectional variation of expected returns even when size and book-to-market factors are included in the model. Perhaps most interestingly, they report that skewness commands a premium of over 3.5% *per annum*. The much earlier paper by Simkowitz and Beedles (1978) pursues a related theme. Using monthly data on 549 U.S. stocks that were listed continuously on the New York Stock Exchange from

January 1945 to December 1965, they argue that investors who seek skewness in portfolio returns should limit the number of assets in their portfolios. A related study from the same era, Simkowitz and Beedles (1980), is based on monthly returns from the Dow-Jones 30 stocks from January 1951 to June 1968 and offers evidence that these stocks exhibit substantial positive skewness.

We draw this short review to a close by weaving together some of the themes revealed in the work discussed. The results of empirical studies are equivocal, a lack of persistence in skewness is reported, and the majority of studies are based on U.S. data. Several different frequencies are covered and, viewed from the present day, the data sets are old. Nonetheless, the theoretical role of skewness continues to attract the attention of researchers, as the paper by Brunnermeier *et al* (2007) shows. It is also perhaps important to report a point made by Peirò (1999), who writes that the detection of skewness may depend on the test used and on the underlying (null hypothesis) distribution assumed for asset returns. He studies daily returns for nine major stock market indices and three spot exchange rates for the period January 1980 (January 1984 for the FTSE100) until September 1993. He reports that if the standard test of skewness is used, symmetry is rejected for eight of the nine indices and for all currencies. Using other distributions and, in particular, nonparametric tests, he concludes that daily returns are symmetric in most markets. Franceschini and Loperfido (2010) report tests for multivariate skewness based on the skew-normal distribution.

The main implications of these works and the many others that have appeared in the literature is that the empirical investigation of skewness remains a current research topic of interest and importance. However, the lack persistence of skewness and the equivocal nature of the some of the findings do not necessarily mean that it is impossible to exploit the third moment for portfolio selection.

3.2 Portfolio Selection

Most researchers would agree that the paper by Kraus and Litzenberger (1976) (K&L) marks the beginning of formal work on portfolio selection and asset pricing. This paper contains an asset-pricing model in which the expected return of an asset is a linear function of beta and the co-skewness of the asset with the market portfolio. This result, however, is based on the assumption that the expected value of the utility of portfolio return is a function of the portfolio expected value, variance, and skewness.

Using the notation defined in Section 1, the skewness of asset returns is defined as

$$\chi_i = E\left\{(X_i - \theta_i)^3\right\}^{1/3}.$$

The parameter known in the skewness literature as gamma is defined for security i with respect to the market portfolio as

$$\gamma_{im} = E\left\{(X_i - \theta_i)(X_m - \theta_m)^2\right\} / \chi_m^3 ,$$

with p replacing m if these quantities are computed with respect to a portfolio. K&L then assume that the expectation of the investor's utility of return is an explicit function $W(.)$ of μ_p, θ_p^2, and χ_p and that this is maximized subject to the budget constraint, with the investor being able lend or borrow at the risk-free rate. Denoting the derivatives of $W(.)$ with respect to μ_p, θ_p^2,and χ_p by W_{μ_p}, $W_{\theta_i^2}$, and W_{χ_p}, the first-order conditions are

$$\tilde{\mu}_i = -\left(W_{\theta_i^2} / W_{\mu_p}\right)\beta_{ip}\theta_p^2 - \left(W_{\chi_p} / W_{\mu_p}\right)\gamma_{ip} .$$

These lead to the equilibrium model $\tilde{\mu}_i = b_1\beta_{im} + b_2\gamma_{im}$, which is discussed below in the section on asset pricing. There are modifications to this model if there is no risk-free rate and if $\gamma_{mm} = 0$. This result implicitly requires that higher moments may be ignored or that they are suitable functions of one or more of the first three moments. It also assumes that differentiation and the expectation operations are interchangeable. This result motivated a number of empirical studies, some of which are summarized in Section 3.1. It also motivated other portfolio selection studies. Levy and Markowitz (1979), for example, suggest that expected utility may be approximated by a function of mean and variance. Other authors, including K&L, suggest that higher moments may be included in portfolio selection and asset pricing by considering the expectation of a Taylor series expansion of the utility and truncating it after a number of terms. This approach has generated some controversy even though it is widely used. Loistl (1976) points out that the Taylor series may not converge and that a truncated version therefore may lead to incorrect results. Hlawitschka (1994) disagrees and argues that the usefulness of a Taylor series approximation is "*strictly an empirical issue unrelated to the convergence properties of the series.*"

More generally, portfolio selection and, in particular, the tradeoff between risk aversion and preference for skewness requires the specification of a utility function $U(X_p)$ and the computation of its expected value. In K&L and related papers, this requires the derivatives with respect to the first three central moments of the distribution of portfolio returns. This task, in turn, requires at least the specification of a suitable probability distribution for portfolio return. A more rigorous foundation would be a suitable multivariate probability distribution for asset returns, from which the distribution of portfolio return may be derived. This still leaves open the choice of utility function and, in particular, the question of whether one utility function is better than another.

The preceding paragraph suggests that both portfolio selection and asset pricing in the presence of skewness should be based on a suitable class of multivariate probability distributions for asset returns. This class would be analogous to the class of elliptically symmetric distributions that underpin standard theory. One such is the subject of Yusif Simaan's Ph.D. thesis (Simaan, 1987, reported in Simaan, 1993). He proposes that the n-vector of returns on financial assets $\mathbf{X}$ may be represented as

$$\mathbf{X} = \mathbf{U} + \lambda V ,$$

where the n-vector $\mathbf{U}$ has an elliptically symmetric distribution, the scalar random variable V is distributed independently of $\mathbf{U}$ and has a nonelliptical distribution, and $\boldsymbol{\lambda}$ is a vector of parameters. Although general, this clearly defines a broad class of skew-elliptical distributions that are different from the Azzalini-style skew-elliptical distributions reviewed in Section 2.

Simaan shows (Simaan, 1993, Corollary 4.1) that under this class of elliptically symmetric distributions, the maximization of the expected value of any concave utility function may be solved by quadratic programming. This leads, in turn, to an asset-pricing model that is equivalent to Kraus and Litzenberger's and to a single mean-variance-skewness efficient surface. In the first article to employ skew-normal distributions in finance, Adcock and Shutes (2001) use the multivariate extended skew-normal distribution for both portfolio selection and asset pricing. They use the convolution method of generating the distribution, assuming that the vector U has the multivariate normal distribution $N_n(\xi, \Sigma)$ and that the scalar random variable V is distributed as $N(\tau,1)$ but is truncated from below at zero. They show that under the negative exponential utility function there is a mean-variance-skewness efficient surface. Adcock (2007) extends Stein's lemma for the multivariate skew-normal distribution to show that the same efficient surface obtains regardless of the utility function. These results are essentially a subset of those proven by Simaan (1987, 1993), which are more general in the sense that although the vector $\mathbf{U}$ and the scalar V are required to be independent, the choice of their respective probability distributions is broad.

Adcock (2007) contains a further extension of Stein's lemma in which returns are represented as

$$\mathbf{X} = \mathbf{U} + \boldsymbol{\Lambda}\mathbf{V},$$

where $\mathbf{V}$ is a k-vector of nonnegative shocks and $\boldsymbol{\Lambda}$ is an $n \times k$ matrix. This results in an extended version of the CSN distribution, which is summarized in Section 2.6. Under this distribution, portfolio selection may be performed by quadratic programming and there is a single mean-variance-skewness hyper surface, with k skewness dimensions. Adcock (2009) provides the analogous results for the multivariate extended skew-Student t distribution. As the discussion in Section 2.6 makes clear, the CSN distributions are immensely complex and are mainly of theoretical interest at present. However, the skew-normal or skew-Student t distribution with one truncated variable is a reasonable model for a single market in that it assumes that there is a single shock, represented by the scalar V, which affects all assets. If there are different types of shocks, then the more general model above is potentially realistic.

Menciá and Sentana (2008) report portfolio selection results based on a model that is a location–scale mixture of a normal vector. Using notation similar to that above, a vector of returns on financial assets is given by

$$\mathbf{X} = \varpi + \varsigma\boldsymbol{\lambda} + \sqrt{\varsigma}\mathbf{W},$$

where $\mathbf{W}$ is distributed as $N_n(\boldsymbol{0}, \boldsymbol{\Omega})$ and ς, which is nonnegative, is independently distributed. Menciá and Sentana (2008) posit that $1/\varsigma$ may have one of a number of distributions, of which the general inverse Gaussian is an example. It is also acceptable in this framework for $1/\varsigma$ to be a discrete random variable, in which case the unconditional distribution of X is a finite mixture of multivariate normals. The distribution of $\mathbf{X}$ is closed under affine transformations and, subject the existence of low-order moments of $1/\varsigma$, there is a mean-variance-skewness surface.

The search for better utility functions, which loosely means more expected return for less risk, means that Stein's lemma is an important component of finance theory. The fact that under elliptical symmetry it is only necessary to select a portfolio on the Markowitz mean-variance efficient frontier results in simplicity and time saving. The correspondence between the results of Simaan (1987, 1993) and Adcock (2007, 2009) suggests that further extensions of Stein's lemma may hold for the class of models in which returns are represented by the equation above with $\mathbf{U}$ having a multivariate elliptically symmetric distribution and V or $\mathbf{V}$ being independently distributed with a non-symmetric but otherwise arbitrary probability distribution. Results like Simaan's Corollary 4.1 or an extension of Stein's lemma for Menciá and Sentana's model are subjects for future research.

Chunhachinda *et al* (1997), Sun and Yan (2003), and Adcock (2005) are all comprehensive studies in which skewness is incorporated in portfolio selection. Meucci (2006) reports a version of the Black Litterman model based on the multivariate skew-Student t distribution. Harvey *et al* (2010) carry out a comprehensive study using the multivariate skew-normal distribution. Interestingly, this latter paper uses the MCMC method for parameter estimation. Kane (1982) is concerned with skewness preference and the implications for portfolio choice. Jondeau and Rockinger (2006a) are concerned with optimal portfolio allocation under higher moments. Wongwachara (2009) reports a study of the estimation error effect of higher moments. These three papers are based on a Taylor series expansion of the utility function. Post *et al* (2008) draw attention to problems with using Taylor's theorem. Benedetti (2004) examines higher-moment portfolio selection for hedge funds. De Athayde and Flôres (2004) present a general solution to the three-moment portfolio choice problem. Low *et al* (2012) introduce a new measure of skewness that they call "skewness aware deviation." This is deployed in a number of portfolio selection formulations. Other works concerned with skewness or with skewness in the context of higher-moment portfolio selection include Lai (1991), Briec *et al* (2007), Liu *et al* (2003), and Shiraishi and Taniguchi (2009).

3.3 Asset Pricing

K&L's formulation of mean-variance skewness leads to the equilibrium model

$$\tilde{\mu}_i = b_1\beta_{im} + b_2\gamma_{im}.$$

Harvey and Siddique (2000) propose a conditional model in which the stochastic discount factor is quadratic in the return on the market portfolio. This results in explicit formulae for b_1 and b_2. It also suggests that the market model takes the form

$$X_i = \beta_{0i} + \beta_{1i}X_m + \beta_{2i}X_m^2 + \varepsilon_i,$$

which is essentially the market-timing model of Treynor and Mazuy (1966). This model implies that the unconditional distribution of asset returns exhibits skewness. Exploiting the property that the extended skew-normal and skew-Student t distributions are closed under conditioning, Adcock (2004, 2010) derives the conditional distribution of X_i given X_m. The result is a market model that contains a nonlinear component in addition to the familiar linear term. This model behaves similarly to that of Treynor and Mazuy's. Positive (negative) skewness results in a concave (convex) curve. In this case, the curvature of the conditional expected value is due to skewness rather than market timing.

Carmichael and Coen (2011) use a bivariate skew-normal distribution to derive an asset pricing model based on the Euler equation and a power function for the gross return on wealth. Other papers cover the CAPM in the presence of skewness. Racine (1998) reports that the risk premia for beta and co-skewness are time varying. Tjetjep and Seneta (2006) propose that higher moments be modeled by a class of distributions referred to as the general normal variance-mean model. This is a mixture distribution in which the mixing variable affects both the mean and variance of the conditional normal distribution. The skew-normal and extended skew-normal distributions are members of this class.

3.4 Time Series Models and Other Applications

The limitations of the normal distribution in describing financial data and the presence of temporal effects in returns series, particularly at high frequency, have undoubtedly motivated the development of many well-known time series models. Many members of these classes of models are well known and are widely used in finance. Many of them are capable of modeling skewness in the conditional and unconditional distribution of returns. It is therefore unsurprising that there are very few applications of skew-normal and skew-Student t distributions to financial time series.
Among the studies that do exist, De Luca and Loperfido (2004) build a skew-in-mean GARCH model using the skew-normal distribution and apply their model to the U.K. FTSE index. De Luca *et al* (2006) describe a multivariate skew-GARCH model. A multivariate skew-GARCH model for financial times series is described in Franceschini and Loperfido (2010). Other papers concerned with skewness in the context of times series include well-known works by Bauwens and Laurent (2005) and Jondeau and Rockinger (2003, 2006b, 2012). Given the debate summarized above about the persistence of skewness, the development and implementation of models in which skewness, as well as volatility, is time varying is both welcome and timely.

An increasing number of other applications employ the skew-normal distribution. Corns and Satchell (2007) present an option pricing model based on skew-Brownian motion. A related work is in Madan *et al* (1998), which employs the variance gamma process for option pricing. Lien and Shrestha (2010) construct the minimum variance hedge ratio using the bivariate skew-normal distribution and then use it to investigate the hedge ratio for more than 20 commodities when futures are used to hedge the spot position. Bortot (2010) studies tail dependence in both bivariate skew-normal and skew-Student distributions. In a substantial monograph, Jurczenko and Maillet (2006) present a collection of works concerned with higher moments.

4. Actuarial Science

In the following, we first generally discuss the role of skewness in actuarial science. We therefore consider three important fields of actuarial science: the collective model of risk theory, actuarial pricing, and solvency measurement. This is followed by a review of papers using the skew-normal distribution and its extensions in an actuarial context.

The collective model of risk theory (see Ambagaspitiya, 1999; Sundt, 1999) considers a random sum of the form

$$S = \sum_{i=1}^{N} X_i \ ,$$

that is, the sum of the elements X_i of a random vector of length *N*. In an insurance context, *N* can be interpreted as the random number of claims in a given period of time and X_i can be interpreted as random size of claim *I*, which is a nonnegative random variable. S then represents the aggregate claim, also called total claims. The insurance literature proposes many distributions for modeling both claim number and claim size. Typically, the claim number is modeled using a Poisson or a negative binomial distribution. For the claim size, gamma, lognormal, or Pareto distributions are often considered (see Kaas *et al*, 2009; Mikosch, 2009). For the compound distribution of the aggregate claim *S*, Cummins and Wiltbank (1983) present the cumulants and Sundt (2002) the distribution function.

In most insurance lines there is a relatively high number of small losses and a relatively low number of extremely high losses. As such, the distribution of total claims is typically highly skewed. Early on, Dickerson *et al* (1961) indicated that the distribution of amounts of loss is usually skewed (see also Lane, 2000). One classical approach to integrate skewed distributions in actuarial analysis is to use a Cornish Fisher expansion in order to transform the quantiles and the tail expectations of the skewed distributions into a standard normal distribution (see Sandström, 2007). In the actuarial literature this is known as the "normal power approximation" (see, for example, Beard *et al*, 1984; Daykin *et al*, 1994; Sandström, 2005; Eling *et al*, 2009). Another motivation for considering skewness is the use of Edgeworth expansions to model insurance claims (Chaubey *et al*, 1998; Albers, 1999; Brito and Freitas, 2008; van Haastrecht *et al*, 2010).

As to claim size distribution, actuarial literature uses the skew-normal to analyze skewness in insurance claims. Vernic (2006) discusses the skew-normal distribution as an alternative to classical distributions in the context of risk measurement and capital allocation. Bolance *et al* (2008) empirically evaluate tail conditional expectation for a set of bivariate claims data from motor insurance (that is, property damage as well as medical expenses costs) using several parametric bivariate distributions (normal, lognormal, and log-skew-normal, among others) and kernel estimation. These two papers show that when skewness is present, the skew-normal and its extensions are promising candidates in both theoretical and empirical work. Both will be discussed in more detail below.

A second important field of insurance literature in which skewness is relevant is actuarial pricing and cost of capital estimation. The traditional CAPM, which requires equity returns to be elliptically distributed, is often used to estimate the cost of equity for insurance firms (Harrington, 1983; Cummins and Harrington, 1988; Cummins and Lamm-Tennant, 1994; Lee and Cummins, 1998). However, property-liability insurance claims at the firm level can be highly skewed and heavy-tailed (Cummins, Dionne, and Pritchett, 1990; McNeil, 1997), implying that equity returns for property-liability insurers will not be normally distributed either. For example, Cummins and Harrington (1988) document significant levels of skewness in insurance stock returns for a sample of property-liability companies for the time period 1970 to 1983 (see also Cummins and Nye, 1980). Similar findings are reported for the life insurance sector (see, for example, Gentry and Pike, 1970; Harrington, 1983).

Insurance companies thus typically face negatively asymmetric or skewed distributions of outcomes. In particular, many insurance and reinsurance portfolios contain exposures to catastrophic risks. Such events are particularly damaging to insurers and reinsurers and should impact risk allocation, pricing, and capital structure decisions (see Froot, 2007). An important set of new studies in the field of insurance thus analyzes return distributions without the assumption of normal distribution in order to incorporate skewness or even higher moments. To incorporate non-normal conditions into the pricing of insurance contracts, the literature suggests use of the three-moment CAPM (following Kraus and Litzenberger, 1976) and an *n*-moment CAPM (Kozik and Larson, 2001). In Kozik and Larson (2001), these two models are considered in the context of the insurance CAPM proposed by D'Arcy and Doherty (1988). More recent studies by Harvey and Siddique (2000), which is discussed in Section 3, and Chung *et al* (2006) empirically examine the effects of co-skewness and higher-order co-moments on determination of the cost of equity. According to Wen *et al* (2008), adoption of the n-moment insurance CAPM could possibly capture the non-normal characteristics of the insurance claims process, but the determination of the optimal moment and its finite nature limits the application of this model.

Froot (2007) develops a new pricing model that explicitly considers the asymmetry in insurance risk. The model is an extension of Doherty (1991) and the Froot and Stein (1998) model and explicitly applicable to insurers and reinsurers. Specifically, Froot (2007) develops a three-factor pricing model for nontradable, negatively skewed insurance risks. In addition to the market systematic risk factor, the model includes a factor for the covariability of a given risk with the firm's other nontraded risks (the "firm-wide" risk factor) as well as a factor that prices the asymmetry of the insurer's return distribution. The predictions of the model are similar to those of Froot and Stein (1998), except that Froot (2007) reports even stronger deviation from prices predicted by models that assume perfect market conditions (for example, the single-factor CAPM), reflecting policyholder risk aversion and asymmetrical returns.

A second study using an alternative modeling approach is Wen *et al* (2008). Using a sample of publicly traded property-liability insurers, they compare the equity betas generated from the Rubinstein/Leland (Leland, 1999) model to those generated by the CAPM. As suggested by Leland (1999), the authors find significantly different equity betas for insurers with a greater degree of non-normal or asymmetric returns, since CAPM assumptions are violated. Furthermore, based on economies of scale, they find that larger insurers, through the insurance pooling process, reinsurance, and/or the use of financial hedging techniques, are more able to mitigate asymmetric risks embedded in their insurance portfolios.

Cummins, Lin, and Phillips (2009) empirically test the theories developed by Froot and Stein (1998) and Froot (2007). Overall, their prediction is that prices of illiquid, imperfectly hedgeable intermediated risk products should depend on the firm's capital structure, the covariability of the risks with the firm's other projects, their marginal effects on the firm's insolvency risk, and negative asymmetries of return distributions. In particular, prices should be higher for insurance lines having a higher covariability with the insurer's overall insurance portfolio and for lines having a greater marginal effect on the insurer's insolvency risk. Cummins, Lin, and Phillips (2009) provide empirical tests of these theoretical predictions. They first estimate the price of insurance for a sample of U.S. property-casualty insurers and then regress insurance prices on variables representing the firm's insolvency risk, capital allocations byline, and other firm-specific characteristics. The Cummins-Lin-Phillips tests support the stated theoretical predictions and provide evidence that insurance prices reflect the asymmetries of return distributions.

The relevance of skewness in actuarial science is also recognized in recent articles on regulation. An important development in this context is Solvency II, which will be the new regulatory framework in the European Union from 2013 onward and has been under development since the late 1990s. Sandström (2007) proposes and discusses an approach to calibrate the Solvency II standard formula for skewness. He notes that if any of the underlying probability distributions are skewed, the Solvency II model must be calibrated for that so as to retain its consistency. Pfeifer and Strassburger (2008) point out problems of the standard formula for calculating solvency capital to cope with skewness in the individual risk distribution.

The presence of negatively asymmetric or skewed distributions of outcomes is thus the main motivation for considering the skew-normal distribution and its extensions in actuarial science. The literature on skew-elliptical distributions is not as broad in actuarial science as it is in finance, but some applications of skew-elliptical distributions in actuarial research have been published in recent years. The skew-normal is used in studies on risk measurement and capital allocation (Vernic, 2006), which are two of the most popular research fields in actuarial science. Furthermore, a number of empirical studies consider the skew-normal distribution, especially in a goodness-of-fit context (Bolance et al, 2008; Eling, 2012). It should be mentioned that though this section is focused on actuarial science, the work on risk measurement and capital allocation is also of great importance for other financial services sectors, such as banking.

4.1. Risk Measurement

Vernic (2006) discusses the skew-normal distribution as an alternative to the classical normal one in the context of risk measurement and capital allocation. Risk measurement and capital allocation have been two of the most popular topics in actuarial science in recent years, garnering a multitude of theoretical and empirical contributions. In risk measurement, a great deal of literature has been published since the work by Artzner *et al* (1999) on value at risk and the subsequent discussion of the drawbacks of value at risk compared to the benefits of tail conditional expectation (or conditional value at risk, see Acerbi and Tasche, 2002; Heyde *et al*, 2007). The popularity of risk measurement in actuarial science is reflected in the number of papers that study particular cases of tail conditional expectation when the loss vector follows a certain multivariate distribution. Panjer (2002) studies the multivariate normal, Valdez and Chernih (2003) the multivariate elliptical, Furman and Landsman (2005) a multivariate gamma, Landsman and Valdez (2005) exponential dispersion models, Chiragiev and Landsman (2007) a multivariate Pareto distribution, and Dhaene *et al* (2008) elliptical distributions.

Vernic (2006) extends this stream of literature by considering the tail conditional expectation (henceforth, TCE) for the multivariate skew-normal distribution. In the following, we reconsider the definition of the TCE for the normal, the skew-normal, and the skew-Student distribution and provide a numerical example. Results for the extended skew-normal and skew-Student distribution are omitted.

Normal

Let X be a random variable of losses with $X \sim N(\zeta, \omega^2)$. The properties of the truncated loss distribution may be found in Johnson and Kotz (1970, page 81). Note that in the insurance literature, the expression for the TCE for normal distribution is usually ascribed to Panjer (2002). If value at risk at probability α is denoted $VaR_\alpha = \zeta + \omega Z_\alpha$, the TCE for the normal distribution is

$$TCE_\alpha(X) = E(X \mid X \geq VaR_\alpha) = \zeta + \omega\varphi(Z_\alpha)/(1-\alpha).$$

Skew-Normal

If X now has the skew-normal distribution $SN(\eta, \omega^2, \varsigma)$ value at risk at probability α is given by $VaR_\alpha(X) = \zeta + \omega\tilde{Z}$, where $\tilde{Z}$ is the value at risk at probability α for an $SN(0, 1, \varsigma)$ variable. The TCE is then as computed by Vernic (2006). In the notation used here it is

$$TCE_\alpha(X) = \zeta + \omega(1-\alpha)^{-1}\left\{2\varphi(\tilde{Z})\Phi(\varsigma\tilde{Z}) + \delta\sqrt{2/\pi}\,\overline{\Phi}\left(\sqrt{1+\varsigma^2}\,\tilde{Z}\right)\right\},$$

where $\delta = \varsigma\big/\sqrt{1+\varsigma^2}$ is the Azzalini skewness coefficient, and $\overline{\Phi}(x) = 1 - \Phi(x)$.

Skew-Student

For the skew-Student distribution, value at risk at probability α is $VaR_\alpha = \zeta + \omega\tilde{Z}$ where $\tilde{Z}$ now denotes the value at risk for the standardized skew-Student distribution. This is the solution of the equation

$$\alpha = \int_{-\infty}^{\tilde{Z}} 2t_\upsilon(z)T_{\upsilon+1}\left\{\varsigma z\sqrt{(\upsilon+1)/(\upsilon+z^2)}\right\}dz\,.$$

For $\upsilon > 1$, the TCE for standardized skew-Student distribution is

$$\upsilon(\upsilon-1)^{-1}(1-\alpha)^{-1}\left[2t_\upsilon(\tilde{Z})\left(1+\tilde{Z}^2/\upsilon\right)T_{\upsilon+1}\left\{\sqrt{\upsilon+1}\,\varsigma\tilde{Z}\Big/\sqrt{\upsilon+\tilde{Z}^2}\right\} + \delta\xi_\upsilon\overline{T}_{\upsilon+1}\left\{\sqrt{\upsilon+1}\sqrt{1+\varsigma^2}\,\tilde{Z}\Big/\sqrt{\upsilon}\right\}\right]$$

where δ is as defined before and

$$\xi_\upsilon = 2\Gamma(\upsilon/2 + 1/2) \big/ \left\{\Gamma(\upsilon/2)\Gamma(1/2)\sqrt{\upsilon}\right\}.$$

For $\upsilon \leq 1$ the TCE does not exist. As the degrees of freedom increase, the TCE tends to that for the skew-normal distribution. Figure 1 shows numerical results for TCE with $\zeta = 0, \omega = 1$ for the standard normal (upper part of the figure), the skew-normal with three different shape parameters (10, 0, –10) (middle part of the figure), and the skew-Student the same three shape parameters (lower part of the figure).

Figure 1 about here

The skew-normal with a shape parameter of zero presented in the middle part of Figure 1 corresponds to the normal one presented in the upper part of the figure. Increasing the shape parameter increases skewness of the loss distribution, which increases the TCE. Vice versa, decreasing the shape parameter leads to a left skewed distribution so that TCE for the loss distribution is, *ceteris paribus*, lower than in the normal case.

For the skew-Student, TCE is higher than with the normal one if the shape parameter is zero since the tails are fatter than in the normal case. With a positive (negative) shape parameter, TCE again increases (decreases) compared to the symmetric case (shape = 0). Note that the absolute value of TCE is much higher in the lower part of Figure 1 than it is in the upper part of the figure, emphasizing the impact of fat tails on TCE.

4.2. Capital Allocation

Capital allocation is an application of risk measures to the task of allocating risk capital to different business segments of a company. Vernic (2006) investigates an allocation formula based on the TCE, which builds on the allocation formula presented by Wang (2002) for the normal distribution. Capital allocation is an important decision, especially since the amount of risk capital that is allocated to a business unit will also determine the importance and profitability of the business segment. The sum of the stand-alone risk capital of different business segments typically is higher than the risk capital of the whole company and this effect represents diversification opportunities inherent in combining different market segments in one company. The critical question for capital allocation is thus to decide to whom this diversification belongs.
Wang (2002) provides a capital allocation formula based on the tail conditional expectation of the normal distribution. The basic idea of this capital allocation is that the capital allocated to business segment *i* is proportional to the stand-alone TCE of business segment i in relation to the TCE of the (insurance) company. Vernic (2006) provides this capital allocation formula for the skew-normal distribution.

4.3. Goodness of Fit

Vernic's (2006) work is theoretical in nature and empirical study of risk measures is still rare in actuarial literature. This situation motivated Bolance *et al* (2008) to empirically evaluate tail conditional expectation for a set of bivariate claims data from motor insurance (property damage as well as medical expenses costs). First, they fit several parametric bivariate distributions (normal, lognormal, and log-skew-normal, among others) and used transformation kernel estimation to analyze a real data set from auto insurance. Second, they numerically illustrated the application of the TCE risk measures. As Table 1 of their paper shows, among the bivariate distributions investigated, the one with the worst fit (bivariate normal) clearly underestimates the empirical TCE. The bivariate lognormal and log-skew-normal offer the best fit to the data among parametric alternatives. This confirms that skew-elliptical distributions might be promising for research in actuarial science.[1] However, neither of these parametric alternatives is as close to empirical data as the nonparametric approximation.

Eling (2012) considers two well-known data sets from actuarial science and fits a number of parametric distributions as well as a nonparametric kernel estimator to these data. He finds that the skew-normal and skew-Student are reasonably competitive compared to other models in the actuarial literature when it comes to describing insurance data. In addition to goodness-of-fit tests, tail risk measures, such as value at risk and tail value at risk, are estimated for the data sets under consideration.

5. Conclusions and Directions for Future Research

The multivariate skew-normal distributions covered in this review are tractable, parsimonious in their parameterization, and capable of interpretation in terms of finance theory. They offer some specific theoretical advantages, notably for portfolio selection and asset pricing. Given that these distributions can model both skewness and kurtosis, there is a wide range of potential empirical applications for them in finance and actuarial science, including many of the topics covered in this review: portfolio selection, volatility modeling, option pricing, risk measurement, and capital allocation, among others.

Use of these distributions is not without challenges, however. The problems of parameter estimation and inference are well covered in the statistics literature. Notwithstanding the flexibility inherent in these distributions, there may be applications that require even greater flexibility in modeling non-normality.

Some specific areas in which further development would be welcome are as follows. First, given the literature about skewness persistence in financial markets and about the intrinsic skewness in insurance data, there is a case for further development of GARCH-type models in which skewness as well as variance is allowed to be time varying. Given the flexibility of skew-normal distributions, it seems clear that such models could be defined in a number of ways.

---

[1] Bolance *et al* (2008) also tried to fit their data to a bivariate Pareto distribution of the first kind and a bivariate Weibull, but the results were so poor compared with the other distributions that they were not included in the paper.

Second, for work with an international dimension it may be necessary to model more than one skewness shock, that is, more than one truncated variable. This raises a number of serious issues: estimation, inference, and efficient computation.

Third, the use of a single truncated variable suggests that the tensors of co-skewness and co-kurtosis of a vector of variables must have specific structures. Whether this is supported empirically is a topic for future research.

Fourth, the skew-normal and skew-Student distributions lead to a single mean-variance-skewness efficient surface. Whether these results hold for other skew-elliptical distributions or for models of the kind summarized in Section 3 is also an open question. Most important of all is the need for greater awareness of the potential of these distributions, which will only occur after they have been employed in more comprehensive empirical applications.

References

Abramowitz, M. and Stegun, I.(1965) Handbook of Mathematical Functions, New York: Dover.

Acerbi, C., and D. Tasche (2002) On the Coherence of Expected Shortfall. *Journal of Banking and Finance,* **26**, 1487–1503.

Adcock, C. J. (2002) Asset Pricing and Portfolio Selection Based on the Multivariate Skew-Student Distribution, *Multi-Moment Capital Asset Pricing Models and Related Topics Workshop*, Paris.

Adcock, C. J. (2004) Capital Asset Pricing in UK Stocks Under the Multivariate Skew-Normal Distribution. In Genton, M. G., editor, *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Boca Raton, Chapman and Hall, 191–204.

Adcock, C. J. (2005) Exploiting Skewness To Build An Optimal Hedge Fund With A Currency Overlay, *The European Journal of Finance*, **11**, 445-462.

Adcock, C. J. (2007) Extensions of Stein's Lemma for the Skew-Normal Distribution. *Communications in Statistics—Theory and Methods*, **36**, 1661–1671.

Adcock, C. J. (2009). The Multivariate Skew-Student Distribution and Extensions to Stein's and Siegel's Lemmas, Working Paper.

Adcock, C. J. (2010) Asset Pricing and Portfolio Selection Based on the Multivariate Extended Skew-Student-t Distribution. *Annals of Operations Research*, **176**, 221–234.

Adcock, C. J., and K. Shutes (2001) Portfolio Selection Based on the Multivariate Skew Normal Distribution. In Skulimowski, A., editor, *Financial Modelling*, Krakow, Progress & Business Publishers.

Akgiray, V., and G. B. Booth (1988) The Stable-Law Model of Stock Returns. *Journal of Business and Economic Statistics*, **61**, 51–57.

Albers, W. (1999) Stop-Loss Premiums Under Dependence. *Insurance: Mathematics and Economics*, **24**, 173–185.

Ambagaspitiya, R. S. (1999) On the Distributions of Two Classes of Correlated Aggregate Claims. *Insurance: Mathematics and Economics*, **24**, 301–308.

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

Arellano-Valle, R. B., and A. Azzalini (2008) The Centred Parametrization for the Multivariate Skew-Normal Distribution. *Journal of Multivariate Analysis*, **99**, 1362–1382. [Correction: vol.100 (2009), 816.]

Arellano-Valle, R. B., and M. G. Genton (2010) Multivariate Extended Skew-t Distributions and Related Families. *Metron*, **68,** 201-234.

Arnold, B. C., and R. J. Beaver (2000) Hidden Truncation Models. *Sankhya, Series A*, **62**, 22–35.

Arnold, B. C., and R. J. Beaver (2002) Skewed Multivariate Models Related to Hidden Truncation and/or Selective Reporting (with discussion). *Test*, **11**, 7–54.

Arnold, B. C., and R. J. Beaver (2004) Elliptical Models Subject to Hidden Truncation or Selective Sampling. In Genton, M. G., editor, *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Boca Raton, Chapman and Hall, 101–112.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999) Coherent Measures of Risk. *Mathematical Finance*, **9**, 203–228.

Azzalini, A. (1985) A Class of Distributions Which Includes the Normal Ones. *Scandinavian Journal of Statistics*, **12**, 171–178.

Azzalini, A. (2005). The Skew-Normal Distribution and Related Multivariate Families (with Discussion by Marc G. Genton and a Rejoinder by the Author). *Scandinavian Journal of Statistics*, **32**, 159–200

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

Azzalini, A., and A. Capitanio (1999) Statistical Applications of the Multivariate Skew-normal Distribution. *Journal of the Royal Statistical Society Series B*, **61**, 579–602.

Azzalini, A., and A. Capitanio (2003) Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t Distribution. *Journal of the Royal Statistical Society Series B,* **65**, 367–389.

Azzalini, A., and M. G. Genton (2008) Robust Likelihood Methods Based on the Skew-t and Related Distributions. *International Statistical Review,* **76**, 106–129.

Azzalini, A., M. G. Genton, and B. Scarpa (2010) Invariance-Based Estimating Equations for Skew-Symmetric Distributions, *Metron* , **68,** 275–298.

Azzalini, A., T. Dal-Cappello and S. Kotz (2003) Log-skew-normal and log-skew-t distributions as models for family income data. *J. Income Distribution*, **11**, no. 3-4, 12-20.

Badrinath S. G., and S. Chatterjee (1988) On Measuring Skewness and Elongation in Common Stock Return Distributions: The Case of the Market Index. *Journal of Business*, **61**, 451–472.

Bauwens, L., and S. Laurent (2005) A New Class of Multivariate Skew Densities, with Application to Generalized Autoregressive Conditional Heteroscedasticity Models. *Journal of Business & Economic Statistics*, **23**, 346–354.

Beedles, W. L. (1979) On the Asymmetry of Market Returns. *Journal of Financial & Quantitative Analysis*, **14**, 653–660.

Bekaert G., C. R. Harvey, C. B. Erb, and T. E. Viskantam (1998) Distributional Characteristics of Emerging Market Returns & Asset Allocation. *Journal of Portfolio Management*, **24**, 102–116.

Benedetti, S. M. (2004) Hedge Fund Portfolio Selection with Higher Moments, MASc Thesis, University of Zurich.

Bolance, C., M. Guillen, E. Pelican, and R. Vernic (2008) Skewed Bivariate Models and Nonparametric Estimation for the TCE Risk Measure. *Insurance: Mathematics and Economics*, **43**, 386–393.

Bortot, P. (2010) Tail Dependence in Bivariate Skew-Normal and Skew-t Distributions, Working Paper.

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

Briec, W., K. Kerstens, and O. Jokung (2007) Mean-Variance-Skewness Portfolio Performance Gauging: A General Shortage Function and Dual Approach. *Management Science*, **53**, 135–149.

Brito, M., and A. Freitas (2008) Edgeworth Expansion for an Estimator of the Adjustment Coefficient. *Insurance: Mathematics and Economics*, **43**, 203–208.

Brunnermeier, M. K., C. Gollier, and J. A. Parker (2007) Optimal Beliefs, Asset Prices, and the Preference for Skewed Returns. *NBER Working Papers 12940*, National Bureau of Economic Research, Inc.

Capitanio, A., A. Azzalini, and E. Stanghellini (2003) Graphical Models for Skew-Normal Variates. *Scandinavian Journal of Statistics,* **30**, 129–144.

Carmichael, B. and A. Coen (2011) Asset Pricing with Skewed Returns, Université Laval Working Paper.

Chaubey, Y., J. Garrido, and S. Trudeau (1998) On the Computation of Aggregate Claims Distributions: Some New Approximations. *Insurance: Mathematics and Economics*, **23**, 215–230.

Chiragiev, A., and Z. Landsman (2007) Multivariate Pareto Portfolios: TCE-Based Capital Allocation and Divided Differences. *Scandinavian Actuarial Journal*, **4**, 261–280.

Chung, Y.P., H. Johnson, and M. Schill (2006) Asset Pricing When Returns Are Nonnormal: Fama-French Factors Versus Higher-Order Systematic Comoments. *Journal of Business*, **79**, 923–940.

Chunhachinda, P., K. Dandapani, S. Hamid, and A. J. Prakash (1997) Portfolio Selection and Skewness: Evidence from International Stock Market. *Journal of Banking & Finance*, **21**, 641–660.

Copas, J. B., and H. G. Li (1997) Inference for Non-Random Samples (with Discussion). *Journal of the Royal Statistical Society Series B,* **59**, 55–95.

Corns, T. R. A., and S. E. Satchell (2007) Skew Brownian Motion and Pricing European Options. *European Journal of Finance*¸**13**, 523–544.

Crocetta, C., and N. Loperfido (2005) The Exact Sampling Distribution of L-Statistics. *Metron*, **63**, 213-223.

Crocetta, C., and N. Loperfido (2009) Maximum Likelihood Estimation of Correlation Between Maximal Oxygen Consumption and the Six-Minute Walk Test in Patients with Chronic Heart Failure. *Journal of Applied Statistics*, **36**, 1101–1108.

Cummins, J. D., and D. J. Nye (1980) The Stochastic Characteristics of Property-Liability Insurance Company Underwriting Profits, *Journal of Risk and Insurance*, 47(1), 61–77.

Cummins, J. D., and J. Lamm-Tennant, (1994) Capital Structure and the Cost of Equity Capital in the Property-liability Insurance Industry, *Insurance: Mathematics and Economics*, 15(2–3), 187–201.

Cummins, J. D., and L. J. Wiltbank (1983) Estimating the Total Claims Distribution Using Multivariate Frequency and Severity Distributions. *Journal of Risk and Insurance* 50, 377–403.

Cummins, J. D., and S. E. Harrington (1988) The Relationship between Risk and Return: Evidence for Property-Liability Insurance Stocks, *Journal of Risk and Insurance*, 55(1), 15–31.

Cummins, J.D., Dionne, G., McDonald, J.B., and Pritchett, B.M., 1990, Applications of the GB2 family of distributions in modeling insurance loss processes, *Insurance Mathematics & Economics,* 9(4), 257-272.

Cummins, J.D., Lin, Y., and Phillips, R.D., 2009, Capital Allocation and the pricing of financially intermediated risks: An empirical investigation, Working Paper, Georgia State University, Atlanta, GA.

D'Arcy, S. P., and N. A. Doherty (1988) *The Financial Theory of Pricing Property-Liability Insurance Contracts*, Heubner Foundation.

de Athayde, G. M., and R. G. Flôres Jr. (2004) Finding a Maximum Skewness Portfolio—A General Solution to Three-Moments Portfolio Choice. *Journal of Economic Dynamics & Control*, **28**, 1335–1352.

de Helguero, F. (1908) Sulla Rappresentazione Analitica Delle Statistiche Abnormali. *Atti Del IV Congresso Internazionale dei Matimatici III*, 288—299.

De Luca, G., and N. Loperfido (2004) A Skew-in-Mean Garch Model. In Genton, M. G., editor, *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Boca Raton, Chapman and Hall, 205–222.

De Luca, G., M. G. Genton, and N. Loperfido (2006) A Multivariate Skew-Garch Model. In Terrell, D., editor, *Econometric Analysis of Economic and Financial Time Series (Advances in Econometrics, Volume 20),* Elsevier, 33–57.

DeFusco, R. A., G. V. Karels, and K. Muralidhar (1996) Skewness Persistence in US Common Stock Return: Results from Bootstrapping Tests. *Journal of Business Banking & Accounting*, **23**, 1183–1195.

Dhaene, J., L. Henrard, Z. Landsman, A. Vandendorpe, and S. Vanduffel (2008) Some Results on the TCE-Based Capital Allocation Rule. *Insurance: Mathematics and Economics*, **42**, 855–863.

Domínguez-Molina, J. A., G. González-Farías, and R. Ramos-Quiroga (2004) Skew-Normality in Stochastic Frontier Analysis. In Genton, M. G., editor, *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Boca Raton, Chapman and Hall, 223–242.

Eling, M. (2008) Performance Measurement in the Investment Industry. *Financial Analysts Journal,* **64**(3), 54–66.

Eling, M. (2012) Fitting Insurance Claims to Skewed Distributions: Are the Skew-Normal and Skew-Student Good Models? *Insurance: Mathematics and Economics*, **51**(2), 235–248.

Eling, M., N. Gatzert, and H. Schmeiser (2009) Minimum Standards for Investment Performance: A New Perspective on Solvency Assessment. *Insurance: Mathematics and Economics*, **45**(1), 113–122.

Embrechts, P., A. McNeil, and D. Straumann (2002) Correlation and Dependence in Risk Management: Properties and Pitfalls. In Dempster, M. A. H., editor, *Risk Management: Value at Risk and Beyond*, Cambridge University Press, 176–223.

Engle. R. F. (1982) Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation, *Econometrica,* **50**, 987-1008.

Fama, E. (1970) Efficient Capital Markets: A Review of Theory and Empirical Work, *Journal of Finance,* **25**, 383-417.

Fama, E. F. (1965) The Behavior of Stock-Market Prices. *Journal of Business*, **38**, 34–105.

Fogler, H. R., and R. C. Radcliffe (1974) A Note on Measurement of Skewness. *Journal of Financial and Quantitative Analysis*, **9**, 485–489.

Franceschini, C., and N. Loperfido (2010) A Skewed GARCH-Type Model for Multivariate Financial Time Series. In Corazza, M. and C. Pizzi, editors, *Mathematical and Statistical Methods for Actuarial Sciences and Finance,* XII, 143–152.

Friedman, M., and L. J. Savage (1948) The Utility Analysis of Choices Involving Risk. *Journal of Political Economy,* **56**, 379–304.

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

Froot, K. A. (2007) Risk Management, Capital Budgeting, and Capital Structure Policy for Insurers and Reinsurers. *Journal of Risk and Insurance*, **74**, 273–299.

Froot, K. A., and J. C. Stein (1998) Risk Management, Capital Budgeting, Capital Structure Policy for Financial Institutions: An Integrated Approach. *Journal of Financial Economics*, **47**, 55–82.

Gonzàlez-Farías, G. Dominguez-Molína, A, Gupta, A. K. (2004) Additive Properties of skew normal random vectors. *Journal of Statistical Planning and Inference,* **126**, 521-534.

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

Harvey, C. R., J. C. Liechty, M. W. Liechty, and P. Müller (2010) Portfolio Selection with Higher Moments. *Quantitative Finance*, **10**, 469–485.

Heckman, J. J. (1979) Sample Selection Bias as a Specification Error. *Econometrica*, **47**, 153–161.

Heyde, C. C., S. Kou, and X. H. Peng (2007) What Is a Good Risk Measure: Bridging the Gaps Between Data, Coherent Risk Measures, and Insurance Risk Measures. *Financial Engineering Report 2007-09*, Center of Financial Engineering, Columbia University.

Hlawitschka, W. (1994) Empirical Nature of Taylor Series Approximations to Expected Utility. *American Economic Review*, **84** , 713–719.

Johnson, N., and S. Kotz (1970) *Continuous Univariate Distributions* 1, Boston, Wiley.

Jondeau, E., and M. Rockinger (2003) Conditional Volatility, Skewness, and Kurtosis: Existence, Persistence, and Comovements*. Journal of Economic Dynamics and Control*, **27**, 1699–1737.

Jondeau, E., and M. Rockinger (2006a) Optimal Portfolio Allocation Under Higher Moments. *European Financial Management*, **12**, 29–55.

Jondeau, E., and M. Rockinger (2006b) The Impact of Shocks on Higher Moments, *Journal of Financial Econometrics*, **7**, 77–105.

Jondeau, E., and M. Rockinger (2012) On the Importance of Time-Variability in Higher Moments for Asset Allocation. *Journal of Financial Econometrics*, **10**(1), 84–123.

Jurczenko, E., and B. Maillet (2006) *Multi-Moment Asset Allocation and Pricing Models,* Chichester, John Wiley & Sons.

Kaas, R., M. Goovaerts, J. Dhaene, and M. Denuit (2009) *Modern Actuarial Risk Theory,* Berlin/Heidelberg, Springer.

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

Kon, S. J. (1984) Models of Stock Returns—A Comparison. *Journal of Finance,* **39**, 147–165.

Kozik, T. J., and A. M. Larson (2001) The n-Moment Insurance CAPM*. Proceedings of the Casualty Actuarial Society,* **87**, 168–169.

Kraus, A., and R. H. Litzenberger (1976) Skewness Preference and the Valuation of Risk Assets. *Journal of Finance*, **31,** 1085–1100.

Kumbhakar, S. C., S. Ghosh, and J. T. McGuckin (1991) A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in US Dairy Farms. *Journal of Business and Economic and Statistics*, **9**, 279–286.

Lai, T. Y. (1991) Portfolio Selection with Skewness: A Multiple-Objective Approach. *Review of Quantitative Finance and Accounting*, **1**, 293–305.

Landsman, Z., and E. A. Valdez (2005) Tail Conditional Expectation for Exponential Dispersion Models. *Astin Bulletin*, **35**, 189–209.

Lane, M. N. (2000) Pricing Risk Transfer Transactions. *ASTIN Bulletin*, **30**, 259–293.

Lau, H.-S., and J. R. Wingender (1989) The Analytics of the Intervaling Effect on Skewness and Kurtosis of Stock Returns. *Financial Review*, **24**, 215–233.

Lau, H.-S., J. R. Wingender, and A. Lau (1989) On Estimating Skewness in Stock Returns. *Management Science*, **35**, 1139–1142.

Leland, H. E. (1999) Beyond Mean-Variance: Performance Measurement in a Non-Symmetrical World. *Financial Analysts Journal*, **55**, 27–36.

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

Lien, D., and K. Shrestha (2010) Estimating Optimal Hedge Ratio: A Multivariate Skew-Normal Distribution Approach. *Applied Financial Economics,* **20**, 626–636.

Lim, K.-G. (1989) A New Test of the Three-Moment Capital Asset Pricing Model. *Journal of Financial & Quantitative Analysis*, **24**, 205–216.

Liseo, B. and N. Loperfido (2003) A Bayesian Interpretation of the Multivariate Skew-Normal Distribution. *Statistics and Probability Letters*, **61**, 395-401.

Liu, S., S. Y. Wang, and W. Qiu (2003) Mean-Variance-Skewness Model for Portfolio Selection with Transaction Costs. *International Journal of Systems Science*, **34**, 255–262.

Loistl, O. (1976) The Erroneous Approximation of Expected Utility by Means of a Taylor's Series Expansion: Analytic and Computational Results. *American Economic Review*, **66**, 904–910.

Loperfido, N. (2004). Generalized Skew-Normal Distributions In Genton, M. G., editor, *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Boca Raton, Chapman and Hall, 65-80.

Loperfido, N. (2008) Modeling Maxima of Longitudinal Contralateral Observations. *Test*, **17**, 370–380.

Loperfido, N. (2010) Canonical Transformations of Skew-Normal Variates. *Test*, **19**, 146–165.

Loperfido, N., and P. Guttorp (2008) Network Bias in Air Quality Monitoring Design *Environmetrics*, **19**, 661–671.

Loperfido, N., J. Navarro, J. M. Ruiz, and C. J. Sandoval (2007) Some Relationships Between Skew-Normal Distributions and Order Statistics from Exchangeable Normal Random Vectors. *Communications in Statistics—Theory and Methods,* **36**, 1719–1733.

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

Ma, Y., and A. A. Tsiatis (2006) On Closed Form Semiparametric Estimators for Measurement Error Models. *Statistica Sinica,* **16** 183–193.

Ma, Y., M. G. Genton, and A. A. Tsiatis (2005) Locally Efficient Semiparametric Estimators for Generalized Skew-Elliptical Distributions. *Journal of the American Statistical Association.* **100,** 980–989.

Madan, D. B., P. P. Carr, and E. C. Chang (1998) The Variance Gamma Process and Option Pricing. *European Finance Review,* **2,** 79–105.

Mardia, K. V. (1970) Measures of Multivariate Skewness and Kurtosis. *Biometrika*, **57**, 519–530.

Mardia, K. V. (1975) Assessment of Multinormality and the Robustness of Hotelling's T-Squared Test. *Applied Statistics*, **24**, 163–171.

McNeil, A. J. (1997) Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory. *ASTIN Bulletin*, **27**, 117–137.

Meucci, A. (2006) Beyond Black-Litterman: Views on Non-Normal Markets. *Risk Magazine*, **19**, 87–92.

Mikosch, T. (2009) *Non-Life Insurance Mathematics*, Heidelberg, Springer.

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

Muralidhar, K. (1993) The Bootstrap Approach for Testing Skewness Persistence. *Management Science*, **39**, 487–491.

O'Hagan, A., and T. Leonard (1976) Bayes Estimation Subject to Uncertainty About Parameter Constraints. *Biometrika*, **63**, 201–202.

Panjer, H. H. (2002) Measurement of Risk, Solvency Requirements and Allocation of Capital Within Financial Conglomerates, *Proceedings of the 27th International Congress of Actuaries,* Cancun.

Peirò, A. (1999) Skewness in Financial Returns. *Journal of Banking & Finance*, **23**, 847–862.

Pewsey, A. (2000). Problems of Inference for Azzalini's Skew-Normal Distribution. *Journal of Applied Statistics*, **27**, 859–870.

Pfeifer, D., and D. Strassburger (2008) Solvency II: Stability Problems with the SCR Aggregation Formula. *Scandinavian Actuarial Journal*, **5**, 61–77.

Post, T., H. Levy, and P. van Vliet (2008) Risk Aversion and Skewness Preference: A Comment. *Journal of Banking and Finance*, **32**, 1178–1187.

Racine, M. D. (1998) Hedging Volatility Shocks to the Canadian Investment Opportunity Set. *Quarterly Journal of Business and Economics*, **37**, 59–79.

Roberts, C. (1966) A Correlation Model Useful In The Study of Twins, *Journal of the American Statistical Association*, **61**, 1184-1190.

Roch, O., and E. A. Valdez (2011) Lower Convex Order Bound Approximations for Sums of Log-Skew Normal Random Variables. *Applied Stochastic Models in Business and Industry*, **27**, 487–502.

Rotnitzky, A., D. R. Cox, M. Bottai, and J. Robins (2000) Likelihood Based Inference with Singular Information Matrix. *Bernoulli*, **6**, 243–284.

Samuelson, P. A. (1970) The Fundamental Application Theorem of Portfolio Analysis in Terms of Means, Variance and Higher Moments. *Review of Economic Studies*, **37**, 537–542.

Sharples, J. J., and J. C. V. Pezzey (2007) Expectations of Linear Functions with Respect to Truncated Multinormal Distributions with Applications for Uncertainty Analysis in Environmental Modelling. *Environmental Modelling & Software*, **22**, 915–923.

Shiraishi, H., and M. Taniguchi (2009) Statistical Estimation of Optimal Portfolios depending on Higher Order Cumulants. *Annales de I'I.S.U.P. Publication de l'Institut de Statistique de l' Université de Paris*, **53**, 3–18.

Siegel, A. F. (1993) A Surprising Covariance Involving the Minimum of Multivariate Normal Variables. *Journal of the American Statistical Association*, **88**, 77–80.

Simaan, Y. (1987) Portfolio Selection and Capital Asset Pricing for a Class of Non-Spherical Distributions of Assets Returns, PhD Thesis, City University of New York.

Simaan, Y. (1993) Portfolio Selection and Asset Pricing-Three Parameter Framework. *Management Science*, **39**, 568–587.
Simkowitz, M. A., and W. L. Beedles (1978) Diversification in a Three Moment World. *Journal of Financial & Quantitative Analysis*, **13,** 927–941.
Simkowitz, M. A., and W. L. Beedles (1980) Asymmetric Stable Distributed Security Returns. *Journal of the American Statistical Association*, **75**, 306–312.
Singleton, J. C., and J. Wingender (1986) Skewness Persistence in Common Stock Returns. *Journal of Financial & Quantitative Analysis*, **21**, 335–341.
Sun, Q., and Y. Yan (2003) Skewness Persistence with Optimal Portfolio Selection. *Journal of Banking & Finance*, **27**, 111–1121.
Sundt, B. (1999) On Multivariate Panjer Recursions. *ASTIN Bulletin*, **29**, 29–45.
Sundt, B. (2002) Recursive Evaluation of Aggregate Claims Distributions. *Insurance: Mathematics and Economics*, **30**, 297–322.
Tjetjep, A., and E. Seneta, E. (2006) Skewed Normal Variance-Mean Models for Asset Pricing and the Method of Moments. *International Statistical Review*, **74**, 109–126.
Treynor, J. L., and K. Mazuy (1966) Can Mutual Funds Outguess the Market? *Harvard Business Review*, **44**, 131–136.
Valdez, E. A., and A. Chernih (2003) Wang's Capital Allocation Formula for Elliptically Contoured Distributions. *Insurance: Mathematics and Economics*, **33**, 517–532.
van Haastrecht, A., R. Plat, and A. Pellser (2010) Valuation of Guaranteed Annuity Options Using a Stochastic Volatility Model for Equity Prices. *Insurance: Mathematics and Economics*, **47**, 266–277.
Vernic, R. (2006) Multivariate Skew-Normal Distributions with Applications in Insurance. *Insurance: Mathematics and Economics*, **38**, 413–426.
von Neumann, J., and O. Morgenstern (1944) *Theory of Games and Economic Behaviour,* Princeton, Princeton University Press.
Wang, S. (2002) A Set of New Methods and Tools for Enterprise Risk Capital Management and Portfolio Optimization. *Casualty Actuarial Society Forum*, Summer, 43–78.
Wen, M. M., A. D. Martin, G. Lai, and T. J. O'Brien (2008) Estimating the Cost of Equity for Property-Liability Insurance Companies, *Journal of Risk and Insurance*, **75**, 101–124.
Wongwachara, W. (2009) Optimal Four-Moment Portfolios: Effects of Higher Moments and Estimation Errors, University of Cambridge Working Paper.

Table 1: Moments of the MESN Distribution Under Both Representations

Moment	MESN	$MESN^{AS}$
Expected value	$\boldsymbol{\eta} + \boldsymbol{\Psi}\xi_1(\tau) \big/ \sqrt{1 + \boldsymbol{\Psi}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\Psi}}$	$\boldsymbol{\zeta} + \boldsymbol{\delta}\tau + \boldsymbol{\delta}\xi_1(\tau)$
Covariance matrix	$\boldsymbol{\Omega} + \boldsymbol{\Psi}\boldsymbol{\Psi}^T \xi_2(\tau) \big/ \left(1 + \boldsymbol{\Psi}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\Psi}\right)$	$\boldsymbol{\Sigma} + \boldsymbol{\delta}\boldsymbol{\delta}^T \{1 + \xi_2(\boldsymbol{\tau})\}$
Co-skewness	$\psi_i \psi_j \psi_k \xi_3(\tau) \big/ \left(1 + \boldsymbol{\Psi}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\Psi}\right)^{3/2}$	$\delta_i \delta_j \delta_k \xi_3(\tau)$
Co-kurtosis	$\psi_i \psi_j \psi_k \psi_l \xi_4(\tau) \big/ \left(1 + \boldsymbol{\Psi}^T \boldsymbol{\Omega}^{-1} \boldsymbol{\Psi}\right)^2$	$\delta_i \delta_j \delta_k \delta_l \xi_4(\tau)$

Figure 1: Numerical results for TCE for the normal, the skew-normal, and the skew-Student