Flexible statistical models: Methods for the ordering and comparison of theoretical distributions

Rigby, Robert and Stasinopoulos, Dimitrios and Voudouris, Vlasios

Statistics, Operational Research and Mathematics Centre, London Metropolitan University, ESCP Europe Business School

13 April 2015

Flexible statistical models: Methods for the ordering and comparison of theoretical distributions

Robert Rigby[1], Dimitrios Stasinopoulos[1] and Vlasios Voudouris[2,*]

[1]Statistics, Operational Research and Mathematics Centre, London Metropolitan University, Holloway Road, London N7 8DB, UK
[2]ESCP Europe Business School, 527 Finchley Road, NW3 7BG, London, UK
[*]email: *vvoudouris@escpeurope.eu*

Abstract

Statistical models usually rely on the assumption that the shape of the distribution is fixed and that it is only the mean and volatility that varies. Although the fitting of heavy tail distributions has become easier due to computational advances, the fitting of the appropriate heavy tail distribution requires knowledge of the properties of the different theoretical distributions. The selection of the appropriate theoretical distribution is not trivial. Therefore, this paper provides methods for the ordering and comparison of continuous distributions by making a threefold contribution. Firstly, it provides an ordering of the heaviness of distribution tails of continuous distributions. The resulting classification of over 30 important distributions is given. Secondly it provides guidance on choosing the appropriate tail for a given variable. As an example, we use the USA box-office revenues, an industry characterised by extreme events affecting the supply schedule of the films, to illustrate how the theoretical distribution could be selected. Finally, since moment based measures may not exist or may be unreliable, the paper uses centile based measures of skewness and kurtosis to compare distributions. The paper therefore makes a substantial methodological contribution towards the development of conditional densities for statistical model in the presence of heavy tails.

1 Introduction

Many observable data are usually characterized by a heavy (right-hand) tail. In recognition of "fat tail" events, applied statisticians usually employ a distribution-free approach. In the presence of heavy tails, nonparametric approaches can lead to very unreliable statistical inferences as discussed by Bahadur and Savage (1956). Although statistical models with parametric theoretical distributions can perform better in the presence of heavy tails, they too can have problems (Davidson, 2012), particularly when the 'right' theoretical distribution is not selected. The important point here is that the modelling and statistics of 'heavy tail' events are tail dependent and much different than classical modelling of business-related data , which give primacy to central moments rather than centile based measures.

Building along lines first proposed by Haavelmo (1943, 1944), the contribution of this work is to provide further support for the development of conditional densities for statistical models by providing methods for the ordering and comparison of theoretical distributions for models of 'fat tail' events. It is important to clarify that knowledge of the properties of the distributions aims to support the selection of the distribution for the statistical model to be fitted. Thus, we distinguish between methods to select a distribution for the statistical model to be fitted and methods to assess the fitted distribution by examining the residuals of the fitted statistical model. We provide methods for the former rather than the latter.

Furthermore, by comparing different distributions (rather than focusing on an individual distribution), the proposed classification of theoretical distributions is a useful guide in developing models from a list of theoretical distributions (some expert systems have more than 80 distributions - see the gamlss package in R) with the appropriate heavy tail distribution when flexible modelling tools are used to analyze processes and/or systems characterized by highly skew and/or kurtic data.

The introduction of the generalised additive models for location scale and shape (GAMLSS) framework (proposed by Rigby and Stasinopoulos, 2005) and vector generalized additive model (VGAM) framework (proposed by Yee and Wild, 1996) have enhanced the ability of applied statisticians to develop models by selecting appropriate distributions from an extensive list of theoretical distributions. While there are flexible modeling frameworks to model 'fat tail' events by selecting an appropriate theoretical distribution, guidelines on how to choose a specific distribution and why some distributions are preferable in a certain application to others is missing. We use a dataset (derived from industry standard data sourced by Nielsen EDI for 13 years from 1988 to 1999) of the North American film market during the 1990s. The market for films in theatrical release is an experimental place in which consumers quest for aesthetic novelty but ex-ante are uncertain about the quality of the films they select to watch. For their part producers are uncertain about how best to satisfy the not perfectly understood preferences of consumers, while the task of the other two agencies in the supply triumvirate - distributors and exhibitors - is to satisfy demand once revealed, making films that consumers wish to see much less scarce than those films to which they are not so attracted. This is an industry that supply adapts to demand dynamics by means of adaptive contracts (e.g., film rental differentials and booking periods). Thus, density forecasts of film performance is important to dynamically manage the emerging demand for blockbusters.

The tail of the distribution of a continuous random variable Y is commonly ordered based on the survivor function $\bar{F}(y) = 1 - F_Y(y)$, for the right tail and the cumulative distribution function of Y, $F_Y(y)$ for the left tail. In Section 2 it is shown that an ordering based on the log of the probability density function $\log f_Y(y)$ results in the same ordering for the probability density function $f_Y(y)$ and the same ordering of $\bar{F}(y)$ for the right tail and $F(y)$ for the left tail. The resulting classification (type I, II and III) of important distributions on the real and positive real line is also given. A film revenue data example is given.

A detailed comparison of distributions on the real line based on centile measures of skewness and kurtosis is given in Section 3. Moment based measures of skewness and kurtosis have traditionally been used to compare distributions. However moment based measures suffer from being affected by an extreme tail of the distribution which may have negligible probability. In particular it has been shown by Ali (1974) that a sequence of random variables can be constructed which converge in distribution (uniformly) to the normal distribution, yet the moment based kurtosis tends to infinity, proving that moment based kurtosis is not a reliable measure of the shape of the distribution. For certain famous distributions, e.g. the Stable distribution and the t distribution with degrees of freedom $d \leq 4$, no finite moment based skewness and/or kurtosis exists. For the above reasons, in this paper we use centile based measures of skewness and kurtosis, which exist for all distributions, to compare distributions. Conclusions are given in Section 4.

2 Ordering heaviness of tails of continuous distributions

2.1 Types of tail

The heaviness of the tail of a continuous distribution is ordered here based on the log of the probability density function. If random variables Y_1 and Y_2 have continuous probability density functions $f_{Y_1}(y)$ and $f_{Y_2}(y)$ and $\lim_{y\to\infty} f_{Y_1}(y) = \lim_{y\to\infty} f_{Y_2}(y) = 0$ then Y_2 has a heavier right tail than $Y_1 \Leftrightarrow \lim_{y\to\infty} [\log f_{Y_2}(y) - \log f_{Y_1}(y)] = \infty$. The resulting ordering of $\log f_Y(y)$ for the right tail of Y results in the same ordering for the probability density function $f_Y(y)$, [where Y_2 has a heavier tail than $Y_1 \Leftrightarrow f_{Y_1}(y) = o\left[f_{Y_2}(y)\right]$ as $y \to \infty$ by Lemma B1 in Appendix B], and also the same ordering as the standard ordering for the survivor function $\bar{F}_Y(y) = 1 - F_Y(Y)$ where $F_Y(y)$ is the cumulative distribution function, [where Y_2 has a heaver tail than $Y_1 \Leftrightarrow \bar{F}_{Y_1}(y) = o\left[\bar{F}_{Y_2}(y)\right]$, by Lemma B2 in Appendix B]. Similarly for the left tail of Y.

There are three main forms for $\log f_Y(y)$ for a tail of Y, i.e. as $y \to \infty$ (for the right tail) or as $y \to -\infty$ (for the left tail), $\log f_Y(y) \sim$

Type I: $-k_2 \, (\log|y|)^{k_1}$,

Type II: $-k_4 \, |y|^{k_3}$,

Type III: $-k_6 \, e^{k_5|y|}$,

in decreasing order of heaviness of the tail. For $-k_2 \, (\log|y|)^{k_1}$, decreasing k_1 results in a heavier tail, while decreasing k_2 for fixed k_1 results in a heavier tail. Similarly for $-k_4 \, |y|^{k_3}$ with (k_3, k_4) replacing (k_1, k_2) and $-k_6 \, e^{k_5|y|}$ with (k_5, k_6) replacing (k_1, k_2). Important special cases are $k_1 = 1$, $k_1 = 2$, $k_3 = 1$ and $k_3 = 2$. To avoid unnecessary cluttering references for the distributions considered in this paper are given in Table A1 in Appendix A.

2.2 Classification Tables

Tables 1 and 2 provide a summary of many important distributions on the real line and positive real line respectively. Many of the distributions in Tables 1 and 2 have important special cases. For example, the generalized beta type 2 distribution, GB2(μ, σ, ν, τ), also known as the generalized beta-prime distribution and the generalized beta of the second kind, includes special cases the Burr III (or Dagum) distribution when $\tau = 1$, the Burr XII (or Singh-Maddala) when $\nu = 1$, (Johnson *et al.*, 1994, p 54), a form of Pearson type VI when $\sigma = 1$, (Johnson *et al.*, 1995, p 248), the generalized Pareto distribution when $\sigma = 1$ and $\nu = 1$ and the log logistic when $\nu = 1$ and $\tau = 1$. The skew exponential power type 3 distribution, SEP3(μ, σ, ν, τ) includes the skew normal type 2 when $\tau = 2$, (Johnson *et al.*, 1994, p 173) .

The parametrizations of the distributions (column 2 of Tables 1 and 2) are those used by the open source R-based GAMLSS tool. This parameterization was chosen for consistency with a highly flexible and open source modelling tool with extensive documentation. The parameters for all distributions (up to four parameters) are defined as μ, σ, ν and τ. Note that μ and σ are (usually location and scale) parameters and not, in general, the mean and standard deviation of the distribution, while ν and τ are usually skewness and kurtosis parameters. Some distributions are parameterized in two different ways, for example JSU and JSUo. For many distributions the left and right tails have the same asymptotic form for $\log f_Y(y)$, otherwise the relevant tail is specified in the table, see e.g. Gumbel distribution. Some distributions have different tail forms dependent on a condition on one (or more) parameters, see e.g. the generalized gamma distribution.

Note, for example, that all distribution tails with $k_1 = 1$ are heavier than those with $k_1 = 2$. Within the $k_1 = 1$ group a smaller k_2 has the heavier tail. Note from Table 1 that the stable distribution and the skew t type 3 distribution with degrees of freedom parameter $0 < \tau < 2$ have the same range for k_2. Distribution tails with $0 < k_3 < \infty$ can be heavier than the Laplace (two sided exponential) if $0 < k_3 < 1$, lighter than the Laplace but heavier than the normal if $1 < k_3 < 2$, or lighter than the normal if $k_3 > 2$. It should be noted that although the tails of two distributions with the same combination of k_1 and k_2 values, are not necessarily equally heavy, a reduction in k_2, no matter how small, for either distribution will make it the heavier tail distribution. Similarly replacing (k_1, k_2) by (k_3, k_4) or (k_5, k_6). Hence the important point is that the k values are dominant in determining the heaviness of the tail of the distribution. [If it is required to distinguish between the two distributions with the same k values the second order terms of $\log f_Y(y)$ can be compared.]

Distribution tails in Tables 1 and 2 can be split into four categories: 'non-heavy' tails ($k_3 \geq 1 \, \text{or} \, 0 < \text{k}_5 < \infty$), 'heavy' tail (i.e. heavier than any exponential distribution) but lighter than any 'Paretian type' tail ($k_1 > 1 \, \text{and} \, 0 < k_3 < 1$), 'Paretian type' tail ($k_1 = 1 \, \text{and} \, k_2 > 1$), and heavier than any 'Paretian type' tail ($k_1 = 1 \, \text{and} \, k_2 = 1$). These four categories correspond closely to mild, slow, wild (pre or proper) and extreme randomness, (Mandlebrot, 1997).

Following Lemma B3 and Corollaries C1 and C2, Tables 1 and 2 also apply to the asymptotic form of the log of the survivor function, $\log \bar{F}_Y(y)$, with the following changes:

(i) when $k_1 = 1$ and $k_2 > 1$ then k_2 is reduced by 1

(ii) when $k_1 = 1$ and $k_2 = 1$ then $\log \bar{F}_Y(y) = o\,(\log |y|)$ and specific asymptotic forms for $\log \bar{F}_Y(y)$ for specific distributions are given in Table 3.

Value of k_1–k_6	Distribution name	Distribution	Condition	Value of k_1–k_6	Parameter range
$k_1 = 1$	Cauchy	$\text{CA}(\mu, \sigma)$		$k_2 = 2$	
	Generalized t	$\text{GT}(\mu, \sigma, \nu, \tau)$		$k_2 = \nu\tau + 1$	$\nu > 0, \tau > 0$
	Skew t type 3	$\text{ST3}(\mu, \sigma, \nu, \tau)$		$k_2 = \tau + 1$	$\tau > 0$
	Skew t type 4	$\text{ST4}(\mu, \sigma, \nu, \tau)$	right tail	$k_2 = \tau + 1$	$\tau > 0$
			left tail	$k_2 = \nu + 1$	$\nu > 0$
	Stable	$\text{SB}(\mu, \sigma, \nu, \tau)$		$k_2 = \tau + 1$	$0 < \tau < 2$
	t	$\text{TF}(\mu, \sigma, \nu)$		$k_2 = \nu + 1$	$\nu > 0$
$k_1 = 2$	Johnson's SU	$\text{JSU}(\mu, \sigma, \nu, \tau)$		$k_2 = 0.5\tau^2$	$\tau > 0$
	Johnson's SU original	$\text{JSUo}(\mu, \sigma, \nu, \tau)$		$k_2 = 0.5\tau^2$	$\tau > 0$
$0 < k_3 < \infty$	Power exponential	$\text{PE}(\mu, \sigma, \nu)$		$k_3 = \nu,\ k_4 = (c_1\,\sigma)^{-\nu}$	$\sigma > 0,\ \nu > 0$
	Power exponential type 2	$\text{PE2}(\mu, \sigma, \nu)$		$k_3 = \nu,\ k_4 = \sigma^{-\nu}$	$\sigma > 0,\ \nu > 0$
	Sinh-arcsinh original	$\text{SHASHo}(\mu, \sigma, \nu, \tau)$		$k_3 = \tau$	$\sigma > 0,\ \nu > 0,\ \tau > 0$
			right tail	$k_4 = \text{e}^{-2\nu}\tau\,\sigma^{-2\tau}$	
			left tail	$k_4 = \text{e}^{2\nu}\tau\,\sigma^{-2\tau}$	
	Sinh-arcsinh	$\text{SHASH}(\mu, \sigma, \nu, \tau)$	right tail	$k_3 = 2\tau,\ k_4 = 2^{2\tau-3}\sigma^{-2\tau}$	$\sigma > 0,\ \tau > 0$
			left tail	$k_3 = 2\nu,\ k_4 = 2^{2\nu-3}\sigma^{-2\nu}$	$\sigma > 0,\ \nu > 0$
	Skew exponential power type 3	$\text{SEP3}(\mu, \sigma, \nu, \tau)$		$k_3 = \tau$	$\sigma > 0,\ \nu > 0,\ \tau > 0$
			right tail	$k_4 = 0.5(\sigma\,\nu)^{-\tau}$	
			left tail	$k_4 = 0.5\sigma^{-\tau}\nu^{\tau}$	
	Skew exponential power type 4	$\text{SEP4}(\mu, \sigma, \nu, \tau)$	right tail	$k_3 = \tau,\ k_4 = \sigma^{-\tau}$	$\tau > 0$
			left tail	$k_3 = \nu,\ k_4 = \sigma^{-\nu}$	$\nu > 0$
$k_3 = 1$	Exponential generalized beta type 2	$\text{EGB2}(\mu, \sigma, \nu, \tau)$	$\sigma > 0$	$k_4 = \tau\,\sigma^{-1}$	$\tau > 0$
			$\sigma < 0$	$k_4 = \nu\ \lvert\sigma\rvert^{-1}$	$\nu > 0$
	Gumbel	$\text{GU}(\mu, \sigma)$	left tail	$k_4 = \sigma^{-1}$	$\sigma > 0$
	Laplace	$\text{LA}(\mu, \sigma)$		$k_4 = \sigma^{-1}$	$\sigma > 0$
	Logistic	$\text{LG}(\mu, \sigma)$		$k_4 = \sigma^{-1}$	$\sigma > 0$
	Reverse Gumbel	$\text{RG}(\mu, \sigma)$	right tail	$k_4 = \sigma^{-1}$	$\sigma > 0$
$k_3 = 2$	Normal	$\text{NO}(\mu, \sigma)$		$k_4 = 0.5\sigma^{-2}$	$\sigma > 0$
$0 < k_5 < \infty$	Gumbel	$\text{GU}(\mu, \sigma)$	right tail	$k_5 = \sigma^{-1},\ k_6 = \text{e}^{-\frac{\mu}{\sigma}}$	$-\infty < \mu < \infty,\ \sigma > 0$
	Reverse Gumbel	$\text{RG}(\mu, \sigma)$	left tail	$k_5 = \sigma^{-1},\ k_6 = \text{e}^{\frac{\mu}{\sigma}}$	$-\infty < \mu < \infty,\ \sigma > 0$

Table 1: Left and right tail asymptotic form of the log of the probability density function for continuous distributions on the real line, where $c_1^2 = \Gamma(\frac{1}{\nu})\left[\Gamma(\frac{3}{\nu})\right]^{-1}$

Value of k_1–k_6	Distribution name	Distribution	Condition	Value of k_1–k_6	Parameter range
$k_1 = 1$	Box-Cox Cole-Green	BCCG(μ, σ, ν)	$\nu < 0$	$k_2 = \lvert\nu\rvert + 1$	
	Box-Cox power exponential	BCPE(μ, σ, ν, τ)	$\nu < 0$	$k_2 = \lvert\nu\rvert + 1$	
	Box-Cox t	BCT(μ, σ, ν, τ)	$\nu \leq 0$	$k_2 = \lvert\nu\rvert + 1$	
			$\nu > 0$	$k_2 = \nu\tau + 1$	$\tau > 0$
	Generalized beta type 2	GB2(μ, σ, ν, τ)	$\sigma > 0$	$k_2 = \sigma\tau + 1$	$\tau > 0$
			$\sigma < 0$	$k_2 = \lvert\sigma\rvert\, \nu + 1$	$\nu > 0$
	Generalized gamma	GG(μ, σ, ν)	$\nu < 0$	$k_2 = \left(\sigma^2 \lvert\nu\rvert\right)^{-1} + 1$	$\sigma > 0$
	Inverse gamma	IGA(μ, σ)		$k_2 = \sigma^{-2} + 1$	$\sigma > 0$
	log t	LOGT(μ, σ, ν)		$k_2 = 1$	
	Pareto Type 2	PA2o(μ, σ)		$k_2 = \sigma + 1$	$\sigma > 0$
$k_1 = 2$	Box-Cox Cole-Green	BCCG(μ, σ, ν)	$\nu = 0$	$k_2 = 0.5\sigma^{-2}$	$\sigma > 0$
	Lognormal	LOGNO(μ, σ)		$k_2 = 0.5\sigma^{-2}$	$\sigma > 0$
	Log Weibull	LOGWEI(μ, σ)	$\sigma > 1$	$k_1 = \sigma,\ k_2 = \mu^{-\sigma}$	
			$\sigma = 1$	$k_1 = 1,\ k_2 = \mu^{-\sigma} + 1$	
			$\sigma < 1$	$k_1 = 1,\ k_2 = 1$	
$1 \leq k_1 < \infty$	Box-Cox power exponential	BCPE(μ, σ, ν, τ)	$\nu = 0,\ \tau > 1$	$k_1 = \tau,\ k_2 = (c_1\sigma)^{-\tau}$	$\sigma > 0$
			$\nu = 0,\ \tau = 1$	$k_1 = 1,\ k_2 = 1 + (c_1\sigma)^{-\tau}$	
			$\nu = 0,\ \tau < 1$	$k_1 = 1,\ k_2 = 1$	$\sigma > 0$
$0 < k_3 < \infty$	Box-Cox Cole-Green	BCCG(μ, σ, ν)	$\nu > 0$	$k_3 = 2\nu,\ k_4 = \left[2\mu^{2\nu}\sigma^2\nu^2\right]^{-1}$	$\mu > 0,\ \sigma > 0$
	Box-Cox power exponential	BCPE(μ, σ, ν, τ)	$\nu > 0$	$k_3 = \nu\tau,\ k_4 = \left[c_1\mu^{\nu}\sigma\nu\right]^{-\tau}$	$\mu > 0,\ \sigma > 0,\ \tau > 0$
	Generalized gamma	GG(μ, σ, ν)	$\nu > 0$	$k_3 = \nu,\ k_4 = \left[\mu^{\nu}\sigma^2\nu^2\right]^{-1}$	$\mu > 0,\ \sigma > 0$
	Weibull	WEI(μ, σ)		$k_3 = \sigma,\ k_4 = \mu^{-\sigma}$	$\mu > 0,\ \sigma > 0$
$k_3 = 1$	Exponential	EX(μ)		$k_4 = \mu^{-1}$	$\mu > 0$
	Gamma	GA(μ, σ)		$k_4 = \mu^{-1}\sigma^{-2}$	$\mu > 0,\ \sigma > 0$
	Generalized inverse Gaussian	GIG(μ, σ, ν)		$k_4 = 0.5c_2\mu^{-1}\sigma^{-2}$	$\mu > 0,\ \sigma > 0$
	Inverse Gaussian	IG(μ, σ)		$k_4 = 0.5\mu^{-2}\sigma^{-2}$	$\mu > 0,\ \sigma > 0$

Table 2: Right tail asymptotic form of the log of the probability density function for continuous distributions on the positive real line, where $c_2 = \left[K_{\nu+1}\left(\frac{1}{\sigma^2}\right)\right]\left[K_{\nu}\left(\frac{1}{\sigma^2}\right)\right]^{-1}$ where $K_{\lambda}(t) = \frac{1}{2}\int_0^{\infty} x^{\lambda-1} \exp\left\{-\frac{1}{2}t\left(x + x^{-1}\right)\right\} dx$

Distribution	**Asymptotic form of** $\log \bar{F}_Y(y)$
LOGT(μ, σ, ν)	$-\nu \log(\log y)$
BCT$(\nu = 0)$	$-\tau \log(\log y)$
LOG WEI(μ, σ) for all $0 < \sigma < \infty$	$-\mu^{-\sigma}(\log y)^{\sigma}$
BCPE$(\nu = 0)$ for all $0 < \tau < \infty$	$-(c_1\sigma)^{-\tau}(\log y)^{\tau}$

Table 3: Asymptotic form of $\log \bar{F}_Y(y)$ as $y \to \infty$.

Note that the distributions having log survivor function upper tails in exactly the forms $-k_2(\log y)^{k_1}$, $-k_4 y^{k_3}$ and $-k_6 e^{k_5 y}$ are the log Weibull (LOGWEI), the Weibull (WEI) and the Gumbel (GU), respectively.

2.3 Methods for choosing the appropriate tail

The substantive practical implications of ordering of distribution tails is in the development and selection of statistical distributions with tails appropriate for observations on a variable. An important way of distinguishing different distribution tails in practice is the complementary cumulative distribution function (CCDF) plot given by plotting $\log[\bar{F}_Y(y)]$ against $\log y$. Another possible way is the complementary log log plot of the survival function. We refer below the sample versions of those two plots as exploratory methods 1 and 2 respectively.

2.3.1 Exploratory Method 1

Note that if the upper tail of the survivor function $\bar{F}_Y(y)$ is asymptotically in the form $-k_2(\log y)^{k_1}$, $-k_4 y^{k_3}$ and $-k_6 e^{k_5 y}$ then the CCDF plot of $\log[\bar{F}_Y(y)]$ against $t = \log y$ will be asymptotically in the form $-k_2 t^{k_1}$, $-k_4 e^{k_3 t}$ and $-k_6 e^{k_5 e^t}$ respectively, i.e. power, exponential and double-exponential respectively. A sample version of CCDF plot is given by plotting $\log\left(1 - \dfrac{i - 0.5}{n}\right)$ against $\log y_{(i)}$, where $y_{(i)}$ is the ith largest value of y in the sample. The sample CCDF plot can be used to investigate the tail form of $\bar{F}_Y(y)$.

2.3.2 Exploratory Method 2

Correspondingly the upper tail of $\log\left\{-\log\left[\bar{F}_Y(y)\right]\right\}$ is asymptotically as $y \to \infty$ in the form $\log k_2 + k_1 \log[\log(y)]$, $\log k_4 + k_3 \log y$ and $\log k_6 + k_5 y$, and hence a plot of $\log\left\{-\log\left[\bar{F}_Y(y)\right]\right\}$ against $\log[\log(y)]$, $\log y$ and y will be asymptotically linear in each case. The corresponding sample plot can be used to investigate the tail form of $\bar{F}_Y(y)$ (although a large sample size may be required especially in the first case).

2.4 Example

As discussed in the introduction, to demonstrate our approach, we use the total USA box office film revenue (`F90`). This is because film revenues are highly skewed, in such a way that a small number of large revenue films coexist alongside considerably greater numbers of smaller revenue films. Moreover, the skewed nature of these distributions appears to be an empirical regularity, with Pokorny and Sedgwick (2010) dating this phenomenon back to at least the 1930s, making it an early example of a mass market long tail. This data was analysed by Voudouris et al. (2012)

	Intercept	slope	Error SS
type I	-28.2422	10.1806	0.12597
type II	-8.91724	0.560959	0.09090
type III	0.75172	5.697e-09	2.82237

Table 4: Estimated coefficients from exploratory method 2.

Figure 1 shows the sample CCDF plot (exploratory method 1) for the largest 10% of revenues together with fitted linear, quadratic and exponential functions. The linear fit appears inadequate, hence $k_1 = 1$ (e.g. a Pareto distribution) appears inappropriate. The quadratic or exponential fits adequately suggesting $k_1 = 2$ (e.g. lognormal tail) or $0 < k_3 < \infty$ (e.g. Weibull tail) respectively may be appropriate.

Figure 2 plots $\log\left\{-\log\left[\bar{F}_Y(y)\right]\right\}$ against $\log(\log(y))$, $\log y$ and y respectively, (exploratory method 2), with the middle graph providing the best linear fit (error sum of squares equal 0.0909, see Table 4), with estimates $\hat{k}_3 = 0.561$ and $\hat{k}_4 = \exp(-8.917) = 0.000134$, suggesting a Weibull tail may be appropriate.

Truncated lognormal and Weibull distributions were fitted to the largest 10% of revenues leading to good fits in each case. Figure 3 provides a normal QQ plot for the normalised quantile residuals Dunn and Smyth (1996) from the truncated Weibull fit to the largest 10% of revenues, indicating a good fit particularly in the upper tail. The estimated Weibull parameters were $\hat{\mu} = 13467053$ and $\hat{\sigma} = 0.6476$. Sequential fits of the truncated Weibull distribution to the largest r revenues, for $r = 4, 5, 6, \ldots, 403$, were followed by a plot of the parameter estimate $\hat{\sigma}$ against r indicating that the fitted parameter $\hat{\sigma}$ is relatively stable (Figure 4) indicating that the Weibull fit to the tail is relatively stable as r changes. This plot is analogous to the Hill plot (Hill, 1975).

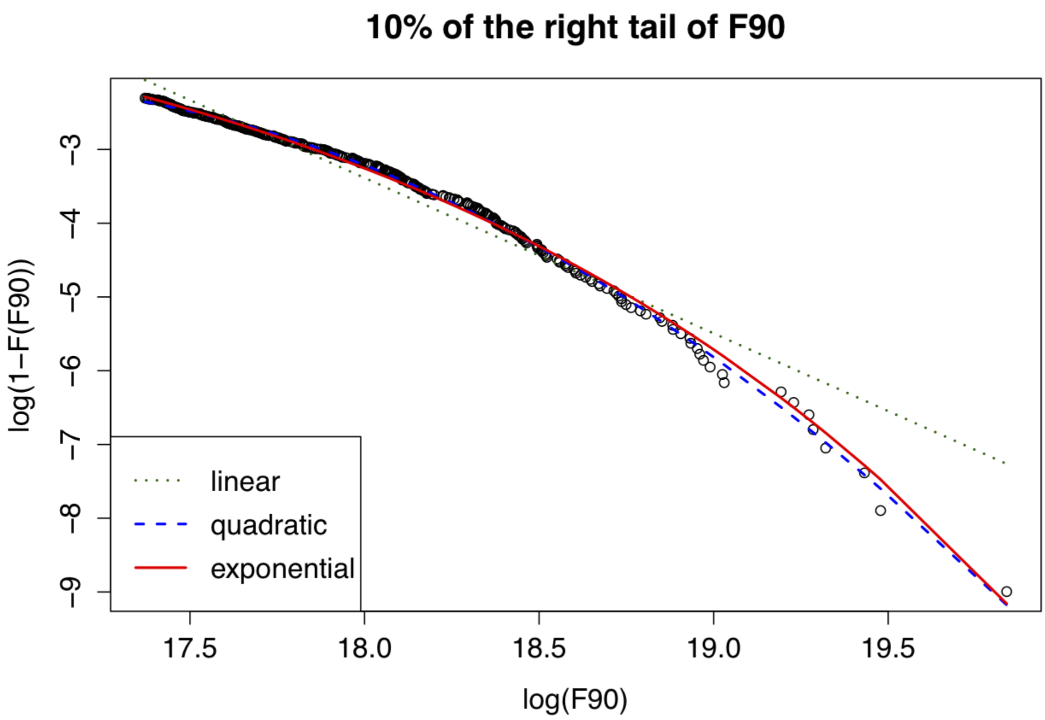


Figure 1: Exploratory Method 1 applied to the 90's film revenues data

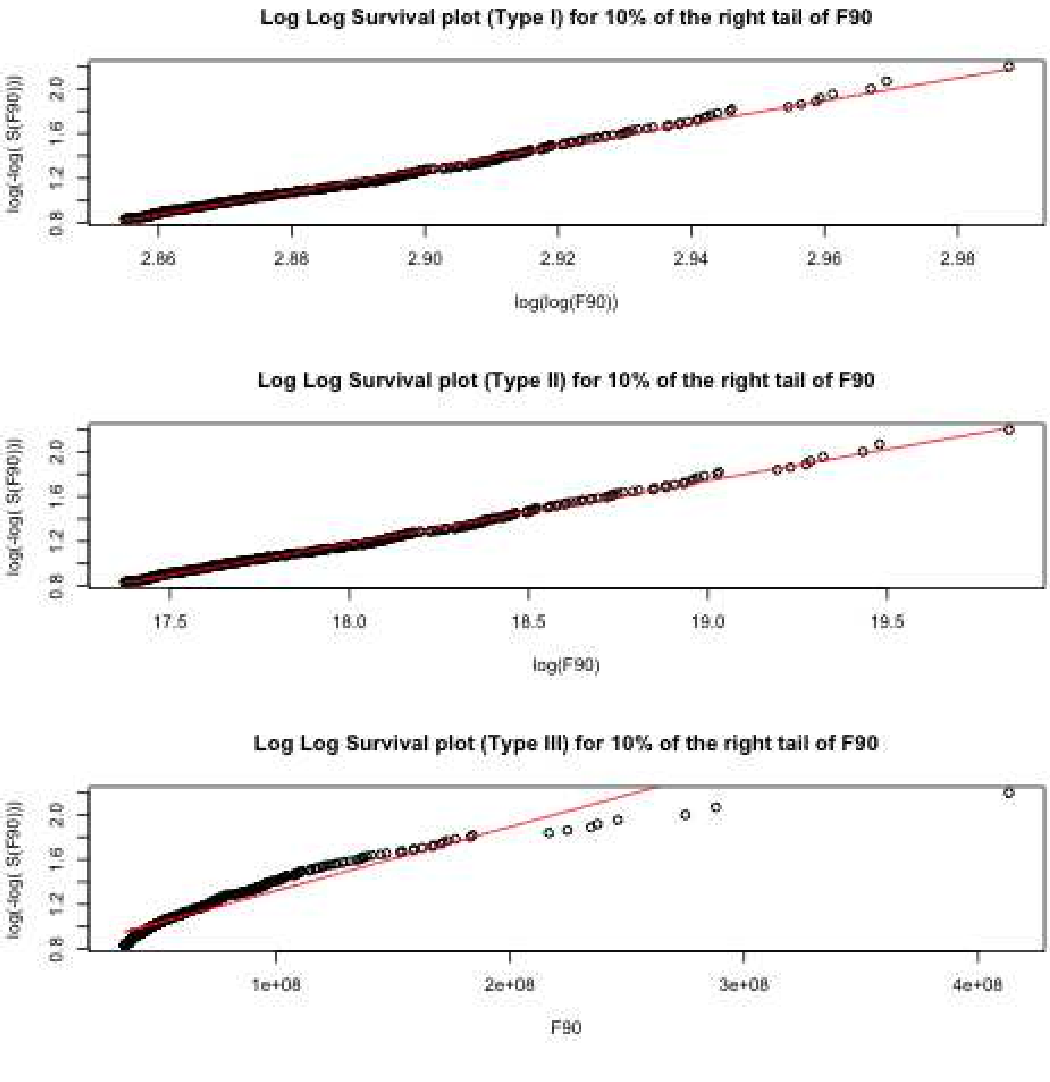


Figure 2: Exploratory Method 2 applied to the 90's film revenues data

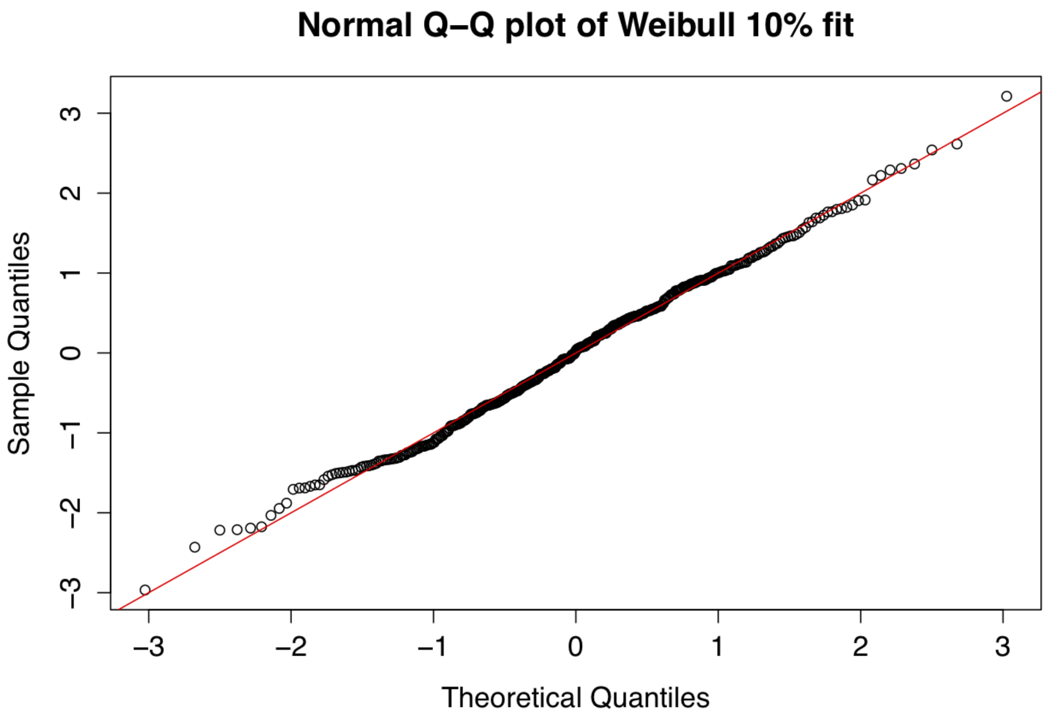


Figure 3: QQ plot for the truncated Weibull.

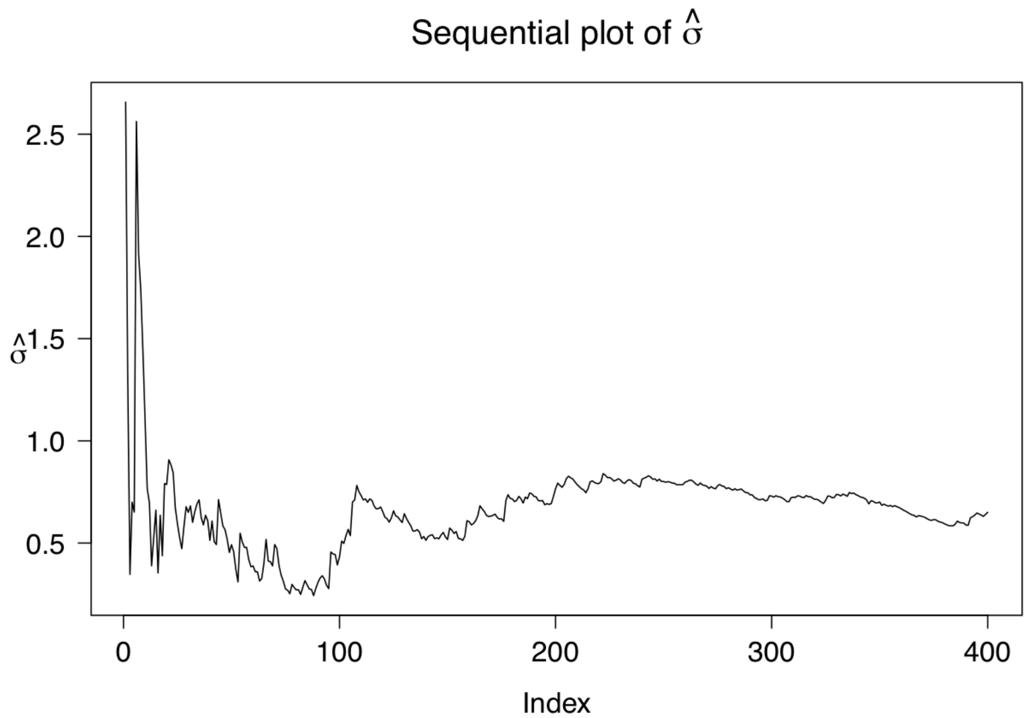


Figure 4: Sequencial plot of $\hat{\sigma}$ for the truncated Weibull

3 Centile based comparisons of distributions

3.1 Introduction

The centile based measures of skewness and kurtosis are defined using the quantile function of the distribution of a random variable Y given by $y_p = F_Y^{-1}(p)$ for $0 < p < 1$, where F_Y^{-1} is the inverse cumulative distribution function of Y.

A general centile based measure of skewness is given by MacGillivray (1986):

$$s_p = \frac{(y_p + y_{1-p})/2 - y_{0.5}}{(y_{1-p} - y_p)/2} \tag{1}$$

for $0 < p < 0.5$, i.e. the midpoint of a central $100(1-2p)\%$ interval for Y minus the median, divided by the half length of the central $100(1-2p)\%$. Note that $-1 \leq s_p \leq 1$.

One important case is $p = 0.25$, giving Galton's measure of skewness:

$$s_{0.25} = \frac{(Q_1 + Q_3)/2 - m}{(Q_3 - Q_1)/2} \tag{2}$$

i.e. the mid quartile $(Q_1 + Q_3)/2$ minus the median divided by the semi-quartile range $(Q_3 - Q_1)/2$, where $Q_1 = y_{0.25}$ and $Q_3 = y_{0.75}$. This can be considered as a measure of central skewness since it focuses on the skewness within the interquartile range for Y.

A second important case is $p = 0.01$, giving

$$s_{0.01} = \frac{(y_{0.01} + y_{0.99})/2 - y_{0.5}}{(y_{0.99} - y_{0.01})/2} \tag{3}$$

i.e. the midpoint of a central 98% interval for Y minus the median, divided by the half length of the central 98% interval for Y. This can be considered as a measure of tail skewness since it focuses on skewness within a central 98% interval for Y. A third important case is $p = 0.001$ which measures extreme tail skewness.

Following Balanda and MacGillivray (1988), a general centile based measure of kurtosis is given by Andrews et al. (1972):

$$k_p = \frac{(y_{1-p} - y_p)}{(Q_3 - Q_1)} \tag{4}$$

for $0 < p < 0.5$, i.e. the ratio of the length of a central $100(1-2p)\%$ interval for Y to its interquartile range. An important case is $p = 0.01$, i.e. $k_{0.01}$ (Andrews et al., 1972). This has been scaled relative to a normal distribution for which $k_{0.01} = 3.49$ giving

$$s\,k_{0.01} = \frac{(y_{0.99} - y_{0.01})}{3.49\,(Q_3 - Q_1)}\ , \tag{5}$$

Rosenberger and Gasko (1983). Hence a normal distribution has $s\,k_{0.01} = 1$. To allow the full range of kurtosis to be plotted it is transformed to

$$t\,s\,k_{0.01} = \frac{s\,k_{0.01} - 1}{s\,k_{0.01}}\ . \tag{6}$$

Note that $t\,s\,k_{0.01} \in (-2.49, 1)$, where $t\,s\,k_{0.01} \to 1$ corresponds to $k_{0.01} \to \infty$ and $t\,s\,k_{0.01} \to -2.49$ corresponds to $k_{0.01} \to 1$. Also $t\,s\,k_{0.01} = 0$ corresponds to $s\,k_{0.01} = 1$, e.g. a normal distribution, while $t\,s\,k_{0.01} = -1$ corresponds to $s\,k_{0.01} = 0.5$. See Balanda and MacGillivray (1988) for a review of kurtosis.

In sections 3.2 and 3.3 we compare plots of transformed centile kurtosis (6) against each of centile central skewness respectively (2) and centile tail skewness (3) for commonly used heavy tailed distributions on the real line. The following distributions on the real line are considered: exponential generalized beta type 2 (EGB2), Johnson's SU (JSU), sinh-arcsinh original (SHASHo), skew exponential power type 3 (SEP3), skew t type 3 (ST3) and stable (SB). See Table A1 in Appendix A for references.

3.2 Transformed (centile) kurtosis against (centile) central skewness

Here we investigate the relationship between transformed kurtosis $t\,s\,k_{0.01}$ given by (6) against positive central skewness $s_{0.25} \in (0, 1)$ given by (3). For each of the six distributions the boundary of central skewness is plotted against transformed kurtosis in Figure 5. The vertical line at central skewness equals zero and the horizontal line at transformed kurtosis equals one form the outer boundaries of each of the six regions of the distributions. The corresponding plot for negative central skewness is a mirror image around the vertical origin axis.

Note that the normal distribution is plotted at the point $(0, 0)$ in Figure 5. Transformed kurtosis below 0 can be considered as 'platykurtic' while above 0 can be considered 'leptokurtic'. Clearly the EGB2, JSU and SB distributions do not allow 'platykurtic' distributions, while SEP3 allows the lowest kurtosis (most 'platykurtic') distributions for a fixed low central skewness $s_{0.25} < 0.05$ and SHASHo allows the lowest kurtosis distributions for a fixed high central skewness $s_{0.25} > 0.05$.

The SHASHo distribution is generally the most versatile covering the largest range of central skewness $s_{0.25}$ for a given value of transformed kurtosis $t\,s\,k_{0.01}$ (provided $t\,s\,k_{0.01} > -0.36$). The SEP3 distribution is most versatile for $t\,s\,k_{0.01} < -0.36$ and second most versatile for $t\,s\,k_{0.01} > -0.36$. The JSU and ST3 distributions have more restricted central skewness for a given transformed kurtosis. The EGB2 distribution is more restricted in central skewness and transformed kurtosis, with the transformed kurtosis at or moderately above that of the normal distribution. The stable distribution is restrictive in central skewness for a given transformed kurtosis, with the transformed kurtosis generally much higher than the normal distribution. The range of possible central skewness increases with the transformed kurtosis for all distributions (except EGB2) .

Figure 6 (a) and (b) show the transformed kurtosis against central skewness for the SHASHo and SB distributions respectively, showing contours for different values of each of the skewness and kurtosis parameters, ν and τ respectively, of the distribution, while keeping the other parameter constant. The SHASHo was chosen because of its flexibility, while SB was chosen because its moment based kurtosis-skewness plot is not possible. For the SHASHo distributions in Figure 6(a) the horizontal contours correspond to $\tau = 0.001, 0.5, 0.75, 1, 1.5, 3$ from top to bottom while the 'vertical' contours correspond to $\nu = 0, 0.1, 0.25, 0.5, 0.75, 1, 1.5, 100$ from left to right. Note that $\tau = 0.001$ and $\nu = 100$ effectively correspond to the limits $\tau = 0$ and $\nu = \infty$ as no change in the contours was observed as τ was decreased below 0.001 and ν increased above 100, respectively. Note also that for a fixed τ, ν affects the centile skewenss only. For the stable SB distribution in Figure 6(b) the 'horizontal' contours correspond to $\tau = 0.001, 0.75, 1, 1.25, 1.5, 1.75$ while the 'vertical' contours correspond to $\nu = 0, 0.1, 0.25, 0.5, 0.75, 1$ from left to right. Note

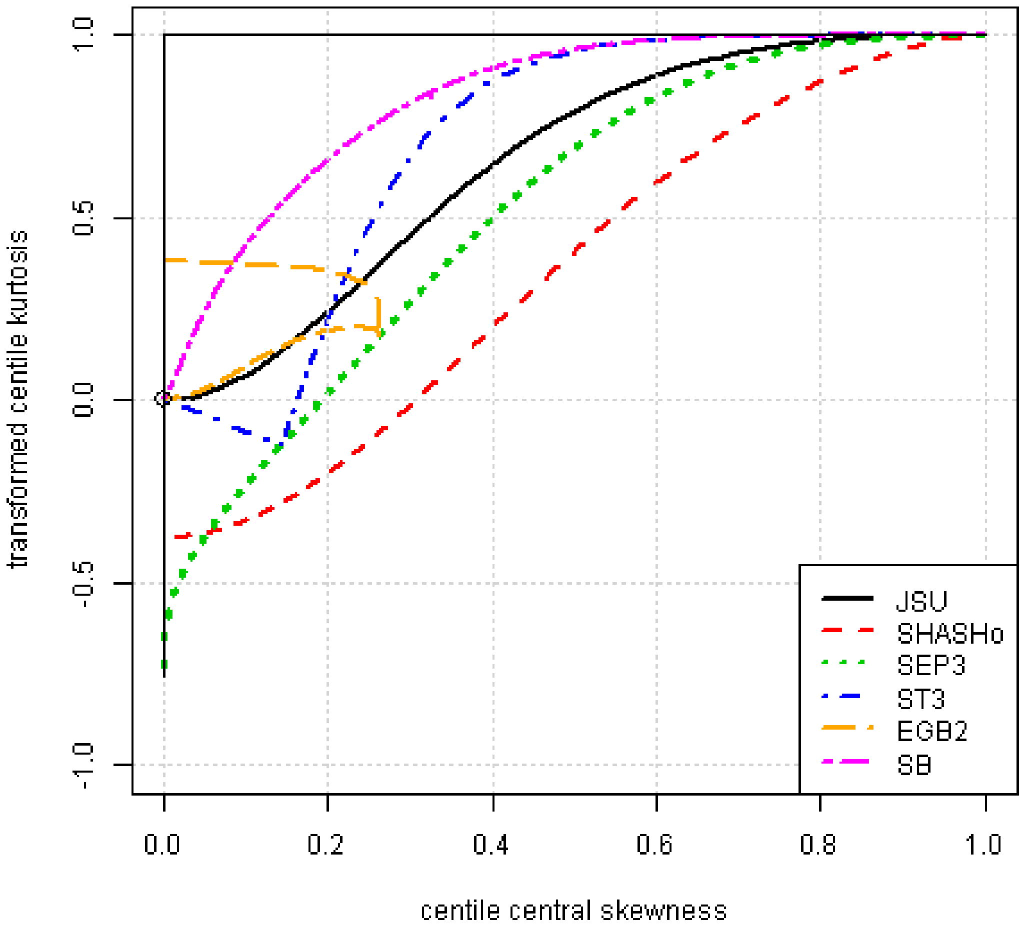
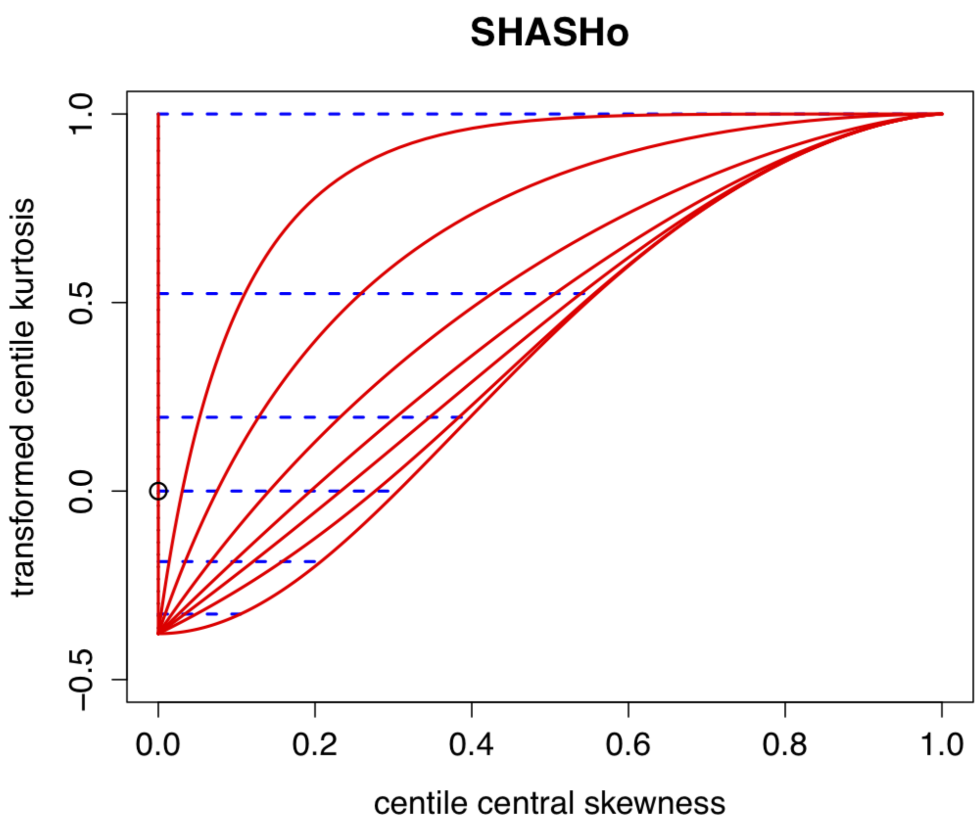


Figure 5: The upper boundary of centile central skewness against the transformed centile kurtosis for six distributions on the real line.

that $\tau = 0.001$ effectively corresponds to the limit $\tau = 0$.

3.3 Transformed (centile) kurtosis against (centile) tail skewness

Section 3.2 is amended to replace the central skewness $s_{0.25}$ given by (3) with the tail skewness $s_{0.01}$ given by (4). Figures 7 and 8 correspond to Figures 5 and 6. The contour values of ν and τ in Figure 4 are the same as used in Figure 6. Note that the range of tail skewness for the six distributions is now more restricted to $(0, 0.5)$ instead of $(0, 1)$ for the central skewness. However the general comments about the kurtosis-skewness relationship for the six distributions still apply.



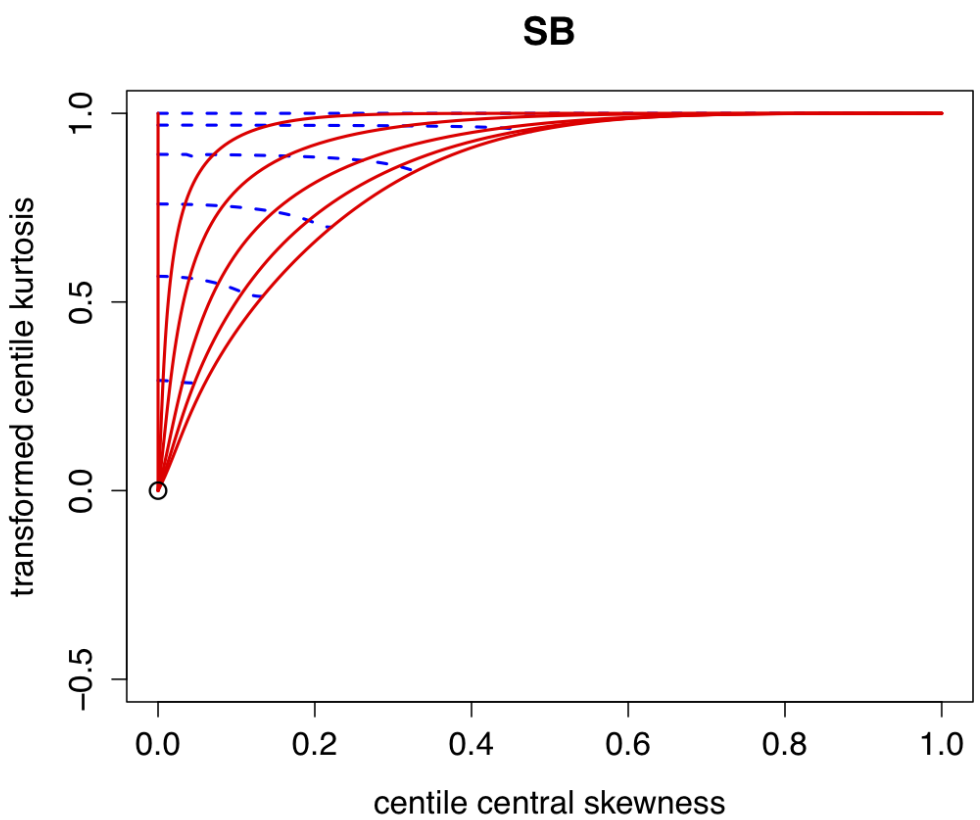


Figure 6: Contours of centile central skewness against the transformed centile kurtosis for constant values of ν and τ for the SHASHo and SB distributions.

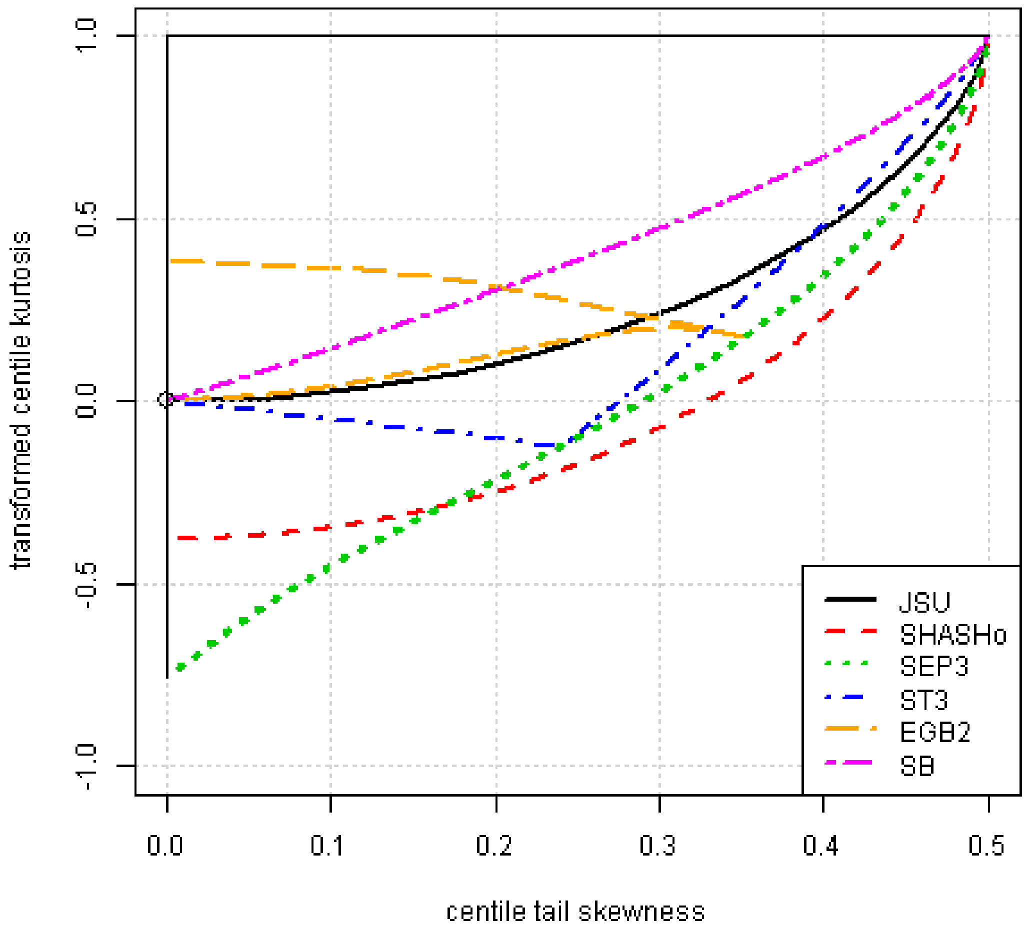
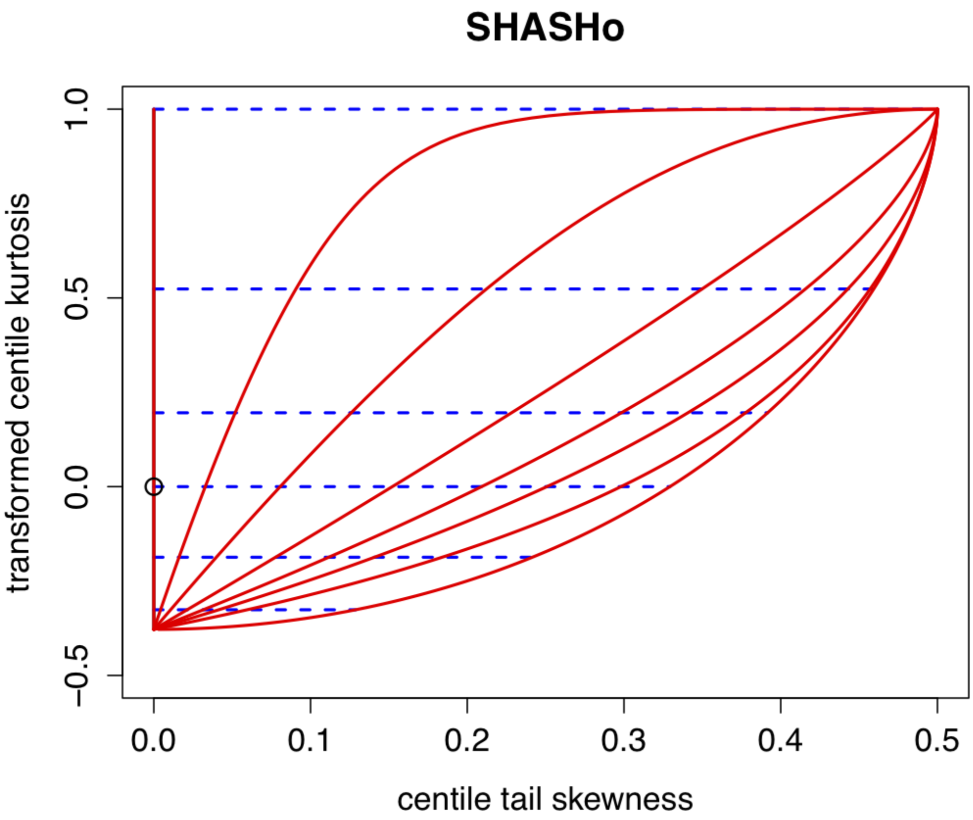


Figure 7: The upper boundary of centile tail skewness against the transformed centile kurtosis for six distributions on the real line.



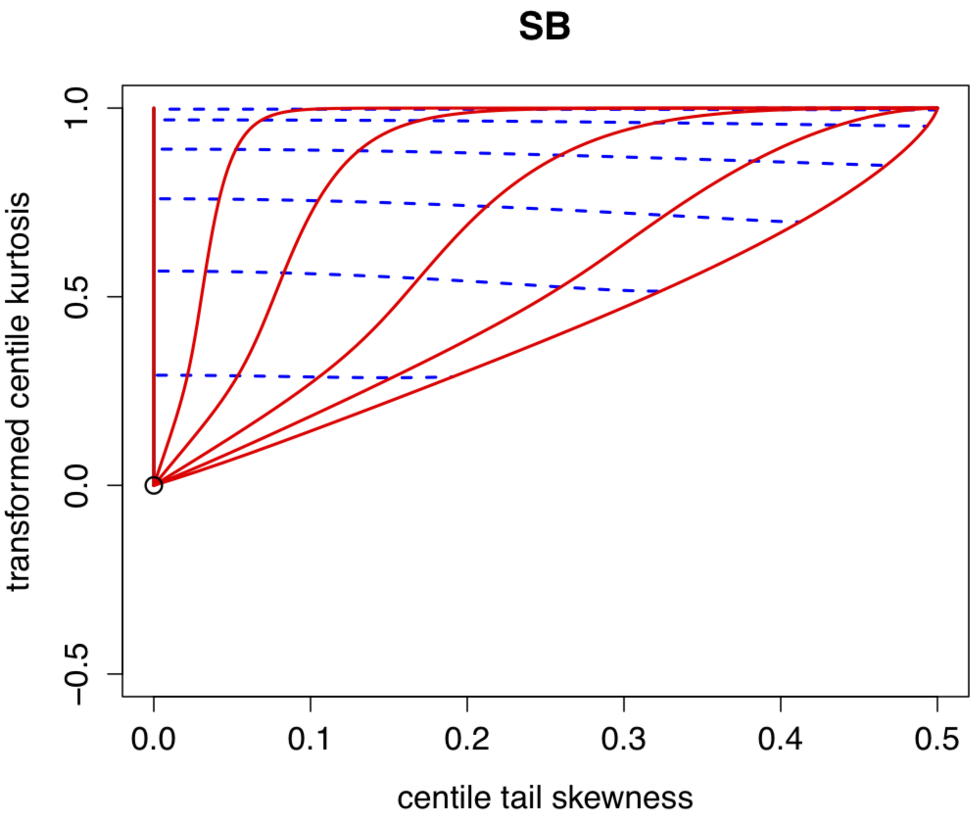


Figure 8: Contours of centile tail skewness against the transformed centile kurtosis for constant values of ν and τ for the SHASHo and SB distributions.

4 Conclusions

We argue that the selection of appropriate theoretical distributions for the development of conditional densities for statistical models is not trivial while methods for the ordering and comparison of theoretical distributions are missing. An ordering of the heaviness of the tail(s) of distributions based on three main asymptotic forms of the log of the probability density function has been shown. Tables 1 and 2 summarise the results for a selection of distributions.

The boundary of (centile) central and tail skewness against the transformed (centile) kurtosis is also given for six important four parameter distributions on the real line. Overall the sinh-arcsinh (SHASHo) is the most flexible distribution in modelling the skewness and kurtosis. However its tails are not as heavy as the stable (SB) or skew t type 3 (ST3). Hence the SHASHo and SEP3 are flexible enough to model business-related data which can exhibit a wide range of skewness and kurtosis, while the SB and ST3 are more appropriate to model data with high kurtosis and low skewness. The EGB2 is only appropriate for mild leptokurtosis and low skewness. New or different distributions can be included in the tail comparison in Tables 1 and 2 and in the kurtosis-skewness plots in Figures 5 and 7.

The substantive applied implications of the comparison of 'heavy tails' given here is that the development of conditional densities for decision analysis need not to be founded upon a very narrow set of distributions and distribution shapes and tails. The classification and centile comparison presented here is a way forward in detecting the flexibility of theoretical distributions when processes and/or systems characterized by highly skew and/or kurtic data are analysed.

Models to explore and investigate past performance in order to drive forward-looking planning can be extended to include distributions that are sufficiently flexible in modelling different distribution shapes and tails. This will enable the better qualification of risks and better evaluation of actions. Consequently when a statistical model is applied to data some outcomes are not misclassified as 'possible' and others as 'practically impossible' under the assumptions of an inflexible (and inappropriate) distribution. When an event misclassified as 'practically impossible' actually occurs, catastrophic errors are likely to be observed.

Appendix A

Box-Cox Cole-Green	Cole and Green (1992)
Box-Cox power exponential	Rigby and Stasinopoulos (2004)
Box-Cox t	Rigby and Stasinopoulos (2006)
Cauchy	Johnson *et al.* (1994)
Exponential	Johnson *et al.* (1994)
Exponential generalized beta type 2	McDonald and Xu (1995); McDonald (1996)
Gamma	Johnson et al. (1994)
Generalized beta type 2	McDonald and Xu (1995); McDonald (1996)
Generalized gamma	Lopatatzidis and Green (2000); Harter (1967)
Generalized inverse Gaussian	Jørgensen (1997); Jørgensen (1982)
Generalized t	McDonald and Newey (1988); McDonald (1991)
Gumbel	Crowder *et al.* 1991
Inverse Gamma	Johnson *et al.* (1994)
Johnson's SU	Johnson *et al.* (1994)
Johnson's SU Original	Johnson *et al.* (1994)
Laplace	Johnson *et al.* (1995)
Lognormal	Johnson*et al.*(1994)
Normal	Johnson *et al.* (1994)
Pareto Type 2	Johnson *et al.* (1994)
Power exponential	Nelson (1991)
Power exponential type 2	Nelson (1991); Johnson *et al.* (1995)
Reverse Gumbel	Johnson *et al.* (1995)
Sinh-arcsinh	Jones (2005)
Sinh-arcsinh original	Jones and Pewsey (2009)
Skew exponential power type 3	Fernandez *et al.* (1995)
Skew exponential power type 4	Jones (2005)
Skew t type 3	Fernandez and Steel (1998)
Skew t type 4	Stasinopoulos *et al.*(2008)
Stable	Nolan (2012)
t	Johnson *et al.* (1995)
Weibull	Johnson *et al.* (1994)

Table 5: References for continuous distributions

Appendix B

B.1 Lemma B1

Let the random variables Y_1 and Y_2 have probability density functions $f_{Y_1}(y)$ and $f_{Y_2}(y)$ respectively, then $f_{Y_1}(y) = o\left[f_{Y_2}(y)\right]$ as $y \to \infty \Leftrightarrow \lim_{y\to\infty} \left[\log f_{Y_2}(y) - \log f_{Y_1}(y)\right] = +\infty$. Similarly replacing $y \to \infty$ by $y \to -\infty$ for the left tail.

Proof B1 $\qquad f_{Y_1}(y) = o\left[f_{Y_2}(y)\right] \qquad$ as $y \to \infty$

$$\Leftrightarrow \quad \lim_{y\to\infty}\left[\frac{f_{Y_1}(y)}{f_{Y_2}(y)}\right] = 0$$
$$\Leftrightarrow \quad \lim_{y\to\infty}\left[\log\frac{f_{Y_2}(y)}{f_{Y_1}(y)}\right] = +\infty$$

B.2 Lemma B2

Let random variables Y_1 and Y_2 have probability density functions $f_{Y_1}(y)$ and $f_{Y_2}(y)$, cumulative distribution functions $F_{Y_1}(y)$ and $F_{Y_2}(y)$ and survivor functions $\bar{F}_{Y_1}(y)$ and $\bar{F}_{Y_2}(y)$ respectively, then

$$f_{Y_1}(y) = o\left[f_{Y_2}(y)\right] \text{ as } y \to \infty \quad \Leftrightarrow \bar{F}_{Y_1}(y) = o\left[\bar{F}_{Y_2}(y)\right] \text{ as } y \to \infty$$
$$f_{Y_1}(y) = o\left[f_{Y_2}(y)\right] \text{ as } y \to -\infty \quad \Leftrightarrow F_{Y_1}(y) = o\left[F_{Y_2}(y)\right] \text{ as } y \to -\infty$$

provided $F_{Y_1}(y)$ and $F_{Y_2}(y)$ are differentiable and, as $y \to \infty$ and as $y \to -\infty$, $\lim f_{Y_1}(y) = \lim f_{Y_2}(y) = 0$ and $\lim\left[\frac{f_{Y_1}(y)}{f_{Y_2}(y)}\right]$ exists.

Proof B2 $\qquad f_{Y_1}(y) = o\left[f_{Y_2}(y)\right] \qquad$ as $\quad y \to \infty$

$$\Leftrightarrow \quad \lim_{y\to\infty}\frac{f_{Y_1}(y)}{f_{Y_2}(y)} = 0$$

$$\Leftrightarrow \quad \lim_{y\to\infty}\frac{\bar{F}_{Y_1}(y)}{\bar{F}_{Y_2}(y)} = 0 \qquad \text{using L'Hopital's rule}$$

$$\Leftrightarrow \quad \bar{F}_{Y_1}(y) = o\left[\bar{F}_{Y_2}(y)\right]$$

The proof follows similarly for the left tail as $y \to -\infty$.

B.3 Lemma B3

Let $f_Y(y)$ and $F_Y(y)$ be respectively the probability density function and cumulative distribution function of a random variable Y. Then

$$\frac{1}{\bar{F}_Y(y)} \sim \left[\frac{1}{f_Y(y)}\right]' \qquad \text{as } y \to \infty$$

provided $\lim_{y\to\infty} f_Y(y) = 0$ and $\lim_{y\to\infty}\frac{f'_Y(y)}{f_Y(y)}$ exists, where $'$ indicates the derivative with respect to y and $\bar{F}_Y(y) = 1 - F_Y(y)$.

Proof B3 $\qquad \lim_{y\to\infty}\frac{f_Y(y)}{\bar{F}_Y(y)} = \lim_{y\to\infty}\frac{f'_Y(y)}{f_Y(y)} \qquad$ using L'Hopital's rule

$$\therefore \qquad \lim_{y\to\infty}\left\{\frac{[f_Y(y)]^2}{\bar{F}_Y(y)f_Y'(y)}\right\}=1$$

$$\therefore \qquad \frac{1}{\bar{F}_Y(y)}\sim\frac{f_Y'(y)}{[f_Y(y)]^2}=\left[\frac{1}{f_Y(y)}\right]'$$

Appendix C

C.1 Corrolary C1

If $\log f_y(y)\sim -g(y)$ as $y\to\infty$ then $\log \bar{F}_Y(y)\sim\begin{cases}-g(y)-\log g^{'}(y) & \text{if } g(y)\nsim -\log g^{'}(y)\\ o\left[g(y)\right] & \text{if } g(y)\sim -\log g^{'}(y)\end{cases}$

Proof C1 $\qquad \log f_Y(y)=-g(y)[1+o\,[1]]$

$$\therefore \qquad \frac{1}{f_Y(y)}=\mathrm{e}^{g(y)[1+o[1]]}$$

$$\therefore \qquad \left[\frac{1}{f_Y(y)}\right]^{'}\sim g^{'}(y)\mathrm{e}^{g(y)}$$

since $\qquad \dfrac{d}{dy}\{g(y)[1+o\,[1]]\}\sim g^{'}(y)$ as $\dfrac{d}{dy}[1+o\,[1]]=o(1)$

$$\therefore \qquad \bar{F}_Y(y)\,g^{'}(y)\mathrm{e}^{g(y)}\to 1 \qquad \text{as } y\to\infty \text{ using Lemma B3}$$

$$\therefore \qquad \log\bar{F}_Y(y)+g(y)+\log g^{'}(y)\to 0 \qquad \text{as } y\to\infty$$

Hence result.

C.2 Corrolary C2

As $y\to\infty$ (or $y\to-\infty$),

(a) If $\log f_Y(y)\sim -k_2(\log|y|)^{k_1}$ then $\log F_Y(y)\sim\begin{cases}-k_2(\log|y|)^{k_1} & \text{if } k_1>1\\ -(k_2-1)\log|y| & \text{if } k_1=1 \text{ and } k_2>1\\ o(\log|y|) & \text{if } k_1=k_2=1\end{cases}$

(b) If $\log f_Y(y)\sim -k_4\,|y|^{k_3}$ then $\log F_Y(y)\sim -k^4\,|y|^{k_3}$

(c) If $\log f_Y(y)\sim -k_6\mathrm{e}^{-k_5|y|}$ then $\log F_Y(y)\sim -k_6\mathrm{e}^{-k_5|y|}$

Proof C2

(a) From Corrolary C1, $\log\bar{F}_Y(y)\sim -k_2(\log|y|)^{k_1}-\log\left[\frac{k_1k_2}{|y|}(\log|y|)^{k_1-1}\right]$ if $k\geq 1$ and $k_2>1$ and $\log\bar{F}_Y(y)\sim o(\log|y|)$ if $k_1=k_2=1$.

(b) (c) From corrolary C1.

References

Ali, M. (1974). Stochastic ordering and kurtosis measure. Journal of the American Statistical Association, 69:543–545.

Andrews, D., Bickel, P., Hampel, F., Huber, P., Rogers, W., and Tukey, J. (1972). Robust estimation of location: Survey and advances. Technical report, Princeton University Press, Princeton, NJ.

Bahadur, R. and Savage, L. (1956). The non-existence of certain statistical procedures in non- the non-existence of certain statistical procedures in nonparametric problems. Annals of Statistics, 27:1115–1122.

Balanda, K. P. and MacGillivray, H. L. (1988). Kurtosis: A critical review. The American Statistician, 42:111–119.

Cole, T. J. and Green, P. J. (1992). Smoothing reference centile curves: the lms method and penalized likelihood. Statistics in Medicine., 11:1305–1319.

Crowder, M. J., Kimber, A. C., Smith R. L. and Sweeting, T. J. (1991). Statistical Analysis of Reliability Data. Chapman and Hall, London.

Davidson, R. (2012). Statistical inference in the presence of heavy tails. The Econometrics Journal, 15:31–53.

Dunn, P. K. and Smyth, G. K. (1996). Randomised quantile residuals. J. Comput. Graph. Statist., 5:236–244.

Fernandez, C. and Steel, M. F. J. (1998). On bayesian modelling of fat tails and skewness. J. Am. Statist. Ass., 93:359–371.

Fernandez, C., Osiewalski, J. and Steel, M. J. F. (1995). Modeling and inference with v-spherical distributions. J. Am. Statist. Ass., 90:1331–1340.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. Econometrica, 11(1):1–12.

Haavelmo, T. (1944). The probability approach in econometrics. Econometrica, 12:1–118.

Harter, H. L. (1967). Maximum-likelihood estimation of the parameters of a four parameter generalized gamma population from complete and censored samples. Technometrics, 9:159–165.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. Ann. Statist., 3:1163–1174.

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. Biometrika, 36:149–176.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). Continuous Univariate Distributions, Volume I, 2nd edn. Wiley, New York.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). Continuous Univariate Distributions, Volume II, 2nd edn. Wiley, New York.

Jones, M. C. (2005). In discussion of Rigby, R. A. and Stasinopoulos, D. M. (2005) Generalized additive models for location, scale and shape,. Applied Statistics, 54:507–554.

Jones, M. C. and Pewsey, A. (2009). Sinh-arcsinh distributions. Biometrika, 96:761–780.

Jørgensen, B. (1982). Statistical Properties of the Generalized Inverse Gaussian Distribution, *Lecture Notes in Statistics* No.**9**. Springer-Verlag, New York.

Jørgensen, B. (1997). The Theory of Dispersion Models. Chapman and Hall: London.

Lopatatzidis, A. and Green, P. J. (2000). Nonparametric quantile regression using the gamma distribution. Private Communication.

MacGillivray, H. (1986). Skewness and asymmetry: measures and orderings. Annals of Statistics, 14:994–1011.

Mandelbrot, B. (1997). Fractals and scaling in finance: discontinuity, concentration, risk: selecta volume E. Springer Verlag.

McDonald, J. B. (1991). Parametric models for partially adaptive estimation with skewed and leptokurtic residuals. Economic Letters, 37:273–278.

McDonald, J. B. (1996). Probability distributions for financial models. In Maddala, G. S. and Rao, C. R., editors, Handbook of Statistics, Vol. 14, pages 427–460. Elsevier Science.

McDonald, J. B. and Newey, W. K. (1988). Partially adaptive estimation of regression models via the generalized t distribution. Econometric Theory, 4:428–457.

McDonald, J. B. and Xu, Y. J. (1995). A generalisation of the beta distribution with applications. Journal of Econometrics, 66:133–152.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. Econometrica, 59:347–370.

Nolan, J. P. (2012). Stable Distributions - Models for Heavy Tailed Data. Birkhauser, Boston. In progress, Chapter 1 online at academic2.american.edu/∼jpnolan.

Pokorny, M. and Sedgwick, J. (2010). Profitability trends in Hollywood: 1929 to 1999: somebody must know something. Economic History Review, 63:56–84.

Rigby, R. A. and Stasinopoulos, D. M. (2004). Smooth centile curves for skew and kurtotic data modelled using the Box-Cox power exponential distribution. Statistics in Medicine, 23:3053–3076.

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape, (with discussion). Appl. Statist., 54:507–554.

Rigby, R. A. and Stasinopoulos, D. M. (2006). Using the Box-Cox t distribution in gamlss to model skewness and kurtosis. Statistical Modelling, 6:209–229.

Rosenberger, J. and Gasko, M. (1983). Comparing location estimators: Trimmed means, medians and trimean. In Hoaglin, D., Mosteller, F., and Tukey, J., editors, Understanding Robust and Exploratory Data Analysis, pages 297–338. John Wiley, New York.

Stasinopoulos, D. M., Rigby, R. A. and Akantziliotou, C. (2008). Instructions on how to use the gamlss package in r, second edition. Technical Report 01/08, STORM Research Centre, London Metropolitan University, London.

Voudouris, V. Gilchristand, R., Rigby, R., Sedgwick, J., and Stasinopoulos, D. (2012). Modelling skewness and kurtosis with the bcpe density in gamlss. Journal of Applied Statistics.

Yee, T. W. and Wild, C. J. (1996). Vector generalized additive models. Journal of the Royal Statistical Society, Series B (Methodological), 58(3):481–493.