

Kernel density estimation for heavy-tailed distributions using the champernowne transformation

TINE BUCH-LARSEN[†], JENS PERCH NIELSEN[‡], MONTSERRAT GUILLÉN^{*§} and CATALINA BOLANCO[¶]

[†]Department of Research, Codan, 60 Gammel Kongevej, DK-1790 Copenhagen V, Denmark

[‡]Royal & SunAlliance, 60 Gammel Kongevej, DK-1790 Copenhagen V, Denmark

[§]Department of Econometrics, RFA-IREA, University of Barcelona, Diagonal, 690, 08034 Barcelona, Spain

[¶]Department of Econometrics, RFA-IREA, University of Barcelona, 08034 Barcelona, Spain

(Received 17 January 2005; in final form 17 October 2005)

When estimating loss distributions in insurance, large and small losses are usually split because it is difficult to find a simple parametric model that fits all claim sizes. This approach involves determining the threshold level between large and small losses. In this article, a unified approach to the estimation of loss distributions is presented. We propose an estimator obtained by transforming the data set with a modification of the Champernowne cdf and then estimating the density of the transformed data by use of the classical kernel density estimator. We investigate the asymptotic bias and variance of the proposed estimator. In a simulation study, the proposed method shows a good performance. We also present two applications dealing with claims costs in insurance.

Keywords: Actuarial loss models; Transformation; Skewness; Champernowne distribution

2000 Mathematics Subject Classifications: 62G07; 62-07; 91B30

1. Introduction

In finance and non-life insurance, estimation of loss distributions is a fundamental part of the business. In most situations, losses are small, and extreme losses are rarely observed, but the number and the size of extreme losses can have a substantial influence on the profit of the company. Standard statistical methodology, such as integrated error and likelihood, does not weigh small and big losses differently in the evaluation of an estimator. These evaluation methods do not, therefore, emphasize an important part of the error: the error in the tail.

Practitioners often decide to analyse large and small losses separately, because no single, classical parametric model fits all claim sizes. This approach leaves some important challenges: choosing the appropriate parametric model, identifying the best way of estimating the parameters and determining the threshold level between large and small losses.

*Corresponding author. Email: mguillen@ub.edu

This work presents a systematic approach to the estimation of loss distributions which is suitable for heavy tailed situations. The proposed estimator is obtained by transforming the data set with a parametric estimator and afterwards estimating the density of the transformed data set using the classical kernel density estimator [1, 2]

$$\hat{f}(y) = \frac{1}{Nb} \sum_{i=1}^N K\left(\frac{y - Y_i}{b}\right),$$

where K is the kernel function, b is the bandwidth and $Y_i, i = \{1, \dots, N\}$ is the transformed data set. The estimator of the original density is obtained by back-transformation of $\hat{f}(y)$. We will call this method a *semiparametric estimation procedure* because a parametrized transformation family is used. We propose to use a transformation based on the little-known Champernowne cdf, because it produces good results in all the studied situations and it is straightforward to apply.

The semiparametric estimator with shifted power transformation was introduced by Wand *et al.* [3] in 1991. They showed that the classical kernel density estimator was improved substantially by applying a transformation and suggested the shifted power transformation family. Bolancé *et al.* [4] improved the shifted power transformation for highly skewed data by proposing an alternative parameter selection algorithm. The semiparametric estimator with the Johnson family transformation function was studied by Yang and Marron [5]. Hjort and Glad [6] advocated a semiparametric estimator with a parametric start, which is closely related to the bias reduction method described by Jones *et al.* [7]. The Möbius-like transformation was introduced by Clements *et al.* [8]. In contrast to the shifted power transformation, which transforms $(0, \infty)$ into $(-\infty, \infty)$, the Möbius-like transformation transforms $(0, \infty)$ into $(-1, 1)$ and the parameter estimation method is designed to avoid boundary problems. Scaillet [9] has recently studied non-parametric estimators for probability density functions which have support on the non-negative real line using alternative kernels.

The original Champernowne distribution has density [10]

$$f(x) = \frac{c}{x((1/2)(x/M)^{-\alpha} + \lambda + (1/2)(x/M)^{\alpha})} \quad x \geq 0, \quad (1)$$

where c is a normalizing constant and α, λ and M are parameters. The distribution was mentioned for the first time in 1936 by D.G. Champernowne when he spoke on ‘The Theory of Income Distribution’ at the Oxford Meeting of the Econometric Society [11, 12] in 1936. Later, he gave more details about the distribution [13], and its application to economics. When λ equals 1 and the normalizing constant c equals $(1/2)\alpha$, the density of the original distribution is simply called the *Champernowne distribution*

$$f(x) = \frac{\alpha M^{\alpha} x^{\alpha-1}}{(x^{\alpha} + M^{\alpha})^2}$$

with cdf

$$F(x) = \frac{x^{\alpha}}{x^{\alpha} + M^{\alpha}}. \quad (2)$$

The Champernowne distribution converges to a Pareto distribution in the tail, while looking more like a lognormal distribution near 0 when $\alpha > 1$. Its density is either 0 or infinity at 0 (unless $\alpha = 1$).

In the transformation kernel density estimation method, if we transform the data with the Champernowne cdf, the inflexible shape near 0 results in boundary problems. We argue that a modification of the Champernowne with an additional parameter can solve this inconvenience.

We did not choose to work with classical extensions of the Pareto distribution such as the generalized Pareto distribution [14], GPD. The reason for doing so is that the GPD often estimates distributions of infinite support to have finite support and hence it cannot be used as a transformation. We carried out a small simulation study of a standard lognormal distribution; more than half the time the GPD suggested a distribution with finite support. Furthermore, the GPD needs a (hard to pick) threshold from where the distribution starts; such that the transformation methodology meets problems also in the beginning of the distribution.

In this paper, we study the transformation kernel density estimation method. The conclusion of the simulation study is that the new approach based on the modified Champernowne distribution is the preferable method, because it is the only estimator which has a good performance in most of the investigated situations. Section 2 describes the transformation family and explains the parameter estimation procedure. Section 3 presents the semiparametric kernel density estimator and its properties. In section 4, the simulation study is presented and section 5 shows two applications. Finally, section 6 outlines the main conclusions.

2. The modified Champernowne distribution function

We generalize the Champernowne distribution with a new parameter c . This parameter ensures the possibility of a positive finite value of the density at 0 for all α .

DEFINITION 2.1 *The modified Champernowne cdf is defined for $x \geq 0$ and has the form*

$$T_{\alpha,M,c}(x) = \frac{(x+c)^\alpha - c^\alpha}{(x+c)^\alpha + (M+c)^\alpha - 2c^\alpha} \quad \forall x \in \mathbb{R}_+ \quad (3)$$

with parameters $\alpha > 0$, $M > 0$ and $c \geq 0$ and density

$$t_{\alpha,M,c}(x) = \frac{\alpha(x+c)^{\alpha-1}((M+c)^\alpha - c^\alpha)}{((x+c)^\alpha + (M+c)^\alpha - 2c^\alpha)^2} \quad \forall x \in \mathbb{R}_+.$$

Corresponding to the Champernowne distribution, the modified Champernowne distribution converges to a Pareto distribution in the tail:

$$t_{\alpha,M,c}(x) \rightarrow \frac{\alpha \left(((M+c)^\alpha - c^\alpha)^{1/\alpha} \right)^\alpha}{x^{\alpha+1}} \quad \text{as } x \rightarrow \infty.$$

The effect of the additional parameter c is different for $\alpha > 1$ and for $\alpha < 1$. The parameter c has some ‘scale parameter properties’: when $\alpha < 1$, the derivative of the cdf becomes larger for increasing c , and conversely, when $\alpha > 1$, the derivative of the cdf becomes smaller for increasing c . When $\alpha \neq 1$, the choice of c affects the density in three ways. First, c changes the density in the tail. When $\alpha < 1$, positive c s result in lighter tails, and the opposite when $\alpha > 1$. Secondly, c changes the density in 0. A positive c provides a positive finite density in 0:

$$0 < t_{\alpha,M,c}(0) = \frac{\alpha c^{\alpha-1}}{(M+c)^\alpha - c^\alpha} < \infty \quad \text{when } c > 0.$$

Thirdly, c moves the mode. When $\alpha > 1$, the density has a mode, and positive c s shift the mode to the left. We therefore see that the parameter c also has a shift parameter effect. When $\alpha = 1$, the choice of c has no effect.

Figure 1 illustrates the role of c : the two graphs on the top show the cdfs and the densities for the modified Champernowne distribution for fixed $\alpha < 1$ and $M = 3$. In the cdf plot, we

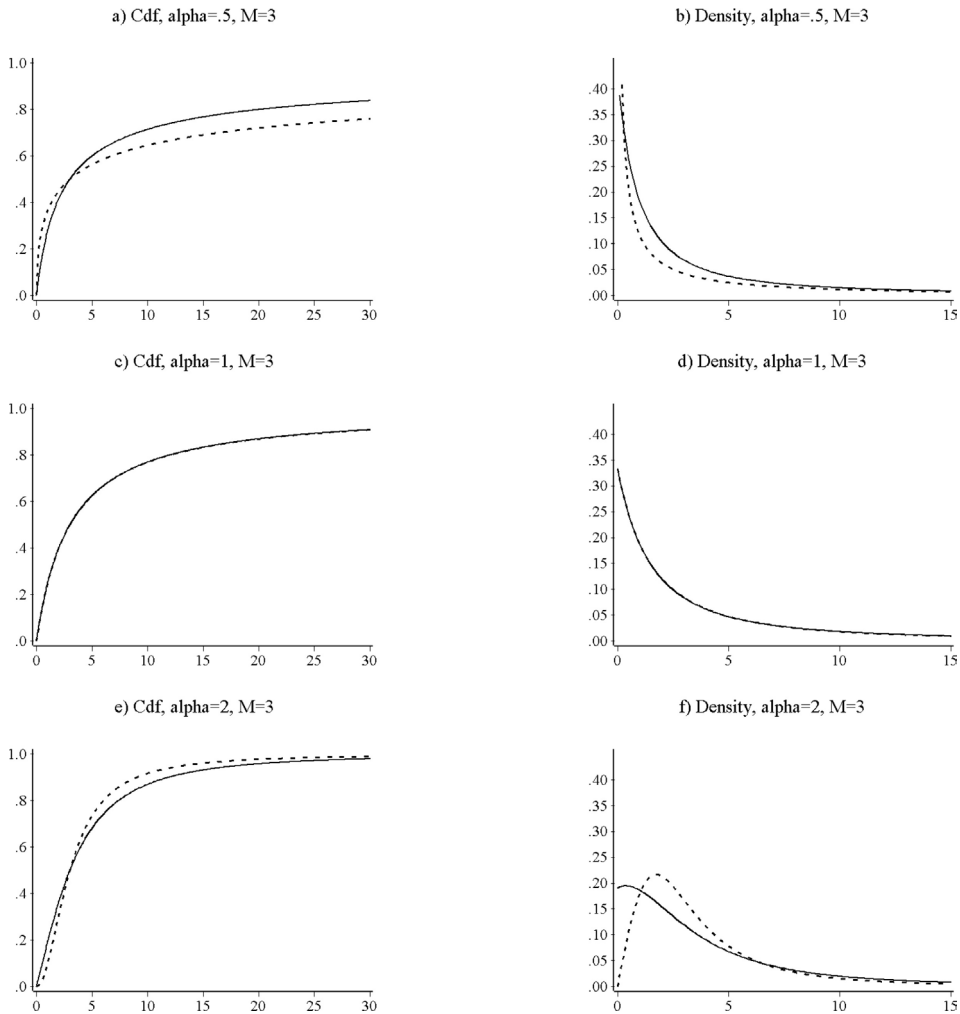


Figure 1. Different shapes of the modified Champernowne distribution with different choices of α , as well as the effect of the parameter c . In all plots $c = 0$, dashed line and $c = 2$, solid line.

see that increasing c results in lower values of the cdf in the interval $[0, M)$ and higher values of the cdf in the interval $[M, \infty)$. In the density plot, we see that increasing c results in a lighter tail and a finite density at 0. In the two graphs in the middle, we have fixed $\alpha = 1$ and $M = 3$. We see that changing c has no effect. The two graphs at the bottom illustrate the effect of increasing c when $\alpha > 1$, for $M = 3$. Notice that the values of the cdf become higher in the interval $[0, M)$ and lower in the interval $[M, \infty)$. The density plot shows that positive c s move the mode to the left and produce a heavier tail.

From a computational point of view, it is simpler to estimate M and then proceed to the other parameters.

In the Champernowne distribution, we notice that $T_{\alpha, M, 0}(M) = 0.5$. The same holds for the modified Champernowne distribution: $T_{\alpha, M, c}(M) = 0.5$. This suggests that M can be estimated as the empirical median of the data set. The empirical median is a robust estimator, especially for heavy-tailed distributions, as shown by Lehmann [15]. He studied the properties of the median and the mean as an estimator of location for the normal distribution and the Cauchy

distribution, and showed that whereas the mean works well as an estimator of location for the normal distribution, it works poorly for the Cauchy distribution due to its heavy tail. Tukey [16] reached the same conclusion when he studied the efficiency of the median and the mean. He showed that the median efficiency increases as the tail becomes heavier. Corresponding models have also been studied for heavy-tailed distributions [17–19]. A similar type of discussion for the variance estimation was done by Hubert [20]. As we are especially concerned about heavy tails, we consider the robustness of the median to be important.

After parameter M has been estimated as described earlier, the next step is to estimate the pair (α, c) which maximizes the log likelihood function:

$$l(\alpha, c) = N \log \alpha + N \log ((M + c)^\alpha - c^\alpha) + (\alpha - 1) \sum_{i=1}^N \log(X_i + c) - 2 \sum_{i=1}^N \log ((X_i + c)^\alpha + (M + c)^\alpha - 2c^\alpha). \quad (4)$$

For a fixed M , this likelihood function is concave and has a maximum.

3. The semiparametric transformation kernel density estimator

In this section, we will make a detailed derivation of the estimator based on the modified Champernowne distribution, which we will call KMCE. The resulting estimator is obtained by computing a non-parametric classical kernel density estimator for the transformed data set and, finally, the result is back-transformed.

3.1 Transformation with the modified Champernowne distributions

Let $X_i, i = 1, \dots, N$, be positive stochastic variables with an unknown cdf F and density f . The following describes in detail the transformation kernel density estimator of f , and figure 2 illustrates the four steps of the estimation procedure for a data set with 1000 observations generated from a Weibull distribution. The resulting transformation kernel density estimator of f based on the Champernowne distribution is denoted by KMCE.

- (i) Calculate the parameters $(\hat{\alpha}, \hat{M}, \hat{c})$ of the modified Champernowne distribution as described in section 2 to obtain the transformation function. In the first plot in figure 2, we see the estimated transformation function and the true Weibull distribution. Notice that the modified Champernowne density has a larger mode and that the tail is too heavy.
- (ii) Transform the data set $X_i, i = 1, \dots, N$, with the transformation function, T :

$$Y_i = T_{\hat{\alpha}, \hat{M}, \hat{c}}(X_i), \quad i = 1, \dots, N.$$

The transformation function transforms data into the interval $(0, 1)$, and the parameter estimation is designed to make the transformed data as close to a uniform distribution as possible. The transformed data are illustrated in the second plot in figure 2.

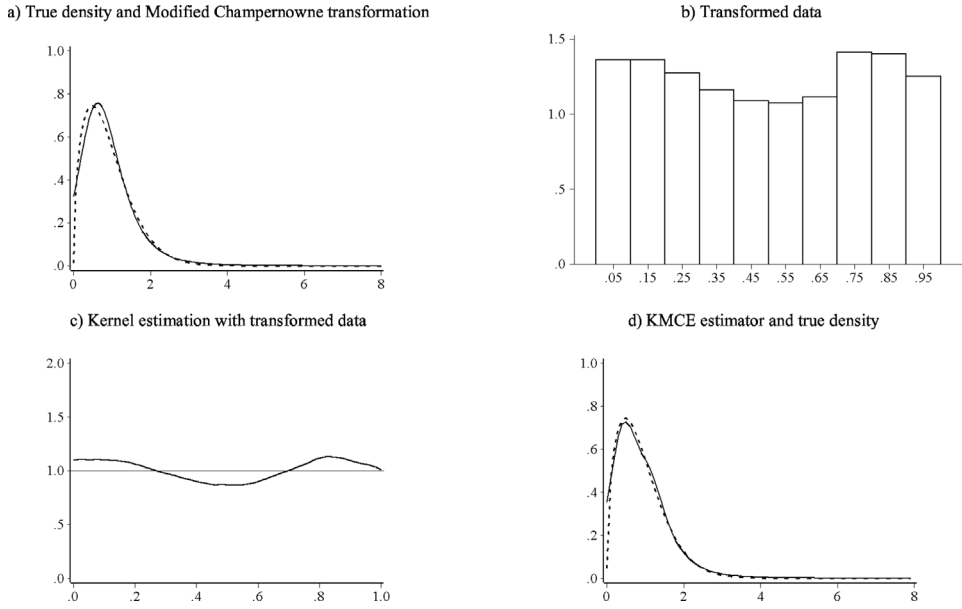


Figure 2. Four steps of the KMCE estimator. a) The estimated transformation function (solid line) and the true density (dashed line). b) The histogram of the transformed data set. c) The estimated classical kernel density estimator of the transformed data set. d) The final KMCE estimator (solid line) and the true density (dashed line).

(iii) Calculate the classical kernel density estimator on the transformed data, $Y_i, i = 1, \dots, N$:

$$\hat{f}_{\text{trans}}(y) = \frac{1}{Nk_y} \sum_{i=1}^N K_b(y - Y_i),$$

where $K_b(\cdot) = (1/b)K(\cdot)$ and $K(\cdot)$ is the kernel function. The boundary correction, k_y , is required because the Y_i are in the interval $(0, 1)$ so that we need to divide by the integral of the part of the kernel function that lies in this interval. The boundary correction k_y is defined as

$$k_y = \int_{\max(-1, -y/b)}^{\min(1, (1-y)/b)} K(u) du.$$

The classical kernel density estimator of the transformed data set is illustrated in the third plot in figure 2.

(iv) The classical kernel density estimator of the transformed data set results in the KMCE estimator on the transformed scale. Therefore, the estimator of the density of the original data set, $X_i, i = 1, \dots, N$ is

$$\hat{f}(x) = \frac{\hat{f}_{\text{trans}}(T_{\hat{\alpha}, \hat{M}, \hat{c}}(x))}{\left| \left(T_{\hat{\alpha}, \hat{M}, \hat{c}}^{-1} \right)' (T_{\hat{\alpha}, \hat{M}, \hat{c}}(x)) \right|}.$$

The KMCE results for the Weibull data set is seen in the last plot in figure 2.

The expression of the KMCE is

$$\hat{f}(x) = \frac{1}{N k_{T_{\hat{\alpha}, \hat{M}, \hat{c}}(x)}} \sum_{i=1}^N K_b(T_{\hat{\alpha}, \hat{M}, \hat{c}}(x) - T_{\hat{\alpha}, \hat{M}, \hat{c}}(X_i)) T'_{\hat{\alpha}, \hat{M}, \hat{c}}(x). \quad (5)$$

3.2 Asymptotic theory for the transformation kernel density estimator

In this section, we investigate the asymptotic theory of the transformation kernel density estimator in general. We derive its asymptotic bias and variance.

THEOREM 3.1 *Let X_1, \dots, X_N be independent identically distributed variables with density f . Let $\hat{f}(x)$ be the transformation kernel density estimator of $f(x)$*

$$\hat{f}(x) = \frac{1}{N} \sum_{i=1}^N K_b(T(x) - T(X_i))T'(x),$$

where $T(\cdot)$ is the transformation function.

Then the bias and the variance of $\hat{f}(x)$ are given by

$$\begin{aligned} \mathbb{E}[\hat{f}(x)] - f(x) &= \frac{1}{2} \mu_2(K) b^2 \left(\left(\frac{f(x)}{T'(x)} \right)' \frac{1}{T'(x)} \right)' + o(b^2), \\ \mathbb{V}[\hat{f}(x)] &= \frac{1}{Nb} R(K) T'(x) f(x) + o\left(\frac{1}{Nb}\right) \end{aligned}$$

as $N \rightarrow \infty$, where $\mu_2(K) = \int u^2 K(u) du$ and $R(K) = \int K^2(u) du$.

Proof We assume that X_1, \dots, X_N are independent identically distributed variables with density f . Let $\hat{f}(x)$ be the transformation kernel density estimator of $f(x)$:

$$\hat{f}(x) = \frac{1}{N} \sum_{i=1}^N K_b(T(x) - T(X_i))T'(x),$$

where $T(\cdot)$ is the transformation function. Let the transformed variable have distribution g :

$$Y_i = T(X_i) \sim g(y) = \frac{f(T^{-1}(y))}{T'(T^{-1}(y))}$$

and let $\hat{g}(y)$ be the classical kernel density estimator of $g(y)$:

$$\hat{g}(y) = \frac{1}{N} \sum_{i=1}^N K_b(y - Y_i).$$

The mean and variance of the classical kernel density estimator is

$$\mathbb{E}[\hat{g}(y)] = g(y) + \frac{1}{2} b^2 \mu_2(K) g''(y) + o(b^2), \quad (6)$$

$$\mathbb{V}[\hat{g}(y)] = \frac{1}{Nb} R(K) g(y) + o\left(\frac{1}{Nb}\right). \quad (7)$$

The transformation kernel density estimator can be expressed by the standard kernel density estimator:

$$\hat{f}(x) = T'(x) \hat{g}(T(x))$$

implying

$$\begin{aligned}\mathbb{E}[\widehat{f}(x)] &= T'(x)\mathbb{E}[\widehat{g}(T(x))] \\ &= T'(x) \left(g(T(x)) + \frac{1}{2}b^2\mu_2(K) \frac{\partial^2 g(T(x))}{\partial(T(x))^2} + o(b^2) \right).\end{aligned}\quad (8)$$

Note that

$$g(T(x)) = \frac{f(x)}{T'(x)}, \quad \frac{\partial g(T(x))}{\partial T(x)} = \left(\frac{f(x)}{T'(x)} \right)' \frac{1}{T'(x)},$$

and

$$\frac{\partial^2 g(T(x))}{\partial(T(x))^2} = \left(\left(\frac{f(x)}{T'(x)} \right)' \frac{1}{T'(x)} \right)' \frac{1}{T'(x)}$$

which are used to find the mean of the transformation kernel density estimator

$$\mathbb{E}[\widehat{f}(x)] = f(x) + \frac{1}{2}b^2\mu_2(K) \left(\left(\frac{f(x)}{T'(x)} \right)' \frac{1}{T'(x)} \right)' + o(b^2).\quad (9)$$

The variance is calculated in a similar way

$$\begin{aligned}\mathbb{V}[\widehat{f}(x)] &= (T'(x))^2 \mathbb{V}[\widehat{g}(T(x))] \\ &= (T'(x))^2 \left(\frac{1}{Nb} R(K) g(T(x)) + o\left(\frac{1}{Nb}\right) \right) \\ &= \frac{1}{Nb} R(K) T'(x) f(x) + o\left(\frac{1}{Nb}\right).\end{aligned}\quad (10)$$

It is known [21] that the classical kernel density estimator follows a normal distribution asymptotically:

$$\sqrt{Nb} (\widehat{g}(y) - \mathbb{E}[\widehat{g}(y)]) \sim \mathbf{N}\left(0, \frac{1}{Nb} R(K) g(y)\right).$$

Then, as $\widehat{f}(x) = T'(x)\widehat{g}(y)$ with $y = T(x)$, then

$$\sqrt{Nb} (\widehat{f}(x) - \mathbb{E}[\widehat{f}(x)]) \sim \mathbf{N}\left(0, \frac{1}{Nb} R(K) T'(x) f(x)\right).$$

For a parametric transformation $T(x) = T_\theta(x)$, if we assume that $\hat{\theta}$ is a square-root- n consistent estimator of θ , then it follows that the asymptotic distribution of $\widehat{f}(x)$ with parametric estimated transformation $T_{\hat{\theta}}(x)$, equals the asymptotic distribution of $\widehat{f}(x)$ with parametric transformation $T_\theta(x)$.

4. Simulation study

This section presents a comparison of our semiparametric method based on the modified Champernowne distributions with two benchmark estimators. We simulate data from four distributions with different tails and different shapes near 0. We measure the error between the estimated density and the true density using four different error measures. In subsection 4.3, we evaluate the performance of the KMCE estimators compared to the estimator described by Clements *et al.* [8], in the following called CHL, and the estimator described by Bolancé *et al.* [4], in the following called BGN.

4.1 The distributions

We have simulated four distributions with different characteristics: *lognormal*, *lognormal–Pareto*, *Weibull* and *truncated logistic*. The lognormal distribution has a moderately light tail, and when we mix the lognormal distribution with the Pareto distribution, which is a heavy-tailed distribution, the resulting distribution is also heavy-tailed. The Weibull distribution is a light-tailed distribution that starts at 0 and has a mode. The truncated logistic is a light-tailed distribution that has a positive finite density at 0. The distributions and the chosen parameters are listed in table 1 and figure 3 plots the densities to show the diversity of shapes.

4.2 Measuring the error

We measure the performance of the estimators by the error measures L_1 , L_2 , WISE and E. Let $\hat{f}(x)$ be the estimated density and $f(x)$ be the true density. The L_1 norm measures the distance between the estimated density and the true density on the whole support.

$$L_1 = \int_0^\infty |\hat{f}(x) - f(x)| \, dx.$$

We also calculate the L_2 norm between the two distributions.

$$L_2 = \left(\int_0^\infty (\hat{f}(x) - f(x))^2 \, dx \right)^{1/2}.$$

Both L_1 and L_2 weigh errors of the estimator near 0 and in the tail equally, although the consequences for some real-world situations of a poor estimation in the tail are much more critical than the consequences of a poor estimation near 0.

Table 1. Distributions used in the simulation study.

Distribution	Density for $x > 0$	Parameters
Lognormal (μ, σ^2)	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2x}} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}$	$(\mu, \sigma^2) = (0, 0.5)$
Mixtures of p lognormal (μ, σ) and $(1-p)$ Pareto (λ, ρ, c)	$f(x) = p \frac{1}{\sqrt{2\pi\sigma^2x}} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} + (1-p)(x-c)^{-(\rho+1)} \rho \lambda^\rho$	$(p, \mu, \sigma, \lambda, \rho, c) = (0.7, 0, 1, 1, 1, -1)$ $= (0.3, 0, 1, 1, 1, -1)$
Weibull (γ)	$f(x) = \gamma x^{(\gamma-1)} e^{-x^\gamma}$	$\gamma = 1.5$
Truncated logistic	$f(x) = \frac{2}{s} e^{\frac{x}{s}} \left(1 + e^{\frac{x}{s}}\right)^{-2}$	$s = 1$

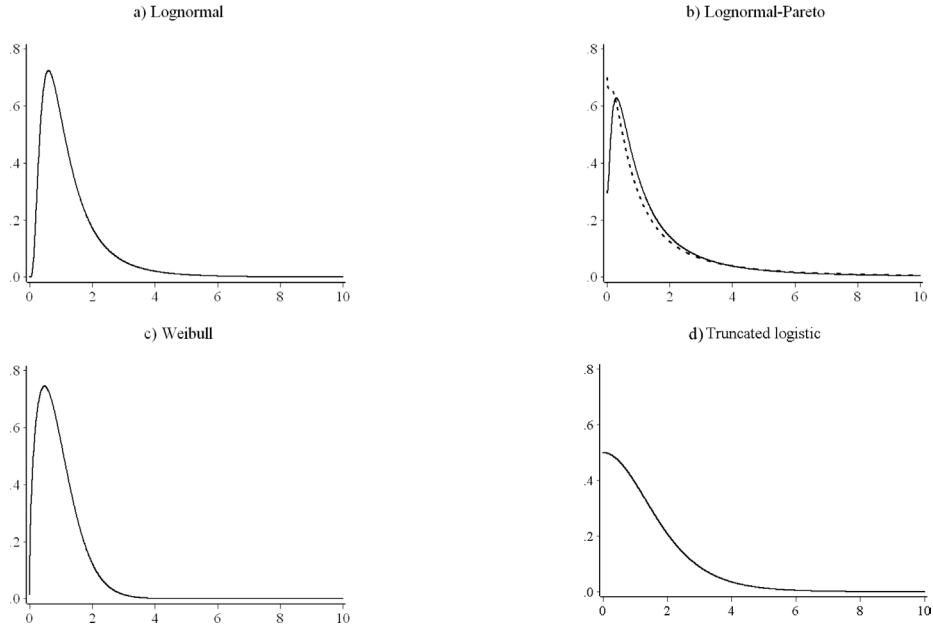


Figure 3. Shape of the different distributions used in the simulation study. In plot b) 30% Pareto (solid line) and 70% Pareto (dashed line).

WISE weighs the distance between the estimated and the true distribution with the squared value of x . This results in an error measure that emphasizes the tail of the distribution, which is very relevant in practice when dealing with income or cost data.

$$\text{WISE} = \left(\int_0^{\infty} (\hat{f}(x) - f(x))^2 x^2 dx \right)^{1/2}.$$

The last error measure, E , calculates the distance between the estimated mean excess function and the true mean excess function. It emphasizes the error in the tail as well.

$$\begin{aligned} E &= \left(\int_0^{\infty} (\hat{e}(x) - e(x))^2 f(x) dx \right)^{1/2} \\ &= \left(\int_0^{\infty} \left(\int_x^{\infty} u (f(u) - \hat{f}(u)) du \right)^2 f(x) dx \right)^{1/2}, \end{aligned}$$

To calculate the error measures, we used the change of variable $y = (x - M)/(x + M)$ proposed by Clements *et al.* [8].

4.3 Comparison of the estimation methods

We compare the performance of the KMCE, the CHL and the BGN estimators. The comparison is based on data simulated from the four distributions described in table 1, and four sample sizes: $N = 50$, $N = 100$, $N = 500$ and $N = 1000$. Each combination of distribution and sample size is replicated 2000 times. In tables 2 and 3, we show the means of the error measures for the 2000 samples. We show the results obtained when using the rule of thumb method [2] for bandwidth selection. We also investigated the Seather and Jones [22] method and the conclusions do not change.

Table 2. The estimated error measures for sample size 50 and 100 based on 2000 repetitions.

				Log-Pareto				
				Log-normal	$p = 0.7$	$p = 0.3$	Weibull	Tr. Logist.
$N = 50$	L_1	KMCE	0.1821	0.1713	0.1664	0.1855	0.1732	
		CHL	0.1811	0.1815	0.1860	0.1955	0.2009	
		BGN	0.1927	0.1860	0.1895	0.1852	0.1892	
	L_2	KMCE	0.1402	0.1130	0.1099	0.1420	0.1065	
		CHL	0.1326	0.1311	0.1617	0.1338	0.1183	
		BGN	0.1456	0.1151	0.1267	0.1391	0.1257	
	WISE	KMCE	0.1391	0.1139	0.1299	0.1178	0.1281	
		CHL	0.1316	0.1385	0.1693	0.1288	0.1546	
		BGN	0.1431	0.1347	0.1588	0.1217	0.1403	
	E	KMCE	0.0373	0.0760	0.1474	0.0313	0.0480	
		CHL	0.0388	0.1109	0.2295	0.0359	0.0574	
		BGN	0.0402	0.0974	0.2024	0.0338	0.0523	
$N = 100$	L_1	KMCE	0.1363	0.1287	0.1236	0.1393	0.1294	
		CHL	0.1381	0.1412	0.1423	0.1533	0.1468	
		BGN	0.1451	0.1383	0.1413	0.1426	0.1578	
	L_2	KMCE	0.1047	0.0862	0.0837	0.1084	0.0786	
		CHL	0.1018	0.1021	0.1244	0.1044	0.1000	
		BGN	0.1100	0.0855	0.0972	0.1079	0.0924	
	WISE	KMCE	0.1039	0.0859	0.0958	0.0886	0.0977	
		CHL	0.1018	0.1068	0.1276	0.1029	0.1078	
		BGN	0.1093	0.1004	0.1191	0.0939	0.1241	
	E	KMCE	0.0268	0.0572	0.1073	0.0224	0.0344	
		CHL	0.0289	0.0816	0.1677	0.0277	0.0391	
		BGN	0.0297	0.0716	0.1572	0.0255	0.0443	

Table 3. The estimated error measures for sample size 500 and 1000 based on 2000 repetitions.

				Log-Pareto				
				Log-normal	$p = 0.7$	$p = 0.3$	Weibull	Tr. Logist.
$N = 500$	L_1	KMCE	0.0786	0.0676	0.0646	0.0831	0.0745	
		CHL	0.070	0.0786	0.0786	0.0869	0.0915	
		BGN	0.0761	0.0703	0.0761	0.0763	0.0791	
	L_2	KMCE	0.0585	0.0480	0.0470	0.0676	0.0437	
		CHL	0.0560	0.0581	0.0685	0.0594	0.0591	
		BGN	0.0590	0.0454	0.0601	0.0588	0.0574	
	WISE	KMCE	0.0585	0.0471	0.0517	0.0530	0.0598	
		CHL	0.0555	0.0587	0.0685	0.0604	0.0718	
		BGN	0.0579	0.0507	0.0686	0.0510	0.0585	
	E	KMCE	0.0125	0.0306	0.0591	0.0111	0.0171	
		CHL	0.0143	0.0405	0.0808	0.0149	0.0235	
		BGN	0.0145	0.0342	0.0974	0.0129	0.0120	
$N = 1000$	L_1	KMCE	0.0659	0.0530	0.0507	0.0700	0.0598	
		CHL	0.0572	0.0606	0.0609	0.0688	0.0730	
		BGN	0.0684	0.0541	0.0584	0.0583	0.0587	
	L_2	KMCE	0.0481	0.0389	0.0393	0.0582	0.0339	
		CHL	0.0435	0.0450	0.0528	0.0476	0.0521	
		BGN	0.0454	0.0360	0.0509	0.0453	0.0434	
	WISE	KMCE	0.0481	0.0384	0.0417	0.0450	0.0501	
		CHL	0.0434	0.0453	0.0524	0.0478	0.0561	
		BGN	0.0448	0.0390	0.0539	0.0394	0.0437	
	E	KMCE	0.0094	0.0251	0.0492	0.0084	0.0126	
		CHL	0.0107	0.0295	0.0588	0.0113	0.0178	
		BGN	0.0108	0.0255	0.0780	0.0096	0.0143	

For the moderately light-tailed lognormal, all three estimators exhibit good performance in general. The KMCE and CHL estimators show the best performance. The CHL estimator outperforms the KMCE estimator for all N , but it seems to outperform the KMCE estimator near 0 rather than in the tail, as seen by the fact that the improvement obtained on L_1 and L_2 is greater than the improvement obtained on WISE and E. The BGN estimator also performs well in this case. The performance of this estimator is only 3–4% worse than the performance of the KMCE estimator.

For the heavy-tailed distributions, the KMCE estimator shows a significantly better performance than the CHL and the BGN estimators. The performance of the CHL estimator is poor compared to the KMCE estimator. For the 70% lognormal–30% Pareto, the KMCE estimator outperforms the CHL estimator by about 15–20%, and the performance gap seems to become larger when N increases. The largest performance gap occurs with WISE and E, which indicates that the performance gap is mainly in the tail. The BGN estimator is also outperformed by the KMCE estimator: the error measures are about 10% better for the KMCE estimator than the BGN estimator for the 70% lognormal–30% Pareto, but the improvement seems to go down when N increases.

The results for the 30% lognormal–70% Pareto are similar to the previous ones. The KMCE estimator still outperforms the CHL and BGN estimators. This indicates that when the tail becomes heavier, the benefits of using the KMCE estimator instead of the CHL and the BGN estimators become greater. For the lognormal–Pareto distribution, the parameter c in the KMCE estimator tends to 0 when N increases. Comparing the estimated α s for the KMCE estimator, we observe that the α s are around 1.4–1.8 for the 70% lognormal–30% Pareto distribution, whereas they are around 1.2–1.3 for the 30% lognormal–70% Pareto distribution. This is due to the fact that the 30% lognormal–70% Pareto distribution has a heavier tail than the 70% lognormal–30% Pareto distribution.

For the light-tailed Weibull distribution, we can see that the KMCE, the CHL and the BGN estimators show good performance. The KMCE estimator is 5–20% worse on L_2 compared to the CHL estimator, and about 10% better with respect to WISE. This means that the KMCE estimator near 0 is worse than the CHL estimator, whereas the KMCE estimator is better than the CHL estimator in the tail. As compared to the BGN estimator, the KMCE estimator also gives a similar performance.

For the truncated logistic distribution, the KMCE and the BGN estimators show good performance. The bad performance of the CHL estimator is due to the fact that the transformation functions in this case always starts at 0 when $\alpha > 1$. The estimator therefore transforms the true distribution, which has a positive value at 0, with a function that is 0 at 0, and this gives a positive value divided by 0, which results in a bad fit near 0. We see that the CHL estimator underestimates the true distribution around 0 for all values of N . The KMCE estimator also underestimates the true density around 0, but when N increases, the error around 0 decreases. We have also seen that the KMCE estimator overestimates the tail, which is because the transformation function has a heavy Pareto tail.

The main conclusion of our simulation study is that the KMCE estimator is recommended for heavy tailed situations.

We have designed the simulation study to be comparable to the simulation study in Clements *et al.* [8]. They compared their estimator to the transformation kernel density estimator with the modified-power transformation function proposed by Wand *et al.* [3], which we call WMR, and the transformation kernel density estimator with the iterated transformation function suggested by Yang and Marron [5], which we called YM. We can therefore use their simulation study to compare our estimators and the BGN estimator with the WMR and the YM estimators, even though the CHL simulation study only compares the L_1 and the L_2 error measures, and not error measures that emphasize the tail.

The CHL simulation study [8] shows that the CHL estimator performs well in general compared with WMR and YM. But for some distributions, the CHL estimator is outperformed by one of the other estimators. For the heavy-tailed 70% lognormal–30% Pareto distribution, the CHL estimator is outperformed with respect to L_2 by the YM estimator, but the performance of the KMCE estimator in our simulation study is even better than that of the YM estimator in the heavy-tail situation. For the Weibull distribution, the WMR estimator still gives very good performance compared with both the CHL and KMCE estimators. On the other hand, we are also able to make a comparison between the BGN estimator and the WMR and the YM estimators: the BGN estimator outperforms the WMR and the YM estimator in all situations investigated in the CHL simulation study.

5. Data study

In this section, we will apply our semiparametric estimation method to two data sets. The first data set contains automobile claims from a Spanish insurance company, and the second data set is about employer's liability from an Irish insurance company. The first data set was analysed in detail by Bolancé *et al.* [4]. It is a typical insurance claims amount data set: it contains a lot of observations and it seems to be heavy-tailed. Unlike the automobile insurance, the liability data set from Ireland is rather light-tailed. The reason is that claims are undeveloped, i.e., large claims are underrepresented in this data set because they take longer to process.

5.1 Automobile claims

We study bodily injury payments from automobile accidents occurring in Spain in 1997. The data are divided into two age groups: claims from policyholders who are less than 30 years old and claims from policyholders who are 30 years old or older. The first group of the data set consists of 1061 observations in the interval $[1; 126,000]$ with mean value 402.7. The second group consists of 4061 observations in the interval $[1; 17,000]$ with mean value 243.1. Estimation of the parameters in the modified Champernowne distribution function is for young drivers $\hat{\alpha}_1 = 1.116$, $\hat{M}_1 = 66$, $\hat{c}_1 = 0.000$ and for older drivers $\hat{\alpha}_2 = 1.145$, $\hat{M}_2 = 68$, $\hat{c}_2 = 0.000$, respectively. The bandwidths are $b_1 = 0.172$ and $b_2 = 0.134$. Figure 4 presents the classical kernel density estimator of the transformed data separated in the two age groups.

We notice that $\alpha_1 < \alpha_2$, which indicates that the data set for young drivers has a heavier tail than the data set for older drivers.

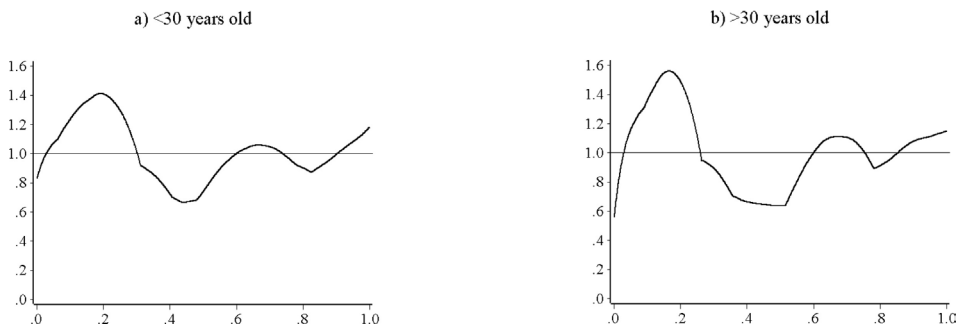


Figure 4. Classical kernel density estimator of the transformed automobile claims separated into policyholders <30 years old and >30 years old from an insurance company.

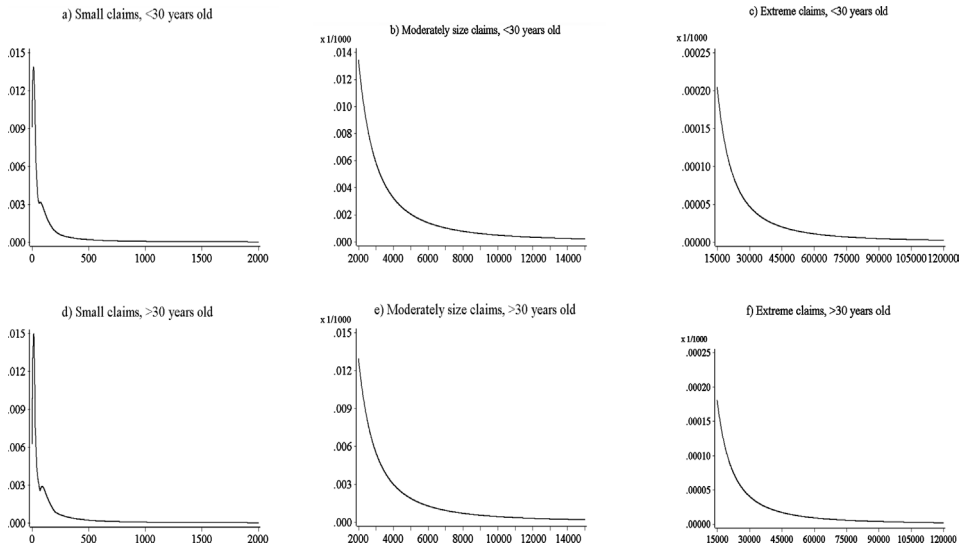


Figure 5. KMCE estimator of automobile claims from an insurance company, claims are separated into policyholders <30 years old and >30 years old, split into three groups.

Figure 5 shows the resulting KMCE estimator for the two groups of policyholders. The claims have been split into three categories: *Small claims* in the interval $(0; 2000)$, *moderately sized claims* in the interval $[2000; 14,000)$, and *extreme claims* in the interval $[14,000; \infty)$. The figure illustrates that the tail in the estimated density of young policyholders is heavier than the tail of the estimated density of older policyholders. This can be taken as evidence that young drivers are more likely to claim a large amount so that they should pay a higher premium than older drivers. Therefore, the method is useful to identify high risk groups, i.e., those having more extreme claims. The usefulness of the methodology is specially interesting in this point. It allows to plot the estimated density in regions where data are scarce. If risk groups (such as young drivers or type of vehicles) are plotted separately, the density estimates inform about the risk orderings (i.e., which type of customers are likely to claim an extreme cost).

5.2 Employer's liability

In this section, we will apply our semiparametric estimation method to the costs of employer's liability from an Irish insurance company. The data set consists of 2522 claims. Here we want to see the effect of not including the additional c parameter in the transformation. The estimation of the parameters in the modified Champernowne distribution is $\hat{\alpha} = 1.955$, $\hat{M} = 32379.307$, $\hat{c} = 64758.614$ and bandwidth $b = 0.147$. When c is assumed equal to 0, then $\hat{\alpha} = 0.954$ while \hat{M} is the same because it corresponds to the sample median. Figure 6 presents the classical kernel density estimator of the transformed data, using two different values of c .

In figure 7, we plot the estimators on the original scale. The estimators are nearly identical for small and moderate claims (low costs), whereas the KMCE with $c = 0$ overestimated the tail. This shows the importance of considering the modified Champernowne distribution.

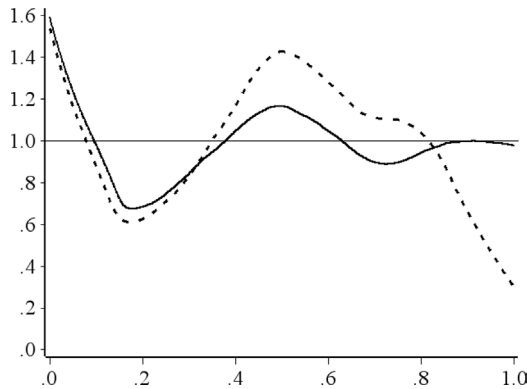


Figure 6. Classical kernel density estimation of the EL data transformed with estimated Champernowne ($c = 0$) distribution (dashed line) and modified Champernowne ($c > 0$) distribution (solid line).

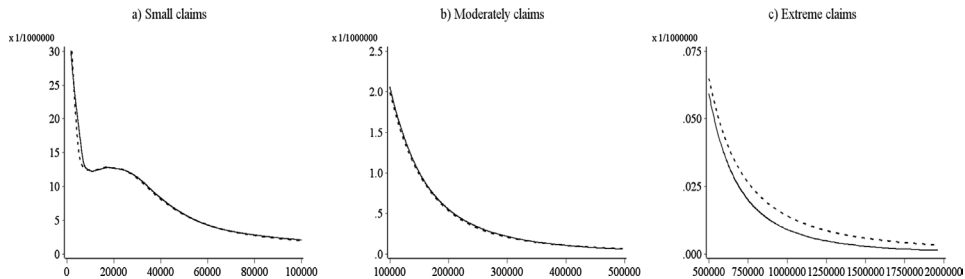


Figure 7. KMCE with $c = 0$ (dashed line) and KMCE with $c > 0$ (solid line) estimation of the EL data, separated into three disjoint intervals.

6. Conclusion

In this work, we have introduced an alternative method for estimating loss distributions. The method, which we have called a semiparametric transformation kernel density estimator, is based on a parametric estimator that is subsequently corrected with a non-parametric estimator. When we have a lot of information, the estimator is close to a non-parametric estimator, whereas it is close to a parametric estimator when we have little information.

The Champernowne distribution has an inflexible shape near 0, and we have generalized the distribution to the modified Champernowne distribution, which is heavy-tailed as well. We have used this modification for the transformation kernel density estimator.

The KMCE estimator turns out to perform very well compared with existing transformation kernel density estimators. The estimators were compared on simulated data. The KMCE estimator is the only estimator that performed well for all distributions. Therefore, the KMCE estimator is a basis for a unified approach that can be used for all kinds of data.

In insurance companies today, many analyses of loss distributions are based on parametric estimation. Our results show that our proposed method can overcome many disadvantages: in parametric estimation, the analyst must decide on a parametric model and a parameter estimation method. Insurance data sets are often large and the true distribution of real data rarely follows a simple known parametric distribution. We claim that there is no need to separate small and large claims. We believe that the use of our unified method results in an estimation of loss distributions that is very straightforward and can be useful in practice.

Acknowledgement

Helpful suggestions made by anonymous referees are acknowledged. The Spanish authors also thank the Ministerio de Educación y Ciencia and FEDER grant SEJ2004-05052.

References

- [1] Wand, M.P. and Jones, M.C., 1995, *Kernel Smoothing* (London: Chapman & Hall).
- [2] Silverman, B.W., 1986, *Density Estimation for Statistics and Data Analysis* (London: Chapman & Hall).
- [3] Wand, P., Marron, J.S. and Ruppert, D., 1991, Transformations in density estimation. *Journal of the American Statistical Association*, **86**, 343–361.
- [4] Bolancé, C., Guillén, M. and Nielsen, J.P., 2003, Kernel density estimation of actuarial loss functions. *Insurance: Mathematics and Economics*, **32**, 19–36.
- [5] Yang, L. and Marron, J.S., 1999, Iterated transformation-kernel density estimation. *Journal of the American Statistical Association*, **94**, 580–589.
- [6] Hjort, N.L. and Glad, I.K., 1995, Nonparametric density estimation with a parametric start. *The Annals of Statistics*, **23**, 882–904.
- [7] Jones, M., Linton, O. and Nielsen, J.P., 1995, A simple bias reduction method for density estimation. *Biometrika*, **82**, 327–338.
- [8] Clements, A.E., Hurn, A.S. and Lindsay, K.A., 2003, Möbius-like mappings and their use in kernel density estimation. *Journal of the American Statistical Association*, **98**, 993–1000.
- [9] Scaillet, O., 2004, Density estimation using inverse and reciprocal inverse Gaussian kernels. *Journal of Nonparametric Statistics*, **16**, 217–226.
- [10] Johnson, N.L., Kotz, S. and Balakrishnan, N., 1994, *Continuous Univariate Distributions*, Vol. 1 (2nd edn) (New York: John Wiley & Sons).
- [11] Champernowne, D.G., 1936, The Oxford Meeting, September 25–29, *Econometrica*, **5**, October 1937.
- [12] Champernowne, D.G., 1937, The theory of income distribution. *Econometrica*, **5**, 379–381.
- [13] Champernowne D.G., 1952, The graduation of income distributions. *Econometrica*, **20**, 591–615.
- [14] Coles, S., 2001, *An Introduction to Statistical Modelling of Extreme Values* (Berlin: Springer).
- [15] Lehmann, E.L., 1983, *Theory of Point Estimation* (New York: Wiley).
- [16] Tukey, J.W., 1960, A survey of sampling from contaminated distributions. In: I. Olkin (Ed.) *Contributions to Probability and Statistics* (Stanford, California: Stanford University Press).
- [17] Stigler, S.M., 1973, Simon Newcomb, Percy Daniell, and the history of robust estimation. *Journal of the American Statistical Association*, **68**, 872–879.
- [18] Newcomb, S., 1882, Discussion and results of observations on transits of Mercury from 1677 to 1881. *Astronomical Papers*, **1**, 363–487.
- [19] Newcomb, S., 1886, A generalized theory of the combination of observations so as to obtain the best results. *American Journal of Mathematics*, **8**, 343–366.
- [20] Hubert, P.J., 1981, *Robust Statistics* (New York: John Wiley & Sons).
- [21] Yang, L., 2000, Root-n convergent transformation-kernel density estimation. *Journal of Nonparametric Statistics*, **12**, 447–474.
- [22] Sheather, S.J. and Jones, M.C., 1991, A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Series B*, **53**, 683–690.