Using Parametric Classification Trees For Model Selection with Applications to Financial Risk Management

by

C J Adcock[1] and N Meade[2]

Sheffield University Management School[1], Imperial College London[2]

Abstract

We describe two parametric classification tree methods, which allow formal selection of a member of a class of generalised distributions. In the paper we consider Generalised Beta distributions for non-negative random variables and the Generalised skew-Student distribution for random variables distributed on the real line. We introduce a class of symmetric generalised multivariate Student distributions, members of which may also be selected using the classification trees. We present two versions of the parametric classification tree: specific to general and general to specific. We apply the classification methods to daily returns on stocks from a selection of 15 major, mid-cap and emerging markets. The results show that the majority of return distributions follow Student's t, but that a non-negligible minority follow a symmetric generalised Student distribution. We confirm a well-known stylised fact about skewness: it tends not to be persistent. By contrast, kurtosis is persistent. Using the symmetric generalised multivariate Student distribution, we present a risk management study based on efficient portfolios constructed from UKFTSE250 stocks and specifically concerned with the computation of value at risk. The case study demonstrates that the model selection procedures based on the classification trees lead to more accurate computation of VaR than those based on the normal distribution or on non-parametric approaches. The study also shows that the normal distribution may be used for VaR computations for larger portfolios when the holding period is longer.

Key words: Finance; Classification; Persistence; Risk-management; Skew-Student

Correspondence Address

C J Adcock
Sheffield University Management School
Conduit Road
Sheffield, S10 1FL UK

Tel: +44 (0)114 222 3402

Email: c.j.adcock@shef.ac.uk

This is the version of the article accepted for publication in European Journal of Operational Research Vol. 259 (2), 746-765. Published version available from Elsevier at:
https://doi.org/10.1016/j.ejor.2016.10.051

1. Introduction

There are many empirical research papers published in both operational research journals and journals in cognate disciplines in which asymmetry is a feature of the probability distribution of the variable under study. There are two general types of application: those in which the variable of interest is without loss of generality non-negative and those in which the variable is defined on the whole real line. Exemplars of the first type include studies of times to failure of engineering components or income distribution. Examples of the second include studies in financial risk management. It is also well known that there is a wide variety of probability distributions that may be used to model asymmetry and other departures from normality. From the perspective of model selection, it is often the case, however, that a probability distribution may be selected arbitrarily for use in an empirical study. It is also often the case that a number of candidate distributions will be estimated. The maximum value of the estimated log-likelihood function often then determines the single model to be used. Alternatively the minimum value of Akaike's information criterion or a similar test statistic is also often used. These approaches can lead to the selection of a distribution whose log-likelihood function is only marginally greater or AIC only marginally smaller than that of the second ranked model. They can also lead to the selection of models which are not parsimonious.

The aim of this paper is fourfold. First, it is to present methods of model selection based on classification trees which may be used in conjunction with the generalised beta distribution [GB henceforth] and its two-sided extension, the generalised skew-Student distribution [GST]. These two families of probability distributions embody flexible parameterisation, which means that they are able to represent different forms of asymmetry, as well as symmetric departures from normality. When used in conjunction with the classification trees they enable parsimonious model selection and, in the case of the selection of an asymmetric distribution, insights into the nature of the asymmetry. Secondly, the paper presents an extension to the GST distribution, which enhances the flexibility of the class. Thirdly, the symmetric special cases of the GST distribution motivate a generalised multivariate Student distribution. This is of potential use for applications in finance, such as portfolio selection and risk management, but is also of general applicability for multivariate problems. Lastly, the paper presents two related empirical studies which demonstrate the application of the classification tree and the importance of selection of an appropriate model. The paper reports the application of the GST distribution and the classification tree to returns on stocks traded in a number of major international stocks markets. It then reports an application of the multivariate Student and generalised multivariate Student distributions to portfolio risk management. These two empirical studies demonstrate the importance to effective financial risk management of appropriately identifying the distribution of returns

The GB and GST distributions each include numerous special cases. For example, the gamma and Student's t are special cases of the GB and GST distribution respectively. As the sections below show, the classification tree allows these and other special cases to be selected. It is beyond the scope of this paper to apply the tree to other families of distributions, but it is clear that the principles of the tree may be applied more generally. Extensions to applications in which there is a given embedded regression model or a given GARCH model for the residuals are straightforward. The tree may for example also be used in conjunction the extended skew-Student distribution[1], which also embodies numerous special cases, as well as extensions of the GB distribution which have been reported in the literature.

The structure of the paper is as follows. Section 2 is a review of background literature. This starts with a brief summary of applications of the GB distribution. To support the two empirical studies, there is then a more detailed review of non-normality in stock returns and risk management applications. Section 3 summarises the GB distribution and presents the extended version of the GST distribution that is used with the classification tree. It also describes the generalised multivariate Student distribution and properties that are required for the study of risk management application. Section 4 presents the parametric classification trees for some of the examples in the previous section. Section 5 describes the data that is used in the two empirical studies. The results of applying the parametric classification trees to the stock return data set and related results are presented in Section 6. Section 7 is concerned with the application of the generalised multivariate Student distribution to portfolio risk management. Section 8 concludes. In keeping with modern practice only key empirical results are presented here, but further details are available from the corresponding author on request. Most notation is that in common use, but is otherwise defined in the text.

2. Literature Review

The theme of this paper is the identification of an appropriate probability density function for a given data set. The paper presents a parametric classification tree to identify the appropriate member of a family of probability densities based on distributions which are generalisations of the Weibull or GB distribution. In addition to parsimony, the overall motivation for identifying the appropriate density for an application is either concerned in a broad sense with efficient risk management or cost minimisation. Applications range from equipment replacement to software maintenance to measuring value at risk in finance.

There are many empirical papers reported in the operational research literature and in journals in cognate disciplines which use the Weibull or GB distribution or which could employ it. Examples of non-negative random variables in different types of risk management application are: times to failure of engineering components (Singla *et al*, 2012) and of software (Okamura and Dohi, 2015); ambulance travel times (Budge *et al*, 2010); stream flow and precipitation volumes in water resource management (Mielke and Johnson, 1974) and costs of misclassification of credit risks (Lessman *et al*, 2015) in financial risk management. Examples of cost minimisation are: timber quantities in forestry management (Ducey and Gove, 2015); excess demand in airline spill analysis (Li and Oum, 2000). Boccanfuso *et al* (2008) investigate different functional forms of distribution in a study of income, poverty and computable general equilibrium modelling. They conclude that for their application no single form of distribution is superior, thus implicitly emphasising the need for appropriate model selection procedures.

In addition to empirical studies, the flexibility of the GB distribution has stimulated the development of analytical tools. Two examples are the development of Bayesian sampling

[1] See for example Azzalini and Capitanio (2003) or Arellano-Valle and Genton (2010).

plans for the Weibull distribution with censoring (Chen *et al*, 2007 and imputation procedures for missing data (Graf and Tillé, 2014). Finally, there are other applications reported in the literature where the random variable of interest exhibits asymmetry but for which the Weibull or GB distributions are not employed, even though they would be appropriate. Two examples are a study of the optimal burn-in decisions for products with a unimodal failure rate using the log-normal distribution (Chang, 2000) and a stochastic frontier model to evaluate regional financial efficiency in China using a truncated normal distribution (Zhang *et al*, 2015).

Examples of applications of GST distributions for which the variables may in principle take any real value are found mainly in finance. It is important to note, none the less, that *a priori* the GST distributions and hence the classification tree described in Section 4 below are candidate models for any application in which the variable of interest takes any real value. The traditional assumption that returns on financial assets are normally distributed has been criticised on empirical grounds for many years. One of the first to express dissatisfaction was Mandelbrot (1963) who suggested the use of stable distributions to capture fat tails in asset returns. Praetz (1972) and Blattberg and Gonedes (1974) pioneered the use of Student's t distribution in finance. Aparicio Acosta and Estrada (2001) applied it to European stock returns. Kon (1984) examines mixtures of normal distributions. Mauleon and Perote (2000) and Mauleon (2006) compared the Student's t with the Edgeworth-Sargan distribution. Corrado and Su (1996) make use of Gram-Charlier series expansions. Hansen (1994) and Harris, Kukukozmen and Yilmaz (2004) specifically investigate skewness using an asymmetric generalisation of the Student t distribution. The skew-normal and skew-Student distributions associated with Azzalini (1985, 1986) and Azzalini and Capitanio (2003) have been applied in finance by Adcock and Shutes (2001), Adcock (2010, 2014) and Harvey *et al* (2010). Fernandez and Steel (1998) present a model based on Student's t that accommodates both skewness and fat tails. McDonald and Nelson (1993) employ a skew-Student distribution to estimate beta in the market model.

Perusal of recently published research indicates that there is a continuing need for methods which can deal with the non-normality found in financial data. For example Dias *et al* (2015) who write "*It has also been recognized that stock market returns and returns of financial assets contain skewness and excessive kurtosis*" present methods based on Markov switching. Yu *et al* (2015) present a value-at-risk model which accommodates transaction costs and short-selling and which is motivated by the well-known stylised facts of returns on financial assets; namely that they exhibit asymmetry and fat tails. Laih (2015) measures the rank correlation coefficients between financial time series using a copula which employs the Skew-t distribution due to Hansen (1994). There are many other examples of the use of non-normal distributions, particularly in finance.

The persistence of non-normal behaviour, particularly skewness, has been considered by several authors. Singleton and Wingender (1986) consider the persistence of skewness using monthly returns data for the period 1961-80. They find that positively skewed assets are as likely to exhibit negative skewness in the next period as positive and vice versa. Wang *et al* (2015) examine different types of asymmetries in stock markets returns. Consistent with the findings of Singleton and Wingender, they conclude that persistence of returns is weak. In a study of emerging markets, Bekaert *et al* (1998) draw attention to the hypothesis that skewness and kurtosis may be time varying. An explanation for such temporal variation is that these forms of non-normality are artefacts of the process of market emergence. An implication of this hypothesis, if correct, is that incidence of significant values of these moments will decrease as time progresses. Theodossiou and Savva (2015) examine the effect of skewness on the relationship between risk and return using the GST distribution. There are interpretations in the finance literature for departures from normally distributed returns. These include: fat tails arise because of changing volatility; peakedness can reflect thin trading; skewness is evidence of the degree of market efficiency. The (symmetric) generalised Student t distribution, which is presented in Section 3.3 of this paper, is able to indicate the presence of peakedness, which may be interpreted as due to thinner trading.

An accurate forecast of the density function of the future return on an asset is important for the management of risk, using measures such as Value at Risk (VaR), Expected Shortfall (ES) or Conditional Value at Risk (CVaR). The results reported in Aldowaisan *et al* (2015) provide a clear reminder that the tail behaviour of the GB and therefore GST distributions differ, thus reinforcing the need for selection of an appropriate model for the density function. The problem of risk management may be addressed by considering either the unconditional density or the conditional density function of the return series. Using unconditional probability estimates, Longin (2000) discusses the applications of extreme value theory to VaR. In Longin (2005), this work is further developed and used to inform the choice of distribution for asset returns. Using US data, he demonstrates that of the Gaussian, Student t and the stable Paretian density functions, only the Student t is acceptable as an unconditional model of returns. Stoyanov *et al* (2013) consider the skewed Student's t and stable distributions while examining the effect of adjusted portfolio weights on CVaR. Studies using a conditional density approach include the following. Zhu and Galbraith (2011) use a generalised asymmetric Student's in conjunction with a non-linear GARCH model to predict expected shortfall on the returns of six stocks and the S&P 500. Bhattacharyya *et al* (2008) use a Pearson type IV distribution, which is a further generalisation of the GB2, in conjunction with a GARCH model for estimation of CVaR using 14 national equity indices. Although they find their results superior to using a normal-GARCH combination, the Pearson type IV failed some goodness of fit tests. Using out of sample density forecasting accuracy as a model selection criterion, Meade (2010) finds that a mixture of two normal distributions captures the behaviour of oil price returns for a horizon of up to two years.

3. The Generalised Beta Skew Student and Generalised Multivariate Student Distributions

Here we describe the generalised distributions for non-negative random variables, random variables on the real line and multivariate random variables.

3.1 Generalised Beta and Gamma Distributions

Generalised beta distributions are attributed to McDonald (1984). Consistent with the well-known beta distributions there are two forms of the distribution, generally referred to as generalised beta distributions of the first and second kind and usually denoted by the abbreviations GB1 and GB2. The respective probability density functions are

$$f(x)=|\omega|x^{\omega\alpha-1}\left\{1-(x/\phi)^{\omega}\right\}^{\beta-1}\Big/\phi^{\omega\alpha}B(\alpha,\beta);\, 0\le x\le\phi;\, \phi,\alpha,\beta>0\,, \tag{1.}$$

and

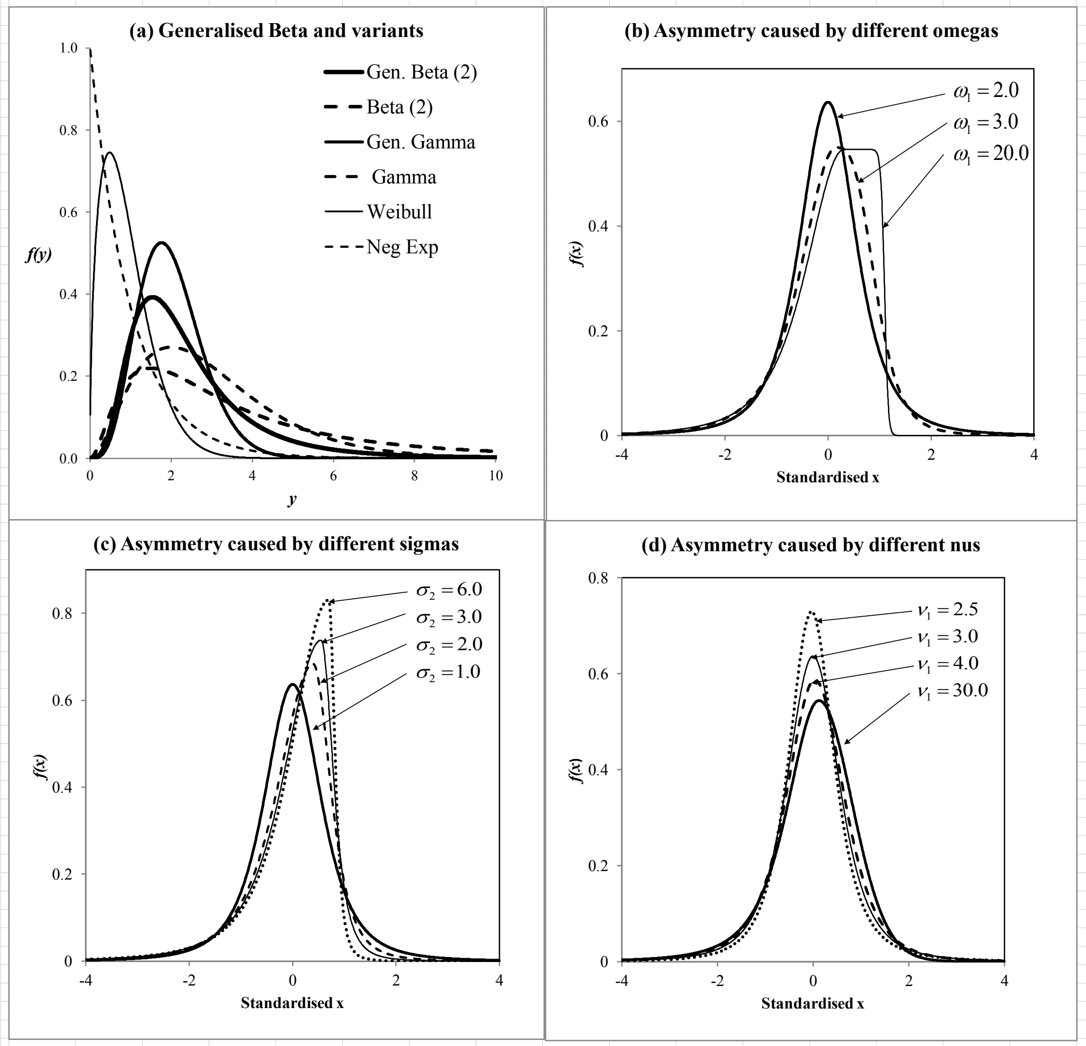
$$f(x)=|\omega|\,x^{\omega\alpha-1}\Big/\phi^{\omega\alpha}\left\{1+(x/\phi)^{\omega}\right\}^{\alpha+\beta}B(\alpha,\beta);\, 0\le x<\infty;\, \phi,\alpha,\beta>0\,. \tag{2.}$$

If $\phi=\psi\beta^{1/\omega}$ the limit as $\beta\to\infty$ gives the generalised gamma distribution, with density function

$$f(x)=|\omega|x^{\omega\alpha-1}e^{-(x/\psi)^{\omega}}\Big/\psi^{\omega\alpha}\Gamma(\alpha);\, 0\le x<\infty;\, \psi,\alpha>0\,, \tag{3.}$$

which is due to Stacy (1962) and Stacy and Mirham (1965). The case $\omega=\alpha^{-1}$ gives the Weibull distribution. Placing constraints on the coefficients of the distributions at (2.) and (3.) (by letting β tend to infinity or setting ω or α to unity) produces a range of different densities, examples of which are shown in Figure 1.

Figure 1. (a) shows examples of the generalised beta (type 2) family. In the general case the coefficients are: ω = 1.5, α = 3, β =3; constraining the coefficients produces the other density functions. (b), (c) and (d) show the effects of asymmetric parameter values on the shape of the GST.


(a) Generalised Beta and variants
Gen. Beta (2)
Beta (2)
Gen. Gamma
Gamma
Weibull
Neg Exp
f(y)
y
(b) Asymmetry caused by different omegas
$\omega_1 = 2.0$
$\omega_1 = 3.0$
$\omega_1 = 20.0$
f(x)
Standardised x
(c) Asymmetry caused by different sigmas
$\sigma_2 = 6.0$
$\sigma_2 = 3.0$
$\sigma_2 = 2.0$
$\sigma_2 = 1.0$
f(x)
Standardised x
(d) Asymmetry caused by different nus
$\nu_1 = 2.5$
$\nu_1 = 3.0$
$\nu_1 = 4.0$
$\nu_1 = 30.0$
f(x)
Standardised x

In Table 1 of their paper, Richards and McDonald (1987) list a number of special cases of both distributions. They also note that these distributions are referred to by different names in some of the literature. In this paper, however, we refer to these two distributions using the names and abbreviations above. In Figure 1 in their paper they also report a tree structure which shows the linkages between the different special cases. McDonald and Xu (1995) introduce a further generalisation of the beta distribution which contains both GB1 and GB2 as special cases. The probability density function is

$$f(x)=|\omega|x^{\omega\alpha-1}\left\{1-(1-\lambda)(x/\varphi)^{\omega}\right\}^{\beta-1}/\varphi^{\omega\alpha}\left\{1+\lambda(x/\varphi)^{\omega}\right\}^{\alpha+\beta}B(\alpha,\beta);\, 0\leq x<\varphi/(1-\lambda). \quad (4.)$$

with $\phi,\alpha,\beta>0, 0\leq\lambda\leq 1$. In the same paper, McDonald and Xu introduce the exponential generalised beta distribution, referred to by the abbreviation EGB, by considering the transformation $y=logx$. There are numerous extensions in the literature. Singla *et al* (2012) report generalisations of both GB1 and GB2 which arise from considering the transformation $x=G(z)$ where $G(.)$ is a distribution function. If the derivative of $G(.)$ exists, the probability density function of z corresponding to the GB1 is

$$f(x)=|\omega|G'(z)\{G(z)\}^{\omega\alpha-1}\left[1-\{G(z)/\varphi\}^{\omega}\right]^{\beta-1}/\varphi^{\omega\alpha}B(\alpha,\beta);\ \varphi,\alpha,\beta>0. \quad (5.)$$

A number of related distributions are reported in Gordy (1998), Ye (2012), Almalki and Yuan (2013) and Ducey and Grove (2015). All of these families contain special cases are so are capable of being used with the type of model selection process that is described in Section 4.

3.2 Generalised Skew-Student Distributions

A two sided distribution that is based on the generalised beta is due originally to McDonald and Newey (1988) and is referred to as a generalised skewed Student t distribution. It is derived from the GB2 distribution, whose density function is given at equation (2.), for the case where $\omega=\alpha^{-1}$ which gives a density function which is non-zero at the origin. It is extended and applied to modelling asset returns in Theodossiou (1998) who points out that this is a rich family of probability distributions. Placing constraints on the parameters generates a wide range of special cases, including the Gaussian, Laplace, Student's t itself and generalised error distributions. Similar distributions are reported in Hansen (1994), Fernandez and Steel (1998) and Zhu and Galbraith (2010). It may be noted that many two-sided distributions may be obtained from suitable combinations of the densities reported at Equations (1.) through (4.)

Greater flexibility in modelling both fat-tails and asymmetry is obtained by extending the parameterisation of the skewed Student t distribution. Specifically, this paper considers a random variable X which has a probability distribution with density function

$$f_X(x)=K\left(1+|x/\sigma_i|^{\omega_i}/v_i\right)^{-v_i/2-1/\omega_i};\, i=1 \text{ for } x\geq 0, i=2 \text{ for } x<0, -\infty<x<\infty, \quad (6.)$$

with the normalising constant K given by

This is the version of the article accepted for publication in European Journal of Operational Research Vol. 259 (2), 746-765. Published version available from Elsevier at: https://doi.org/10.1016/j.ejor.2016.10.051

$$K^{-1} = K_1^{-1} + K_2^{-1};\ K_i = \omega_i \Gamma(v_i/2 + 1/\omega_i)/\Gamma(v_i/2)\Gamma(1/\omega_i)\sigma_i v_i^{1/\omega_i};\ \omega_i, v_i, \sigma_i > 0, i = 1,2.$$

This is referred to as the generalised skewed t or GST distribution. The two ν parameters are referred to as degrees of freedom. The two ω parameters measure Studentness: deviations from 2 measure the departure of the distribution from Student's t. The σs are scaling parameters. Theodossiou's model is a special case of the GST; when the degrees of freedom and Studentness parameters are equal, i.e. $\nu_1 = \nu_2$ and $\omega_1 = \omega_2$. In this model, skewness is driven only by differences in (σ_1, σ_2). The contribution of the GST distribution at (6.) is that both skewness and kurtosis can be generated by differences between the values of one or more of the pairs of parameters, (ν_1,ν_2), (ω_1, ω_2), (σ_1, σ_2). Examples of the effect of variation in each of these parameters are shown in Figure 2. Details of the derivation of (6.) are available on request.

As well as the general case of the distribution, which are denoted GS in this paper, there are 7 special cases in which one or more of the pairs of parameters are restricted to take equal values. In addition, two closely related families of distributions are obtained by imposing fixed values on the degrees of freedom or the Studentness parameters. First, as ν_1 and ν_2 both increase without limit, a skewed version of the generalised error distribution, denoted GE, is obtained, (see Nelson, 1991 for further details). Secondly, if $\omega_1 = \omega_2 = 2$ the generalised Student distribution, GT, is obtained (this is the density function used by Zhu and Galbraith, 2010). There are 4 general cases for GE distributions and are 4 general cases for the GT distributions, plus the special cases of the normal and Student's t. This gives an overall total of 18 models.

At first sight, it may appear that this distribution is over-parameterised. In the tails of the distribution, the probability density function is of the form

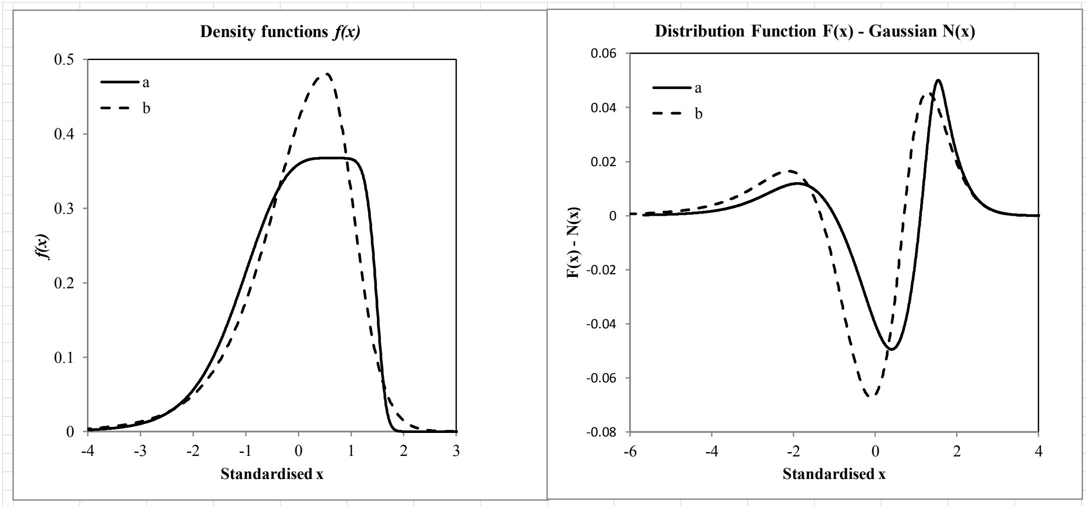
$$f_X(x) \approx K' \left| x/\sigma \right|^{-\upsilon\omega/2-1} \qquad (7.)$$

where K' is a constant and the subscript i has been omitted. Since the tail values dominate the likelihood function, a more appropriate parameterisation might be in terms of the products $\nu_i\omega_i$, $i = 1,2$. Furthermore, in the symmetric case, the implication of (7.) is that Student t distribution may be a more suitable model with degrees of freedom equal to $\nu\omega/2$. However, Figure 2 shows two density functions, GS (a) and GS (b), with parameters as listed in the display below.

GST	ν_1	ω_1	σ_1	ν_2	ω_2	σ_2	Skewness	Excess Kurtosis
a	9.5	8.0	1.5	4.5	3.0	2.4	-1.6	7.9
b	25.0	2.0	0.6	6.0	2.0	1.2	-1.6	7.3

Both distributions have mean zero, unit variance, a skewness of –1.6 and a small difference in excess kurtosis. The asymmetry depicted is noticeably different and GS (b) is visually more peaked, although having slightly lower kurtosis. Sufficient empirical evidence to justify the use of up to seven parameters will clearly depend on the application in question. However, for the data set studied in this paper, Section 6 shows empirical evidence to support the GST distribution defined at (6.) in preference to the more restricted versions that appear in earlier literature.

Figure 2. Two GSTs both with skewness of –1.6 but displaying different forms of asymmetry



The notation for the different cases of the GST distribution has the general form *GSαβγ*. *GSωυσ* means that the three pairs of parameters take equal values; that is the distribution is symmetric. In general, replacement of the Greek letter by a hyphen means that asymmetry may be generated by variation of the suppressed parameter. For example, *GSων-* means that asymmetry is generated only through variation in the suppressed parameter σ. The most general case is *GS---*. For other restricted cases *GS* is replaced by *GT* or *GE*. The normal distribution is denoted NORML. Moments about the origin can be derived from equation (6.). As shown in the appendix the n^{th} moment exists if $min(v_i \omega_i) > 2n$. As shown in Theodossiou (1998), analytical evaluation of the moments of this distribution is complicated. However, numerical evaluation of mean, variance, skewness and kurtosis is straightforward and cumulative probabilities may be computed using the incomplete beta function.

3.3 Multivariate Symmetric Generalised Student Distributions

For the symmetric GST distribution, it is natural to consider multivariate versions. This section presents a multivariate form of the symmetric *GSωνσ* distribution. Such distributions have the potential to be applicable for multivariate problems for which the variables exhibit kurtosis. The multivariate extension of the Weibull distribution is a special case of the symmetric Kotz type distribution, which is described in Fang, Kotz and Ng (1990, page 76) and is a member of the elliptically symmetric class. If $\mathbf{X}$ denotes an *n*-vector of random variables with location parameter vector $\boldsymbol{\mu}$ and positive definite scale matrix $\mathbf{\Sigma}$, the probability density function of the Kotz distribution is

$$f(\boldsymbol{x}) = C_n |\mathbf{\Sigma}|^{-1/2} q^{\alpha-1} e^{-rq^{\beta}}; -\infty < q < \infty; r, \beta > 0; 2\alpha + n > 2, \quad (8.)$$

where the quadratic form q and the normalising constant are, respectively

$$q = (\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}), C_n = \beta\Gamma(n/2) r^{\gamma} / \pi^{n/2} \Gamma(\gamma); \gamma = (2\alpha + n - 2)/2\beta .$$

If it is assumed that vr with $v > 0$ is independently distributed as $\chi^2_{(v)}$, integration over the mixing distribution with $\alpha = 1$ and $\beta = \omega/2$ shows that the probability density function of a symmetric generalised Student distribution is

$$f(\boldsymbol{x}) = K_n |\boldsymbol{\Sigma}|^{-1/2} \left(1 + q^{\omega/2}/v\right)^{-(n/\omega + v/2)},$$
$$K_n = \omega\Gamma(n/2)\Gamma(n/\omega + v/2)/2\pi^{n/2} v^{n/\omega} \Gamma(n/\omega)\Gamma(v/2). \tag{9.}$$

This is also an elliptically symmetric distribution. When $\omega = 1$ it is a multivariate generalisation of the Laplace distribution. The multivariate Student distribution, $\omega = 2$, is well known and its properties are described in Johnson and Kotz (1972, page 132 *et sec*). The detailed properties of this distribution are beyond the scope of this article, although they are special cases of general results for elliptically symmetric distributions as described in Fang, Kotz and Ng (1990). An important property that is used in Section 7.3 below is that the distribution is closed under affine transformations. In particular, if the n-vector $\mathbf{X}$ has the distribution at (9.), the scalar $\boldsymbol{a}^{\mathrm{T}}\mathbf{X}$ has a symmetric $GS\omega v\sigma$ distribution with location parameter $\mu = \boldsymbol{a}^T \boldsymbol{\mu}$ and scale $\sigma = \sqrt{\boldsymbol{a}^T \boldsymbol{\Sigma} \boldsymbol{a}}$. As we describe in Section 7.3, closure under affine transformations is a useful property for applications in portfolio theory for which the vector $\boldsymbol{a}$ represents a vector of investment proportions. Further, if multivariate elliptically symmetric distributions are used as a model for portfolio selection, the extension to Stein's lemma by Landsman and Nešlehová (2008) means that efficient portfolios lie on Markowitz' mean variance efficient frontier. Elliptically symmetric distributions also have useful properties from the perspective of asset pricing and the empirical modelling of returns. This is because conditional distributions have expected values that are linear in the conditioning variables and thus provide general foundations for the market model of Sharpe (1964), as well as the Capital Asset Pricing Model, and the types of factor model due to Carhart (1997) and Fama and French (1993, 2015). It is useful to note that the classification tree presented in this paper may be used with minor modifications to select an appropriate member of the class of multivariate symmetric generalised Student distributions.

A univariate form of (9.) is described in McDonald and Newey (1988). The general form of the Kotz distribution at (8.) is used in Arslan (2004, 2005) to derive another representation of a Student-esque multivariate distribution. In the 2005 paper, the result class of distributions is termed a "*family of t type distributions*" by Arslan. For applications in finance, which are the subject of Section 6, the case $\alpha > 1$ would not arise in empirical work as it implies a zero density at $\boldsymbol{x} = \boldsymbol{\mu}$. There may be a case for $\alpha < 1$, but this implies an infinite density at $\boldsymbol{x} = \boldsymbol{\mu}$.

This is the version of the article accepted for publication in European Journal of Operational Research Vol. 259 (2), 746-765. Published version available from Elsevier at:
https://doi.org/10.1016/j.ejor.2016.10.051

4. Hypothesis Testing Procedures Based on Parametric Classification Trees

The identification of a density function for a given data set involves a sequence of hypothesis tests. We describe our procedure as a parametric classification tree as our objective is to identify the appropriate parameterisation of the density function describing the data set. Our terminology also serves to distinguish our procedure from the use of the term classification tree, which is applied to a non-parametric type of discriminant analysis used in machine learning and data mining (see Breiman et al, 1984). In this case, the objective is to sub-divide multivariate data into homogeneous subsets, constructing a tree by recursively partitioning the data in a training set; for examples, see Harper and Winslett (2006), Razi and Athappilly (2005) and Chou (1991).

The precise properties of the hypothesis testing tree in terms of the probabilities of correct and incorrect classification are very difficult to pin down. These probabilities depend on the number of available observations, the parameter values of the data generating processes and the proportions of different members of the GB or GST families which may be present in a given data set. Furthermore, perusal of the literature shows that there are few papers which are concerned with testing nested hypotheses. In addition, for simple tests of hypothesis the power of the test will depend on the true value of the parameter(s) under consideration. For example, where the choice is between the normal and Student t distributions, the power of the test to reject normality will decrease as the true value of the degree of freedom parameter ν increases. This is because the asymptotic standard error of the maximum likelihood estimator of the degrees of freedom is approximately proportional to υ^2.

The following subsections present classification trees for increasingly general random variables. Considering non-negative variables, Section 4.1 presents the classification trees for the GB2 distribution and notes other families of distributions for which similar model selection procedures could be developed. Considering variables on the whole real line, Section 4.2 presents a detailed description of the parametric classification trees for the GST distribution. Considering multivariate symmetric variables, Section 4.3 describes classification trees for the generalised Student distribution. Section 4.2 contains more analysis than the other two sections because the GST is used for the empirical studies reported in Sections 6 and 7.

4.1 Classification Trees for the Generalised Beta and Other Distributions

Faced with a data set of non-negative observations, it is straightforward to construct parametric classification trees. There are two separate sequences of tests: first, the general to specific test sequence [*GtoS* henceforth] goes from the most general case, the GB2 distribution, to the most specific; secondly, the specific to general [*StoG*] sequence goes from the most specific distribution, the negative exponential, to more general cases. For the general to specific sequence, each test examines a simple constraint on the parameters. The testing sequence is depicted as a tree in Figure 3. For the specific to general sequence, S*toG*, each test removes a restriction on the parameters. The parametric classification tree for the *StoG* tests is shown in Figure 4. In the figures, if the null hypothesis cannot be rejected, the analysis follows the left hand branch, if it can be rejected the right hand branch is followed.

Using the same principles, *GtoS* and *StoG* classification trees may be constructed for the GB1 distribution and for the extensions of the generalised beta which are summarised in Section 3.

Figure 3 - A general to specific parametric classification tree to identify members of the generalised beta (type 2) family of distributions.

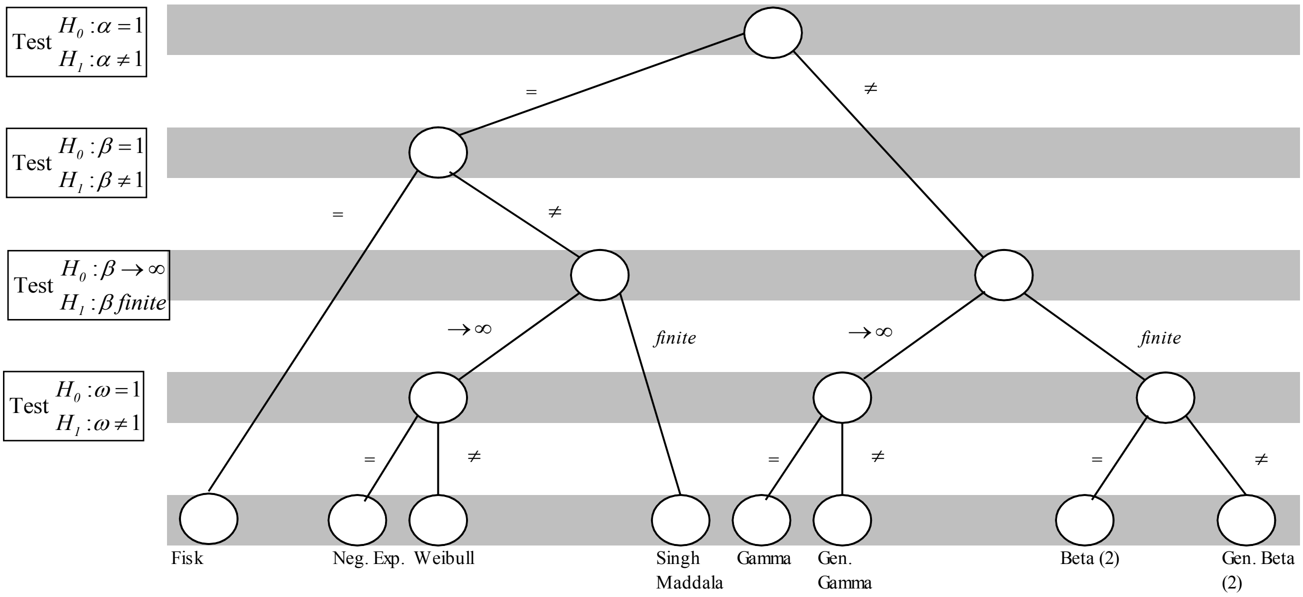
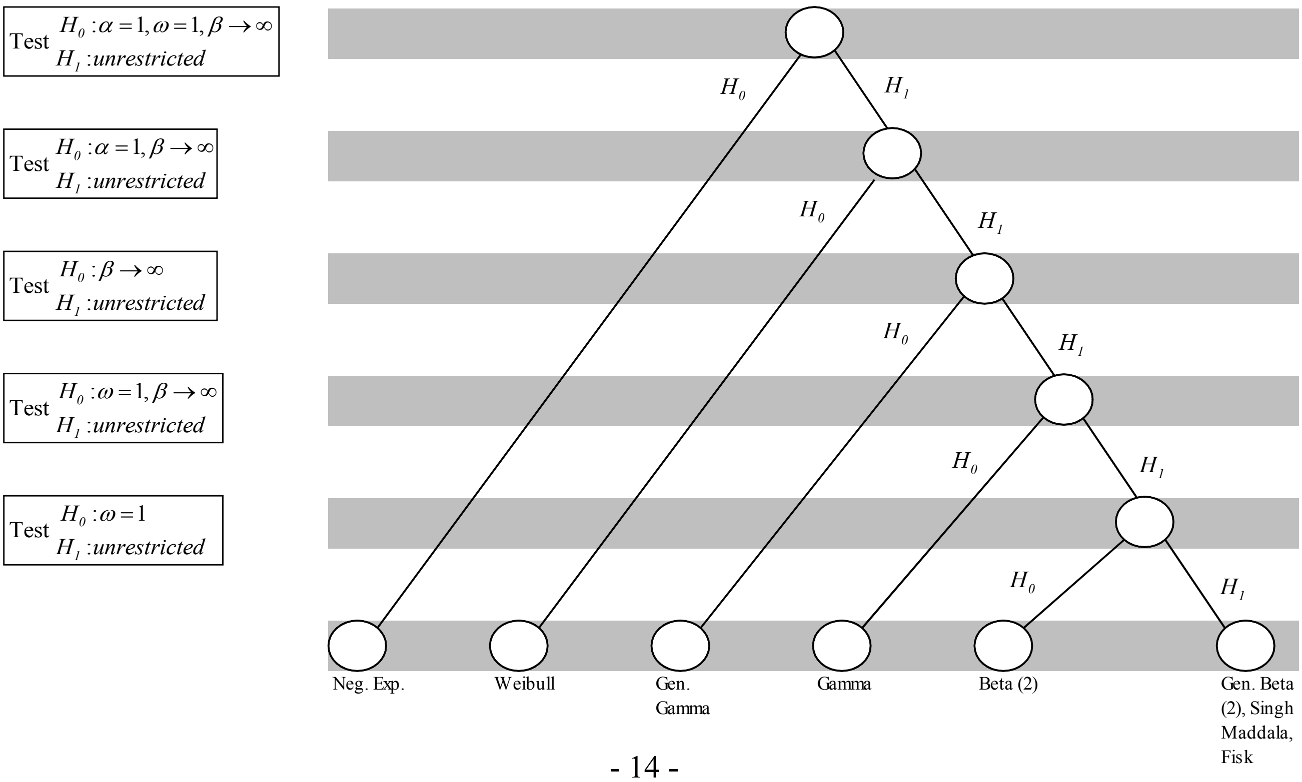


Figure 4 - A specific to general parametric classification tree to identify members of the generalised beta (type 2) family of distributions.



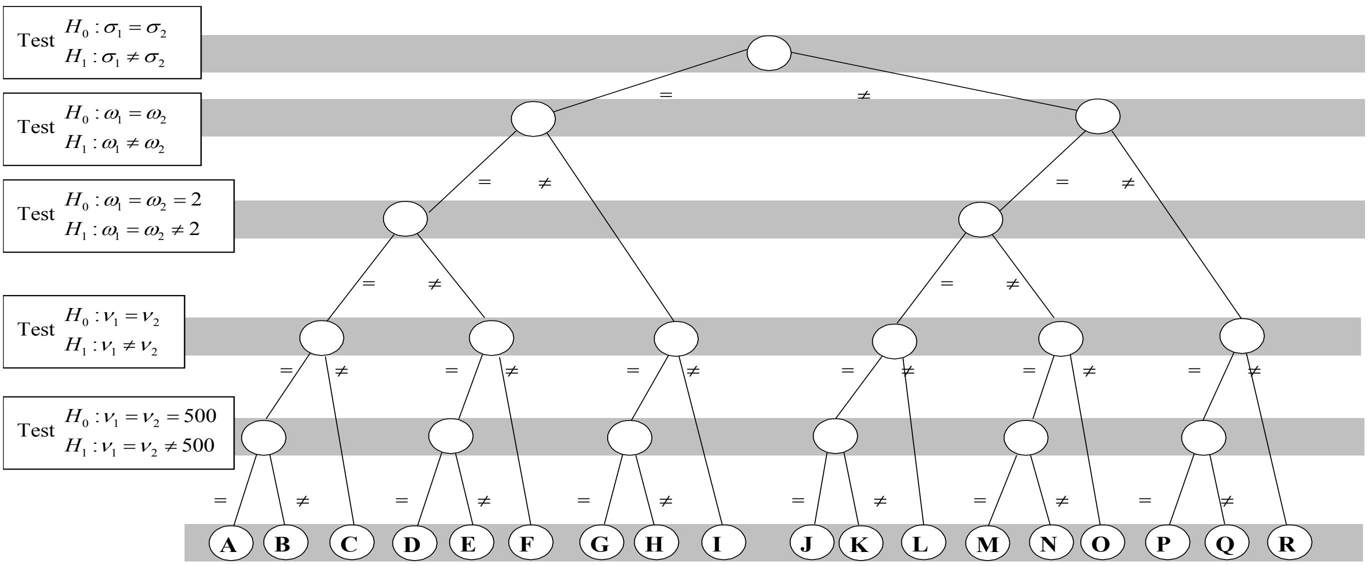
This is the version of the article accepted for publication in European Journal of Operational Research Vol. 259 (2), 746-765. Published version available from Elsevier at:
https://doi.org/10.1016/j.ejor.2016.10.051

4.2 Classification Trees for the GST Distribution

In this case the data may take positive or negative values, as above we consider both testing sequences. First, the general to specific test sequence [*GtoS* henceforth] goes from the most general case, the *GS*--- distribution, to the most specific. Secondly, the specific to general [*StoG*] sequence goes from the most specific distribution, the normal, to more general cases. The *GtoS* testing sequence is depicted as a tree in Figure 5. The analysis commences with the estimation of the unconstrained GST model, *GS*---. Likelihood ratios are used to test parameter restrictions one at a time. If asymmetry is not caused by a particular parameter, for example if there is no reason to doubt $\omega_1 = \omega_2$, then a further check for Studentness is carried out. Similarly, if there is no reason to doubt $\nu_1 = \nu_2$, then the magnitude of the degrees of freedom is examined (with 500 representing normality). The 18 specific cases are identified alphabetically, with a key below Figure 5. Special cases include: A, the normal distribution; B, Student's t; D, the generalised error distribution; E, the symmetric generalised Student's t. K is Theodossiou's model.

The second sequence of tests, S*toG*, starts with the normal density. Thereafter, each test removes a restriction on the parameters. The parametric classification tree for the *StoG* tests is shown in Figure 6. In this test sequence, there are fewer tests and the final classification is coarser. The justification for the loss of granularity compared to the test sequence in Figure 5 is the expectation that fewer of the more exotic density functions will be identified. This is justified, for the data set used in this paper at least, by the empirical results that are described in Section 6. In the following sections in this paper, the more detailed results from the *GtoS* tree are consolidated so that they are directly comparable with the *StoG* tree. The details are available on request.

Figure 5 - A general to specific parametric classification tree to identify members of the GST family



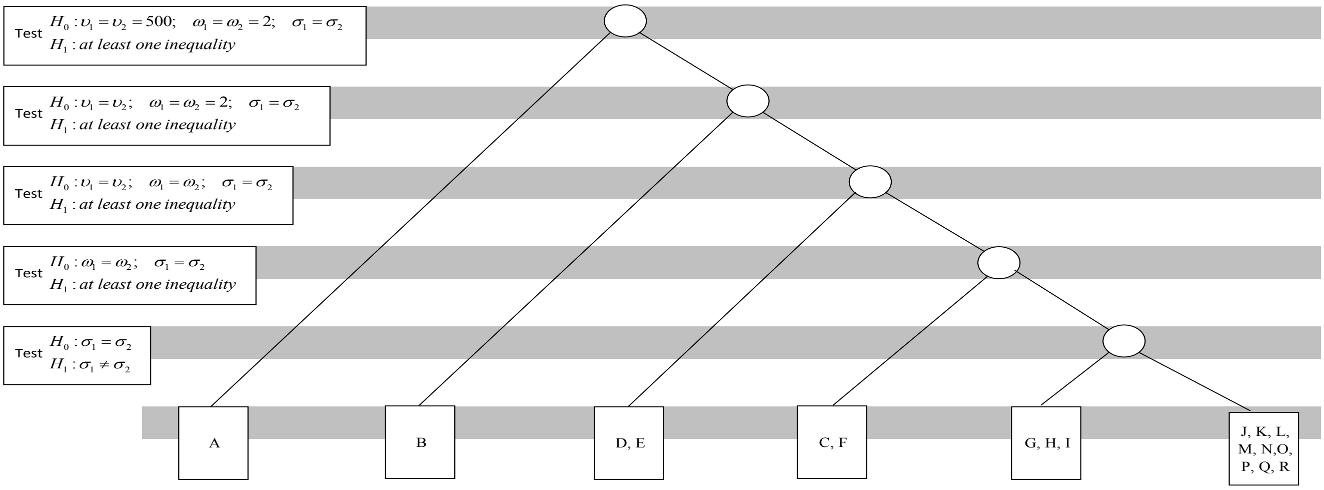
Key:

Bin	A	B	C	D	E	F	G	H	I
Model	*Normal*	*Student's t*	*GTω-σ*	*GEωυσ*	*GSωυσ*	*GSω-σ*	*GE-υσ*	*GT-υσ*	*GS--σ*

Bin	J	K	L	M	N	O	P	Q	R
Model	*GEωυ-*	*GTωυ-*	*GTω--*	*GEωυ-*	*GSωυ-*	*GSω--*	*GE-υ-*	*GS-υ-*	*GS---*

Note: In some tables *Student's t* is denoted *GTωυσ* for brevity.

Figure 6 - A specific to general parametric classification tree to identify members of the GST family



In order to assess the accuracy of the parametric classification trees shown in Figures 6 and 7, a simulation exercise was carried out. Given the well-established role that Student's t distribution plays in both theory and in empirical finance, 100 simulated return series of 5000 observations were generated from Student's t distribution with degrees of freedom set to 1, 3, 5, 8, 12, 15, 20, 35 and 50. The location parameter was set to zero and the scale parameter to 0.003 and 0.032. If the data is assumed to be daily returns, these values correspond to annual scales of about 5% and 50% respectively. For the resulting 18 sets of data, estimation was carried out using sample sizes of 500, 1000, 2500 and 5000.

Table 1 – Simulated Performance of the Classification Trees for Student's t and the Laplace Distribution

	Reduced *GtoS* Classification, Tree D oF υ									*StoG* Classification Tree, D oF υ								
m	*1*	*3*	*5*	*8*	*12*	*15*	*20*	*35*	*50*	*1*	*3*	*5*	*8*	*12*	*15*	*20*	*35*	*50*
(i) Student, σ=0.003																		
500	100	100	100	77	37	21	12	2	2	100	100	95	37	9	6	2	0	0
5000	100	99	99	98	99	100	92	47	23	100	100	100	100	100	99	68	13	1
(ii) Student, σ=0.032																		
500	100	100	99	78	35	20	14	7	0	100	100	90	47	12	5	5	1	0
5000	100	100	100	99	100	100	95	48	14	100	100	100	100	100	98	67	14	1
(iii) GSωνσ, ω = 1 [Laplace] and σ=0.032																		
500	39	57	53	46	36	32	33	38	31	11	46	29	24	17	15	15	18	13

5000	100	97	98	97	99	99	99	100	98	100	100	100	100	100	100	99	94	94

The cells in the table show the number of stocks (out of 100) that are correctly classified by the tree. Vertical section 1 reports results for the *GtoS* tree, Section 2 the *StoG* tree. Results are reported for the range of degrees of freedom shown as column headings and sample sizes shown as row headings. Panel *(i)* shows results For Student's t distribution when the scale parameter equals 0.003, Panel *(ii)* shows the corresponding results for scale equal to 0.032. Panel *(iii)* shows the results for the $GS\omega\nu\sigma$ for the Studentness parameter $\omega = 1$, that is the Laplace distribution, and scale equal to 0.032.

Panels *(i)* and *(ii)* of Table 1 show the percentage of stocks which were correctly classified by each version of the classification tree. The table has two vertical sections, one for each version of the tree. Horizontal panel (i) shows the results when the scale is 0.003 for the four different sample sizes for the *GtoS* tree. As the table shows, the ability of the tree to classify correctly falls as the degrees of freedom increase but rises with sample size. Thus when the degrees of freedom equal 20, 12 out of the 100 stocks are correctly classified when the sample size equals 500. For a sample of 5000, 92 are correctly classified. Panel (ii) shows the corresponding results when the scale equals 0.032. As the panel shows, the results are not substantially affected by increased scale. The second vertical section shows the corresponding results for the *StoG* classification tree. As the results indicate, the *StoG* tree is more conservative. For the *GtoS* tree and degrees of freedom equal to 8 or less, the tree correctly classifies at least 75% of the simulated time series for a sample size equal to 500.

The empirical results reported later in the paper point to the use of the symmetric *GSwnv* distribution for a non-negligible number of securities. Accordingly the simulation exercise was also carried out for the case where the underlying distribution has the Studentness parameter ω set equal to one; that is a generalisation of the Laplace distribution. Scale is set to 0.032 and the degrees of freedom are as above. The results are shown in panel *(iii)* of Table 1.

The results in panel *(iii)* show that for sample size 500 and smaller values of the degrees of freedom, the classification trees do not perform as well as they do when the underlying distribution is Student's t. However, as the degrees of freedom increase, the opposite is true; classification under the Laplace distribution is more accurate than that under Student's t. As the sample size increases, the percentage of correct classifications increases. Furthermore, under the Laplace distribution the results for both versions of the classification tree are more robust to increasing degrees of freedom. In a separate paper, Adcock and Meade (2015) show that this is due at least in part to the fact that the asymptotic variance of the MLE of the Studentness parameter increases relatively slowly with the true value of ω and is orders of magnitude smaller than the asymptotic variance of the MLE of the degrees of freedom υ. Similar to the results in the two Student t panels, the *GtoS* tree tends to be more conservative than the *StoG* tree.

It is accepted that the results shown in Table 1 would render the classification trees unsuitable for some applications. These would be those for which the sample sizes are small and/or for some cases when the degrees of freedom are large. In financial applications, by contrast, the sample sizes are often large and empirical evidence from numerous sources suggests that, where Student's t is an appropriate model, the degrees of freedom are small. Furthermore, as noted above, the power of standard approaches may in any case be low. In the empirical study that is described in the following sections, the data is daily and the sample size is 500, which corresponds to about two years of data. The results in Table 1 imply that that there will be a number of securities that are misclassified. However, as is shown in the results below, the effects of such misclassification is acceptably low.

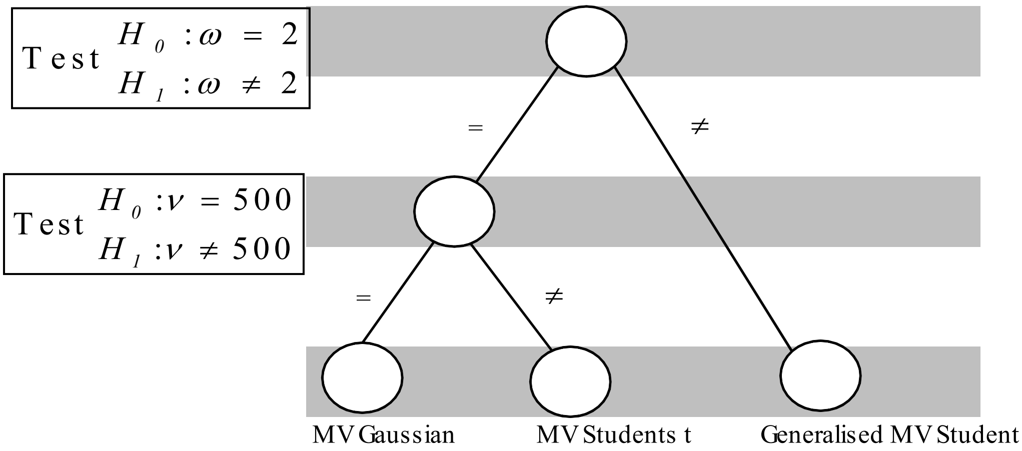
Similar principles may be employed to develop trees for other families of distributions which embody special cases. One example is the extended skew-Student distribution (Azzalini and Capitanio, 2003; Adcock, 2010; Arellano-Valle and Genton, (2010). When residuals follow a distribution which is a member of any of the families referred to in this section, classification trees may also be employed for regression models or models with GARCH effects as long as these are given.

4.3 Classification trees for Multivariate Symmetric Generalised Student Distributions

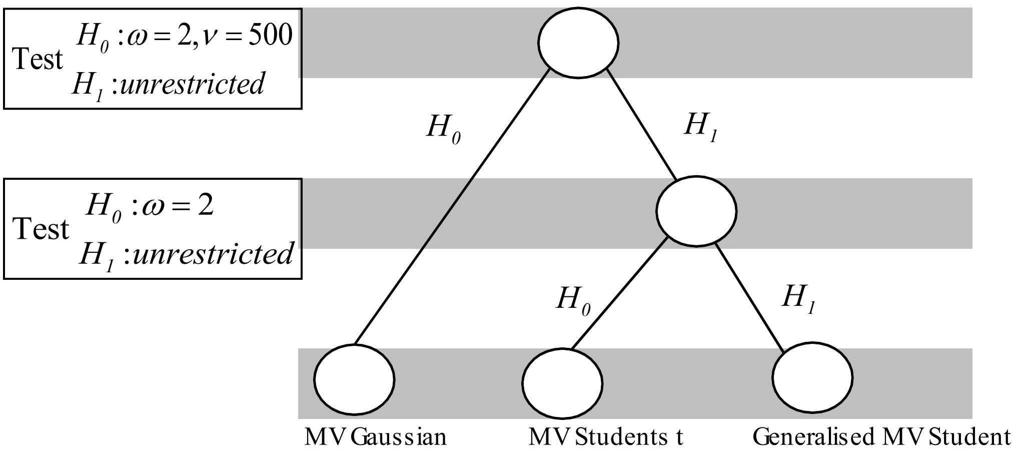
Trees may also be developed for some multivariate distributions. We show *GtoS* and *StoG* parametric classification trees for the case of the generalised Student summarised in Section 3.3 in Figure 7.

Figure 7 - Parametric classification trees for the choice of members of the generalised multivariate student distribution

General to Specific



Specific to General



5. Description of Data

Following the findings of Bekaert *et al* (1998) that skewness and kurtosis are more prevalent in less mature markets, the data set used in this paper was chosen to include sets of equity returns from both mature and emerging markets. Equities from a number of mid-capitalisation indices were also included. This is to allow the comparison of the occurrence of non-normal behaviour in large and mid-capitalised companies in the same national market. Table 2 shows the indices used in the study. In the text of the paper and other tables the names used in the *Country* column of Table 2 are used. The data set consists of time series of daily logarithmic returns for the equities that are members of the indices listed.

Table 2 - Indices used in the study

Country	Mnemonic	Country	Mnemonic
(i) Developed Markets		(ii) Mid-cap Indices	
Australia	ASX50	Australia(MC)	ASXM50
Canada	TSX60	Canada(MC)	TSX Midcap
France	CAC40	UK(MC)	FTSE250
Germany	DAX30	(iii) Emerging and Other Market Indices	
Japan	Nikkei 225	Estonia	OMX Tallinn
UK	FTSE100	Hungary	BUX
USA	S&P500	Poland	WIG
		Kenya	NSE
		South Africa	JSE

Note: In the text of the paper and other tables the names used in the Country column are used.

The data set is divided into three sets using a sample size of 500 days. The three sets used non-overlapping estimation windows, referred to below as *period0, period1* and *period2*, ending on 9th May 2007, 8th April 2005 and 9th May 2003 respectively[2]. In each window, a stock was included in the estimation process if 500 days of return data were available.

Basic descriptive statistics for all the stocks included in the study are available on request. A short summary is in Table 3 for *period0*. For each of the stocks in the 15 indices included in the estimation process for *period0*, Table 3 shows the average values of average return and volatility. It also shows the average values of the Jarque-Bera Test and the average p-values. Also shown are the average values of the components of the test and the corresponding p-values. As the table indicates, there is strong evidence of non-normality. However, on average and according to the Jarque-Bera test, it is due more to kurtosis than skewness.

Table 3 – Summary Statistics for period 0

Based on 500 days of daily data

	Avg	Vol	JBskew	JBskewp	JBkurt	JBkurtp	JBtest	JBprob
	(i) Developed Markets Main Indices							
Australia	0.0009	0.0126	144.0719	0.2166	5853.8235	0.0516	5997.8954	0.0494
Canada	0.0010	0.0173	43.2746	0.2441	1307.6196	0.0146	1350.8942	0.0241
France	0.0006	0.0148	15.0689	0.2479	366.1759	0.0137	381.2448	0.0202
Germany	0.0007	0.0163	21.2923	0.2111	455.0548	0.0001	476.3471	0.0002

[2] These are the dates for UK data. There is some minor variation of dates in some markets due to differences in the number of trading days per year and the number of public holidays.

Japan	0.0010	0.0199	17.7954	0.2572	421.4348	0.0009	439.2301	0.0016
UK	0.0007	0.0146	123.3804	0.1716	5154.0485	0.0174	5277.4289	0.0199
USA	0.0007	0.0161	85.4788	0.1581	3534.5521	0.0135	3620.0309	0.0178
	(ii) Mid-cap Indices							
Australia(MC)	0.0010	0.0160	150.0609	0.1189	5957.9296	0.0314	6107.9906	0.0127
Canada(MC)	0.0008	0.0164	68.1874	0.1457	2202.4620	0.0001	2270.6494	0.0001
UK(MC)	0.0010	0.0158	118.3327	0.0986	4193.1952	0.0004	4311.5279	0.0001
	(iii) Emerging and Other Market Indices							
Estonia	0.0016	0.0249	49.9408	0.0396	1329.2795	0.0000	1379.2203	0.0000
Hungary	0.0012	0.0193	308.7950	0.2210	14575.4574	0.0000	14884.2524	0.0000
Poland	0.0011	0.0179	6.8661	0.1522	84.8408	0.0104	91.7068	0.0092
Kenya	0.0010	0.0243	430.8342	0.0833	39120.5411	0.0000	39551.3753	0.0000
South Africa	0.0014	0.0195	151.9749	0.1465	14095.2913	0.0046	14247.2663	0.0038

The descriptive statistics shown in the column titles are averages over the corresponding values for each stock in the index and used in the study for the *period0* data set. Table entries are shown to 4 decimal places. The index names are listed in Table 2. JB refers to the Jarque-Bera test. The skewness and kurtosis components are shown as well as the test statistics, along with the corresponding p-values.

Corresponding results for *period1* and *period2* are similar and are omitted but are available on request.

6. Empirical Study of the Classification of Stock Returns

This section has 3 sub-sections. The results of using both versions of the parametric classification tree are summarised in Section 6.1. This is followed by an analysis of persistence and then by an analysis of the errors in computed probabilities and quantiles that arise if the model identified by the parametric classification tree is not used. These two sub-sections are based on the *GtoS* tree.

The 18 different models that arise with the GST distributions described at equation (6.) were estimated by the method of maximum likelihood for each stock in each of the 15 indices listed in Table 2. This was done using BHHH algorithm, Berndt *et al* (1974). To avoid the possibility of negative estimates of ω_i, v_i and σ_i^2; $i = 1,2$, a logarithmic transformation was used. The derivatives of the log-likelihood with respect to the parameters were computed numerically.

6.1 Basic Classification

Tables 4 and 5 below show the results of using the two versions of the classification procedure for the data sets for *period0*. The tables show results for a significance level of 1%. Similar results for *period1* and *period2* are omitted but are available in the separate appendix. Results for significance levels of 0.1% and 5% are also available on request. Table 4 shows results for the *GtoS* procedure for a total of 1510 stocks. Although not identified explicitly in Table 4, none of the GE distributions are selected for any stock.

Table 4 - Classification of empirical distributions using the general to specific tree (1% Critical Region)

Based on 500 days of daily data for *period0.*

	NORML	*GTωνσ*	*GEωνσ&GSωνσ*	*GTω−σ & GSω−σ*	*GE−νσ & GS−νσ & GS−−σ*	*All others*	*Total*
	(i) Developed Markets Main Indices						
Australia	11	35	1	2	0	0	49
Canada	6	48	1	0	2	0	57
France	4	37	0	0	0	0	41
Germany	1	30	0	0	0	0	31
Japan	2	210	4	4	5	1	226
UK	5	85	0	4	1	0	95
USA	47	423	0	17	2	0	489
Sub-total	76	868	6	27	10	1	988
As %'age	*7.7*	*87.9*	*0.6*	*2.7*	*1*	*0.1*	
	(ii) Mid-cap Indices						
Australia(MC)	4	36	2	2	1	0	45
Canada(MC)	0	54	0	1	3	0	58
UK(MC)	1	136	47	8	20	3	215
Sub-total	5	226	49	11	24	3	318
As %'age	*1.6*	*71.1*	*15.4*	*3.5*	*7.5*	*0.9*	
	(iii) Emerging & Other Market Indices						
Estonia	0	5	1	1	5	1	13
Hungary	0	7	0	2	2	0	11
Poland	2	13	0	0	0	0	15
Kenya	0	8	2	0	6	2	18
South Africa	6	56	49	6	23	7	147
Sub-total	8	89	52	9	36	10	204
As %'age	*3.9*	*43.6*	*25.5*	*4.4*	*17.6*	*4.9*	
	(iv) Totals						
All	89	1183	107	47	70	14	1510
As %'age	*5.9*	*78.3*	*7.1*	*3.1*	*4.6*	*0.9*	

The index names are listed in Table 2. The table shows the number of stocks classified under the distributions shown in the columns using the General to Specific (*GtoS)* procedure. Specific

For large capitalisation stocks in the developed market indices, 88% of the return series were classified as Student's t distribution and 8% as being normally distributed. Less than 1% were classified as being (symmetric) *GSwnv*. Of the remaining series, less than 4% showed evidence of skewness. Some variation in the percentage classifications may be observed. For Student's t distribution, the percentage varies from 92% for Japan

This is the version of the article accepted for publication in European Journal of Operational Research Vol. 259 (2), 746-765. Published version available from Elsevier at:
https://doi.org/10.1016/j.ejor.2016.10.051

(NIKKEI225) to about 70% for Australia (ASX50). For the normal distribution, the corresponding percentages are less than 1% and 22%.

For the series from mid-capitalisation indices, the frequency of stocks classified as being normally or Student t distributed drops to 2% and 72% respectively. As with the large capitalisation stocks, there is some variation in the percentages in the three markets in this category. About 15% of the stocks are classified as the (symmetric) GSwnv distribution. As the table shows, the vast majority of stocks in this category are members of the UK Mid-Cap FTSE250 or South African JSE indices. There are 12% of stocks which are classified as having skewed return distributions. For the five emerging markets in the study, 4% and 44% are classified as being normally and Student t distributed respectively. For these five indices, 36% of the series are classified as GSwnv and 27% are classified with skewed distributions. Also overall, at *period0* more than ten times as many stocks are classified as Student's t as normally distributed.

Overall, using the *GtoS* classification, about 9% of stocks have skewed return distributions. The detailed breakdown between the 14 skewed versions of the GST distribution that are selected at least once is shown below.

	GSω–σ	*GSωv–*	*GS–vσ*	*GSω—*	*GS—σ*	*GS–v–*	*GS---*	*GTω–σ*	*GTωv–*	*GTω—*
No. of stocks	37	12	82	10	111	17	12	108	5	2
As %'age	0.9	0.3	1.9	0.2	2.6	0.4	0.3	2.5	0.1	0

As this display shows, the most common skewed models for *period0* are *GTω–σ*, *GS—σ* and *GS-νσ*. The *GSων–* model described in Theodossiou (1998) is selected only rarely. The implication is that returns above or below the location parameter do not have different scales, but can have different degrees of freedom or Studentness.

Table 5 shows the corresponding results based on the specific to general procedure. As the table shows, at the 1% level of significance, there are two main differences in the results. First, there is a decrease in the number of stocks classified with skewed distributions, less than 2%. Secondly, the percentages classified as Student t or symmetric GS are similar to those for the *GtoS* procedure, but there is an increase in the number of stocks classified as normally distributed. Overall therefore, for the stocks present in the *period0* data set, the *StoG* method is more conservative than *GtoS*. In the interests of brevity, the corresponding results for the two other periods are omitted. However Table 6, below, implies that the results are similar. Also omitted in the interests of space are the results for significance levels of 5% and 0.1%. The former significance level results in more series being classified with skewed distributions. The latter results in fewer, indeed at the 0.1% level of significance, less than 1% of the 1510 series are classified with skewed distributions using the *GtoS* tree.

Table 5 - Classification of empirical distributions using the specific to general tree (1% Critical Region)

Based on 500 days of daily data for *period0.*

	NORML	$GT\omega v\sigma$	$GE\omega v\sigma \& GS\omega v\sigma$	$GT\omega{-}\sigma$ & $GS\omega{-}\sigma$	$GE{-}v\sigma$ & $GS{-}v\sigma$ & $GS{-}{-}\sigma$	*All others*
	(i) Developed Markets Main Indices					
Australia	17	30	1	1	0	0
Canada	15	41	1	0	0	0
France	12	29	0	0	0	0
Germany	5	26	0	0	0	0
Japan	17	203	3	3	0	0
UK	17	77	0	1	0	0
USA	98	386	0	5	0	0
Sub-total	181	792	5	10	0	0
As %'age	*18.3*	*80.2*	*0.5*	*1.0*	*0.0*	*0.0*
	(ii) Mid-cap Indices					
Australia(MC)	9	32	3	0	1	0
Canada(MC)	5	52	0	1	0	0
UK(MC)	6	150	56	1	1	1
Sub-total	20	234	59	2	2	1
As %'age	*6.3*	*73.6*	*18.6*	*0.6*	*0.6*	*0.3*
	(iii) Emerging & Other Market Indices					
Estonia	0	9	4	0	0	0
Hungary	0	8	2	1	0	0
Poland	2	13	0	0	0	0
Kenya	0	11	5	1	0	1
South Africa	11	68	61	2	4	1
Sub-total	13	109	72	4	4	2
As %'age	6.4	53.4	35.3	2.0	2.0	1.0
	(iv) Totals					
All	214	1135	136	16	6	3
As %'age	14.2	75.2	9.0	1.1	0.4	0.2

The index names are listed in Table 2. The table shows the number of stocks classified under the distributions shown in the columns using the Specific to General to (S*toG)* procedure.

6.2 Persistence

Table 6 reports the persistence of the classification of stocks. The table has two panels; (i) for the *GtoS* tree; (ii) for the *StoG* tree. The classifications are based on those stocks which were included in the estimation and classification for all three periods.

Table 6 – Persistence of the classification of estimated distributions under the general to specific and specific to general trees

Based on 500 days of daily data.

	NORML	*GTωνσ*	*GEωνσ&GSωνσ*	*GTω−σ & GSω−σ*	*GE−νσ & GS−νσ & GS−−σ*	*All others*
	(i) General to Specific Tree					
(a) Period 2 to period 1						
NORML	6	44	0	0	0	0
GTωνσ	52	879	25	19	17	12
GEωνσ&GSωνσ	1	75	73	5	17	5
GTω−σ & GSω−σ	3	51	9	2	3	0
GE−νσ & GS−νσ & GS−−σ	0	27	16	3	16	3
All others	0	8	7	1	4	1
(b) Period 1 to period 0						
NORML	5	54	1	0	2	0
GTωνσ	77	915	27	31	27	7
GEωνσ&GSωνσ	0	70	32	5	20	3
GTω−σ & GSω−σ	0	18	9	2	1	0
GE−νσ & GS−νσ & GS−−σ	0	20	19	3	13	2
All others	0	5	8	2	4	2
	(ii) Specific to General Tree					
(a) Period 2 to period 1						
NORML	33	76	0	0	0	0
GTωνσ	122	888	44	3	4	2
GEωνσ&GSωνσ	4	85	95	1	2	1
GTω−σ & GSω−σ	3	14	1	0	0	0
GE−νσ & GS−νσ & GS−−σ	0	1	4	0	0	0
All others	0	0	1	0	0	0
(b) Period 1 to period 0						
NORML	45	116	1	0	0	0
GTωνσ	152	843	52	10	4	3
GEωνσ&GSωνσ	4	73	62	4	2	0
GTω−σ & GSω−σ	0	2	2	0	0	0
GE−νσ & GS−νσ & GS−−σ	0	1	5	0	0	0
All others	0	0	2	1	0	0

The cell entries are based on stocks which were included in the estimation and classification for all three periods. Sub-panel (a) shows the extent of persistence from *period2* to *period1*. Sub-panel (b) shows the corresponding results for *period1* to *period0*.

In each panel, sub-panel (a) shows the extent of persistence from *period2* to *period1*. Sub-panel (b) shows the corresponding results for *period1* to *period0*. In the first three rows and columns of each sub-panel of panel (i) the data shows that most stocks which are classified as symmetric at the start of the period remain symmetric. For stocks which are classified as Student's t, the majority remain so classified. For stocks which are normal or symmetric GS at the start of the period, the majority remain symmetric, but there are changes in the distribution. For stocks which are skewed at the start of the period, the table suggests that in general skewness does not persist. This finding is consistent with that reported in the skewness literature: the majority of stocks will not exhibit skewness but that for those that do it will be a transient phenomenon. For panel (ii), and consistent with Table 5, the number of stocks with skewed returns is small. The behaviour of stocks with symmetric returns, the changes in classification are similar to the corresponding parts of panel (i): stocks classified as Student's t generally remain so classified. For stocks which are normal or symmetric GS at the start of the period, the majority remain symmetric, but there are changes in distribution. In all sub-panels, the data for the symmetric distributions suggest that kurtosis is a persistent phenomenon.

6.3 Errors in Computed Probabilities and Quantiles

Both the methods described in this paper will result in mis-classifications. The robustness of the procedure to mis-classification is therefore of importance. Here we look at tail probabilities as these values are important in risk management tools such as Value at Risk. Table 7 reports the average errors in computed probabilities corresponding to a nominal probability of 0.01. The table has two panels. Panel (i) shows the average absolute error. Panel (ii) shows the 95% point of the empirical distribution of absolute errors. The rows in the panel correspond to the model as chosen by the general to specific classification. Models which have not been selected are omitted from the table. In the rows of the table, the distributions are identified using the mnemonics defined in Section 3. After each mnemonic the number in parentheses serves to identify the model in each column of the table. The entries in panel (i) are computed as follows. For every stock in the study (at all three periods), the 1% point of the distribution is computed using the model selected by the *GtoS* method and the corresponding estimated parameters. The probability is then recomputed for the stock using all the other 17 distributions. Panel (i) shows the average absolute error. Thus, all entries in the diagonal cells of panel (i) are exactly zero. In row 1, for stocks classified as normally distributed by the GtoS procedure the average absolute errors based on use of the corresponding estimated Student t distribution is 0.002. If the symmetric GS distribution is used, the average absolute error is 0.003. The *NORML* row of panel (i) shows that for stocks classified as normal, the average absolute error is never worse than 0.003.

In general, panel (i) of Table 7 shows that the computation of quantiles at 1% probability is robust to the choice of distribution if the selected model is symmetric and the normal distribution is avoided. For skewed distributions, however, the results are more complex. The normal distribution results in large average absolute errors. Skewed members of the GS class are robust to use of other members of the same class, including the symmetric distribution. Similarly, but only to an extent, skewed members of the GT class are robust to other skewed members of the GT class, but not to Student's t. Panel (ii) of Table 7 shows the results for the 19^{th} vigintile, the 95% percent point of the empirical distribution of absolute errors. The table entries, although larger than those in panel (i), lead to a similar set of comments. When the nominal probability is 0.01, it is inevitable that percentage absolute errors will be large. and the assessment is somewhat different from that of the errors in probabilities. For normally distributed returns, the quantiles are robust to the choice of distribution. For the other two symmetric distributions, correct choice of distribution is more important. For skewed members of the GS class, use of a symmetric distribution or a skewed member of the GT class will result in larger errors of quantiles. For the (skewed) $GT\omega{-}\sigma$ and $GT\omega{-}{-}$ distributions, the 1% quantiles are robust to the use of any skewed GT or GS distribution. By contrast, the $GT\omega\nu{-}$ distribution is not robust. When the 19^{th} vigintile is considered, the conclusions are the same qualitatively.

Table 7 – Absolute errors in computed probabilities for nominal value of 0.01

	1	2	3	4	5	6	7	8	9	10	11	12	13
	(i) Average												
NORML(1)	0.000	0.002	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.002	0.003
GTωνσ(2)	0.006	0.000	0.001	0.001	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.003	0.003
GSωνσ(3)	0.009	0.003	0.000	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
GSων–(4)	0.009	0.004	0.002	0.000	0.002	0.002	0.002	0.001	0.002	0.002	0.004	0.005	0.003
GSω–σ(5)	0.008	0.007	0.004	0.002	0.000	0.002	0.001	0.002	0.001	0.002	0.006	0.005	0.004
GS–νσ(6)	0.007	0.005	0.004	0.001	0.002	0.000	0.002	0.001	0.002	0.002	0.005	0.006	0.005
GSω--(7)	0.008	0.004	0.003	0.003	0.002	0.002	0.000	0.001	0.002	0.000	0.008	0.010	0.006
GS-ν–(8)	0.009	0.004	0.003	0.002	0.001	0.001	0.001	0.000	0.002	0.001	0.008	0.009	0.005
GS--σ(9)	0.009	0.005	0.003	0.003	0.001	0.003	0.001	0.002	0.000	0.001	0.008	0.007	0.005
GS---(10)	0.008	0.006	0.004	0.003	0.001	0.002	0.001	0.002	0.002	0.000	0.009	0.006	0.006
GTων–(11)	0.010	0.010	0.007	0.008	0.009	0.008	0.008	0.008	0.008	0.008	0.000	0.004	0.003
GTω–σ(12)	0.009	0.011	0.012	0.005	0.001	0.002	0.001	0.002	0.001	0.001	0.002	0.000	0.001
GTω--(13)	0.011	0.009	0.010	0.004	0.002	0.000	0.001	0.001	0.004	0.001	0.001	0.001	0.000
	(ii) Empirical 95% point												
NORML(1)	0.000	0.003	0.004	0.004	0.005	0.005	0.005	0.005	0.005	0.005	0.007	0.005	0.006
GTωνσ(2)	0.008	0.000	0.001	0.003	0.005	0.005	0.005	0.005	0.005	0.006	0.005	0.006	0.006
GSωνσ(3)	0.010	0.010	0.000	0.003	0.005	0.004	0.005	0.005	0.005	0.006	0.005	0.006	0.006
GSων–(4)	0.010	0.010	0.004	0.000	0.004	0.004	0.004	0.003	0.005	0.006	0.016	0.016	0.009
GSω–σ(5)	0.010	0.010	0.011	0.009	0.000	0.006	0.003	0.005	0.004	0.005	0.014	0.022	0.010
GS–νσ(6)	0.010	0.010	0.009	0.003	0.005	0.000	0.005	0.003	0.005	0.005	0.024	0.027	0.020
GSω--(7)	0.010	0.010	0.005	0.006	0.004	0.007	0.000	0.004	0.006	0.001	0.027	0.037	0.020
GS-ν–(8)	0.010	0.009	0.006	0.004	0.003	0.004	0.003	0.000	0.004	0.004	0.021	0.028	0.013
GS--σ(9)	0.010	0.010	0.007	0.009	0.004	0.008	0.003	0.004	0.000	0.003	0.025	0.029	0.018
GS---(10)	0.010	0.011	0.010	0.008	0.004	0.007	0.004	0.004	0.003	0.000	0.024	0.017	0.017
GTων–(11)	0.010	0.010	0.009	0.009	0.010	0.009	0.010	0.009	0.010	0.009	0.000	0.004	0.003
GTω–σ(12)	0.018	0.016	0.017	0.009	0.001	0.004	0.001	0.004	0.002	0.003	0.006	0.000	0.003
GTω--(13)	0.015	0.015	0.016	0.006	0.002	0.001	0.001	0.001	0.007	0.001	0.001	0.002	0.000

For every stock in the study (at all three periods), the 1% point of the distribution is computed using the distribution selected by the *GtoS* method and the corresponding estimated parameters. The probability is then recomputed for the stock using all the other 17 distributions. Panel (i) shows the average absolute error. Panel (ii) shows the 95% of the empirical distribution of absolute errors.

Table 8 shows summary statistics for the estimated parameters for Student's t and the GSwnv distributions. The table was constructed from the parameter estimates for all stocks in each of the three estimation windows which were classified by the *StoG* tree as Student's t and *GSωνσ* respectively. The results for the degree of freedom parameter ν and Studentness parameter ω are shown rounded to one decimal place. Results for location and scale are shown to four decimal places. As the number of stocks classified as having a skewed distribution is small even under the *StoG* tree, the corresponding summary of the parameters is omitted, as is the summary for normally distributed stocks.

Table 8 – Summary statistics for the parameter estimates for Student's t and *GSωνσ* distribution.

	Student's t			*GSωνσ*			
	μ	ν	σ	μ	ω	ν	σ
	(i) Summary statistics						
Avg	-0.0001	5.3	0.0160	-0.0001	1.9	5.8	0.0155
Vol	0.0011	4.5	0.0084	0.0010	0.2	6.3	0.0084
Skew	-1.1521	9.7	1.2560	-1.1995	-1.9	14.3	1.1338
Kurt	6.8980	173.9	6.5393	7.3360	4.9	339.8	5.8083
(ii) Percentiles of empirical distribution							
Min	-0.0073	1.0	0.0000	-0.0073	0.7	0.6	0.0010
5%	-0.0020	1.0	0.0039	-0.0019	1.3	1.5	0.0028
10%	-0.0014	2.0	0.0073	-0.0013	1.4	2.9	0.0061
20%	-0.0008	3.0	0.0096	-0.0007	1.9	3.5	0.0089
80%	0.0007	6.5	0.0214	0.0006	2.0	6.9	0.0210
90%	0.0011	8.5	0.0255	0.0010	2.0	8.9	0.0253
95%	0.0014	10.5	0.0311	0.0013	2.0	11.5	0.0303
Max	0.0036	112.0	0.0715	0.0036	2.0	203.6	0.0627

The results in this table are based on all stocks in the study and all three periods; that is a total of 4346 stocks with each estimated parameters based on 500 days of data.

Table 8 has two vertical sections. The first shows summary statistics for stocks classified as having a Student's t distribution. The second shows the corresponding statistics for stocks classified as *GSωνσ*. The first horizontal panel of Table 8 shows the conventional summary statistics. The average values of the location and scale parameters are similar under both distributions. Under Student's t the average degrees of freedom is 5.3. Under the *GSwnv* distribution, the average degrees of freedom parameter and Studentness parameters are 1.9 and 5.8 respectively. Note that the equivalent degrees of freedom suggested by the tail behaviour of the *GSwnv* is 1.9*5.8/2 = 5.5. The lower panel of Table 8 shows the lower and upper vigintiles of the empirical distribution for each parameter estimate. As the panel shows, for stocks classified as Student's t about 20% of stocks have degrees of freedom equal to 3 or less and only 5% have degrees of freedom greater than 10.5. For the same percentiles, the ranges of ω and υ for stocks classified as *GSωνσ.* are 1.9 to 2 and 3.5 to 11.5, respectively.

7. Portfolio Risk Management under Generalised Multivariate Student Distributions

Student's t distribution is a well-established model for asset returns. In addition to empirical modelling, the multivariate Student and other multivariate elliptically symmetric distributions play an important role in financial theory, see for example Chamberlain (1983) and Owen and Rabinovitch (1983). As noted in Section 3.3, if multivariate elliptically symmetric distributions are used as a model for portfolio selection, the extension to Stein's lemma by Landsman and Nešlehová (2008) means that efficient portfolios lie on Markowitz' mean variance efficient frontier. As also noted in Section 3.3, closure under affine transformations means that it is also straightforward to compute quantiles of the distribution of portfolio returns. Two computations that are widely used in finance for risk management purposes are Value at Risk (*VaR*) and Conditional Value at Risk (*CVaR*). The former is a quantile of the distribution of portfolio returns corresponding to a specified probability α and is thus of general applicability. For applications in finance, the specified probability is invariably small, in the left tail of the distribution, and the quantile thus represents a loss that would be exceeded with probability α. *CVaR* is a closely related concept and is defined as the expected value of return given that it is less than the *VaR*. Using the notation defined at equation (6.), X^* is the α-quantile of the distribution and is defined as the solution to

$$\alpha = \int_{-\infty}^{X^*} K\left(1+|(x-\mu)/\sigma|^{\omega}/v\right)^{-\upsilon/2-1/\omega} dx;\ K = \omega\Gamma(v/2+1/\omega)/2\Gamma(v/2)\Gamma(1/\omega)\sigma v^{1/\omega}. \quad (10.)$$

Quantiles in the left tail of the distribution are invariably negative and so *VaR* is defined as being equal to $-X^*$. The corresponding value of *CVaR* is defined as

$$CVaR = -\alpha^{-1}\int_{-\infty}^{X^*} Kx\left(1+|(x-\mu)/\sigma|^{\omega}/v\right)^{-\upsilon/2-1/\omega} dx. \quad (11.)$$

As well as being used in its own right, CVaR is a constituent of the Rachev ratio (Rachev *et al*, 2007), which is also a well-known measure of portfolio performance. It is clear that accurate computation of both CVaR and Rachev ratios depend in turn on accurate computation of value at risk itself. The effect of departures from a normal distribution on commonly used risk measures, such as value at risk or conditional value at risk (also called expected shortfall) have been considered by several operational researchers including Stoyanov *et al* (2013), Goh *et al* (2012), Natarajan *et al* (2008) and Bhattacharyya *et al* (2008). In addition there is a large finance literature covering these two risk measures.

7.1 Variation in VaR and CVaR under Different GST Distributions

Suppose that a portfolio is constructed under the assumption of multivariate normality and is designed to have a specified mean μ_0 and standard deviation or volatility σ_0. If the multivariate Student distribution is used to achieve the same mean and variance, the scale of portfolio returns must be set to $\sqrt{(\upsilon-2)/\upsilon}\sigma_0$. In general, for the $GS\omega v\sigma$ distribution the scale is reset to

$$\left\{2\Gamma(\upsilon/2)\Gamma(1/\omega)/\upsilon^{2/\omega}\,\Gamma(3/\omega)\Gamma(\upsilon/2-2/\omega)\right\}^{1/2}\sigma_0.$$

The implication of the extension to Stein's lemma is that under this approach the resulting portfolios will be the same, because they are located at the same point on the efficient frontier. The tail probabilities will, however, differ depending on the model employed.

Table 9 shows the quantiles (*VaR)* and *CVaR* for a number of symmetric univariate $GS\omega\nu\sigma$ distributions. The Table has two vertical sections. The first reports results for *VaR* and the second the corresponding values of *CVaR*. Each vertical section has 6 columns corresponding to the degrees of freedom shown. The Table has four horizontal panels. Panels (i) and (iii) show results for the generalised Laplace and Student's t distributions. Panels (ii) and (iv) show results for the $GS\omega\nu\sigma$ distributions with ω set to 1.5 and 2.5 respectively. In panel (i) the column for 3 degrees of freedom has NAs because the variance does not exist.

Table 9 –Examples of Value at Risk and Conditional Value at Risk

	Value at Risk						Conditional Value at Risk					
Prob'y	3	5	15	40	100	Inf	3	5	15	40	100	Inf
(i) Generalised Laplace (ω = 1.0)												
0.0005	NA	9.093	6.392	5.395	5.081	4.885	NA	15.563	8.026	6.367	5.884	5.592
0.0010	NA	6.743	5.455	4.765	4.539	4.394	NA	11.647	6.944	5.704	5.331	5.101
0.0050	NA	3.251	3.585	3.386	3.309	3.256	NA	5.827	4.787	4.252	4.076	3.963
0.0100	NA	2.316	2.895	2.825	2.791	2.766	NA	4.268	3.991	3.662	3.548	3.473
0.0250	NA	1.417	2.076	2.113	2.117	2.118	NA	2.770	3.046	2.912	2.861	2.825
0.0500	NA	0.926	1.519	1.596	1.616	1.628	NA	1.951	2.404	2.368	2.349	2.335
0.1000	NA	0.553	1.012	1.096	1.122	1.138	NA	1.331	1.818	1.842	1.845	1.845
(ii) Generalised Student ω = 1.5												
0.0005	7.914	7.018	4.622	4.095	3.930	3.826	14.310	9.749	5.389	4.599	4.363	4.218
0.0010	5.788	5.735	4.134	3.744	3.618	3.538	10.493	8.016	4.869	4.249	4.060	3.942
0.0050	2.759	3.489	3.063	2.914	2.862	2.828	5.073	5.006	3.738	3.429	3.329	3.266
0.0100	1.980	2.761	2.623	2.546	2.517	2.498	3.689	4.039	3.277	3.069	3.000	2.956
0.0250	1.245	1.957	2.051	2.042	2.037	2.033	2.396	2.987	2.687	2.582	2.546	2.522
0.0500	0.845	1.442	1.619	1.643	1.649	1.653	1.705	2.326	2.249	2.202	2.185	2.173
0.1000	0.536	0.986	1.180	1.218	1.230	1.238	1.186	1.755	1.812	1.805	1.802	1.799
(iii) Student's t												
0.0005	7.462	5.321	3.792	3.461	3.356	3.291	11.233	6.749	4.262	3.789	3.644	3.554
0.0010	5.897	4.565	3.475	3.223	3.142	3.090	8.897	5.821	3.939	3.559	3.441	3.367
0.0050	3.372	3.123	2.743	2.636	2.599	2.576	5.146	4.067	3.200	2.999	2.933	2.892
0.0100	2.622	2.606	2.423	2.362	2.340	2.326	4.043	3.449	2.883	2.742	2.695	2.665
0.0250	1.837	1.991	1.984	1.970	1.964	1.960	2.910	2.728	2.456	2.380	2.355	2.338
0.0500	1.359	1.561	1.632	1.641	1.644	1.645	2.237	2.239	2.123	2.085	2.072	2.063
0.1000	0.946	1.143	1.248	1.270	1.277	1.282	1.681	1.783	1.772	1.762	1.758	1.755
(iv) Generalised Student ω = 2.5												
0.0005	5.972	4.373	3.320	3.089	3.016	2.969	8.169	5.271	3.652	3.329	3.229	3.166
0.0010	4.945	3.864	3.090	2.911	2.853	2.816	6.776	4.678	3.422	3.160	3.077	3.025
0.0050	3.154	2.838	2.537	2.457	2.430	2.412	4.366	3.499	2.880	2.736	2.690	2.660
0.0100	2.573	2.446	2.285	2.237	2.221	2.210	3.594	3.058	2.639	2.536	2.502	2.480

0.0250	1.925	1.956	1.927	1.913	1.908	1.905	2.752	2.521	2.305	2.247	2.228	2.215
0.0500	1.500	1.593	1.627	1.631	1.632	1.632	2.218	2.137	2.033	2.003	1.993	1.986
0.1000	1.103	1.217	1.283	1.297	1.302	1.305	1.748	1.761	1.737	1.727	1.723	1.721

The first vertical section reports results for VaR for a selection of degrees of freedom (columns) and probabilities (rows). The second section reports the corresponding values of CVaR. The four horizontal panels show then results for four specific members of the class of symmetric $GS\omega\nu\sigma$ distributions. In panel (i) the column for 3 degrees of freedom has NAs because the variance does not exist.

Table 9 makes it clear that there can be substantial differences in both the quantiles and the CVaR values, particularly for small probabilities and degrees of freedom. In panel (iii) the columns headed "*Inf*" correspond to the normal distribution. It is interesting to note that the quantiles and CVaR values for Student's t with 100 degrees of freedom differ from the corresponding values for the normal. Overall, the table implies that model selection using the classification tree can make a material difference to computation of tail probabilities and hence to the risk assessment of a given portfolio.

7.2 The Effect of Using Generalised Multivariate Student Distributions

As is well known, and as the results in this paper confirm, the estimated degrees of freedom for stocks classified as Student's t are not the same for all such stocks. Similarly, for stocks classified as $GS\omega\nu\sigma$, there is variation in the estimated values of the degrees of freedom and ω the Studentness parameter. The usefulness of the multivariate Student and multivariate $GS\omega\nu\sigma$ distribution at (9.) for portfolio selection and therefore for computations like VaR and CVaR depends in practice on the extent of such variation. To investigate this, the following computations were carried out. For each stock classified as Student's t using the *GtoS* procedure, the 0.1%, 1% and 5% quantiles were computed using the estimated parameters. For each stock so classified, and with estimated degrees of freedom greater than two, the estimated scale parameter was reset using the formula

$$\tilde{\sigma} = \sqrt{(\hat{\nu}-2)\nu_m/(\nu_m-2)\hat{\nu}}\,\hat{\sigma},$$

where ν_m is the median value of the estimated degrees of freedom for all such stocks. This procedure ensures that the estimated variance of the stock is preserved. Using the revised scale $\tilde{\sigma}$ and the median degrees of freedom, the probability was recomputed corresponding to the 0.01%, 1% and 5% quantiles based on the original estimated parameters.

Panel I of Table 10 shows a selection of the percentiles of the empirical distributions of the recomputed probabilities. The table has two vertical panels, labelled I and II. Sub-panel I-i presents results for stocks which were classified as Student's t using the *GtoS* procedure. The sub-panel has three columns, corresponding to nominal probabilities of 0.1%, 1% and 5%. In the 1% (5%) column of Sub-panel I-i, the median recomputed probability is 0.97% (5.18%). At the rows of the table indicate, at nominal values of 1% and 5%, the majority of recomputed probabilities are close to their nominal values. For example, when the nominal probability is 5%, 90% of the recomputed probabilities lie in the interval 4.38 to 5.62. For a nominal probability of 0.1%, the median recomputed probability is 0.08%, but the variability in the results is greater. Sub-panel I-ii presents the corresponding results when the procedure described above is carried out for all stocks in the study regardless of their classification based on the *GtoS* procedure. As the table shows, when computed to two decimal places the median recomputed probabilities are the same and the empirical distributions are similar to those in Sub-panel I-i. Panel II of Table 10 shows the results of analogous computations for the $GS\omega\nu\sigma$ distribution. That is, the degrees of freedom and Studentness parameters are set to the median values and the scale adjusted so that the estimated variance is preserved. We conclude that for applications which are intrinsically multivariate in nature the multivariate Student and generalised multivariate Student distributions may be used as long as the focus of interest is not in the extreme tails of the distribution.

Table 10 –Comparison of tail probabilities for different GST distributions

	I - Student's t						II - $GS\omega\nu\sigma$					
	(i) Student's t			(ii) All			(i) $GS\omega\nu\sigma$			(ii) All		
	0.10%	1%	5%	0.10%	1%	5%	0.10%	1%	5%	0.10%	1%	5%
Min	0.01	0.67	3.80	0.00	0.44	3.80	0.00	0.11	0.38	0.00	0.11	0.38
5%	0.02	0.77	3.80	0.01	0.71	3.80	0.02	0.66	2.42	0.00	0.50	4.57
10%	0.03	0.83	4.38	0.02	0.77	4.38	0.04	0.75	3.02	0.01	0.56	5.05
20%	0.05	0.88	4.75	0.04	0.85	4.75	0.06	0.84	3.78	0.02	0.64	5.16
50%	0.08	0.97	5.18	0.08	0.97	5.18	0.11	0.95	4.80	0.04	0.77	5.68
80%	0.12	1.03	5.50	0.12	1.03	5.56	0.15	1.01	5.55	0.07	0.84	6.17
90%	0.15	1.03	5.62	0.15	1.03	5.74	0.16	1.03	5.82	0.09	0.86	6.35
95%	0.18	1.03	5.74	0.18	1.03	5.84	0.16	1.05	6.17	0.12	0.95	6.49
Max	0.18	1.03	5.90	0.18	1.03	6.17	0.17	1.08	6.49	2.12	5.90	12.62

For each stock classified as Student's t using the *GtoS* procedure, the 0.1%, 1% and 5% quantiles were computed. For the stocks so classified and with estimated degrees of freedom greater than two, the estimated scale parameter was reset using the formula

$$\tilde{\sigma} = \sqrt{(\hat{\nu}-2)\nu_m/(\nu_m-2)\hat{\nu}}\,\hat{\sigma}\,,$$

where ν_m is the median value of the estimated degrees of freedom for all such stocks. Using the revised scale $\tilde{\sigma}$ and the median degrees of freedom, the probability was computed corresponding to the 0.01%, 1% and 5% quantiles. Panel I-i of Table 10 shows a selection of percentiles of the empirical distributions of the recomputed probabilities. Panel I-ii shows the results of the same procedure applied to all stocks in the study, regardless of their classification by the *GtoS* tree. Panel II of Table 10 shows the results of analogous computations for the $GS\omega\nu\sigma$ distribution.

This is certainly the case for portfolio selection, because the extension to Stein's lemma makes it clear that the necessary input parameters are the vector of expected returns and the covariance matrix. For risk management applications, the implication of Table 10 is that computations based on the tails of the distribution will be sufficiently accurate for practical purposes at both the 5% and 1% levels. Furthermore, the results in the case study described below suggest that use of the multivariate Student and generalised multivariate Student distributions is preferable to procedures based on the multivariate normal even at the 0.1% level of probability.

7.3 Case Study under the Generalised Multivariate Student Distribution

The aim of the empirical study reported in this section is to investigate the effect of the use of multivariate elliptically symmetric distributions for portfolio selection and in particular for its effect on risk management. The use of such a distribution does mean that some information about the marginal distributions is lost, but from the perspective of portfolio selection the fact that all efficient portfolios are located on Markowitz' efficient frontier brings major benefits. It is no longer necessary to devote resources to the search for a 'better' utility function and the time thus liberated may be devoted to other activities. That elliptically symmetric distributions are closed under affine transformations means that it is straightforward to compute VaR and CVaR parametrically. As noted in Section 3.3, the use of multivariate elliptically symmetric distributions also provides theoretical foundations for the use of linear models and further support for the CAPM.

Our empirical study is based on the stocks in the UKFTSE250 index. As shown in Tables 4 and 5, the majority of stocks in the FTSE250 index are classified as either Student's t or $GS\omega\nu\sigma$. The data used for the case study is daily returns from 2nd June 2001 to 9th March 2007 and comprises the 200 stocks which were constituents of the index throughout that period. To get a cross-section of results, portfolios are chosen using 20 sets of 10 stocks plus sets of 20, 50, 100 and a single set of 200 stocks. The parameters of multivariate Student (MVS) and generalised multivariate Student (GMVS) distributions are estimated using the method of maximum likelihood (ML) using 500 observations from 2nd June 2001 to 30th May 2003, with the remaining 985 observations used for out of sample testing. In addition, the sample mean vector and sample covariance matrix are estimated. This is equivalent to ML estimation assuming a multivariate normal (MVN) distribution.

Three sets of minimum variance portfolios based on (i) the sample mean and covariance matrix and (ii) the corresponding estimates from the MVS and GMVS distributions are constructed. No short positions are allowed; that is the portfolio weights satisfy the standard non-negativity conditions and budget constraint

$$w_i \geq 0; \sum_{i=1}^{n} w_i = 1.$$

Three further sets of portfolio are constructed with all weights restricted to being no greater than $2/n$, thus ensuring that at least $n/2$ stocks are selected to achieve a greater degree of portfolio diversification. Finally, three sets of global minimum variance portfolios are also constructed, but with no restrictions on the weights other than the budget constraint. For each portfolio and each estimation method, the weights are used in conjunction with the 985 observations to compute time series of portfolio returns out of sample.

For the purpose of risk management, the estimated parameters from the MVN, MVS and GMVS models are used to compute estimates of Value at Risk (VaR) and Conditional Value at Risk (CVaR) at six specified levels of probability. In percentages these are 5, 2.5, 1, 0.5, 0.1 and 0.05. In addition, the in-sample time series of returns is used with the estimated portfolio weights to compute in-sample portfolio returns. The resulting time series is used to compute empirical estimates, denoted EMP, of VaR and CVaR as at the date of portfolio construction. In the performance measurement literature this procedure is often referred to historical simulation. The assessment of the estimates of VaR is conducted by computing the number of occasions on which the out of sample portfolio return is less than the VaR at a specified probability. Returns which are less than the VaR are known as exceedances. For a time series of length 100, for example, one would expect about five values of the return time series to be less than the 5% VaR. The extent to which the number of exceedances differs from its expected value provides a measure of the adequacy of the assumed multivariate model and the corresponding parameter estimates. Following Christoffersen (1998) the number of exceedances may be tested by assuming that their incidence follows a binomial distribution. The estimation process follows a sequence of likelihood ratio tests using the *StoG* classification procedure described in Section 4.3 and illustrated in Figure 7. Our results found that the MVS distribution is always selected in preference to the MVN and the GMVS in preference to the MVS. The average estimated degrees of freedom and the Studentness parameter are 2.78 and 1.85, respectively. Due to the comparatively large standard errors, we conclude that the null hypothesis of common values for υ and ω would not be rejected.

Table 11 - Summary of VaR exceedances for portfolios chosen from the sets of FTSE250 stocks for a 1 day Holding Period (Out of sample computations from 2nd June 2003 to 9th March 2007)

Model	Probability						Probability					
	0.05%	0.10%	0.50%	1.00%	2.50%	5.00%	0.05%	0.10%	0.50%	1.00%	2.50%	5.00%
	(i) 10 stocks											
MVN	**0.69**	**0.82**	**1.27**	1.61	2.34	3.47	**0.88**	**1.02**	**1.41**	1.72	2.36	**3.18**
MVS	0.01	0.05	0.7	1.27	2.86	5.49	0	0.1	0.84	1.34	2.6	4.22
GMVS	0	0.02	0.45	0.98	2.43	5.07	0	0.02	0.68	1.27	2.61	4.59
EMP	NA	NA	0.27	0.54	1.69	3.43	NA	NA	0.27	0.65	1.39	**3.07**
	(ii) 20 stocks											
MVN	**0.82**	**0.93**	**1.36**	1.7	2.4	3.44	**0.87**	**1.1**	**1.33**	1.62	2.51	3.28
MVS	0	0.05	0.78	1.33	2.79	5.23	0	0.09	0.8	1.23	2.68	4.09
GMVS	0	0	0.52	1.05	2.48	5.01	0	0	0.64	1.22	2.57	4.28
EMP	NA	NA	0.31	0.61	1.58	3.36	NA	NA	0.33	0.74	1.44	**3.12**
	(iii) 50 stocks											
MVN	**0.91**	**1.05**	**1.4**	1.71	2.36	3.27						
MVS	0	0.08	0.85	1.37	2.6	4.72						
GMVS	0	0.01	0.66	1.21	2.51	4.91						
EMP	NA	NA	0.27	0.58	1.57	**3.04**						

Table presents a summary of the VaR exceedances for the UKFTSE250 portfolios. If the number of exceedances is outside either the lower and upper 99% confidence limits based on the normal approximation to the binomial distribution, this denoted by **. VaR exceedances are computed using parametrically computed values of VaR based on each of the MVN, MVS and GMVS distributions. The EMP row shows results obtained when VaR is computed non-parametrically using the empirical data.

Table 11 presents a summary of the VaR exceedances for each of the sets of portfolios described. The time series of out of sample returns is computed. VaR exceedances are computed using parametrically computed values of VaR based on each of the MVN, MVS and GMVS distributions. In each panel, the first three rows show the percentage exceedances for the MVN, MVS and GMVS distributions. The fourth row, EMP, shows the results obtained when VaR is computed non-parametrically using the empirical data. Note that in the empirical rows, the time series length of 985 is not sufficient to allow the computation of VaR (or CVaR) at nominal probabilities set to 0.1% and 0.05%. The numerical values in each cell should be approximately equal to the probabilities shown at the top of the table. Computed exceedances which fall outside the 99% confidence limits are indicated using bold type face in the table. The MVN and the empirical method both produce poor results. For smaller nominal probabilities, MVN and EMP are conservative: the exceedances probabilities are greater than the corresponding nominal values, or the estimated value at risk is too low. For VaR computed using the MVS and GMVS distributions, the values in the same columns are numerically similar and all fall within the 99% confidence limits. Although numerically similar the values in the GMVS row are, however, numerically closer to their nominal probabilities than those of the MVS

This is the version of the article accepted for publication in European Journal of Operational Research Vol. 259 (2), 746-765. Published version available from Elsevier at:
https://doi.org/10.1016/j.ejor.2016.10.051

distribution. Results for CVaR are similar to those for VaR but are omitted to save space. The results are consistent across the sizes of the portfolios considered. Neither the *MVN* distribution nor the empirical computations produce results that are close to the nominal probabilities. As noted above, the estimated value of the Studentness parameter is 1.85. This might be considered to be numerically close to 2; that is, for this data set the differences between the MVS and GMVS distributions are not great. Nonetheless, use of the GMVS distribution leads to VaR values which are consistently more accurate.

The results presented in Table 11 are appropriate for organisations that need to compute Value at Risk on a daily basis. There are however other organisations which have longer planning horizons and would need to compute VaR for longer holding periods. In the following paragraphs we report the results of an investigation into the computation of VaR for holding periods of 5 days (that is, equivalent to one working week) and 21 days (equivalent to one month). A standard approach to a study of longer holding periods would be to estimate the expected return vector and the covariance matrix using holding period returns and then to use the critical values of the standard normal distribution. An alternative method would be to use estimates based on daily returns, but to scale the standard normal critical values by $\sqrt{h}$ where h is the length of the holding period. Both approaches would be motivated by the assumption that the central limit theorem could be relied on. If, however, portfolio returns follow a Student t or generalised Student t distribution, then it is not clear that the assumption of normality could be justified unless the holding period is long. In addition, as the papers by Ghosh (1975) and Walker and Saw (1978) make clear, the distribution of a convolution of Student t variables, and therefore generalised Student t variables, is not straight forward to compute. To address the computation of VaR and related risk management measures for longer holding periods, we have computed the distribution of convolutions of symmetric generalised Student t variables numerically using the trapezoidal rule. As the underlying distribution is unimodal and vanishes at the end points, a large grid allows computations to any specified level of accuracy. The results for a range values of υ and ω around those reported above indicate that in practice critical values of the normal distribution may be used for convolutions of daily returns over 30 days and longer.

For an assessment of VaR computed over 5 and 21 days, the critical values of convolutions of the generalised Student t and Student t distributions computed numerically have been used. To achieve comparability with the results for a holding period of one day, standard bootstrapping has been used to generate out of sample time series of length 985 of 5 and 21 day holding period returns for each of the 200 stocks in the FTSE250 dataset. Table 12 presents a summary of VaR exceedances for the five sets of portfolios for the five day holding period. As above, computed exceedances which fall outside the 99% confidence limits are indicated in bold type face. At the 1% level of significance, there are 2 exceedances computed for the MVS distribution which fall outside the confidence limits. Values for the MVN and GMVS distributions all fall within the 99% confidence limits. At the 2.5 and 1% levels of significance, all values for the MVN and GMVS distributions are actually within 95 % limits. In general the percentage exceedances for the MVN distribution exhibit a smaller absolute error relative to the nominal probability. However, the results indicate that GMVS distribution may be preferable. For example, in the 2.5% column in Table 12 for a 20 stock portfolios the computed exceedances for the MVN and

GMVS distributions are 2.61% and 2.06% respectively. For cautious organisations, VaR computed using the GMVS distribution embodies a element of "safety first". At the 1% level of significance, the GMVS and MVS distributions might also be preferred to the MVN for the same reason. However, it is appropriate to point out that the computed exceedances for MVN distribution generally show a lower absolute error.

Table 12 - Summary of VaR exceedances for portfolios chosen from the sets of FTSE250 stocks for a 5 day Holding Period (Out of sample computations from 2nd June 2003 to 9th March 2007)

	Probability						Probability					
Model	0.05%	0.10%	0.50%	1%	2.50%	5%	0.05%	0.10%	0.50%	1%	2.50%	5%
	(i) 10 stocks						(iv) 100 stocks					
MVN	0.16	0.22	0.69	1.13	2.43	4.73	0.11	0.15	0.57	1.15	2.79	5.08
MVS	0	0.01	0.3	0.99	3.51	**7.63**	0	0	0.19	0.54	3.16	6.26
GMVS	0	0	0.14	1.3	2.05	5.67	0	0	0.02	0.68	1.93	5.48
EMP	NA	NA	0.3	0.58	1.95	4.46	NA	NA	0.3	0.67	2.32	4.49
	(ii) 20 stocks						(v) 200 stocks					
MVN	0.12	0.2	0.66	1.18	2.61	4.94	0.05	0.18	0.74	1.12	2.59	5.13
MVS	0	0	0.21	0.89	3.2	6.64	0	0	0.1	0.61	2.74	5.82
GMVS	0	0	0.08	0.99	2.06	5.24	0	0	0.03	0.68	1.56	4.97
EMP	NA	NA	0.28	0.57	2.38	4.75	NA	NA	0.7	0.97	2.49	4.96
	(iii) 50 stocks											
MVN	0.1	0.18	0.62	1.04	2.54	4.92						
MVS	0	0	0.21	0.69	2.99	**6.89**						
GMVS	0	0	0.03	0.78	1.69	5.46						
EMP	NA	NA	0.25	0.53	2.1	4.72						

See Notes for Table 11

Exceedances based on the empirically computed values of VaR are similar, but have larger absolute errors. At 0.1% and 0.05% probability, there is relatively little to choose between the three parametric distributions. Overall the results in Table12 suggests that the normal distribution could be assumed for VaR computations for a holding period of 5 days and portfolios of 50 stocks or more. Detailed results for the 21 day holding period are omitted but are similar to those in Table 12, although in this case the normal distribution may be assumed for portfolios of more than 10 stocks.

8. Discussion & Conclusions

This paper describes two parametric classification tree methods, specific to general and general to specific, which allow the formal selection of a member of a family of distributions. In the case of non-negative random variables, we have discussed the generalised beta distribution. For random variables on the real line, we have discussed the generalised skew-Student distributions. For multivariate random variables we have discussed the symmetric generalised Student distribution. This is derived as a scale mixture of a special case of a symmetric Kotz distribution. As a member of the elliptically symmetric class, this enjoys the same properties for asset pricing and portfolio selection as the multivariate Student distribution.

The classification methods are applied to daily returns on stocks from a selection of 15 major, mid-cap and emerging markets. Three non-overlapping estimation windows each of 500 days are used. For the most recent window, which is reported in detail in the paper, the two methods produce generally similar results, with the specific to general method being more conservative. The results of the study show under the general to specific tree that the majority of return distributions, about 78%, follow Student's t, but that a non-negligible minority, 7%, follow a symmetric generalised Student distribution. Only about 6% are normally distributed, with the majority belonging to the major market indices.

About 8.5% of stocks have skewed distributions, with the majority following the $GS{-}\nu\sigma$ or $GS{-}{-}\sigma$ distributions. Under the specific to general tree, the number of stocks with skewed distributions is less than 2.0%. For the $GS\omega\nu\sigma$ distribution and both parametric classification trees, the majority of stocks belong either to the UK FTSE250 or South African JSE indices. Generalised error distributions are never selected and Student type distributions, that is distributions for which $\omega = 2$, occur only rarely. Skewness is possibly evidence of market inefficiency which may be found in emerging or smaller markets. Peakedness, when the (symmetric) generalised Student t has $\omega < 2$, may be interpreted as being due to thinner trading. The study confirms a well-known stylised fact about skewness, namely that it tends not to be persistent. By contrast, kurtosis is persistent. The results for the data used in this paper suggest that there is some degree of robustness to the choice of distribution, but that there are exceptions. The implications for financial risk management are that the methodology proposed here enables the classification of assets into the categories described. The parametric tree will pick the best model according to the procedure. This may or may not include skewness, but a major implication is that the nature of non-normality can manifest itself in various ways. It further indicates that there are a significant minority of stocks which exhibit skewness and other forms of non-normality. This classification allows risk measures to be calculated under appropriate assumptions.

Computation of quantiles of the $GS\omega\nu\sigma$ distribution and related quantities shows that there can be differences in tail probabilities even when the mean and variance are the same. The results from a case study of portfolios constructed from UK FTSE250 stocks shows that the multivariate Student and generalised multivariate Student distributions lead to more accurate VaR computations than the multivariate normal distribution or empirical methods, neither of which leads to satisfactory results for the data set considered. We suggest that for organisations that are concerned with daily risk management portfolio selection and subsequent risk management based on the MVS or GMVS distributions may be more satisfactory in practice than procedures based implicitly or explicitly on the multivariate normal. For organisations with a planning horizon of 5 working days, the GMVS distribution is more appropriate for risk management for portfolio with a small number of securities, for example an asset allocation portfolio. For longer planning horizons and for portfolios with a larger number of securities the central limit theorem comes into play and critical values of the normal distribution may be used for risk management.

As noted in the introduction, the GST distribution and parametric classification trees presented in this paper may be used in conjunction with models that have more complex mean effects and time series properties including heterogeneity of variance. The principles of the classification trees may be applied to other families of distributions.

References

Adcock, C. J. (2010) Asset Pricing and Portfolio Selection Based on the Multivariate Extended Skew-Student-t Distribution. *Annals of Operations Research*, **176**, no. 1:221–234.

Adcock, C. J. (2014) Mean-Variance-Skewness Efficient Surfaces, Stein's Lemma and the Multivariate Extended Skew-Student Distribution, *European Journal of Operational Research*, **234**, no. 2:392-401.

Adcock, C. J. and N. Meade (2015) The Symmetric Generalised Student Distribution and its Properties, *Working Paper*.
Adcock, C. J., and K. Shutes (2001) Portfolio Selection Based on the Multivariate Skew Normal Distribution. In Skulimowski, A., editor, *Financial Modelling*, 167-177, Krakow:Progress & Business Publishers.
Aldowaisan, T., M. Nourelfath and J. Hassan (2015) Six Sigma performance for non-normal processes, *European Journal of Operational Research,* **247**, no. 3:968–977.
Almalki, S. J. and J. Yuan (2013) A new modified Weibull distribution, *Reliability Engineering and System Safety*, **111**, March:164–170.
Aparicio, F. and J. Estrada (2001) Empirical Distributions for Stock Returns: European Securities Markets 1990-95, *The European Journal of Finance,* **7**, no. 1:1-21.
Arellano-Valle, R. B., and M. G. Genton (2010) Multivariate Extended Skew-t Distributions and Related Families. *Metron*, **68,** no. 3**:**201-234.
Arslan, O. (2004) Family of multivaraite generalized *t* distributions, *Journal of Multivariate Analysis*, **89**, no. 2:329-337.
Arslan, O. (2005) A new class of multivariate distributions: Scale mixture of Kotz-type distributions, *Statistics & Probability Letters*, **75**, no. 1:18–28.
Azzalini, A. (1985) A Class of Distributions Which Includes the Normal Ones. *Scandinavian Journal of Statistics*, **12**, no. 2:171–178.
Azzalini, A. (1986) Further Results on a Class of Distributions Which Includes The Normal Ones, *Statistica*, **46**, no. 2:199-208.
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**, no. 2:367–389.
Bekaert G., C. R. Harvey, C. B. Erb and T. E. Viskanta (1998) Distributional Characteristics of Emerging Market Returns & Asset Allocation. *Journal of Portfolio Management*, 24, no. 2:102-116.
Berndt, E. K., H. B. Hall, R. E. Hall and J. A. Hausman (1974) Estimation and Inference in Non-linear Structured Models, *Annals of Economic and Social Measurement*, **3**, no. 4:653-666.
Bhattacharyya, M., A. Chaudhary and G. Yadav (2008) Conditional VaR estimation using Pearson's type IV distribution, *European Journal of Operational Research*, **191**, no. 2:386–397.
Blattberg, R. and N. Gonedes (1974) A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices, *Journal of Business*, 47, no. 2:244-80.
Boccanfuso, D., B. Decalué and L. Savard (2008) Poverty, Income Distribution and CGE Modelling: Does the functional Form of Distribution Matter? *Journal of Income Inequality,* **6**, no. 2:149-184.
Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen (1984) *Classification and regression trees*. NewYork:CRC press.
Budge, S., A. Ingolfsson and D. Zerom (2010) Empirical Analysis of Ambulance Travel Times: The Case of Calgary Emergency Medical Services, *Management Science*, **56**, no. 4:716-723.
Carhart, M. M. (1997) On persistence in mutual fund performance. *The Journal of Finance,* **52**, no.1:57-82.
Chamberlain, G. (1983) A characterization of the distributions that imply mean-variance utility functions, *Journal of Economic Theory*, **29**, no. 1:185-201
Chang, D. S. (2000) Optimal Burn-in Decisions for products with an unimodal failure rate function, *European Journal of Operational Research*, **126**, no. 3:534-540
Chen, J. W., K. H. Li and Y. Lam (2007). Bayesian single and double variable sampling plans for the Weibull distribution with censoring. *European Journal of Operational Research*, **177**, no. 2:1062–1073.
Chou, P.A. (1991) Optimal partitioning for classification and regression trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**, no.4:340-354.
Christoffersen, P. F. (1998), Evaluating Interval Forecasts, *International Economic Review,* **39**, no.4:841-862.
Corrado, C. J. and T. Su (1996) Skewness and Kurtosis in S&P500 Index Returns Implied by Option Prices, *Journal of Financial Researc*h, 19, no. 2:175-192.
Dias, J. G., J. K. Vermunt and S. Ramos (2015) Clustering financial time series: New insights from an extended hidden Markov model, *European Journal of Operational Research*, **243**, no. 3:852–864.
Ducey, M.J. and J. H. Gove (2015) Size-biased distributions in the generalised beta distribution family, with applications to forestry, *Forestry*, **88**, 143 - 151.
Fama, E. F. And K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, **33**, no.1: 3–56.
Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics*, **116**, no.1:1-22.

Fang, K.-T., S. Kotz & K.-W. Ng, (1990). *Symmetric multivariate and related distributions*. Monographs on Statistics and Applied Probability, Vol. 36. London:Chapman & Hall, Ltd.
Fernandez, C. and M. F. J. Steel (1998) On Bayesian Models of Fat Tails and Skewness, *Journal of the American Statistical Association*, **93**, no. 441:359-371.
Ghosh, B. K. (1975) On the Distribution of the Difference of Two t-Variables, *Journal of the American Statistical Association*, **70**, No. 350:463-467,
Goh, J.W., K. G. Lim, M. Sim and W. Zhang (2012) Portfolio value-at-risk optimization for asymmetrically distributed asset returns, *European Journal of Operational Research*, **221**, no. 2:397–406.
Gordy, M. B. (1998) Finance and Economics Discussion Series: A Generalization of Generalized Beta Distributions, *Finance and Economics Discussion Papers of the Federal Reserve.*
Graf, E. and Y. Tillé (2014) Imputation of income data with generalized calibration procedure and GB2 law: illustration with SILC data, *Université de Neuchâtel Working Paper*.
Hansen, B. E. (1994) Auto-regressive Conditional Density Estimation, *International Economic Review*, **35**, no. 3:705-730.
Harper, P.R., D. J. Winslett (2006) Classification trees: A possible method for maternity risk, *European Journal of Operational Research*, **169**, no.1:146-156.
Harris, R. D. F., C. C. Kukukozmen and F. Yilmaz (2004) Skewness in the Conditional Distribution of Equity Returns, *Applied Financial Economics*, **14**, no. 3:195-202.
Harvey, C. R., J. C. Liechty, M. W. Liechty, and P. Müller (2010) Portfolio Selection with Higher Moments. *Quantitative Finance*, **10**, no. 5:469–485.
Johnson N L and S Kotz (1972) *Distributions in Statistics, Continuous Multivariate Distributions*, New York: John Wiley and Sons Inc.
Kon, S. J. (1984) Models of Stock Returns- A Comparison, *Journal of Finance*, **39**, no. 1:147-165.
Laih, Y-W. (2015) Measuring rank correlation coefficients between financial time series: A GARCH-copula based sequence alignment algorithm, *European Journal of Operational Research* , **232** no. 2:375–382.
Landsman, Z. & J. Nešlehová (2008) Stein's Lemma for elliptical random vectors, *Journal of Multivariate Analysis*, **99**, no. 5: 912 – 927
Lessman, S., B. Baesens, H. V. Seow and L. C. Thomas (2015) Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research, *European Journal of Operational Research*, **247,** no.1:124–136
Li, M.Z.F. and T. H. Oum (2000) Airline spill analysis - beyond the normal demand, *European Journal of Operational Research* **125**, no. 1: 205-215.
Longin, F. M. (2000) From value at risk to stress testing: The extreme value approach, *Journal of Banking and Finance*, **24**, no. 7:1097 – 1130.
Longin, F. M. (2005) The choice of the distribution of asset returns: How extreme value theory can help, *Journal of Banking and Finance*, **29**, no. 4:1017 – 1035.
Mandelbrot, B. (1963) The Variation of Certain Speculative Prices, *Journal of Business*, **36**, no. 4:394-419.
Mauleon, I. (2006) Modelling Multivariate Moments in European Stock Returns, *The European Journal of Finance*, **12**, no. 3:241-263.
Mauleon, I. and J. Perote (2000) Testing Densities with Financial Data: An Empirical Comparison of the Edgeworth Sargan Density to the Student's t, *The European Journal of Finance*, 6, no. 2:225-239.
McDonald, J. B. (1984) Some Generalized Functions for the Size Distribution of Income, *Econometrica,* **52**, no. 3:647-665
McDonald, J. B. and W. K. Newey (1988) Partially Adaptive Estimation of Regression Models Via The Generalized T Distribution, *Econometric Theory*, **4**, no. 3:428-457.
McDonald, J. B. and R. D. Nelson (1993) Beta Estimation in the Market Model: Skewness and Leptokurtosis, *Communications in Statistics – Theory and Method,* **22**, no. 10:2843-2862.
McDonald, J. B. and Y. J. Xu (1995) A Generalization of the Beta Distribution with Applications, *Journal of Econometrics*, **66**, no. 1:133-152.
Meade, N. (2010) Oil prices — Brownian motion or mean reversion? A study using a one year ahead density forecast criterion, *Energy Economics*, **32**, no. 10:1485 - 1498.
Mielke, P.W. and E. S. Johnson (1974) Some Generalized Beta Distributions of the Second Kind Having Desirable Application Features in Hydrology and Meteorology, *Water Resources Research*, **10**, no. 2:223 - 226.
Natarajan, K., D. Pachamanova and M. Sim (2008) Incorporating Asymmetric Distributional Information in Robust Value-at-Risk Optimization, *Management Science*, **54**, no. 3:573–585.
Nelson, D. B. (1991) Conditional Heteroscedasticity in Asset Returns: A New Approach, *Econometrica*, **59**,

no. 2:347-370.
Okamura, H. and T. Dohi (2015) Phase-type software reliability model: parameter estimation algorithms with grouped data, *Annals of Operations Research*, DOI 10.1007/s10479-015-1870-0
Owen, J. and R. Rabinovitch (1983) On the Class of Elliptical Distributions and their Applications to the Theory of Portfolio Choice, *The Journal of Finance*, **38**, no. 3: 745-752.
Praetz, P. (1972) The Distribution of Share Price Changes, *Journal of Business*, 45, no. 1:49-55.
Rachev, S., T. Jasic, S. Stoyanov and F. J. Fabozzi (2007) Momentum Strategies Based on Reward-Risk Stock Selection Criteria. *Journal of Banking and Finance*, **31**, no.8:, 2325-2346.
Razi, M.A. and K. Athappilly (2005) A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. *Expert Systems with Applications*, **29**, no.1:65-74.
Richards, D.O. and J. B. McDonald (1987) A general methodology for determining distributional forms with applications in reliability, *Journal of Statistical Planning and Inference,* **16**, no. 3:365-376.
Sharpe, W. F. (1964) Capital Asset Prices: A theory of Market Equilibrium, *Journal of Finance*. **19**, no-3:425-442.
Singla, N., K. Jain and S. K. Sharma (2012) The Beta Generalized Weibull distribution: Properties and applications, *Reliability Engineering and System Safety*, **102**, June:5–15.
Singleton, J. C. and J. Wingender (1986) Skewness Persistence in Common Stock Returns, *Journal of Financial and Quantitative Analysis*, **21**, no. 3335-341.
Stacy, E. W. (1962) A Generalization of the Gamma Distribution, T*he Annals of Mathematical Statistics,* 33, no. 3:1187-1192.
Stacy, E. W. and G. A. Mirham (1965) Parameter Estimation for a Generalized Gamma Distribution, *Technometrics*, **7**, no. 3:349-358.
Stoyanov, S.V., S. T. Rachev and F. J. Fabozzi (2013) Sensitivity of portfolio VaR and CVaR to portfolio return characteristics, *Annals of Operations Research*, **205**, no. 1:169–187.
Theodossiou, P. (1998) Financial Data and the Skewed Generalized T Distribution, *Management Science*, **44,** no. 12:1650-1661.
Theodossiou, P. and C. S. Savva (2015) Skewness and the Relation Between Risk and Return, *Management Science,* http://dx.doi.org/10.1287/mnsc.2015.2201
Walker, G.A. and J. G. Saw (1978) The Distribution of a Linear Combination of t-Variables, *Journal of the American Statistical Association*, , **73**, no.364:876-878.
Wang, P., B. Zhang and Y. Zhu (2015) Asymmetries in stock markets, *European Journal of Operational Research,* **241** , no. 3: 749–762
Ye, Y. (2012) Properties of Weighted Generalized Beta Distributions of the Second Kind, MS Thesis, Georgia Southern University
Yu, J-R., W-J. P. Chiou and D-R, Mu (2015) A linearized value-at-risk model with transaction costs and shortselling, *European Journal of Operational Research*, **247,** no. 3:872–878.
Zhang, Q., Z. Xu, T. Feng and J. Jiao (2015) A dynamic stochastic frontier model to evaluate regional financial efficiency: evidence from Chinese county-level panel data, *European Journal of Operational Research,* **241**, no. 3:907-916.
Zhu, D. and J. W. Galbraith (2010) A Generalized Asymmetric Student-t Distribution with Application to Financial Econometrics, *Journal of Econometrics,* **157**, no. 2:297-305.
Zhu, D. and J.W. Galbraith (2011) Modeling and forecasting expected shortfall with the generalized asymmetric Student-t and asymmetric exponential power distributions, *Journal of Empirical Finance*, **18,** no. 4:765-778.