European Journal of Operational Research 234 (2014) 392–401

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Mean–variance–skewness efficient surfaces, Stein's lemma and the multivariate extended skew-Student distribution

C.J. Adcock *

The University of Sheffield, Sheffield University Management School, 9, Mappin Street, Sheffield S1 4DT, UK

ARTICLE INFO

Article history:
Received 5 October 2012
Accepted 10 July 2013
Available online 20 July 2013

Keywords:
Finance
Multivariate statistics
Utility theory

ABSTRACT

Recent advances in Stein's lemma imply that under elliptically symmetric distributions all rational investors will select a portfolio which lies on Markowitz' mean–variance efficient frontier. This paper describes extensions to Stein's lemma for the case when a random vector has the multivariate extended skew-Student distribution. Under this distribution, rational investors will select a portfolio which lies on a single mean–variance–skewness efficient hyper-surface. The same hyper-surface arises under a broad class of models in which returns are defined by the convolution of a multivariate elliptically symmetric distribution and a multivariate distribution of non-negative random variables. Efficient portfolios on the efficient surface may be computed using quadratic programming.

1. Introduction

If $\mathbf{X}$ is a random vector which has a full rank multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $h(\cdot)$ is a scalar valued function of $\mathbf{X}$ which satisfies certain regularity conditions, then $\mathrm{cov}\{\mathbf{X}, h(\mathbf{X})\} = \boldsymbol{\Sigma}E\{\nabla h(X)\}$, where $\nabla h(\mathbf{X})$ is the vector first derivatives of $h(\cdot)$ with respect to the elements of $\mathbf{X}$. This result is the multivariate generalisation of Stein's lemma, Stein (1973, 1981) and is reported in Liu (1994). Liu also points out that similar results hold for other distributions and gives some examples. In particular, he provides an expression for $\mathrm{cov}\{X, h(X)\}$ for the case where the scalar random variable X has Student's t distribution. Landsman (2006) and Landsman and Nešlehová (2008) extend these results, with the latter paper providing an extension of Stein's lemma for multivariate elliptically symmetric distributions. (For general background see Fang, Kotz, & Ng, 1990.) This is an important class of distributions for portfolio theory. If the multivariate distribution of asset returns is member of the elliptically symmetric class, then the distributions of returns on a portfolio of the assets, which is an affine transformation of the vector of returns, is a member of the same class. Of arguably greater importance is the implication of Landsman and Nešlehova's extension of Stein's lemma. For portfolio selection $h(\mathbf{X}) = U(\boldsymbol{w}^T\mathbf{X})$, where $U(\cdot)$ is a utility function and w is the vector of portfolio weights. Under elliptical symmetry, and subject only to regularity conditions, the portfolios of expected utility maximisers will be located on Markowitz' mean–variance efficient frontier. Since Markowitz' original paper, mean–variance portfolio selection has been the subject of great interest and many articles, books and monographs. The method has been the subject of both praise and criticism. Nonetheless, the fact that the efficient frontier arises under conditions which are far more general than quadratic utility or normally distributed returns is one of many the reasons for its longevity and a tribute to the robustness of the original theory.

As well as non-normality, almost always in the form of fat tails, it has long been accepted that returns on some financial assets are not symmetrically distributed. Indeed, there is a long literature about skewness in asset returns. This dates back at least to the foundation papers of Samuelson (1970), Arditti and Levy (1975) and Kraus and Litzenberger (1976). In addition to these and other theoretical papers, there are numerous articles which report empirical studies of skewness. There are also numerous papers which are concerned with portfolio selection in the presence of skewness. Well known works include, but are not limited to, papers by Chunhachinda, Dandapani, Hamid, and Prakash (1997), Sun and Yan (2003), de Athayde and Flôres (2004), Briec, Kerstens, and Jokung (2007), Li, Qin, and Kar (2010), Goh, Lim, Sim, and Zhang (2012) and Matmoura and Penev (2013). Many of these have been summarised recently in the review paper by Adcock, Eling, and Loperfido (2012). Some papers (Jondeau & Rockinger, 2006, for example) employ Taylor series methods to justify the inclusion of a cubic term in an approximation to a utility function. It is natural therefore to enquire to what extent there is a mean–variance–skewness extension of Markowitz' efficient frontier; that is, a single mean–variance–skewness surface on which the portfolios of all expected utility maximisers will be located.

* Tel.: +44 (0)114 222 3402.
E-mail address: c.j.adcock@shef.ac.uk

0377-2217/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2013.07.011

The view taken in this paper is that the starting point for the development of portfolio selection theory is a coherent multivariate probability distribution for asset returns. In principle, this allows the computation of expected utility and thus makes explicit the nature of the relationship between a set of efficient portfolio weights, the parameters of the distribution and the utility function itself. For symmetrically distributed returns, members of the elliptically symmetric class are attractive because they are tractable with respect to the requirements of efficient portfolio selection. The same requirements imply that skewed multivariate distributions should also be selected for portfolio selection, at least in part, because of their tractability. A well-known skewed multivariate distribution which meets the requirement of tractability is the multivariate skew-normal distribution (MSN henceforth). This was introduced in its original form by Azzalini and Dalla-Valle (1996). The derivation considers a random vector of length $(n + 1)$, $(\mathbf{U}^T, V)$, $\mathbf{U}$ an n-vector and V a scalar, which has a multivariate normal distribution with mean vector $(\boldsymbol{\mu}^T,0)$ and full rank covariance matrix which is arbitrary except for the diagonal entry corresponding to V, which is unity. The multivariate skew-normal arises as the conditional distribution of $\mathbf{U}$ given that $V > 0$. The MSN distribution was also reported as a conditional distribution in Azzalini and Capitanio (1999).

A modified version of this distribution, which has an additional parameter, was reported by Arnold and Beaver (2000). For this version of the distribution, before truncation the mean of the scalar variable V is arbitrary. This distribution was also reported independently in Adcock and Shutes (2001), who were the first to employ it in finance. This modification, which is generally known as the multivariate extended skew-normal (MESN) distribution, is attractive for applications in finance. The additional parameter offers more flexibility in modelling higher moments than the MSN. Explicit formulae for the moments are in Adcock and Shutes (2001).

From the perspective of portfolio selection the MESN distribution also admits an extension to Stein's lemma. Adcock (2007) shows that under the MESN distribution there is a single mean–variance–skewness efficient surface. As is well known, the derivation of both the MSN and MESN distributions employs a single unobserved or hidden variable which is non-negative. From a finance perspective, this implies that there is a single source of asymmetry in returns; a one factor model for skewness. A more complex skew-normal model in which there is more than one hidden non-negative variable is described in Sahu, Dey, and Branco (2003), Gonzàlez-Farías, Dominguez-Molína, and Gupta (2004) and Arellano-Valle and Azzalini (2006). This is generally known as the closed skew-normal (CSN) distribution. Adcock (2007) describes an extended version of this distribution, denoted CESN, and shows that there is an extension of Stein's lemma for it. Under this distribution, skewness is generated by multiple factors and there is a single mean–variance–skewness hyper-surface for expected utility maximisers, with a dimension for each skewness factor. The CSN distribution is employed in portfolio selection by Harvey, Liechty, Liechty, and Müller (2010), who extend Sahu et al.'s (2003) model and thus deal with co-skewness between assets.

Both the MESN and CESN distributions are tractable from the perspective of portfolio theory. However, even though the additional parameter(s) give flexibility in modelling moments, it is debatable whether these distributions will always adequately deal with the fat tails effects observed in empirical data. The first objective of this paper, therefore, is to present results for portfolio selection based on multivariate skew-Student-t distributions. These are increasingly well-known extensions of the multivariate skew-normal. Initial papers by Branco and Dey (2001) and Azzalini and Capitanio (2003) have been followed by articles by several authors including Azzalini and Genton (2008) and Arellano-Valle, Branco, and Genton (2006). A multivariate extended skew-Student-t, MEST, distribution and its properties are described in Adcock (2002, 2010) and Arellano-Valle and Genton (2010). Like its skew-normal counterpart, the MEST distribution employs one truncated variable.

The notable paper by Sahu et al. (2003), already referred to above, presents a general multivariate skew-elliptical distribution. This assumes that a $2n$-vector $(\mathbf{U}^T, \mathbf{V}^T)$ has a multivariate elliptically symmetric distribution and then conditions on all elements of $\mathbf{U}$ being positive. The expected values of the elements of $\mathbf{U}$ are assumed to be equal to zero. Their general results are exemplified using the skew-normal and skew-Student cases. Arellano-Valle and Genton (2010) also describe a version of the skew-Student t distribution, CEST henceforth, which has more than one truncated variable. In this case, the expected values of the elements of $\mathbf{U}$ are not restricted. In their paper, the CEST distribution extends related earlier work by Arellano-Valle and Azzalini (2006).

The second objective of this paper is to present corresponding results for multivariate models of asset returns which are members of the class defined by Simaan (1987, 1993). He proposes that the n-vector of returns on financial assets should be represented as $\mathbf{X} = \mathbf{U} + \lambda V$. The n-vector $\mathbf{U}$ has a multivariate elliptically symmetric distribution and is independent of the non-negative univariate random variable V, which has an unspecified skewed distribution. The n-vector λ, whose elements may take any real value, induces skewness in the return of individual assets. Adcock and Shutes (2012) describe multivariate versions of the normal-exponential and normal-gamma distributions. These are specific cases of the model proposed by Simaan (1987, 1993), which have not appeared explicitly in the literature before.

The paper presents extensions to Stein's lemma for both the MEST and CEST distributions. This work extends Adcock (2007), which presents versions of Stein's lemma for the MESN and CESN distributions. Similar to the results for these skew-normal distributions, a consequence of the single truncated variable employed in the MEST distribution is that the lemma offers results which are not complicated to compute for a given function $h(\cdot)$. As is shown below, the extension to Stein's lemma for the MEST distribution leads to a single mean–variance–skewness surface for expected utility maximisers. The lemma also leads to insights into the nature of skewness preference under this distribution. For the CEST distribution, the extension to the lemma leads to a single mean–variance–skewness hyper-surface. For general applications, this particular extension to the lemma is mainly of theoretical interest. This is because of the necessity to compute integrals of multivariate Student-t distributions and because parameter estimation will not be a trivial task. However once parameter estimates are available, it is shown in Section 7 that portfolio selection may be performed using quadratic programming. Notwithstanding its complexities, and as discussed in Section 5, a case may be made for a useful role for the CEST distribution in finance. Accordingly use of this distribution is a topic for future research. The paper then presents some analogous results for Simaan-type models. Given the more general nature of the models that Simaan proposes, these results are less elegant mathematically than those for the MEST and CEST distributions. Nonetheless, they show that, even under more general conditions, there is a single mean–variance–skewness surface for one factor models. When the models are extended to include more than one non-negative variable, the single mean–variance–skewness hyper-surface arises. Thus, this papers shows that there is a comprehensive set of multivariate probability distributions which incorporate asymmetry and which lead to straightforward extensions to Markowitz' efficient frontier. It is shown in Section 7 that portfolios on these surfaces may also be determined by quadratic programming.

The structure of the paper is as follows. Section 2 summarizes the MEST and CEST distributions. A number of basic properties of the CEST distribution and specific results that are required in the rest of the paper are summarised in Appendices A and B. Many of these results may also be found in a different notation in Arellano-Valle and Genton (2010). Section 3 contains two lemmas which show that the MEST and CEST distributions may be expressed as scale mixtures of the MESN and CESN distributions, respectively. These results, which are believed to be new, are used in the subsequent discussion of skewness preference referred to above. Section 4 presents the main result of the paper, Stein's lemma for the CEST distribution. This section also contains several corollaries to the result, including Stein's lemma for the MEST distribution. Section 5 contains a discussion of portfolio selection and skewness preference under the MEST and CEST distributions. Section 6 presents the analogous results for the models of Simaan (1987, 1993). Section 7 describes how portfolio selection may be performed using quadratic programming and presents an example. Section 8 concludes. Technical results and proofs are in appendices at the end of the paper. The notations $\mathbf{0}_m$ and $\mathbf{0}_{mn}$ denote respectively an m-vector and $m \times n$ matrix of zeros. For m-vectors $\boldsymbol{v}$ and $\boldsymbol{a}$ the notation $\boldsymbol{v} \geqslant \boldsymbol{a}$ means that $v_i \geqslant a_i$ for $i = 1,\ldots,m$. The standard normal distribution function evaluated at x is $\Phi(x)$. Other notation is either that in standard use or is defined in the paper.

2. Multivariate skew-Student distributions

The multivariate skew-Student-t distribution was introduced by Azzalini and Capitanio (2003) and is an extension to the MSN distribution first described in Azzalini and Dalla-Valle (1996). The standard form may be obtained by considering random vector of length $(n+1)$, $(\mathbf{U}^T, V)$, $\mathbf{U}$ an n-vector and V a scalar, which has a multivariate Student distribution with υ degrees of freedom, location parameter vector $(\boldsymbol{\mu}^T, 0)$ and full rank scale matrix which is arbitrary except for the diagonal entry corresponding to V, which is unity. The multivariate skew-Student distribution is obtained by considering the conditional distribution of $\mathbf{U}$ given $V > 0$. Arellano-Valle and Genton (2010) and Adcock (2002, 2010) present a multivariate extended skew-Student-t distribution by allowing the location parameter for the scalar variable V to be equal to τ. For applications in finance, the MEST distribution in this paper is derived in a somewhat different manner, which facilitates interpretation in finance and, from the perspective of the results that follow in Section 4, simplifies the notation. Let $(\mathbf{U}^T, V)$ have a multivariate Student distribution with υ degrees of freedom, location parameter vector $(\boldsymbol{\mu}^T, \tau)$, full rank scale matrix

$$\begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0}_n \\ \mathbf{0}_n^{\mathrm{T}} & 1 \end{bmatrix},$$

with V being truncated from the left at zero. The random vector $\mathbf{X} = \mathbf{U} + \lambda V$ has the MEST distribution with probability density function

$$f(\mathbf{x}) = t_n(\mathbf{x}; \boldsymbol{\mu} + \lambda\tau, \boldsymbol{\Sigma} + \lambda\lambda^T, \upsilon) \times \frac{T\left\{\sqrt{\frac{\upsilon+n}{\upsilon}} \frac{\tau + (\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}\lambda}{\sqrt{\{1 + Q_1(\mathbf{x}, \boldsymbol{\mu}, \lambda, \tau, \boldsymbol{\Sigma})/\upsilon\}(1 + \lambda^T\boldsymbol{\Sigma}^{-1}\lambda)}}, \upsilon\right\}}{T(\tau, \upsilon)}. \tag{1}$$

The quadratic form $Q_1(\mathbf{x}, \boldsymbol{\mu}, \lambda, \tau, \boldsymbol{\Sigma})$ is defined as

$$Q_1(\mathbf{x}, \boldsymbol{\mu}, \lambda, \tau, \boldsymbol{\Sigma}) = (\mathbf{x} - \boldsymbol{\mu} - \lambda\tau)^T (\Sigma + \lambda\lambda^T)^{-1} (\mathbf{x} - \boldsymbol{\mu} - \lambda\tau).$$

The function $t_n(\mathbf{x}; \varsigma, \boldsymbol{\Omega}, \upsilon)$ denotes the probability density function of an n-variate multivariate Student-t distribution with location parameter vector ς, matrix of symmetry $\boldsymbol{\Omega}$ and υ degrees of freedom evaluated at $\mathbf{x}$ and $T(\tau, \upsilon)$ denotes the distribution function for the (univariate) Student-t distribution with υ degrees of freedom evaluated at τ. The notation $MEST(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda, \tau, \upsilon)$ is used to describe the distribution at (1). The notation $MESN(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda, \tau)$ refers to the corresponding MESN distribution, which is the limiting form of the MEST as υ increases without limit. The notation $MT(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \upsilon)$ describes the multivariate Student-t distribution with location parameter vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and υ degrees of freedom.

The CEST distribution may be derived using the same approach. Let $\mathbf{X}$ now be a random vector of length n, which is defined as $\mathbf{X} = \mathbf{U} + \boldsymbol{\Lambda}\mathbf{V}$, where $(\mathbf{U}^T, \mathbf{V}^T)$ is an $(n+m)$ vector which has a multivariate Student-t distribution with location parameter vector $(\boldsymbol{\mu}^T, \boldsymbol{\tau}^T)$, full rank scale matrix

$$\begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0}_{nm} \\ \mathbf{0}_{nm}^{\mathrm{T}} & \boldsymbol{\Psi} \end{bmatrix},$$

and υ degrees of freedom with each element of $\mathbf{V}$ truncated from the left at zero. The matrix $\boldsymbol{\Lambda}$ is $n \times m$. For estimation purposes, it may be assumed without loss of generality $\boldsymbol{\Psi}$ is a correlation matrix. For the sake of generality in most of the results that follow, this assumption is not made. The joint probability density function of $\mathbf{U}$ and $\mathbf{V}$ is

$$f(\boldsymbol{u}, \boldsymbol{v}) = t_{n,m}(\boldsymbol{u}, \boldsymbol{v}; \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}, \upsilon)/T_m(0; -\boldsymbol{\tau}, \boldsymbol{\Psi}, \upsilon), \tag{2}$$

where

$$t_{n,m}(\boldsymbol{u}, \boldsymbol{v}; \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}, \upsilon) = K\left\{1 + (\boldsymbol{u}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{u}-\boldsymbol{\mu})/\upsilon + (\boldsymbol{v}-\boldsymbol{\tau})^T\boldsymbol{\Psi}^{-1}(\boldsymbol{v}-\boldsymbol{\tau})/\upsilon\right\}^{-(\upsilon+m+n)/2},$$

for $\boldsymbol{v} > \mathbf{0}_m$, K being the normalising constant. The function $T_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Omega}, \upsilon)$ denotes the distribution function of a full-rank p-variate multivariate Student-t distribution with location parameter vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Omega}$ and υ degrees of freedom evaluated at $\boldsymbol{x}$. The probability density function of $\mathbf{X}$ is

$$f(\boldsymbol{x}) = t_n(\boldsymbol{x}; \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\tau}, \boldsymbol{\Sigma} + \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}^T, \upsilon)T_m(\boldsymbol{0}; -\boldsymbol{\omega}, \widehat{\boldsymbol{\Pi}}, \upsilon)/T_m(\boldsymbol{0}; -\boldsymbol{\tau}, \boldsymbol{\Psi}, \upsilon), \tag{3}$$

where

$$\widehat{\boldsymbol{\Pi}} = \upsilon\boldsymbol{\Pi}\{1 + Q_m(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \tau, \boldsymbol{\Sigma}, \boldsymbol{\Psi})/\upsilon\}/(\upsilon + n),$$
$$\boldsymbol{\Pi} = (\boldsymbol{\Psi}^{-1} + \boldsymbol{\Lambda}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda})^{-1}, \quad \boldsymbol{\omega} = \{\boldsymbol{\tau} + \boldsymbol{\Pi}\boldsymbol{\Lambda}^T\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu} - \boldsymbol{\Lambda}\boldsymbol{\tau})\},$$
$$Q_m(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \tau, \boldsymbol{\Sigma}, \boldsymbol{\Psi}) = (\boldsymbol{x} - \boldsymbol{\mu} - \boldsymbol{\Lambda}\boldsymbol{\tau})^T(\boldsymbol{\Sigma} + \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}^T)^{-1}(\boldsymbol{x} - \boldsymbol{\mu} - \boldsymbol{\Lambda}\boldsymbol{\tau}). \tag{4}$$

The notation $CEST(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\tau}, \boldsymbol{\Psi}, \upsilon)$ is used to describe the distribution at (3) Apart from a change of notation, the density function is the same as that in Arellano-Valle and Genton (2010). A number of properties of these distributions, which are required in Section 4, are summarised in Appendices A and B.

3. Representation of the MEST and DEST distributions as scale mixtures

For symmetrically distributed returns, the multivariate Student-t distribution may be represented as a scale mixture of the multivariate normal. For a vector time series of asset returns, this representation may be interpreted as a manifestation of changing volatility and contributes to an explanation of why Student's t distribution is often a better model for returns than the normal. This section of the paper presents two results which show that the MEST and CEST distributions may also be expressed as scale mixtures of the MESN and CESN distributions, respectively.

Scale mixtures of skew-normal distributions are an active area of research. There is a comprehensive summary in the recent paper by Kim and Genton (2011). Arellano-Valle and Genton (2010), show that for $\tau = 0$ the MEST distribution may be represented as

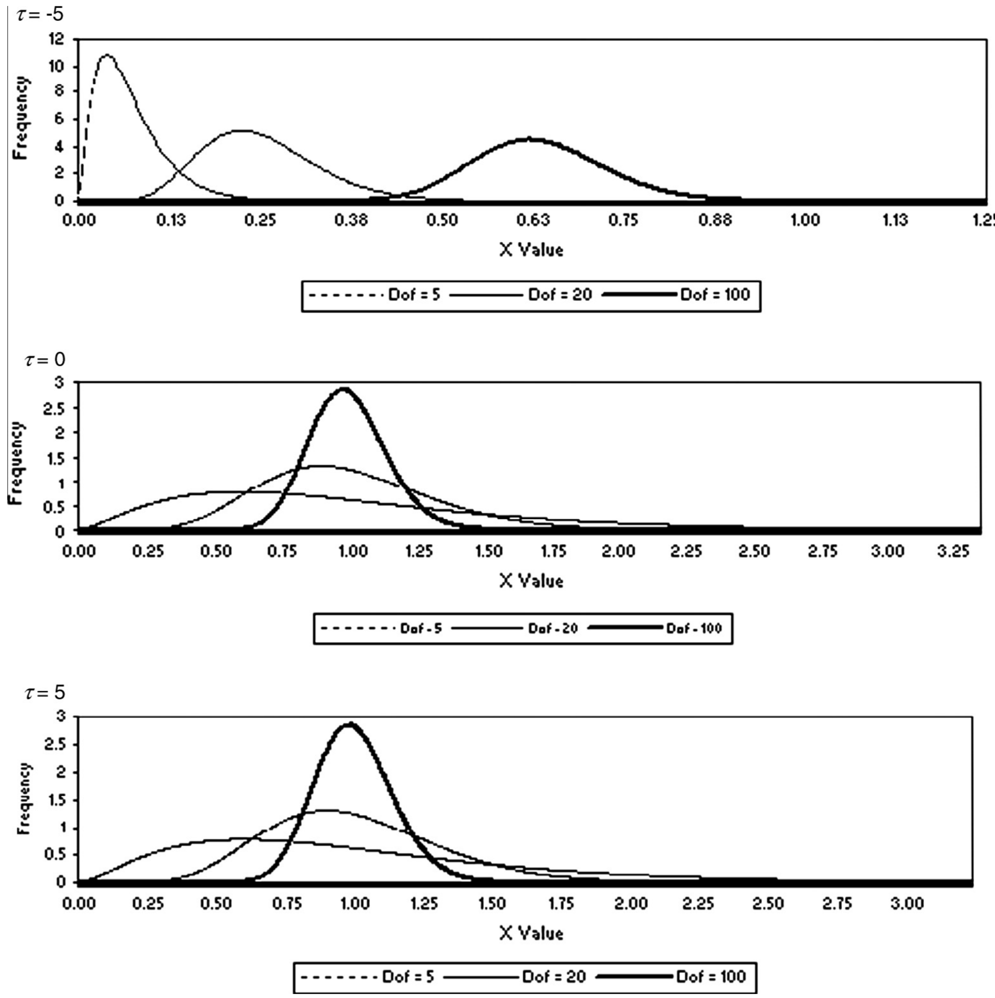

Fig. 1. Sketches of the skew chi-squared density function.

a scale mixture of the MESN distribution. The mixing distribution is proportional to the Chi-squared distribution. The MEST and CEST distributions as parameterised in this paper may be expressed as scale mixture of the MESN and CESN distributions for general values of τ and $\boldsymbol{\tau}$ respectively. In this case, the mixing distributions are generalisations of the Chi-squared distribution.[1]

For the MEST distribution, the following result is a special case of Lemma 1 of Azzalini and Capitanio (2003) and motivates what follows.

Lemma 1 (Azzalini and Capitanio, 2003). *Let Z be distributed as χ^2_υ/υ and let the distribution functions $\Phi(\cdot)$ and $T_\upsilon(\cdot)$ be as defined above. The following holds for any real τ*

$$E\{\Phi(\tau\sqrt{z})\} = T_\upsilon(\tau).$$

This lemma motivates a modified version of the Chi-squared distribution with the density function

$$f(z) = g_{\upsilon/2}(z, 2/\upsilon)\Phi\left(\tau\sqrt{z}\right)/T_\upsilon(\tau), \quad z \geqslant 0;\ \ \upsilon > 0,\ \ -\infty < \tau < \infty, \tag{5}$$

where $g_\upsilon(x, \alpha)$ is the density function of a γ (υ, α) variable, namely

$$g_\upsilon(x, \alpha) = x^{\upsilon-1}e^{-x/\alpha}/\alpha^\upsilon\Gamma(\upsilon), \quad x \geqslant 0;\ \ \upsilon, \alpha > 0.$$

Some sketches of the density function are shown in Fig. 1. As the figure shows, the density becomes symmetric as the degrees of freedom increase. For τ = 0, the density is that of χ^2_υ/υ. The skewing factor $\Phi\left(\tau\sqrt{z}\right)/T_\upsilon(\tau)$ rapidly approaches unity with increasing positive values of τ, thus giving the χ^2_υ/υ density. For negative values of τ, standard asymptotics show that the density becomes progressively more concentrated around values of Z which are close to zero and is infinite at the origin if υ < 3.

Lemma 2, which may be proved by transformation of the variables, shows, the density at (5) is an example of a hidden truncation model.

Lemma 2. *Let Z be distributed as χ^2_υ/υ and let Y be independently distributed as N(0, 1), then the probability density function of Z given that $Y/\sqrt{Z} < \tau$ is given by (5).*

Two basic properties of this distribution are as follows. First, all moments of positive order exist, specifically

$$E(Z^\varpi) = \Gamma(\upsilon/2 + \varpi)(2/\upsilon)^\varpi T_{\upsilon+2\varpi}\left(\tau\sqrt{(\upsilon + 2\varpi)/\upsilon}\right)/\Gamma(\upsilon/2)T_\upsilon(\tau);$$
$$\varpi > -\upsilon/2.$$

Secondly, the moment generating function is

$$M(t) = (1 - 2t/\upsilon)^{-\upsilon/2}T_\upsilon\left(\tau/\sqrt{1 - 2t/\upsilon}\right)/T_\upsilon(\tau), \quad t < \upsilon/2.$$

The scale mixture result is the following. The proof is in Appendix C.

[1] These results were first reported in Adcock (2009).

Lemma 3. *Conditional on $Z \geqslant 0$ which has the density at* Eq. (5), *let* $\mathbf{U}$ *be an n-vector with the full-rank multivariate normal distribution* $N(\mu, \mathbf{\Sigma}/Z)$ *and V be a scalar random variable which is independent of* $\mathbf{U}$ *and which has a normal distribution with mean* τ, *variance* $1/Z$ *and is left truncated at zero. The following result holds: the vector* $\mathbf{X} = \mathbf{U} + \lambda V$ *has the MEST*(μ, $\mathbf{\Sigma}$, λ, τ, υ) *distribution.*

Note that the conditional extended skew-normal distribution may also be represented as $\mathbf{X} = \mathbf{U} + \left(\lambda/\sqrt{Z}\right)W$ *where the distribution of* $\mathbf{U}$ *is as defined for the lemma and W has a normal distribution with mean* $\tau\sqrt{Z}$, *variance 1 and is left truncated at zero.*

For the CEST distribution, the mixing distribution is

$$f(\boldsymbol{x}) = g_{\upsilon/2}(z, 2/\upsilon)\Phi_m\left(\mathbf{0}; -\tau\sqrt{z}, \mathbf{\Psi}\right)/T_m(\mathbf{0}; -\boldsymbol{\tau}, \mathbf{\Psi}, \upsilon), \tag{6}$$

where $\Phi_m(\mathbf{x}; \mu, \mathbf{\Sigma})$ *is the distribution function of an m-variate* $N(\mu, \mathbf{\Sigma})$ *vector evaluated at* $\mathbf{x}$. *This leads to the following, which may be proved in the same way as* Lemma 3.

Lemma 4. *Conditional on $Z \geqslant 0$ which has the density at* Eq. (6), *let* $\mathbf{U}$ *be an n-vector with the full-rank multivariate normal distribution* $N(\mu, \mathbf{\Sigma}/Z)$ *and* $\mathbf{V}$ *be an m-vector which has the full-rank multivariate normal distribution* $N(\tau, \mathbf{\Psi}/Z)$ *with each element of* $\mathbf{V}$ *truncated from the left at zero. The following result holds: the vector* $\mathbf{X} = \mathbf{U} + \mathbf{\Lambda V}$ *has the CEST*(μ, $\mathbf{\Sigma}$, $\mathbf{\Lambda}$, τ, $\mathbf{\Psi}$, υ) *distribution.*

The distribution at (5) *may be used as a scale mixture when the underlying variables have a multivariate normal distribution rather than the MESN. In such a case the resulting distribution is the symmetric modification of the multivariate Student-t distribution which is reported in* Arellano-Valle and Genton (2010) *and* Adcock (2010). *When* (6) *is employed as the mixing distribution, the result is the symmetric distribution defined in* Property 1 *of* Appendix A.

4. Extensions of Stein's lemma

This section presents the main results of the paper, the extensions to Stein's lemma. Lemmas 5 and 6 are provided for background and for completeness. The main result is in Lemma 7 and its corollaries. Lemma 5 is a special case of the results in Landsman and Nešlehová (2008) and an extension to that in Liu (1994). As reported in Liu, the lemma may be established using integration by parts. Lemma 6 is believed to be a new result. The proofs of Lemmas 6 and 7 are in Appendix D.

Lemma 5 (*Stein's lemma for the multivariate Student-t distribution*). *Let* $\mathbf{X}$ *be an n-vector distributed as* $MT(\mu, \mathbf{\Sigma}, \upsilon)$, $\upsilon > 2$. *For any scalar valued function* $h(\mathbf{x})$ *such that* $\partial h(\mathbf{x})/\partial x_i$ *is continuous almost everywhere and* $E[(\partial/\partial x_i)h(\mathbf{X})] < \infty$, $i = 1, \ldots, n$, *the following is true*

$$co\upsilon\{\mathbf{X}, h(\mathbf{X})\} = \{\upsilon\mathbf{\Sigma}/(\upsilon-2)\}E_{MS}\{\nabla h(\mathbf{X})|\boldsymbol{\mu}, \upsilon\mathbf{\Sigma}/(\upsilon-2), \upsilon-2\},$$

where $E_{MS}\{\cdot|\mu, \mathbf{\Sigma}, \upsilon\}$ *denotes expectation taken over the Multivariate Student-t distribution with location parameter vector* μ, *scale matrix* $\mathbf{\Sigma}$ *and* υ *degrees of freedom. The vector* $\nabla h(\mathbf{X})$ *is as defined in* Section 1.

Lemmas 6 and 7 *and the properties of marginal and conditional distributions described in the appendices require the following notation. Let* $\widetilde{\mathbf{V}}_i$ *be the random vector of length* $m-1$ *obtained from* $\mathbf{V}$ *by deleting the ith element* V_i. *Let* $\tilde{\boldsymbol{\tau}}_i$ *and* τ_i *be defined in the same way. Similarly, let* $\widetilde{\mathbf{\Lambda}}_i$ *be the matrix of order* $n \times m-1$ *obtained by deleting the ith column of* $\mathbf{\Lambda}$. *Let* $\widetilde{\mathbf{\Psi}}_{ii}$ *be the matrix formed from* $\mathbf{\Psi}$ *by deleting the ithrow and column,* $\widetilde{\mathbf{\Psi}}_i$ *be the vector formed by taking the ith column of* $\mathbf{\Psi}$ *and deleting the ith element and let* Ψ_{ii} *be the (i, i)th element of* $\mathbf{\Psi}$. *Standard manipulations of the density of* $\mathbf{V}$ *yields the following. It is also convenient to define (when* $\upsilon > 1$)

$$\boldsymbol{\tau}_i' = \tilde{\boldsymbol{\tau}}_i - \widetilde{\mathbf{\Psi}}_i\tau_i,$$

$$\mathbf{\Psi}_{ii}' = \upsilon\left(1 + \tau_i^2/\Psi_{ii}\upsilon\right)\left(\widetilde{\mathbf{\Psi}}_{ii} - \widetilde{\mathbf{\Psi}}_i\widetilde{\mathbf{\Psi}}_i^T/\Psi_{ii}\right)/(\upsilon-1),$$

$$\mathbf{\Sigma}' = \upsilon\left(1 + \tau_i^2/\Psi_{ii}\upsilon\right)\mathbf{\Sigma}/(\upsilon-1).$$

Lemma 6 (*Stein's lemma for the truncated multivariate Student distribution*). *Let* $\mathbf{V}$ *be an m-vector which has the multivariate Student distribution with location parameter vector* τ, *scale matrix* Ψ *and* $\upsilon > 2$ *degrees of freedom and with each element of* $\mathbf{V}$ *truncated from below at zero. This distribution is denoted* $TMT(\tau, \mathbf{\Psi}, \upsilon; \mathbf{0}_m)$. *For a scalar valued function* $h(\mathbf{x})$ *as defined for* Lemma 2, *the following is true*

$$\begin{aligned} co\upsilon\{\mathbf{V}, h(\mathbf{V})\} &= \alpha(\boldsymbol{\tau}, \mathbf{\Psi}, \upsilon)\mathbf{\Psi}E_{TMT}\{\nabla h(\mathbf{V})|\boldsymbol{\tau}, \upsilon\mathbf{\Psi}/(\upsilon-2), \upsilon-2; \mathbf{0}_m\} \\ &\quad + \upsilon\mathbf{\Psi\Gamma}[\mathbf{K} - \mathbf{I}_n E_{TMT}\{h(\mathbf{V})|\boldsymbol{\tau}, \mathbf{\Psi}, \upsilon; \mathbf{0}_m\}]\xi/(\upsilon-1), \end{aligned}$$

where $E_{TMT}\{\cdot|\tau, \mathbf{\Psi}, \upsilon; \mathbf{0}_m\}$ *denotes expectation taken over the truncated multivariate Student-t distribution* $TMT(\tau, \mathbf{\Psi}, \upsilon; \mathbf{0}_m)$, $\mathbf{\Gamma}$ *and* ξ *are as defined above,*

$$\alpha(\boldsymbol{\tau}, \mathbf{\Psi}, \upsilon) = \upsilon T_m\{0, -\boldsymbol{\tau}, \upsilon\mathbf{\Psi}/(\upsilon-2), \upsilon-2\}/[(\upsilon-2)T_m\{0, -\boldsymbol{\tau}, \mathbf{\Psi}, \upsilon\}],$$

and $\mathbf{K}$ *is an* $m \times m$ *diagonal matrix with (i, i)th element given by*

$$\kappa_i = E_{TMT}\left\{h(\widetilde{\mathbf{V}}_i)|\tau_i', \mathbf{\Psi}_{ii}', \upsilon-1\right\}.$$

The function $h(\widetilde{\mathbf{V}}_i)$ *is interpreted* $h(\mathbf{V})$ *with* $V_i = 0$. *Note that when* $m = 1$, $\kappa_i = h(0)$.

Lemma 7 (*Stein's lemma for the CEST Distribution*). *Let* $\mathbf{X}$ *be an n-vector that has the distribution* $CEST(\mu, \mathbf{\Sigma}, \mathbf{\Lambda}, \tau, \mathbf{\Psi}, \upsilon)$ *with* $\upsilon > 2$. *For a scalar valued function* $h(\mathbf{x})$ *as defined for Lemma* 1, *cov*$\{\mathbf{X}, h(\mathbf{X})\}$ *is given by the following expression*

$$\begin{aligned} &\alpha(\boldsymbol{\tau}, \mathbf{\Psi}, \upsilon)\{\mathbf{\Sigma} + \mathbf{\Lambda\Psi\Lambda}^{\mathrm{T}}\}E_{CEST}\{\nabla h(\mathbf{X})|\boldsymbol{\mu}, \upsilon\mathbf{\Sigma}/(\upsilon-2), \mathbf{\Lambda}, \boldsymbol{\tau}, \upsilon\mathbf{\Psi}/(\upsilon-2), \\ &\quad \upsilon-2\} + \upsilon\mathbf{\Lambda\Psi\Gamma}[\mathbf{K}_X - \mathbf{I}_n E_{CEST}\{h(\mathbf{X})|\boldsymbol{\mu}, \mathbf{\Sigma}, \mathbf{\Lambda}, \boldsymbol{\tau}, \mathbf{\Psi}, \upsilon\}]\xi/(\upsilon-1), \end{aligned}$$

where $E_{CEST}\{.|\mu, \mathbf{\Sigma}, \mathbf{\Lambda}, \tau, \mathbf{\Psi}, \upsilon\}$ *denotes expectation taken over the* $CEST(\mu, \mathbf{\Sigma}, \mathbf{\Lambda}, \tau, \mathbf{\Psi}, \upsilon)$ *distribution, and* $\mathbf{K}$ *is an* $m \times m$ *diagonal matrix with (i, i)th element given by*

$$\kappa_{Xi} = E_{CEST}\left\{h(\widetilde{\mathbf{X}}_i)|\boldsymbol{\mu}, \mathbf{\Sigma}', \widetilde{\mathbf{\Lambda}}_i, \boldsymbol{\tau}_i', \mathbf{\Psi}_{ii}', \upsilon-1\right\}.$$

Other terms are as defined above.

The following corollary reduces to Lemma 5 *when* $\tau = \mathbf{0}_m$

Corollary 1 (*Stein's lemma for the distribution of* $\boldsymbol{U}$). *Let* $\mathbf{U}$ *be an n-vector that has the (symmetric) distribution* $CEST(\mu, \mathbf{\Sigma}, \mathbf{0}_{nm}, \tau, \mathbf{\Psi}, \upsilon)$ *with* $\upsilon > 2$. *For a scalar valued function* $h(\mathbf{u})$ *as defined for* Lemma 1, *cov*$\{\mathbf{U}, h(\mathbf{U})\}$ *is given by the following expression*

$$\alpha(\boldsymbol{\tau}, \mathbf{\Psi}, \upsilon)\Sigma E_{CEST}\{\nabla h(\mathbf{U})|\boldsymbol{\mu}, \upsilon\mathbf{\Sigma}/(\upsilon-2), \mathbf{0}_{mn}, \boldsymbol{\tau}, \upsilon\mathbf{\Psi}/(\upsilon-2), \upsilon-2\}.$$

It is useful to specify the result for the MEST distribution, that is $m = 1$. *In this result it is assumed that* $\Psi = 1$.

Corollary 2 (*MEST distribution* $m = 1$).

$$\begin{aligned} co\upsilon\{\mathbf{X}, h(\mathbf{X})\} &= \beta\{\mathbf{\Sigma} + \boldsymbol{\lambda\lambda}^{\mathrm{T}}\}E_{MEST}\{\nabla h(\mathbf{X})|\boldsymbol{\mu}, \upsilon\mathbf{\Sigma}/(\upsilon-2), \boldsymbol{\lambda}, \boldsymbol{\tau}, \upsilon/(\upsilon-2), \\ &\quad \upsilon-2\} + \upsilon\boldsymbol{\lambda}[\kappa_X - E_{MEST}\{h(\mathbf{X})|\boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\lambda}, \tau, 1, \upsilon\}]\xi_\upsilon(\tau)/(\upsilon-1), \end{aligned}$$

where

$$\beta = \upsilon T\{\tau\sqrt{(\upsilon-2)/\upsilon}, \upsilon-2\}/\{(\upsilon-2)T(\tau, \upsilon)\},$$

and

$$\kappa_X = E_{MT}\{h(\mathbf{X})|\boldsymbol{\mu}, \boldsymbol{\Sigma}', \upsilon - 1\}.$$

A final corollary of interest is an extension to Siegel's formula, Siegel (1993), and follows Liu (1994) who defines a function $h(\mathbf{x}) = x_{(i)}$, where $x_{(i)}$ is the ith largest among the elements of $\mathbf{x}$. *The function $h(\cdot)$ is piecewise linear and differentiable almost everywhere. The partial derivatives of $h(\cdot)$ with respect to the elements of* $\mathbf{x}$ *satisfy*

$$\frac{\partial h}{\partial x_j} = \begin{cases} 1, & \text{if } x_j = x_{(i)}, \\ 0, & \text{otherwise}. \end{cases}$$

A straightforward application of Lemma 7 gives

Corollary 3 (*Extension to Siegel's Formula*). *Let* $\mathbf{X}$ *be an n-vector that has the distribution CEST(μ, Σ, Λ, τ, Ψ, υ) with $\upsilon > 2$ and let X_j be an arbitrary element of* $\mathbf{X}$. *The following holds*

$$\begin{aligned}\operatorname{cov}(X_j, X_{(i)}) = \alpha(\boldsymbol{\tau}, \boldsymbol{\Psi}, \upsilon)\sum_{k=1}^{n}\left(\sigma_{jk} + \sum_{l,p=1}^{m}\boldsymbol{\Lambda}_{jl}\Psi_{lp}\boldsymbol{\Lambda}_{kp}\right)Pr(X_k \\ = X_{(i)}) + \{\upsilon/(\upsilon-1)\}\sum_{k,l=1}^{m}\boldsymbol{\Lambda}_{jk}\Psi_{kl}\zeta_l,\end{aligned}$$

where

$$\zeta_l = \gamma_l\left[E_{CEST}\left\{X_{(i)}|\boldsymbol{\mu}, \boldsymbol{\Sigma}', \widetilde{\boldsymbol{\Lambda}}_i\boldsymbol{\tau}_i', \boldsymbol{\Psi}_{ii}', \upsilon - 1\right\} - E_{CEST}\{X_{(i)}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}, \boldsymbol{\tau}, \boldsymbol{\Psi}, \upsilon\}\right]\xi_{l.},$$

with the expression Pr( ) above computed using the CEST{μ, $\upsilon\mathbf{\Sigma}/(\upsilon - 2)$, $\mathbf{\Lambda}$, τ, $\upsilon\mathbf{\Psi}/(\upsilon - 2)$, $\upsilon - 2$} distribution.

A simplified version of this corollary for the MEST distribution is omitted. This lemma is of interest because, as Siegel (1993) reports, it was originally motivated by problems in optimal hedging with futures contracts.

5. Portfolio selection and skewness preference

In this section, Lemma 7 and Corollary 2 are used to derive the first order conditions for portfolio selection. Results are presented first for the MEST distribution and then for the more complicated CEST model. Henceforth, the general notations $\boldsymbol{\delta}$ and $\boldsymbol{\Theta}$ are used to denote the vector of expected returns and covariance matrix respectively.

Under the MEST distribution, the first order conditions for portfolio selection depend on the expression

$$\theta_1\boldsymbol{\delta} + \theta_2\{\boldsymbol{\Sigma} + \boldsymbol{\lambda}\boldsymbol{\lambda}^T\}\boldsymbol{w} + \theta_3\boldsymbol{\lambda}.$$

The scalars $\theta_{1,2,3}$ are defined as

$$\theta_1 = E_{MEST}\{U'(X_p)|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, 1, \upsilon\} > 0,$$

$$\theta_2 = \beta E_{MEST}\{U''(X_p)|\boldsymbol{\mu}, \upsilon\boldsymbol{\Sigma}/(\upsilon-2), \boldsymbol{\lambda}, \tau, \upsilon/(\upsilon-2), \upsilon-2\} < 0,$$

$$\theta_3 = \upsilon[\kappa_X - E_{MEST}\{U'(X_p)|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \tau, 1, \upsilon\}]\xi_\upsilon(\tau)/(\upsilon-1),$$

where $X_p = \boldsymbol{w}^T\mathbf{X}$ denotes portfolio return and β and κ_X are as defined for Corollary 2, with $h(\mathbf{X}) = U(\boldsymbol{w}^T\mathbf{X})$. This equation in $\boldsymbol{w}$ is the same for all investors, except for the three scalar quantities $\theta_{1,2,3}$. The portfolios of all investors who are expected utility maximisers are located on a single mean–variance–skewness efficient surface, the equations of which are essentially the same as those in Appendix B of Adcock (2010). As the degrees of freedom υ increase without limit, the first order conditions are the same as those reported in Adcock (2007) for the MESN distribution.

Adcock and Shutes (2012) provide the following result for the MESN distribution.

Lemma 8 (*Preference for skewness (Adcock & Shutes, 2012)*). *Let* $\mathbf{X} \sim$ *MESN(μ, Σ, λ, τ) Preference for skewness is positive (negative) if portfolio skewness $\lambda_p = w^T\lambda > 0(<0)$.*

Given the scale mixture property reported in Lemma 3 of Section 3, it is clear that Lemma 8 also holds for the MEST distribution. Adcock and Shutes (2012) also draw attention to the fact that the first order conditions may be may be written in terms of the covariance matrix $\boldsymbol{\Theta}$. *For the MEST distribution this gives*

$$\theta_1\boldsymbol{\delta} + \theta_2\boldsymbol{\Theta}\boldsymbol{w} + \theta_4\lambda,$$

where the scalar θ_4 is of indeterminate sign. It is argued that this reflects the fact that portfolio variance $w^T\boldsymbol{\Theta}w$ is smaller than the measure of risk implied by the first order conditions above which is $w^T(\boldsymbol{\Sigma} + \lambda\lambda^T)w$. An overall negative preference for skewness represents the price paid for the smaller risk measure.

For the CEST distribution, Lemma 7 shows that the first order conditions depend on the expression

$$\theta_1\boldsymbol{\delta} + \theta_2\{\boldsymbol{\Sigma} + \boldsymbol{\lambda}\boldsymbol{\lambda}^T\}\boldsymbol{w} + \sum_{j=1}^{m}\theta_{3j}\boldsymbol{\lambda}_j,$$

where $\boldsymbol{\lambda}_j$ denotes the jth column of $\boldsymbol{\Lambda}$ and the θ_{3j} are scalars. Using the same approach as in Appendix B of Adcock (2010) leads to the mean–variance hyper-surface. It is straightforward to show that preference for the portfolio skewness $\boldsymbol{w}^T\boldsymbol{\lambda}_j$ is positive or negative depending on its sign. This set of first order conditions may, as above, be expressed in terms of the covariance matrix. In this case the skewness preferences are of indeterminate sign.

The complexities of the CEST distribution have already been noted in Section 1. However, a case for a model with more than one truncated variable may be made as follows. For the MEST distribution, the single truncated variable V may be thought of as a shock which represent the source of a departure from market efficiency, the effect on returns being described by the values of the vector $\boldsymbol{\lambda}$. For a single market, or perhaps for a homogeneous set of securities, it may be acceptable to specify a model with one such shock. For a more heterogeneous set of securities, for example for international portfolio selection, it is equally reasonable to posit that there is more than one source of asymmetry in returns.

Lemma 7 requires that the degrees of freedom parameter be greater than 2; that is, the variance exists.[2] An implication of this is that expected utility maximisation under this distribution when $\upsilon \leqslant 2$ may not be possible or at best the choice of utility function is severely restricted. This is a topic for further investigation. It should also be noted that even for $\upsilon > 2$ the fact that not all moments of the CEST and MEST distributions exist imposes some limitations on the choice of utility function. From a practical perspective, if the scalars collectively denoted using θ are specified then portfolio selection may be performed using quadratic programming.

6. Portfolio selection using Simaan type models

As noted in the introduction, Simaan (1987, 1993) proposes that the n-vector of returns on financial assets should be represented as $\mathbf{X} = \mathbf{U} + \lambda V$, where $\mathbf{U}$ has a multivariate elliptically symmetric distribution and is independent of the non-negative univariate random variable V, which has an unspecified skewed distribution. If the scale matrix for the distribution of $\mathbf{U}$ is now denoted by $\boldsymbol{\Sigma}$, the independence of $\mathbf{U}$ and V means that the first order conditions for portfolio selection depend on

$$\theta_1\boldsymbol{\delta} + \tilde{\theta}_2\boldsymbol{\Sigma}\boldsymbol{w} + \tilde{\theta}_3\boldsymbol{\lambda},$$

[2] Given the definition of both MEST and CEST vectors in Section 2, it is clear that $\upsilon > 2$ is a necessary condition for the variance to exist.

where the positive scalar θ_1 equals the expected value of $U'(X_p)$ taken over the distribution of $\mathbf{X}$ as defined above. The negative scalar $\tilde{\theta}_2$ is the expected value of $U''(X_p)$ taken over the distribution of V and, appealing to Landsman and Nešlehová (2008), a multivariate elliptically symmetric distribution. The scalar $\tilde{\theta}_3$ is defined as

$$\tilde{\theta}_3 = cov\{V, U'(\boldsymbol{w}^T\mathbf{U} + \lambda_p V)\}, \quad \boldsymbol{\lambda}_p = \boldsymbol{w}^T\boldsymbol{\lambda}.$$

The existence of a version of Stein's lemma thus depends entirely on the distribution of V. Adcock and Shutes (2012), for example, show in their Lemma 8 that there is a version of the lemma if V has a gamma distribution. There is a similar result for the Weibull distribution. Other possibilities include distributions for which the density function includes powers of the random variable. However, it is clear that the first order conditions lead to the same mean–variance efficient surface as that for the MEST distribution, but that the sign of the skewness preference parameter $\tilde{\theta}_3$ may be indeterminate. As noted in Section 5, if the preference parameters are specified, portfolio selection may be performed using quadratic programming (QP). Corollary 4.1 of Simaan (1993) or Simaan (1987, p. 63) give an alternative proof to support the use of QP.

The mean–variance skewness efficient surface also arises if a Simaan type model is posited and the parameters are estimated non-parametrically. A sketch of such an efficient surface is shown in Fig. 2. This shows a section of the efficient surface in which it is clear that skewness preference may take negative values. This surface, which is provided solely as an exemplar, is based on annual returns for a number of UK regional commercial property indices. Further details of the example are available on request.

Simaan's approach may be extended in two ways. First, returns may be defined as $\mathbf{X} = \mathbf{U} + \boldsymbol{\Lambda}\mathbf{V}$, where $\mathbf{V}$ is an m-vector of non-negative variables which are distributed independently of $\mathbf{U}$ and $\boldsymbol{\Lambda}$ is an $n \times m$ matrix. The first order conditions are similar to those for the CEST distribution and the same mean–variance hyper-surface arises. Secondly, it is clear that this approach may be further extended by the use of scale-mixture distributions as for example in Mencía and Sentana (2009).

7. First order conditions and practical portfolio selection

When asset returns follow a multivariate normal distribution and noting that they are scale invariant, the first order conditions for portfolio selection, ignoring terms required to deal with constraints, are

$$\theta\boldsymbol{\delta} - \boldsymbol{\Theta}\boldsymbol{w},$$

with the scalar θ defined as

$$\theta_1 = -E\{U'(X_p)\}/E\{U''(X_p)\},$$

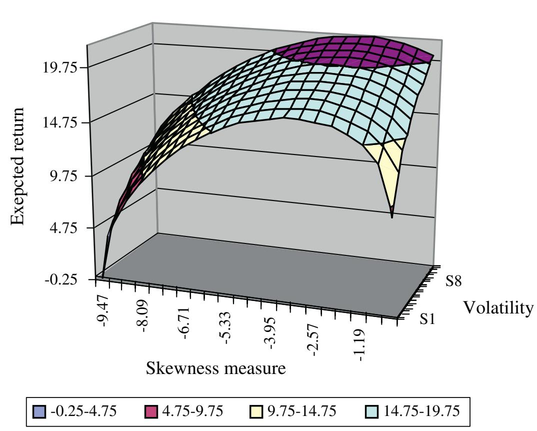

Fig. 2. An example of the mean–variance–skewness efficient surface.

where denotes expectation over the multivariate normal distribution $N(\boldsymbol{\delta}, \boldsymbol{\Theta})$. In general θ is a non-linear function of $\boldsymbol{w}$ the vector of portfolio weights. However, the consequence of Stein's lemma is that all solutions lie on the efficient frontier. Thus, in practice, portfolio selectors will vary values of θ, solving the resulting quadratic programme and selecting an efficient frontier portfolio which meets their specific requirements. This approach is standard, see for example Kallberg and Ziemba (1983).

For the MEST distribution, the first order conditions in Lemma 8 may be written as

$$(\theta\boldsymbol{\delta} + \rho\boldsymbol{\lambda}) - \boldsymbol{\Theta}\boldsymbol{w}.$$

The scalars θ and ρ are also non-linear functions of $\boldsymbol{w}$. The single mean–variance–skewness surface, which is a consequence of the extensions to Stein's lemma for skew-normal and skew-Student distributions, means that a portfolio selector may obtain a suitable portfolio on the surface by varying the scalars θ and ρ. Furthermore, as the first order conditions imply, this task may be performed using quadratic programming. For the CEST distribution, the first order conditions are

$$\theta\boldsymbol{\delta} + \boldsymbol{\Theta}\boldsymbol{w} + \sum_{j=1}^{m}\rho_j\boldsymbol{\lambda}_j.$$

Thus QP may also be used, although in this case it is necessary to specify the values of $(m + 1)$ scalar variables. The first order conditions for Simaan's (1987, 1993) model and the extension with m non-negative variables described in Section 6 may also be implemented using QP.

The use of QP in conjunction with the MESN distribution is illustrated using the same data set as in Adcock and Shutes (2012). This data set consists of daily returns on 13 of the constituent securities in the main stock market index in The Czech Republic for 500 consecutive trading days from 4th June 2003 to 3rd May 2005. The data is from Datastream and prices are in local currency. Model parameters for the MESN distribution were estimated by the method of maximum likelihood. In the interests of brevity, readers are referred to Adcock and Shutes (2012) for details. Portfolio selection was performed subject to the budget constraint and non-negativity condition. That is, subject to

$$\sum_{i=1}^{n} w_i = 1; \quad w_i \geqslant 0, \; i = 1, \ldots, n.$$

The range of values of $\theta \geqslant 0$ and $\rho \geqslant 0$ were set so that the maximum expected return and maximum skewness portfolios each had a single holding. Fig. 3 shows the mean–variance–skewness surface for this set of securities.

The surface shows the non-linear relationship between expected return, volatility and skewness. At zero skewness, the relationship between expected return and volatility is concave as expected. At high values of skewness, concavity does not hold. This is because of the functional relationship between the parameters. Fig. 4 shows the mean percentage holding in assets for which the weight is greater than zero. At the minimum variance and minimum skewness point, the portfolio is well diversified with holdings in all 13 assets. As the volatility or skewness increases, the number of non-zero holdings declines to one. The surface shown in Fig. 4 illustrates the effect of including skewness in portfolio selection. For this data set, the interaction between volatility and skewness leads to a non-linear weight surface.

8. Concluding remarks

It is tempting to write a very short conclusion to this paper: Markowitz rules! The theory developed by him and reported ini-

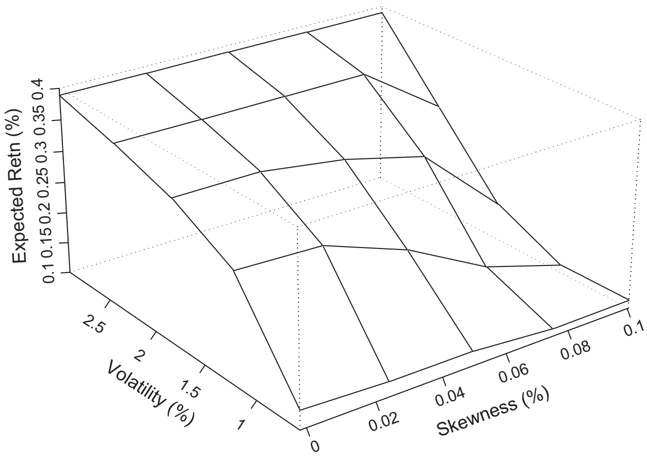

Fig. 3. Mean–variance–skewness for portfolio based on 13 Czech securities. This efficient surface is for portfolios based on 13 Czech stocks as described in Adcock and Shutes (2012). Model parameters were estimated using maximum likelihood. The usual budget and non-negativity constraints were imposed. The values of the scalars preferences were set so that the maximum expected return and maximum skewness portfolios, each with a single holding, were obtained.

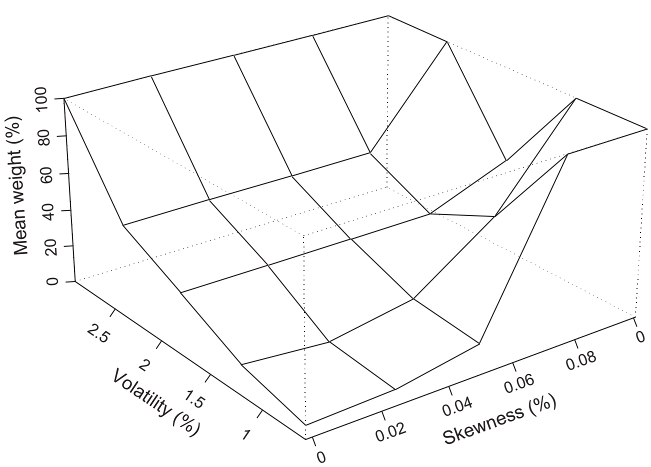

Fig. 4. Average non-zero holding for portfolio based on 13 Czech securities. This surface shows the average value (as a percentage) for non-zero holdings on the mean–variance–skewness efficient surface. Details of the surface are as described in Fig. 3.

tially in Markowitz (1952) is robust to numerous departures from multivariate normality. Elliptically symmetric distributions lead to the efficient frontier. As reported in this paper, several multivariate skew-elliptical distributions lead to an extension to the frontier, which includes an extra dimension or dimensions for skewness. Both the general nature of Simaan's models and the use of scale mixtures suggest that there are many other possible models to describe the distribution of returns, all of which would lead to efficient surfaces that would be invariant to an investor's choice of utility function.

This implies that it is not worthwhile for portfolio selectors to seek utility functions which are better in some sense than others. Given a set of parameter estimates, many models for skewness lead to the efficient surface or hyper-surface. Furthermore, the same situation obtains in principle for agents with short-term horizons who may use suitable GARCH type models, or other representations with time varying parameters.

As shown in Section 7, and importantly from a practical perspective, portfolio selection under the models described in this paper may be performed using quadratic programming.

Estimation and inference for the parameters of multivariate skew-elliptical distribution are not without difficulty. The recent paper by Hallin and Ley (2012) presents numerous new results and gives a bibliography of related earlier works. As indicted in the introduction, models with more than one truncated variable present substantial computational and statistical difficulties. These issues are topics for future work.

Acknowledgements

Thanks are due to Marc Genton for reading a previous version of this paper. They are also due to Zinoviy Landsman for stimulating correspondence about developments of Stein's lemma and to Frank Schuhmacher for discussions of asymmetric location scale distributions. Catherine Jackson kindly provided the data used to generate the example shown in Fig. 2. Very grateful thanks to the two referees who reviewed the previous version of this paper with great care. They pointed out errors and made numerous suggestions for improvement.

Appendix A. Properties of the MEST and CEST distributions

As already noted in the introduction, many of these results may also be found in a different notation in Arellano-Valle and Genton (2010)

Property 1 (*Distribution of* ***U***). *This may be obtained from* (3) *by setting* $\mathbf{\Lambda} = \mathbf{0}_{nm}$. *The density function of* $\mathbf{U}$ *is*

$$\begin{aligned} f(\mathbf{u}) &= t_n(\mathbf{u};\boldsymbol{\mu},\mathbf{\Sigma},\upsilon)T_m(\mathbf{0};-\boldsymbol{\tau},\widehat{\mathbf{\Psi}},\upsilon+n)/T_m(\mathbf{0};-\boldsymbol{\tau},\mathbf{\Psi},\upsilon),\\ \widehat{\mathbf{\Psi}} &= \upsilon\mathbf{\Psi}\{\mathbf{1}+(\mathbf{u}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{u}-\boldsymbol{\mu})/\upsilon\}/(\upsilon+n). \end{aligned} \tag{A1}$$

This elliptically symmetric distribution is an extension to Eq. (3) *of* Arellano-Valle and Genton (2010), *who consider the case where* $m = 1$. *The same distribution is also reported in* Adcock (2010). *When* τ *is the zero vector,* $\mathbf{U}$ *has the multivariate Student-t distribution* $MT(\mu, \mathbf{\Sigma}, \upsilon)$. *Depending on the values in* τ, *the distribution of* $\mathbf{U}$ *may be more or less peaked than the Student distribution.*

The following result is a by-product of this distribution and extends Lemma 1 *of* Adcock (2010).

Lemma A1. *Let* $Y \sim MT(\mu, \mathbf{\Sigma}, \upsilon)$, *let* $\mathbf{\Omega}$ *be a positive definite symmetric matrix and let* $\widehat{\mathbf{\Omega}} = \upsilon\mathbf{\Omega}\{1+(\mathbf{u}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{u}-\boldsymbol{\mu})/\upsilon\}/(\upsilon+n)$. *The following result holds*

$$E\{T_m(\mathbf{0};-\boldsymbol{\tau},\widehat{\mathbf{\Omega}},\upsilon+n)\} = T_m(\mathbf{0};-\boldsymbol{\tau},\mathbf{\Omega},\upsilon).$$

Property 2 (*Distribution of* ***V***). *The vector* $\mathbf{V}$ *is distributed as* $MT(\tau, \mathbf{\Psi}, \upsilon)$ *with each element of* $\mathbf{V}$ *truncated from the left at zero. The probability density function of* $\mathbf{V}$ *is*

$$f(\mathbf{v}) = t_m(\mathbf{v};\boldsymbol{\tau},\mathbf{\Psi},\upsilon)/T_m(\mathbf{0};-\boldsymbol{\tau},\mathbf{\Psi},\upsilon), \quad \mathbf{v} > \mathbf{0}_m.$$

The notation $TMT(\tau, \mathbf{\Psi}, \upsilon; \mathbf{a})$ *is used to denote the truncated multivariate Student-t distribution with* $\mathbf{v} > \mathbf{a}$.

Property 3 (*Conditional Distributions of* ***U*** *given* ***V*** *and* ***X*** *given* ***V***). *The conditional distribution of* $\mathbf{U}$ *given* $\mathbf{V} = \mathbf{v}$ *is* $MT(\boldsymbol{\mu}, \widehat{\mathbf{\Sigma}}, \upsilon+m)$, *where*

$$\widehat{\mathbf{\Sigma}} = \upsilon\mathbf{\Sigma}\{1+(\mathbf{v}-\boldsymbol{\tau})^T\mathbf{\Psi}^{-1}(\boldsymbol{\nu}-\boldsymbol{\tau})/\upsilon\}/(\upsilon+m).$$

The conditional distribution of $\mathbf{X}$ *is the same, except that the location parameter vector is* $\mu + \mathbf{\Lambda}\mathbf{v}$.

Property 4 (*Conditional Distribution of* $\mathbf{V}$ *given* $\mathbf{U}$ *or given* $\mathbf{X}$). *The conditional distribution of* $\mathbf{V}$ *given* $\mathbf{U}$ *is* $TMT(\boldsymbol{\tau}, \widehat{\boldsymbol{\Psi}}, \upsilon + n;\ \mathbf{0}_m)$ *where* $\widehat{\boldsymbol{\Psi}}$ *is defined at (A1). The conditional distribution of* $\mathbf{V}$ *given* $\mathbf{X}$ *is* $TMT(\boldsymbol{\omega}, \widehat{\boldsymbol{\Pi}}, \upsilon + n;\ \mathbf{0}_m)$ *where* ω *and* $\widehat{\boldsymbol{\Pi}}$ *are defined at (4).*

Other properties of the distribution of the truncated Student vector $\mathbf{V}$ *are similar in a number of respects to those of the truncated multivariate normal, which are described in Horrace (2005). In particular, the marginal distributions are not truncated multivariate Student-t, even though some conditional distributions are.*

Using the notation defined in Section 4:

Property 5 (*Expected value of* $\mathbf{V}$). *Using the result in Appendix B, for* $\upsilon > 1$ *the expected value of* $\mathbf{V}$ *is*

$$E(\mathbf{V}) = \boldsymbol{\tau} + \upsilon\boldsymbol{\Psi}\boldsymbol{\Gamma}\xi/(\upsilon - 1),$$

where ξ *is an m-vector with ith element equal to* $t_1(0, -\tau_i, \Psi_{ii}, \upsilon)/T_1(0, -\tau_i, \Psi_{ii}, \upsilon)$ *and* $\boldsymbol{\Gamma}$ *is an* $m \times m$ *diagonal matrix with (i, i)th element given by*

$$\gamma_i = (1 + \tau_i^2/\Psi_{ii}\upsilon)T_1(\mathbf{0}; -\tau_i, \Psi_{ii}, \upsilon)T_{m-1}(\mathbf{0}; -\boldsymbol{\tau}_i', \boldsymbol{\Psi}_{ii}', \upsilon - 1)/T_m(\mathbf{0}; -\boldsymbol{\tau}, \boldsymbol{\Psi}, \upsilon),$$

when $m > 1$ *and* $\gamma_i = 1 + \tau_i^2/\Psi_{ii}\upsilon$ *otherwise.*

Property 6 (*Marginal Distribution of* V_i). *For* $m > 1$*, the probability density function of* V_i *is*

$$f(v_i) = t_1(v_i; \tau_i, \Psi_{ii}, \upsilon)T_{m-1}\{\mathbf{0}; -\tilde{\boldsymbol{\tau}}_i - \widetilde{\boldsymbol{\Psi}}_i(v_i - \tau_i)/\Psi_{ii}, (\upsilon - 1)\boldsymbol{\Psi}_{ii}'/(\upsilon + 1), \upsilon + 1\}/T_m(\mathbf{0}; -\boldsymbol{\tau}, \boldsymbol{\Psi}, \upsilon).$$

When $m = 1$*, the function* $T_0(\cdot)$ *is interpreted as unity, in which case, the marginal distribution of* V *is the truncated Student-t, with density function equal to* $t_1(v, \tau_i, \Psi_{ii}, \upsilon)/T_1(0, -\tau_i, \Psi_{ii}, \upsilon)$*. Note that the definition of* $\boldsymbol{\Psi}_{ii}'$ *means that the result above and Properties 7–9 are valid for any value of* $\upsilon > 0$*.*

Property 7 (*Conditional Distribution of* $\widetilde{\mathbf{V}}_i$ *given* V_i). *Conditional on* $\mathbf{V}_i = v_i, \widetilde{\mathbf{V}}_i$ *has the truncated multivariate Student-t distribution* $TMT\left(\tilde{\boldsymbol{\tau}}_i + \widetilde{\boldsymbol{\Psi}}_i(v_i - \tau_i)/\Psi_{ii}, (\upsilon - 1)\boldsymbol{\Psi}_{ii}'/(\upsilon + 1), \upsilon + 1; \mathbf{0}_{m-1}\right)$.

The two following results are presented for the case where $\mathbf{V}_i = 0$*. Similar results for non-zero values of the conditioning variable are not required in* Section 4 *and so are omitted.*

Property 8 (*Joint distribution of* $\mathbf{U}$ *and* $\widetilde{\mathbf{V}}_i$ *given* $\mathbf{V}_i = 0$). *Conditional on* $\mathbf{V}_i = v_i$ *the vector* $\left(\mathbf{U}^{\mathrm{T}}, \widetilde{\mathbf{V}}_i^{\mathrm{T}}\right)$ *has a multivariate Student-t distribution with* $\upsilon + 1$ *degrees of freedom, location parameter vector* η *and scale matrix* $\boldsymbol{\Theta}$ *defined respectively as*

$$\boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{\mu} \\ \tilde{\boldsymbol{\tau}}_i - \widetilde{\boldsymbol{\Psi}}_i(v_i - \tau_i)/\Psi_{ii} \end{bmatrix}, \quad \boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\Sigma}' & \mathbf{0}_{nm-1} \\ \mathbf{0}_{nm-1}^T & \boldsymbol{\Psi}_{ii}' \end{bmatrix}\{(\upsilon - 1)/(\upsilon + 1)\},$$

with each element of $\widetilde{\mathbf{V}}_i$ *truncated from the left at zero.*

Property 9 (*The conditional distribution of* $\widetilde{\mathbf{X}}_i = \mathbf{U} + \widetilde{\boldsymbol{\Lambda}}_i\widetilde{\mathbf{V}}_i$ *given* $\mathbf{V}_i = 0$). *Using (3) and (4) the conditional distribution of is* $\widetilde{\mathbf{X}}_i$ *is*

$$CEST\left\{\boldsymbol{\mu}, \upsilon\boldsymbol{\Sigma}'/(\upsilon + 1), \widetilde{\boldsymbol{\Lambda}}_i, \boldsymbol{\tau}_i', \upsilon\boldsymbol{\Psi}_{ii}'/(\upsilon + 1), \upsilon + 1\right\}.$$

Appendix B. An integral of the multivariate student distribution

Let $\mathbf{X} \sim MT(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \upsilon)$, $\upsilon > 1$ and let $\boldsymbol{x}^*$ be an n-vector all elements of which are non-negative. The following notation is required. Let $\widetilde{\mathbf{X}}_i$ be the random vector of length $n - 1$ obtained from $\mathbf{X}$ by deleting the ith element, let $\tilde{\boldsymbol{\mu}}_i$ and $\tilde{\boldsymbol{x}}_i^*$ be defined in the same way. Similarly, $\widetilde{\boldsymbol{\Sigma}}_{ii}$ be the matrix formed from $\boldsymbol{\Sigma}$ by deleting the ith row and column and let $\widetilde{\boldsymbol{\Sigma}}_i$ be the vector formed by taking the ith column of $\boldsymbol{\Sigma}$ and deleting the ith element. Finally, let $x_i^*,\ \mu_i$ be respectively the ith elements of $\boldsymbol{x}^*$ and $\boldsymbol{\mu}$ and Σ_{ii} the corresponding diagonal element of $\boldsymbol{\Sigma}$. Define the vector $\boldsymbol{a}$ by the n dimensional vector integral as follows

$$\int_{x^*}^{\infty} (\boldsymbol{x} - \boldsymbol{\mu})t_n(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma, \upsilon)d\boldsymbol{x} = \boldsymbol{\Sigma}\boldsymbol{a}.$$

The ith element of $\boldsymbol{a}$ is given by

$$\{\upsilon/(\upsilon - 1)\}\left\{1 + (x_i^* - \mu_i)^2/\Sigma_{ii}\upsilon\right\}t_1(x_1^*; \mu_i, \Sigma_{ii}, \upsilon)b_i,$$

where b_i is the $(n - 1)$ dimensional integral

$$\int_{\tilde{x}_i^*}^{\infty} t_{n-1}(\tilde{\boldsymbol{x}}_i; \hat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\Sigma}}_{ii}, \upsilon)d\tilde{\boldsymbol{x}}_i = T_{n-1}\left(-\tilde{\boldsymbol{x}}_i^*; -\hat{\boldsymbol{\mu}}_i, \widehat{\boldsymbol{\Sigma}}_{ii}, \upsilon\right),$$

and

$$\hat{\boldsymbol{\mu}}_i = \tilde{\boldsymbol{\mu}}_i + \widetilde{\boldsymbol{\Sigma}}_i(x_i^* - \mu_i)/\Sigma_{ii}, \quad \widehat{\boldsymbol{\Sigma}}_{ii} = \upsilon\left(\widetilde{\boldsymbol{\Sigma}}_{ii} - \widetilde{\boldsymbol{\Sigma}}_i\widetilde{\boldsymbol{\Sigma}}_i^T/\Sigma_{ii}\right)\left\{1 + (x_i^* - \mu_i)^2/\Sigma_{ii}\upsilon\right\}/(\upsilon - 1).$$

When $n = 1$, b_i is empty and is interpreted as unity.

Appendix C. Proof of Lemma 3

Ignoring constants, the conditional density function of $\mathbf{U}$ and V is

$$f(\mathbf{U}, V|Z = z) \propto z^{(n+1)/2}e^{-zW/2}/\Phi(\tau\sqrt{z}),$$

where

$$W = \{(v - \tau)^2 + (\boldsymbol{u} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{u} - \boldsymbol{\mu})\}.$$

Integration over the distribution of Z using the density function at (3) gives the joint distribution of $\mathbf{Y}$ and X which is proportional to

$$[1 + \{(\boldsymbol{u} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{u} - \boldsymbol{\mu}) + (v - \tau)^2\}/\upsilon]^{-\frac{1}{2}(\upsilon+n+1)}/T_\upsilon(\tau).$$

Defining $\mathbf{X} = \mathbf{U} + \lambda V$ and integrating out V gives the MEST distribution at (2). An alternative proof is as follows. The conditional density function of $\mathbf{X}$ and V is

$$f(\mathbf{X}, V|Z = z) \propto z^{(n+1)/2}e^{-z\widetilde{W}/2}/\Phi(\tau\sqrt{z}),$$

where

$$\widetilde{W} = \{(\boldsymbol{x} - \boldsymbol{\mu} - \lambda\tau)^T(\boldsymbol{\Sigma} + \boldsymbol{\lambda}\boldsymbol{\lambda}^T)^{-1}(\boldsymbol{x} - \boldsymbol{\mu} - \lambda\tau) + (v - \tilde{\rho})^2/\tilde{\kappa}^2\},$$

and $\tilde{\kappa}^2 = (1 + \lambda^T\Sigma^{-1}\lambda)^{-1},\ \tilde{\rho} = \tau + \lambda^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu} - \lambda\tau)/(1 + \lambda^T\Sigma^{-1}\lambda)$. Integration over V gives the distribution of $\mathbf{X}$ given $Z = z$. This is MESN with density given by (1). Ignoring constants, the probability density function is

$$z^{n/2} \times \exp\left[-\frac{1}{2}z\{(\boldsymbol{x} - \boldsymbol{\mu} - \lambda\tau)^T(\boldsymbol{\Sigma} + \boldsymbol{\lambda}\boldsymbol{\lambda}^T)^{-1}(\boldsymbol{x} - \boldsymbol{\mu} - \lambda\tau)\}\right]\Phi(\sqrt{z}\tilde{\rho}/\tilde{\kappa})/\Phi(\sqrt{z}\tau).$$

Integration over the distribution of Z at (3) gives the density function of $\mathbf{X}$ directly.

Appendix D. Proof of Lemmas 6 and 7

Lemma 6. *Using property 5, the covariance of* $\mathbf{V}$ *with* $h(\mathbf{V})$ *is*

$$cov\{\mathbf{V}, h(\mathbf{V})\} = E\{(\mathbf{V} - \boldsymbol{\tau})h(\mathbf{V})\} - \upsilon E\{h(\mathbf{V})\}\boldsymbol{\Psi}\boldsymbol{\Gamma}\xi/(\upsilon - 1).$$

Ignoring the normalisation constant, the expected value of $(\mathbf{V}-\tau)h(\mathbf{V})$ *is proportional to the vector*

$$\int(\boldsymbol{v}-\boldsymbol{\tau})h(\boldsymbol{v})t_m(\boldsymbol{v};\boldsymbol{\tau},\boldsymbol{\Psi},\upsilon)d\boldsymbol{v}=-\{\upsilon\boldsymbol{\Psi}/(\upsilon-2)\}\int[(\partial/\partial\boldsymbol{v})t_m\{\boldsymbol{v};\boldsymbol{\tau},\upsilon\boldsymbol{\Psi}/(\upsilon-2),\upsilon-2\}]h(\boldsymbol{v})d\boldsymbol{v}.$$

Integrating by parts and ignoring the constant $\upsilon\boldsymbol{\Psi}/(\upsilon-2)$, *the ith element is*

$$\int_{\tilde{\upsilon}_i}[t_m\{\boldsymbol{v},\boldsymbol{\tau},\upsilon\boldsymbol{\Psi}/(\upsilon-2),\upsilon-2\}]_{\upsilon_i=0}h(\tilde{\boldsymbol{v}}_i)d\tilde{\boldsymbol{v}}_i+\int_{\upsilon}\{\partial h(\boldsymbol{v})/\partial\upsilon_i\}t_m\{\boldsymbol{v},\boldsymbol{\tau},\upsilon\boldsymbol{\Psi}/(\upsilon-2),\upsilon-2\}d\boldsymbol{v}.$$

Gathering constants, the second integral gives the first term in the expression above. In the first integral, the term in [] is proportional to the density function of $\widetilde{\mathbf{V}}_i$ *given* $V_i = 0$*; property 7. Gathering constants gives* $\upsilon\boldsymbol{\Sigma}\boldsymbol{\Gamma}\mathbf{K}\xi/(\upsilon-1)$ *as required.*

Lemma 7. *The setup and method of proof follow* Adcock (2007). *The* cov[$\mathbf{X}$, $h(\mathbf{X})$] *may be written as the expected value of*

$$\{(\mathbf{U}-\boldsymbol{\mu})+\boldsymbol{\Lambda}(\mathbf{V}-\boldsymbol{\tau})-\upsilon\boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Gamma}\xi_1/(\upsilon-1)\}h(\mathbf{U}+\boldsymbol{\Lambda}\mathbf{V}),$$

taken over the joint distribution of **U** *and* **V** *at* Eq. (2). *The integration is performed term-by-term using the same method employed for* Lemma 6.

References

Adcock, C. J. (2002). Asset pricing and portfolio selection based on the multivariate skew-student distribution. In *Multi-moment capital asset pricing models and related topics workshop*, Paris.

Adcock, C. J. (2007). Extensions of Stein's lemma for the skew-normal distribution. *Communications in Statistics – Theory and Methods, 36*(9), 1661–1672.

Adcock, C. J. (2009). Moments and distribution of quadratic forms under the multivariate extended skew-normal and skew-student-*t* distributions. *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*(1), 221–234.

Adcock, C. J., Eling, M., & Loperfido, N. (2012). Skewed distributions in finance and actuarial science: A review. *The European Journal of Finance*, http://dx.doi.org/10.1080/1351847X.2012.720269.

Adcock, C. J., & Shutes, K. (2001). Portfolio selection based on the multivariate skew normal distribution. In A. Skulimowski (Ed.), *Financial modelling* (pp. 167–177). Krakow: Progress and Business Publishers.

Adcock, C. J., & Shutes, K. (2012). On the multivariate extended skew-normal, normal-exponential and normal-gamma distributions. *The Journal of Statistical Theory and Practice, 6*(4), 636–664.

Arditti, F. D., & Levy, H. (1975). Portfolio efficiency in three moments: The multiperiod case. *Journal of Finance, 30*(3), 797–809.

Arellano-Valle, R. B., & Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics, 33*(3), 561–574.

Arellano-Valle, R. B., Branco, M. D., & Genton, M. G. (2006). A unified view on skewed distributions arising from selections. *Canadian Journal of Statistics, 34*(4), 581–601.

Arellano-Valle, R. B., & Genton, M. G. (2010). Multivariate extended skew-*t* distributions and related families. *Metron, 68*(3), 201–234.

Arnold, B. C., & Beaver, R. J. (2000). Hidden truncation models. *Sankhya, Series A, 62*(1), 22–35.

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

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

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

Azzalini, A., & Genton, M. G. (2008). Robust likelihood methods based on the skew-*t* and related distributions. *International Statistical Review, 76*(1), 106–129.

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

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

Chunhachinda, P., Dandapani, K., Hamid, S., & Prakash, A. J. (1997). Portfolio selection and skewness: Evidence from international stock market. *Journal of Banking, and Finance, 21*(2), 641–660.

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

Fang, K.-T., Kotz, S., & Ng, K.-W. (1990). *Symmetric multivariate and related distributions. Monographs on statistics and applied probability* (Vol. 36). London: Chapman & Hall, Ltd..

Goh, J. W., Lim, K. G., Sim, M., & Zhang, W. (2012). Portfolio value-at-risk optimization for asymmetrically distributed asset returns. *European Journal of Operational Research, 221*(2), 397–406.

Gonzàlez-Farías, G., Dominguez-Molína, A., & Gupta, A. K. (2004). Additive properties of skew normal random vectors. *Journal of Statistical Planning and Inference, 126*(2), 521–534.

Hallin, M., & Ley, C. (2012). Skew-symmetric distributions and fisher information – A tale of two densities. *Bernoulli, 18*(3), 747–763.

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

Horrace, W. C. (2005). Some results on the multivariate truncated normal distribution. *Journal of Multivariate Analysis, 94*(1), 209–221.

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

Kallberg, J. G., & Ziemba, W. T. (1983). Comparison of alternative functions in portfolio selection problems. *Management Science, 11*(11), 1257–1276.

Kim, H.-M., & Genton, M. G. (2011). characteristic functions of scale mixtures of multivariate skew-normal distributions. *Journal of Multivariate Analysis, 102*(7), 1105–1117.

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

Landsman, Z. (2006). On the generalization of Stein's lemma for elliptical class of distributions. *Statistics and Probability Letters, 76*(10), 1012–1016.

Landsman, Z., & Nešlehová, J. (2008). Stein's lemma for elliptical random vectors. *Journal of Multivariate Analysis, 99*(5), 912–927.

Li, X., Qin, Z., & Kar, S. (2010). Mean–variance–skewness model for portfolio selection with fuzzy returns. *European Journal of Operational Research, 202*(1), 239–247.

Liu, J. S. (1994). Siegel's formula via Stein's identities. *Statistics and Probability Letters, 21*(3), 247–251.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance, 7*(1), 77–91.

Matmoura, Y., & Penev, S. (2013). Multistage optimization of option portfolio using higher order coherent risk measures. *European Journal of Operational Research, 227*(1), 190–198.

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

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

Samuelson, P. A. (1970). The fundamental application theorem of portfolio analysis in terms of means, variance and higher moments. *Review of Economic Studies, 37*(4), 537–542.

Siegel, A. F. (1993). A surprising covariance involving the minimum of multivariate normal variables. *Journal of the American Statistical Association, 88*(421), 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*(5), 568–587.

Stein, C. M. (1973). Estimation of the mean of a multivariate normal distribution. In *Proceedings of the prague symposium on asymptotic statistics* (pp. 345–381).

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Annals of Statistics, 9*(6), 1135–1151.

Sun, Q., & Yan, Y. (2003). Skewness persistence with optimal portfolio selection. *Journal of Banking and Finance, 27*(6), 1111–1121.