Available online at www.sciencedirect.com
SCIENCE DIRECT®

Computational Statistics & Data Analysis 50 (2006) 733 – 759

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Robust measures of tail weight

Guy Brys[a,*], Mia Hubert[b], Anja Struyf[c]

[a] *Faculty of Applied Economics, Universiteit Antwerpen (UA), Prinsstraat 13, B-2000 Antwerpen, Belgium*
[b] *Department of Mathematics, Katholieke Universiteit Leuven (KULeuven), W. de Croylaan 54, B-3001 Leuven, Belgium*
[c] *Department of Mathematics and Computer Science, Universiteit Antwerpen (UA), Middelheimlaan 1, B-2020 Antwerpen, Belgium*

Received 2 March 2004; received in revised form 27 September 2004; accepted 27 September 2004
Available online 19 October 2004

Abstract

The kurtosis coefficient is often regarded as a measure of the tail heaviness of a distribution relative to that of the normal distribution. However, it also measures the peakedness of a distribution, hence there is no agreement on what kurtosis really estimates. Another disadvantage of the kurtosis is that its interpretation and consequently its use is restricted to symmetric distributions. Moreover, the kurtosis coefficient is very sensitive to outliers in the data. To overcome these problems, several measures of left and right tail weight for univariate continuous distributions are proposed. They can be applied to symmetric as well as asymmetric distributions that do not need to have finite moments. Their interpretation is clear and they are robust against outlying values. The breakdown value and the influence functions of these measures and the resulting asymptotic variances are discussed and used to construct goodness-of-fit tests. Simulated as well as real data are employed for further comparison of the proposed measures.

Keywords: Tail weight; Kurtosis; Goodness-of-fit test; Robustness

* Corresponding author. Tel.: +3233224095; fax: +3233224817.
E-mail addresses: guy.brys@ua.ac.be (G. Brys), mia.hubert@wis.kuleuven.ac.be (M. Hubert), anja.struyf@ua.ac.be (A. Struyf).

0167-9473/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2004.09.012

1. Introduction

The classical kurtosis coefficient is introduced in many textbooks and is often regarded as a measure of the tail heaviness of a distribution relative to that of the normal distribution. For any distribution F with finite moments, it is defined as

$$\gamma_2(F) = \frac{\mu_4(F)}{\mu_2(F)^2}, \tag{1.1}$$

where μ_4 and μ_2 denote the fourth and the second central moment of F, i.e.

$$\mu_2 = E_F(X - E_F(X))^2 \quad \text{and} \quad \mu_4 = E_F(X - E_F(X))^4$$

with $X \sim F$. The finite-sample kurtosis will be denoted by b_2.

But the kurtosis coefficient has several drawbacks. As been stated in Bickel and Lehmann (1975), although its wide-spread use, there is no agreement on what kurtosis really measures. Many authors confirm, including Ruppert (1987), that kurtosis measures both peakedness and tail weight, because if one moves probability mass from the flanks to the center of a distribution, then to keep scale fixed one must also move mass from the flanks to the tail. Moreover, theoretical considerations of kurtosis often have been restricted to symmetric distributions, because of its intrinsic comparison with the symmetric normal distribution (e.g. Gleason, 1993). Finally, because the kurtosis is based on moments of the data, b_2 is very sensitive to outlying values. This is reflected in the unbounded influence function of γ_2 (Ruppert, 1987).

In this paper we introduce several measures of left and right tail weight for univariate continuous distributions. They can be applied to symmetric as well as asymmetric distributions. Their interpretation is clear and they are robust against outlying values. Note that several other authors have proposed robust measures of kurtosis, but they were only defined for symmetric distributions or merely measured peakedness instead of tail weight (Groeneveld and Meeden, 1984; Moors et al., 1996; Groeneveld, 1998; Schmid and Trede, 2003).

In Section 2 we define the new tail weight measures and derive some elementary properties. We will prove that they satisfy the anti-skewness ordering of MacGillivray and Balanda (1988). This is used to compare distributions according to the fatness or weakness of their tails (Oja, 1981). In Section 3 we compute the breakdown value, the influence function and the asymptotic variance of the proposed measures. Section 4 studies the finite-sample behavior of the estimators, both at contaminated and uncontaminated data sets. Some examples of real data sets are analyzed in Section 5, and Section 6 contains some conclusions. Some of the proofs of the theorems are collected in the appendix.

2. Measures of tail weight

We will define tail weight measures for continuous univariate distributions F. Their finite-sample versions for an i.i.d. sample $X_n = \{x_1, \ldots, x_n\}$ from F then follow in a straightforward way from the population definitions.

We define left and right tail measures as measures of skewness that are applied to the half of the probability mass lying to the left, respectively the right, side of the median of

F, denoted as $m_F = F^{-1}(0.5)$. In Brys et al. (2003) a comparison is made between several robust skewness measures. The three most interesting skewness measures (considering accuracy, robustness and computational complexity) were the octile skewness (OS), the quartile skewness (QS) and the medcouple (MC). The octile and the quartile skewness originate from Hinkley's (1975) class of skewness measures:

$$\gamma_1(F, p) = \frac{(Q(1-p) - Q(0.5)) - (Q(0.5) - Q(p))}{Q(1-p) - Q(p)}$$

for $0 < p < 1/2$ and $Q(p) = Q_F(p) = F^{-1}(p)$ the quantile function. The octile skewness takes $p = 1/8$ whereas the quartile skewness is defined as $\gamma_1(F, 1/4)$. When we apply $-\gamma_1(F, p)$ to the left half ($x < m_F$), and $\gamma_1(F, 1-p) = \gamma_1(F, q)$ to the right half ($x > m_F$) of F, we obtain the Left Quantile Weight (LQW) and the Right Quantile Weight (RQW):

$$\mathrm{LQW}_F(p) = -\frac{Q((1-p)/2) + Q(p/2) - 2Q(0.25)}{Q((1-p)/2) - Q(p/2)}$$

and

$$\mathrm{RQW}_F(q) = \frac{Q((1+q)/2) + Q(1-q/2) - 2Q(0.75)}{Q((1+q)/2) - Q(1-q/2)}$$

in which $0 < p < \frac{1}{2}$ and $\frac{1}{2} < q < 1$. To retain a reasonable amount of robustness we will study $\mathrm{LQW}_F(0.125)$, $\mathrm{LQW}_F(0.25)$, $\mathrm{RQW}_F(0.875)$ and $\mathrm{RQW}_F(0.75)$. Note that the sample versions $\mathrm{LQW}_n(0.125)$, etc. are easily found by using the quantiles of F_n, the empirical distribution function of X_n.

The medcouple is extensively studied in Brys et al. (2004). It is defined as

$$\mathrm{MC}(F) = \operatorname*{med}_{x_1 \leqslant m_F \leqslant x_2} h(x_1, x_2)$$

with x_1 and x_2 sampled from F and the kernel function h given by

$$h(x_i, x_j) = \frac{(x_j - m_F) - (m_F - x_i)}{x_j - x_i}.$$

As with the QW alternatives, we can easily apply the MC to one side of the distribution, leading to the Left Medcouple (LMC) and to the Right Medcouple (RMC), defined as

$$\mathrm{LMC}_F = -\mathrm{MC}(x < m_F) \quad \text{and} \quad \mathrm{RMC}_F = \mathrm{MC}(x > m_F).$$

Let $q_{F,1} = q_1 = Q(0.25)$ be the first quartile of F, denote $q_{F,3} = q_3 = Q(0.75)$ the third quartile of F and let I be the indicator function, then with

$$H_{F,l}(u) = 16 \int_{q_1}^{m_F} \int_{-\infty}^{q_1} I\left(-\frac{x_2 + x_1 - 2q_1}{x_2 - x_1} \leqslant u\right) \mathrm{d}F(x_1)\,\mathrm{d}F(x_2) \tag{2.1}$$

we obtain the shorter formulation

$$\mathrm{LMC}_F = H_{F,l}^{-1}(0.5).$$

Analogously, we have with

$$H_{F,r}(u) = 16 \int_{q_3}^{+\infty} \int_{m_F}^{q_3} I\left(\frac{x_2 + x_1 - 2q_3}{x_2 - x_1} \leqslant u\right) \mathrm{d}F(x_1)\,\mathrm{d}F(x_2) \tag{2.2}$$

the expression

$$\mathrm{RMC}_F = H_{F,r}^{-1}(0.5).$$

Similar as in Brys et al. (2004), we can simplify (2.1) to

$$H_{F,l}(u) = 16 \int_{q_1}^{m_F} F\left(\frac{x_2(-u-1) + 2q_1}{-u+1}\right) \mathrm{d}F(x_2)$$

and (2.2) to

$$H_{F,r}(u) = 16 \int_{m_F}^{q_3} F\left(\frac{x_2(u+1) - 2q_3}{u-1}\right) \mathrm{d}F(x_2).$$

By using MC_n, the finite-sample version of MC, we obtain the finite-sample versions LMC_n and RMC_n. Note that they can be seen as estimators of LMC_F and RMC_F, respectively. Their computation can be performed in $\mathrm{O}(n \log n)$ time due to the fast algorithm described in Brys et al. (2004).

Remark that both the quantile and the medcouple tail weight measures only depend on quantiles and consequently can be computed at any distribution, even without finite moments. Moreover, the LMC and RMC measure do not require the choice of any additional parameter, whereas for the LQW (resp. RQW) measures only p (resp. q) has to be fixed in advance, depending on the degree of robustness one is willing to attain. A different approach to estimate the tail behavior of a distribution is studied in the field of extreme value analysis. In this field, the key quantity is the extreme-value index γ, which allows to classify distributions into a Fréchet–Pareto type ($\gamma > 0$), Gumbel type ($\gamma = 0$) or Extremal Weibull type ($\gamma < 0$). Many estimators have been proposed to estimate γ, see, e.g. chapter 5 in Beirlant et al. (2004) for a recent overview. These estimators can yield very accurate results, but as a disadvantage they always require the selection of the tail fraction on which the estimates are based. This is often done by minimizing the asymptotic mean-squared error, which itself can be hard to estimate. The computational complexity of these techniques is thus much harder than the tail weight measures proposed in this paper. Moreover the tail index is not able to distinguish Gumbel-type distributions such as the log-normal, the logistic, the exponential or the Weibull distributions. Note that robust estimators for the tail index in the Pareto case ($\gamma > 0$) have been recently proposed in Vandewalle et al. (2004a, 2004b), whereas Hsieh (1999) addresses the problem of robustly selecting the number of extreme order statistics for the Hill estimator.

Before discussing the robustness properties of our proposed measures, we look at some general tail weight properties. Let the random variable X have a continuous distribution F_X and let W be any of the defined tail weight measures, let LW be a left tail weight measure and RW be a right tail weight measure.

Proposition 1. *W is location and scale invariant, i.e.*

$$W(F_{aX+b}) = W(F_X)$$

for any $a > 0$ *and* $b \in \mathbb{R}$.

Proposition 2. *If we invert a distribution, we have*

$$\mathrm{LW}(F_{-X}) = \mathrm{RW}(F_X).$$

Proposition 3. *If F is symmetric, then* $\mathrm{LW}(F) = \mathrm{RW}(F)$.

Proposition 4. $W \in [-1, 1]$.

Properties 1 and 2 follow immediately from the definitions, and imply Property 3. Property 4 is not really required for a tail weight measure, but at least it gives a lower and an upper bound for any of the tail weights under consideration. A distribution will have a positive (resp. negative) RW measure if its upper half (half the probability mass larger than the median) is skewed to the right (resp. the left). Right skewness occurs more frequently than left skewness, so our RW measures are typically positive. Similarly we obtain a positive (resp. negative) LW measure if the lower half is skewed to the left (resp. the right). The measures become zero when the lower half or the upper half is symmetric, as for example at the uniform distribution. The extreme situation $W = -1$ or 1 will only occur at degenerate situations. Consider for example a distribution that has half of its mass at zero, 25% of its mass to the left and 25% to the right of zero, which is then also the median of the distribution. Then both LW and RW will equal one. Analogously we can construct a distribution with $\mathrm{LW} = \mathrm{RW} = -1$ by setting 25% probability mass at the left and at the right endpoints of the distribution.

The next property tells us whether *W* respects the anti-skewness ordering of distributions as defined by MacGillivray and Balanda (1988), which is inspired on the kurtosis ordering of van Zwet (1964). Let *F* and *G* be continuous distributions with interval support. Because of the location invariancy, we can assume that the medians of both distributions collapse, or $m_F = m_G$. Then it is said that *G* is at least as fat tailed to the left as *F* if and only if

$$F \leqslant_a^l G \quad \Leftrightarrow \quad G^{-1}(F(x)) \text{ is concave for } x < m_F = m_G \tag{2.3}$$

and *G* is at least as fat tailed to the right as *F* if and only if

$$F \leqslant_a^r G \quad \Leftrightarrow \quad G^{-1}(F(x)) \text{ is convex for } x > m_F = m_G$$

both on the support of *F*.

Proposition 5. *If* $F \leqslant_a^l G$, *then* $\mathrm{LW}(F) \leqslant \mathrm{LW}(G)$, *and if* $F \leqslant_a^r G$, *then* $\mathrm{RW}(F) \leqslant \mathrm{RW}(G)$.

The proof of Property 5 is similar to the proof of Property 4 in Brys et al. (2004). For the left tail weight measures, it should be noted that because of the scale invariancy we can also transform *F* and *G* such that $Q_F(0.25) = Q_G(0.25)$. Analogously for the right tail weight measures we rescale them such that $Q_F(0.75) = Q_G(0.75)$.

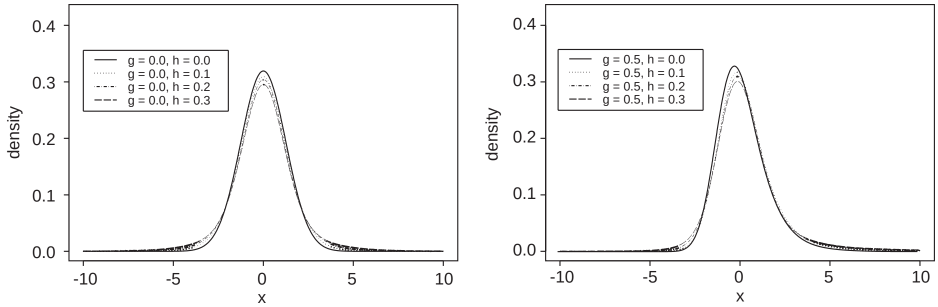

Fig. 1. Nonparametric density estimates of the $G_{0,h}$ (left panel) and $G_{0.5,h}$ (right panel) distributions for several values of h between 0 and 0.3.

More specifically, we can consider Tukey's class of gh-distributions (Hoaglin et al., 1985). When a random variable Z is standard gaussian distributed, then

$$Y_{g,h} = \begin{cases} \dfrac{(\mathrm{e}^{gZ} - 1)}{g} \mathrm{e}^{hZ^2/2}, & g \neq 0, \\ Z\mathrm{e}^{hZ^2/2}, & g = 0 \end{cases}$$

is said to follow a gh-distribution $G_{g,h}$ with parameters $g \in \mathbb{R}$ and $h \geqslant 0$. The parameter g controls the skewness of the distribution, whereas h effects the tail weight. For $g = 0$, the variable $Y_{0,h}$ defines a symmetric distribution with zero skewness, but with increasing tails as h increases. It is also clear that $G_{-g,h}(x) = 1 - G_{g,h}(-x)$, hence we will only consider the symmetric and the right-skewed distributions for which $g \geqslant 0$. The densities of the gh-distributions can only be computed numerically (Rayner and MacGillivray, 2002). In Fig. 1 we have drawn a nonparametric density estimate of $G_{0,h}$ and $G_{0.5,h}$ distributions with h varying between 0 and 0.3.

It is easy to show that these gh-distributions follow the anti-skewness ordering, as $G_{g,h_1} \leqslant_a^l G_{g,h_2}$ and $G_{g,h_1} \leqslant_a^r G_{g,h_2}$ for any $h_1 \leqslant h_2$. To illustrate this, we have drawn in Fig. 2 the left and right tail weight measures for $G_{0,h}$ and $G_{0.5,h}$ with h ranging from 0 to 0.3. We see that all the curves are monotone increasing, which is an obvious consequence of the validity of Property 5 for the studied measures.

Also in Gleason (1993) these gh-distributions with $g = 0$ are explored to study tail weight. In that paper, a graphical tool (the d_F plot) is introduced in order to study elongation (Hoaglin et al., 1985). A distribution is elongated, in some region, if its quantiles change more rapidly there than do the Gaussian quantiles. The d_F-curve allows to compare graphically symmetric distributions. As for the $G_{0,h}$ distributions it holds that $d_F(z) = hz^2$, the curve for any G_{0,h_1} lies completely below the curve for G_{0,h_2} if $h_1 \leqslant h_2$ (see also Fig. 1 in Gleason, 1993). This implies that G_{0,h_2} is more elongated than G_{0,h_1}, which is in line with the anti-skewness ordering based on our measures of tail weight. A drawback of this plot is that it is only defined for continuous, unimodal distributions which are symmetric (about zero). Moreover it does not yield a finite-sample measure of tail weight.

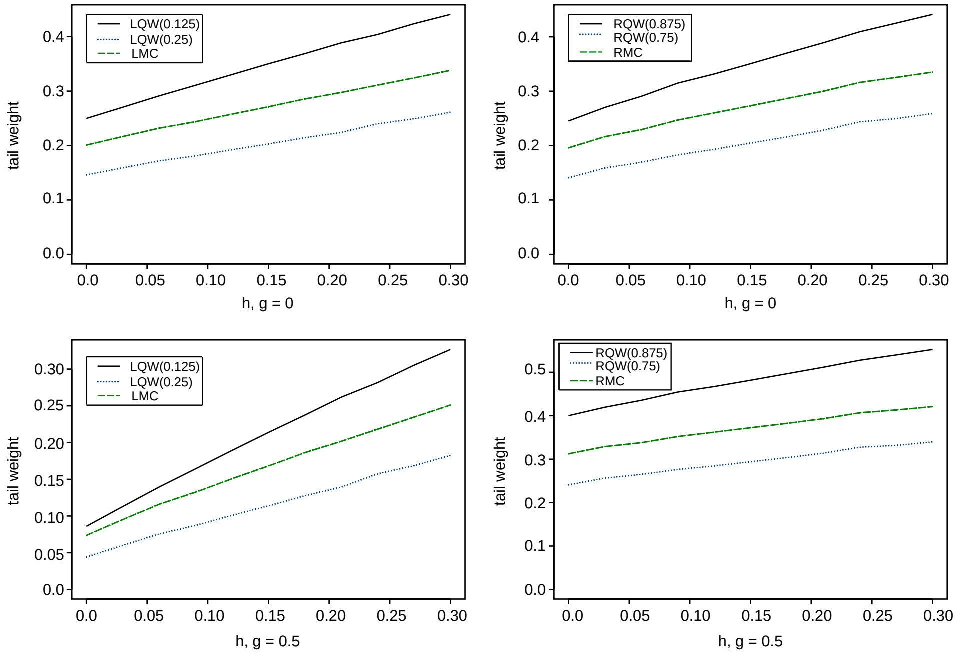

Fig. 2. Monotone behavior of the left and right tail weight measures for $G_{0,h}$ and $G_{0.5,h}$ with h ranging from 0 to 0.3.

3. Robustness properties

In this section we study the robustness properties of the proposed measures of tail weight. In particular, we derive their breakdown value, and their influence function from which the asymptotic variance follows.

3.1. Breakdown value

The breakdown value of an estimator T_n at a sample X_n measures how many observations of X_n need to be replaced to make the estimate worthless (Rousseeuw and Leroy, 1987). For a univariate location estimator, e.g. this means that the absolute value of the estimate becomes arbitrarily large, whereas we say that a scale estimator breaks down if the estimate becomes arbitrarily large or close to zero. Since our tail weight measures are bounded by $[-1, 1]$, we define their finite-sample breakdown value as

$$\varepsilon_n^*(W_n; X_n) = \min\left\{\frac{m}{n}; \sup_{X_n'} |W_n(X_n')| = 1\right\},$$

where the data set X_n' is obtained by replacing m observations from X_n by arbitrary values. The asymptotic breakdown value $\varepsilon^*(W; F)$ is then defined as $\lim_{n\to\infty} \varepsilon_n^*(W_n; X_n)$ with the x_i sampled from F.

Theorem 1. *If the data set X_n is in general position, i.e. no two data points coincide, then*

$$\frac{1}{n}\left(\left\lceil\frac{n}{8}\right\rceil - 1\right) \leqslant \varepsilon_n^*(\mathrm{RMC}_n; X_n) \leqslant \frac{1}{n}\left(\left\lceil\frac{n}{8}\right\rceil + 1\right).$$

Here, $\lceil x \rceil$ denotes the smallest integer larger than or equal to x. The same result holds for LMC_n. The medcouple tail weight measures can thus resist up to 12.5% outliers in the data. In a similar way, it can be shown that the left and right quantile tail weight measures $\mathrm{LQW}_F(0.25)$ and $\mathrm{RQW}_F(0.75)$ have an asymptotic breakdown value of 12.5%, while the $\mathrm{LQW}_F(0.125)$ and $\mathrm{RQW}_F(0.875)$ can withstand at most 6.25% of outliers.

3.2. Influence function

The influence function of an estimator T at some distribution F measures the effect on T when adding a small probability mass at the point x (Hampel et al., 1986). If Δ_x is the point mass in x, then the influence function is defined as

$$\mathrm{IF}(x, T, F) = \lim_{\varepsilon \downarrow 0} \frac{T((1-\varepsilon)F + \varepsilon\Delta_x) - T(F)}{\varepsilon}.$$

The following theorems give the influence functions of the tail weight measures under study for any continuous distribution F with density f. To simplify the conditions, we assumed that $f(x) > 0$ for all x. Note that Huber (1981) showed that the IF of $Q_F(p) = F^{-1}(p)$ is given by

$$\mathrm{IF}(x, Q_F(p), F) = \frac{p - I(x < Q_F(p))}{f(Q_F(p))}.$$

Theorem 2.

$$\begin{aligned}\mathrm{IF}(x, \mathrm{LQW}(p), F) = 2[&(\mathrm{IF}(x, q_1, F) - \mathrm{IF}(x, Q((1-p)/2), F))(q_1 - Q(p/2))\\ &- (\mathrm{IF}(x, Q(p/2), F) - \mathrm{IF}(x, q_1, F))\\ &\times (Q((1-p)/2) - q_1)]/(Q((1-p)/2) - Q(p/2))^2.\end{aligned}$$

Theorem 3.

$$\begin{aligned}\mathrm{IF}(x, \mathrm{RQW}(q), F) = 2[&(\mathrm{IF}(x, Q(1-q/2), F) - \mathrm{IF}(x, q_3, F))(Q((1+q)/2)\\ &- q_3) - (\mathrm{IF}(x, q_3, F) - \mathrm{IF}(x, Q((1+q)/2), F))\\ &\times (q_3 - Q(1-q/2))]/(Q((1+q)/2) - Q(1-q/2))^2.\end{aligned}$$

To derive the influence functions of LMC and RMC, we denote

$$g_{1,l}(x) = \frac{x(-\mathrm{LMC} - 1) + 2q_1}{-\mathrm{LMC} + 1}, \tag{3.1}$$

$$g_{2,l}(x) = \frac{x(-\mathrm{LMC} + 1) - 2q_1}{-\mathrm{LMC} - 1} \tag{3.2}$$

and analogously $g_{1,r}(x) = (x(\mathrm{RMC}-1)+2q_3)/(\mathrm{RMC}+1)$ and $g_{2,r}(x) = (x(\mathrm{RMC}+1)-2q_3)/(\mathrm{RMC}-1)$. Then it holds that

$$H'_{F,l}(\mathrm{LMC}) = \int_{q_1}^{m_F} \frac{32(q_1-x_2)}{(-\mathrm{LMC}+1)^2} f(g_{1,l}(x_2))\,\mathrm{d}F(x_2)$$

and

$$H'_{F,r}(\mathrm{RMC}) = \int_{m_F}^{q_3} \frac{32(q_3-x_2)}{(\mathrm{RMC}-1)^2} f(g_{2,r}(x_2))\,\mathrm{d}F(x_2).$$

Theorem 4. *Assume that* $H'_{F,l}(\mathrm{LMC}) \neq 0$. *Then*

$$\begin{aligned}
&\mathrm{IF}(x, \mathrm{LMC}, F)\\
&\quad= \frac{1}{H'_{F,l}(\mathrm{LMC})}\Bigg[1 - 16F(g_{1,l}(x))(I(x>q_1) - I(x>m_F))\\
&\qquad - 16\left(F(g_{2,l}(x)) - \frac{1}{4}\right) I(x > g_{1,l}(m_F))I(x<q_1) - 4I(x<g_{1,l}(m_F))\\
&\qquad - 8\mathrm{sgn}(x-m_F)F(g_{1,l}(m_F)) + \left(\frac{1}{4} - I(x<q_1)\right)\Bigg(4 - \frac{32}{f(q_1)(-\mathrm{LMC}+1)}\\
&\qquad \times \int_{q_1}^{m_F} f(g_{1,l}(x_2))\,\mathrm{d}F(x_2)\Bigg)\Bigg].
\end{aligned}$$

Theorem 5. *Assume that* $H'_{F,r}(\mathrm{RMC}) \neq 0$. *Then*

$$\begin{aligned}
&\mathrm{IF}(x, \mathrm{RMC}, F)\\
&\quad= \frac{1}{H'_{F,r}(\mathrm{RMC})}\Bigg[1 - 16F(g_{2,r}(x))(I(x>m_F) - I(x>q_3))\\
&\qquad - 16\left(F(g_{1,r}(x)) - \frac{1}{2}\right) I(x < g_{2,r}(m_F))I(x>q_3) - 4I(x<q_3)\\
&\qquad + 8\mathrm{sgn}(x-m_F)F(g_{2,r}(m_F)) - \left(\frac{3}{4} - I(x<q_3)\right)\Bigg(4 - \frac{32}{f(q_3)(\mathrm{RMC}-1)}\\
&\qquad \times \int_{m_F}^{q_3} f(g_{2,r}(x_2))\,\mathrm{d}F(x_2)\Bigg)\Bigg].
\end{aligned}$$

All these influence functions are bounded. Fig. 3 shows the influence functions of the left and right tail weight measures for the standard gaussian distribution $G_{0,0}$, for the right skewed $G_{0.5,0}$ and for the logistic distribution with density function:

$$f(x) = \frac{\mathrm{e}^{-x}}{(1+\mathrm{e}^{-x})^2}.$$

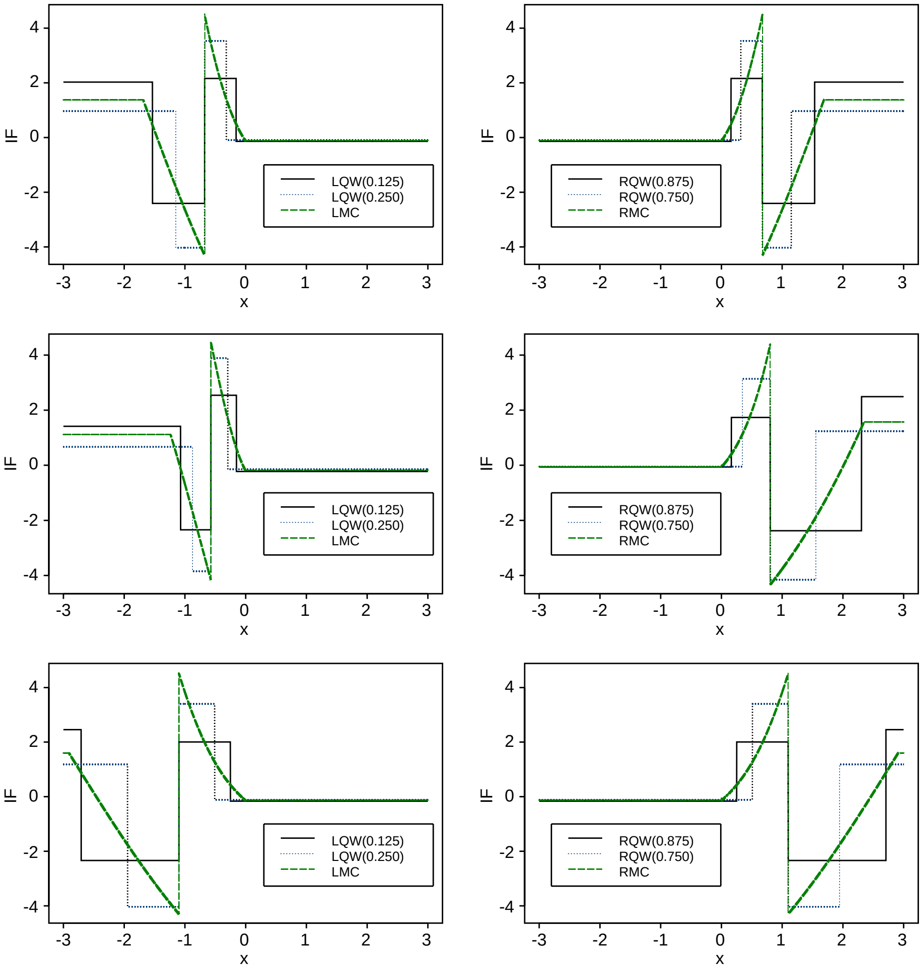

Fig. 3. Influence functions of the left and right tail weight measures for the standard gaussian distribution (upper), for $G_{0.5,0}$ (middle) and for the logistic distribution (lower).

We use the logistic distribution here instead of any $G_{0,h}$ distribution because its density function can be explicitly formulated. The logistic distribution is symmetric around zero and has fatter tails than the standard gaussian distribution. It will also be considered in Section 4 where its density is plotted in Fig. 4. We see that the influence functions of the QW alternatives are step functions, whereas the influence functions of the MC alternatives are continuous, except in q_1 for LMC and q_3 for RMC. They can be seen as smoothed versions of the QW alternatives.

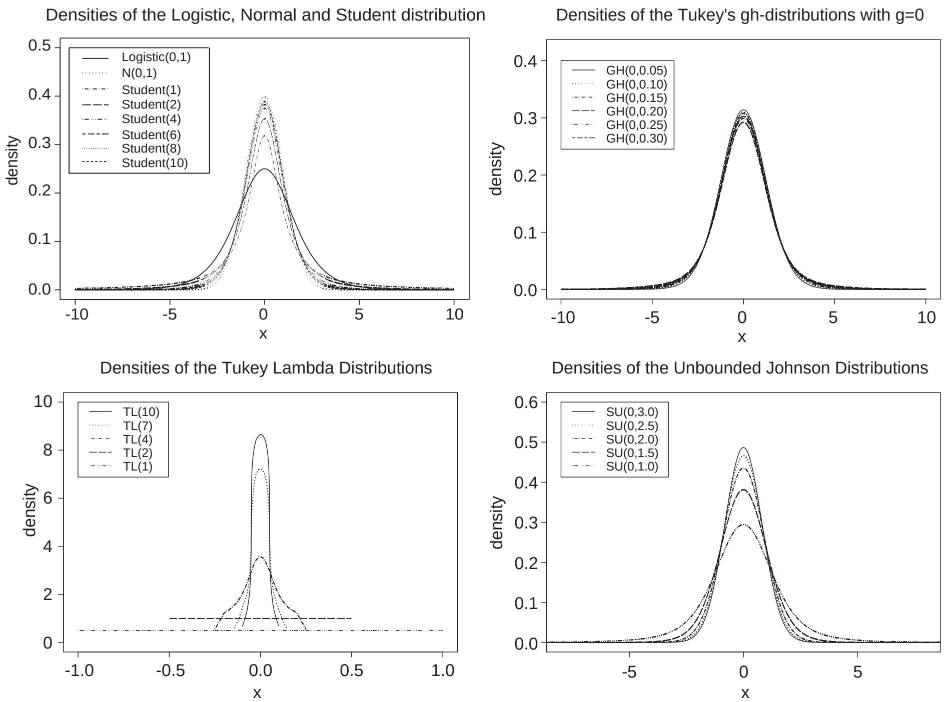

Fig. 4. Nonparametric density estimates of 24 symmetric distributions.

3.3. Asymptotic variance

If an estimator T is asymptotically normal at a distribution F, its asymptotic variance $V(T, F)$ (Hampel et al., 1986) is given by

$$V(T, F) = \int \mathrm{IF}(x, T, F)^2 \, \mathrm{d}F(x). \tag{3.3}$$

For the QW tail weight measures we have the following result:

Theorem 6. *Let $F = \Phi$ (the standard normal distribution), then the QW alternatives are asymptotically normal, i.e.*

$$\sqrt{n}(\mathrm{QW}_n - \mathrm{QW}) \rightarrow_{\mathscr{D}} \mathrm{N}(0, \sigma^2_{\mathrm{QW}})$$

with asymptotic variances $\sigma^2_{\mathrm{LQW}(0.125)} = \sigma^2_{\mathrm{RQW}(0.875)} = 2.23$ *and* $\sigma^2_{\mathrm{LQW}(0.25)} = \sigma^2_{\mathrm{RQW}(0.75)} = 3.71$.

The asymptotic normality of the MC-based measures has not been proven yet, but we constructed QQ-plots based on 1000 samples of size 1000 which suggest their asymptotic normal behavior. Further on we assume for all measures the validity of the asymptotic

Table 1
Asymptotic variances at the standard normal $G_{0,0}$ distribution, at the $G_{0.5,0}$ distribution and at the logistic distribution

	$G_{0,0}$	$G_{0.5,0}$	Logistic
LQW(0.125)	2.23	2.39	2.17
RQW(0.875)	2.23	2.02	2.17
LQW(0.25)	3.71	3.81	3.68
RQW(0.75)	3.71	3.58	3.68
LMC	2.62	2.63	2.60
RMC	2.62	2.55	2.60

normality, such that we may calculate $V(T, F)$ from (3.3). Using numerical integration we obtain the values listed in Table 1 for the $G_{0,0}$, $G_{0.5,0}$ and the logistic distribution.

4. Finite-sample behavior

4.1. Performance at uncontaminated distributions

As in the sequel we will mainly consider the finite-sample versions of the proposed measures, we will from now on omit the subscript n in their abbreviation (e.g. LMC_n is written as LMC).

First we conducted a similar study as in Groeneveld (1998). We considered 24 symmetric distributions: the standard normal distribution N(0,1) ($=G_{0,0}$), the standard logistic distribution, the Student t-distribution with $n = 1, 2, 4, 6, 8, 10$ degrees of freedom; the Tukey gh-distribution with $g = 0$, $h = 0.05, 0.10, 0.15, 0.20, 0.25, 0.30$; the Tukey Lambda distribution with $\lambda = 10, 7, 4, 2, 1$ and the Unbounded Johnson distribution with parameters $\gamma = 0$ and $\delta = 3, 2.5, 2, 1.5, 1$. Note that the Tukey Lambda distribution (Hastings et al., 1947) with parameter λ (say TL(λ)) is defined as

$$
\begin{aligned}
Y &= \frac{X^{\lambda} - (1 - X)^{\lambda}}{\lambda}, \quad \lambda \neq 0,\\
Y &= \text{Logistic}, \quad \lambda = 0
\end{aligned}
$$

with $X \sim \text{Uniform}(0, 1)$. TL(1) corresponds with the uniform distribution on $[-1, 1]$, whereas TL(2) is uniformly distributed on $[-0.5, 0.5]$. The Unbounded Johnson distribution (Johnson, 1949) with parameters γ and δ (say SU(γ, δ)) is given by

$$
Y = \frac{\mathrm{e}^{X} - \mathrm{e}^{-X}}{2}
$$

with $X \sim \mathrm{N}(-\gamma, 1/\delta^2)$. Nonparametric density estimates of all these 24 distributions are drawn in Fig. 4. In each box we consider a group of distributions with increasing tail weight.

We considered the classical kurtosis b_2, the right tail weight measures RQW(0.875), RQW(0.75) and RMC and the measure of peakedness P which was recently proposed in Schmid and Trede (2003). It is defined as

$$P = \frac{Q(0.875) - Q(0.125)}{Q(0.75) - Q(0.25)}$$

hence it is like the QW alternatives only defined on certain quantiles of the data. Its breakdown value is 12.5% just as for LMC, RMC, LQW(0.25) and RQW(0.75). Although P can be defined on any distribution, it specifically measures peakedness for symmetric distributions. Note that Schmid and Trede (2003) also introduced

$$T = \frac{Q(0.975) - Q(0.025)}{Q(0.875) - Q(0.125)}$$

as a measure of fat tails. However, due to its low breakdown value of 2.5% we did not include it in our comparison. We also did not have to consider our left tail weight measures as we now only focus on symmetric distributions. Next, we computed the empirical power of the Shapiro–Wilk test at each of the 24 symmetric distributions based on 1000 samples of size 20. The Shapiro–Wilk test (Shapiro and Wilk, 1965) is a well-known test for normality which has a high power against long- or short-tailed distributions. We thus expect that tail weight measures will adequately detect nonnormality if they differ a lot (in absolute value) from their value at the normal distribution at distributions where Shapiro–Wilk attains high power. Table 2 lists for the 24 distributions the resulting empirical power values of the Shapiro–Wilk test together with the average absolute deviation of the measure at the given distribution compared to its value at the normal distribution. Fig. 5 shows the scatter plots of the values found in Table 2 for b_2 and RMC. Figures of P, RQW(0.875) and RQW(0.75) are comparable to that of RMC. It can be seen that at the Tukey Lambda Distribution with $\lambda = 10$ the power of the Shapiro–Wilk test is very high but the kurtosis is hardly different from 3. This problem does not occur with any of the robust measures. On the contrary they clearly detect the nonnormality of the TL-distributions.

Note that also the test proposed in Bonett and Seier (2002) is very powerful to detect normality but it is also not robust as it is based on moments of the data. Hence, we did not include this test (among many others) in our comparison.

Another way of testing the performance of the proposed measures at uncontaminated distributions is by using goodness-of-fit tests. Assume that we want to test whether the data are sampled from a certain distribution G or whether they come from a distribution with longer tails. We could then formulate the null hypothesis as

$$\begin{aligned} &\mathrm{H}_0 : \omega(F) = \omega(G), \\ &\mathrm{H}_1 : \omega(F) > \omega(G) \end{aligned} \tag{4.1}$$

with ω any tail weight measure. Under H_0 it holds that

$$z_n = \sqrt{n}\,\frac{\omega_n - \omega(G)}{\sqrt{V(\omega, G)}} \approx_{\mathrm{H}_0} \mathrm{N}(0, 1). \tag{4.2}$$

The p-value (significance) of this test then equals $p = P(Z > z) = 1 - \Phi(z)$. We considered the null hypothesis (4.1) for the normal distribution $G = G_{0,0}$ and a skewed distribution $G = G_{0.5,0}$. As alternative distributions, we used the $G_{0,h}$ and $G_{0.5,h}$ distributions with

Table 2
Empirical power of the Shapiro–Wilk test at 24 symmetric distributions together with the average absolute deviation of the different tail weight measures compared to their values at the normal distribution

	Power	b_2	P	RQW(0.875)	RQW(0.75)	RMC
N(0,1)	0.051	0	0	0	0	0
TL(4)	0.029	0.553	0.170	0.048	0.062	0.052
SU(0,3.0)	0.073	0.529	0.027	0.027	0.016	0.019
SU(0,2.5)	0.089	0.821	0.038	0.039	0.022	0.027
GH(0,0.05)	0.094	0.820	0.038	0.037	0.021	0.026
Student(10)	0.095	1.002	0.040	0.039	0.022	0.028
Logistic(0,1)	0.109	1.197	0.065	0.061	0.035	0.044
Student(8)	0.117	1.472	0.049	0.049	0.027	0.034
SU(0,2.0)	0.127	1.492	0.061	0.058	0.034	0.041
Student(6)	0.148	2.843	0.068	0.067	0.037	0.047
GH(0,0.10)	0.156	2.506	0.076	0.072	0.041	0.051
TL(1)	0.192	1.200	1.206	0.249	0.145	0.200
TL(2)	0.195	1.200	0.206	0.249	0.145	0.199
SU(0,1.5)	0.217	4.161	0.110	0.099	0.058	0.071
GH(0,0.15)	0.245	6.742	0.114	0.105	0.061	0.074
Student(4)	0.251	∞	0.109	0.103	0.591	0.073
GH(0,0.20)	0.319	19.39	0.154	0.136	0.080	0.096
TL(7)	0.338	0.879	1.236	0.369	0.310	0.321
GH(0,0.25)	0.375	60.09	0.195	0.165	0.098	0.117
SU(0,1.0)	0.407	30.36	0.251	0.197	0.121	0.142
GH(0,0.30)	0.452	144.2	0.236	0.193	0.115	0.137
Student(2)	0.553	∞	0.258	0.214	0.125	0.149
TL(10)	0.806	2.383	2.957	0.545	0.481	0.485
Student(1)	0.872	∞	0.709	0.418	0.269	0.300

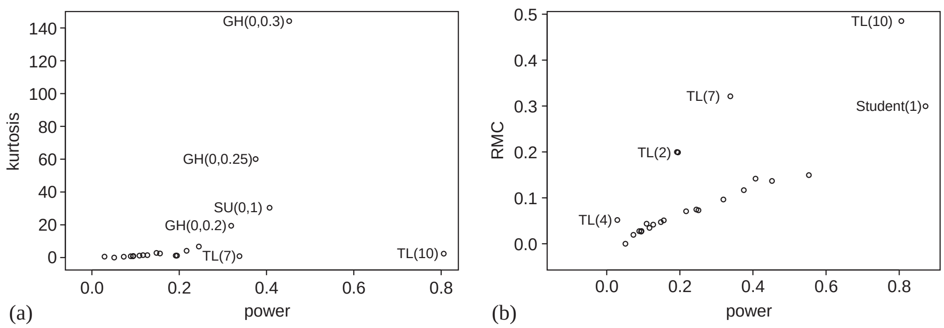

Fig. 5. Scatter plots of the average absolute deviation of (a) the kurtosis b_2 and (b) the RMC from their value at the normal distribution measured at 24 symmetric distributions, against the power of the Shapiro–Wilk test.

h ranging from 0 to 0.3. For the test of normality we also included the Tukey Lambda distribution with $\lambda = 5.2$ because it has a kurtosis $\gamma_2 = 3$. From these distributions, 10,000 samples of size $n = 100$ and 1000 were drawn. The results are summarized in Tables 3 and

Table 3
Fraction of 10,000 samples of different data sizes n from several distributions $G_{g,h}$ on which the null hypothesis H_0: $\omega(F) = \omega(G_{0,0})$ was rejected at the 5% significance level

	$n = 100$					$n = 1000$				
	$G_{0,0.0}$	$G_{0,0.1}$	$G_{0,0.2}$	$G_{0,0.3}$	TL(5.2)	$G_{0,0.0}$	$G_{0,0.1}$	$G_{0,0.2}$	$G_{0,0.3}$	TL(5.2)
SW	0.052	0.456	0.828	0.962	0.469	0.052	0.999	1.000	1.000	1.000
b_2	0.059	0.619	0.910	0.981	0.023	0.058	1.000	1.000	1.000	0.009
P	0.035	0.083	0.173	0.298	0.673	0.040	0.370	0.825	0.980	1.000
LQW(0.125)	0.029	0.086	0.183	0.308	0.341	0.042	0.448	0.890	0.992	0.995
RQW(0.875)	0.028	0.083	0.182	0.311	0.343	0.042	0.445	0.891	0.992	0.992
LQW(0.25)	0.034	0.054	0.080	0.114	0.210	0.047	0.166	0.379	0.612	0.885
RQW(0.75)	0.034	0.055	0.081	0.115	0.199	0.047	0.162	0.374	0.599	0.880
LMC	0.033	0.069	0.118	0.182	0.266	0.046	0.264	0.607	0.855	0.961
RMC	0.029	0.065	0.120	0.182	0.259	0.048	0.261	0.608	0.848	0.959

Table 4
Fraction of 10,000 samples of different data sizes n from several distributions $G_{g,h}$ on which the null hypothesis $H_0 : \omega(F) = \omega(G_{0.5,0})$ was rejected at the 5% significance level

	$n = 100$				$n = 1000$			
	$G_{0.5,0.0}$	$G_{0.5,0.1}$	$G_{0.5,0.2}$	$G_{0.5,0.3}$	$G_{0.5,0.0}$	$G_{0.5,0.1}$	$G_{0.5,0.2}$	$G_{0.5,0.3}$
SW	—	—	—	—	—	—	—	—
b_2	0.001	0.020	0.075	0.154	0.020	0.388	0.814	0.962
P	0.036	0.091	0.159	0.259	0.043	0.296	0.723	0.946
LQW(0.125)	0.041	0.122	0.267	0.430	0.050	0.588	0.967	0.999
RQW(0.875)	0.015	0.039	0.084	0.161	0.029	0.271	0.716	0.936
LQW(0.25)	0.044	0.066	0.112	0.147	0.051	0.204	0.470	0.740
RQW(0.75)	0.032	0.047	0.059	0.094	0.048	0.125	0.290	0.487
LMC	0.042	0.095	0.182	0.267	0.050	0.387	0.790	0.963
RMC	0.026	0.051	0.078	0.135	0.046	0.184	0.442	0.689

4 by computing the fraction of the 10,000 samples on which the null hypothesis (4.1) was rejected in favor of the alternative at the 5% significance level. This is an approximation of the power of the goodness-of-fit test against several alternative hypotheses. We also added the Shapiro–Wilk test (SW) which is a two-sided test of normality, together with the test (4.2) based on $\omega = b_2$ and on $\omega = P$. The asymptotic variances of b_2 and P are respectively 24 and 2.80 at $G_{0,0}$ and 3.6×10^4 and 3.98 at $G_{0.5,0}$.

In the columns $G_{0,0}$ and $G_{0.5,0}$ we expect to find the nominal level $\alpha = 5\%$. As the other columns satisfy the alternative hypothesis, the reported values should be as close to 1 as possible. We clearly observe an increasing trend as h increases. The SW test adequately detects deviations from normality, but it cannot be applied to the skewed distributions $G_{0.5,h}$. The test based on the kurtosis b_2 is powerful to test normality, but it fails at the TL(5.2) distribution. This is due to the fact that $\gamma_2 = 3$ at TL(5.2), hence it cannot be distinguished from the normal distribution. To test for the skewed $G_{0.5,0}$ against other skewed distributions, we see that b_2 is very conservative at $n = 100$. Of our six proposed measures, LQW(0.125) and RQW(0.875) are superior in almost all situations, followed by LMC and RMC. The measures LQW(0.25) and RQW(0.75) are clearly the most conservatives. The power of the P measure at symmetric distributions is comparable with that of LQW(0.125) and RQW(0.875), whereas at the asymmetric $G_{0.5,0}$, P behaves similar to LMC. The results being much better for $n = 1000$ than for $n = 100$, we see that the power of these robust tests is rather low. This could be improved by constructing a test which is not solely based on a tail weight measure, but also on a robust measure of skewness. This is the idea behind the Jarque–Bera statistic (Bera and Jarque, 1981) based on the classical skewness and kurtosis and the test statistics developed in Moors et al. (1996). We are currently investigating this approach for the quantile and the medcouple measures of skewness and tail weight.

Note that it is also possible to construct tables consisting of critical values for different sample sizes n. Hereby it suffices to replace the asymptotic variance $V(\omega, G_{g,0})$ with $nV_n(\omega, G_{g,0})$, with V_n the finite-sample variance of ω. This variance can be approximated through extensive simulations, as e.g. done in Schmid and Trede (2003). We have done this for the situations of Tables 3 and 4, but we found no impressive improvements. Hence we prefer to work with the asymptotic variance which does not depend on the sample size.

4.2. Performance at contaminated distributions

We now compare the robustness of the tail weight measures using contaminated $G_{g,h}$ distributions. To this end, we generated 1000 samples of size $n=1000$ from $G_{0,h}$ (symmetric distributions) and from $G_{0.5,h}$ (asymmetric distributions) with h ranging from 0 to 0.3 and computed the uncontaminated value of the measures by averaging the estimates over these samples. Next, contaminated samples were created by taking samples of size $1000(1 - \varepsilon)$, and adding a normal sample $\mathrm{N}(a, \sigma^2=0.1)$ of size 1000ε with $a=40$ (right contamination), $a = -40$ (left contamination) and $a = 0$ (central contamination), for $\varepsilon = 0.05$. Fig. 6 shows the differences between the average estimate at these contaminated samples and the value at the uncontaminated samples for the right tail weight measures. The figures for the left tail weight measures were similar and are therefore not included.

In all figures, the bias caused by the outliers remains rather stable for increasing h. From the middle pictures we see that the right tail measures are hardly influenced by left contamination. As we would expect, LQW(0.25) and RQW(0.75) are the most robust against several types of contamination, followed by LMC and RMC. Again we thus see that LMC and RMC make a good compromise between the more adequate LQW(0.125) and RQW(0.875) and the more robust LQW(0.25) and RQW(0.75).

Let us now investigate how the goodness-of-fit tests are effected by outliers. As been done in Schmid and Trede (2003) we report in Tables 5–7 the proportion of rejections of the null hypothesis of normality ($g=0$, $h=0$) for various fractions ε of contaminated data. We tested the null hypothesis on 10,000 samples of size 100 and 1000 at the 5% significance level. Outliers were generated from a $\mathrm{N}(0, 5)$ distribution, yielding symmetric contamination, or from a $\mathrm{N}(0, 0.1)$, which is central contamination, or from a $\mathrm{N}(40, 0.1)$ distribution (right contamination).

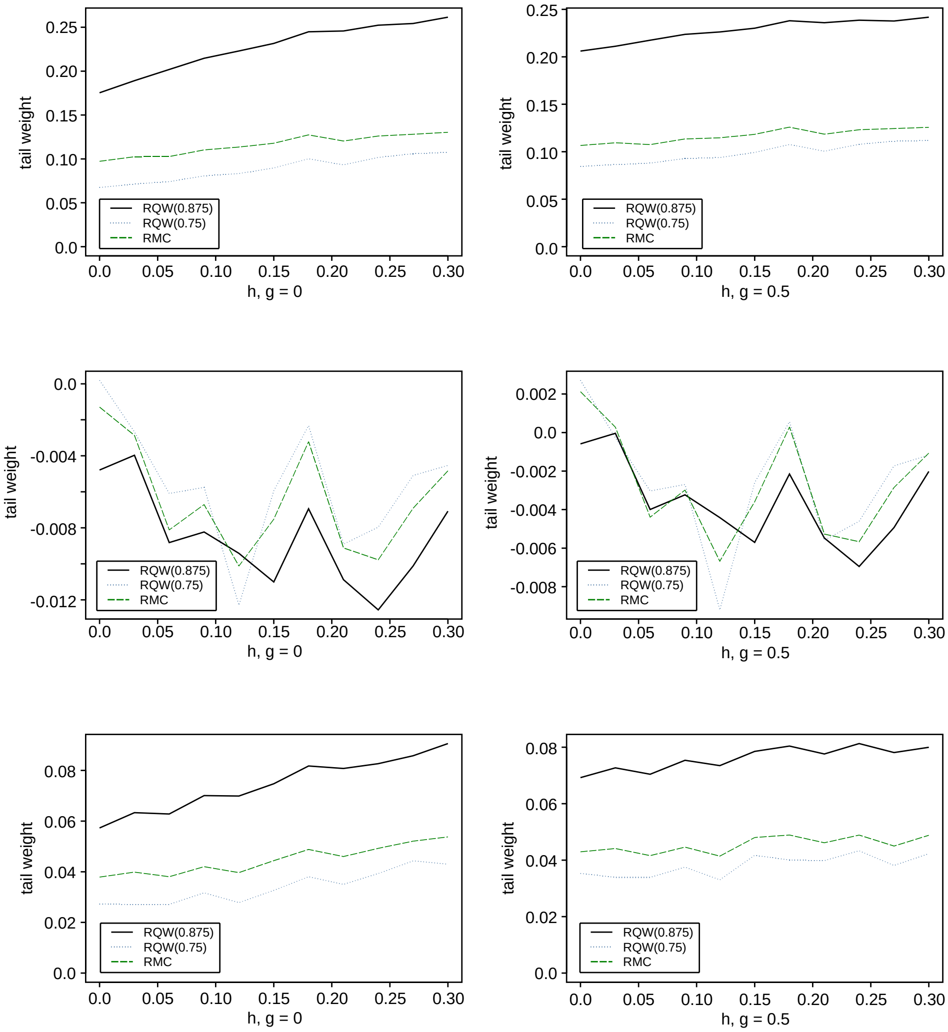

Fig. 6. Right tail weight difference between right contaminated (upper), left contaminated (middle) and central contaminated (lower) samples and uncontaminated samples, with 5% contamination.

As could be seen in Tables 5–7 the power of the SW and b_2 test are heavily influenced by adding contamination. This effect is smaller when adding contamination with small variance in the center of the distribution. In this situation, the *P* test performs worse than the QW and MC alternatives, which is not surprising as *P* measures peakedness. Also at right contamination *P* is more sensitive than our left tail weight measures. Here we see that the right tail weight measures perform not very well. It thus seems that comparing the results of a test based on a left and right tail weight measure gives more information. It remains true that the MC alternatives can be considered as a good compromise for the QW alternatives.

Table 5
Fraction of 10,000 samples of different data sizes n on which the null hypothesis of normality was rejected at the 5% significance level. The samples are drawn from a normal distribution with varying percentage of symmetric contamination

Symmetric	$n=100$				$n=1000$			
	0%	1%	2%	4%	0%	1%	2%	4%
SW	0.052	0.456	0.669	0.879	0.052	0.994	1.000	1.000
b_2	0.059	0.507	0.730	0.917	0.058	0.996	1.000	1.000
P	0.035	0.038	0.040	0.051	0.040	0.051	0.063	0.093
LQW(0.125)	0.029	0.034	0.035	0.047	0.042	0.063	0.075	0.132
RQW(0.875)	0.028	0.033	0.038	0.043	0.042	0.058	0.082	0.128
LQW(0.25)	0.034	0.034	0.033	0.042	0.047	0.053	0.055	0.072
RQW(0.75)	0.034	0.035	0.039	0.042	0.047	0.059	0.057	0.069
LMC	0.033	0.032	0.033	0.043	0.046	0.056	0.065	0.094
RMC	0.029	0.035	0.035	0.044	0.048	0.052	0.070	0.093

Table 6
Fraction of 10,000 samples of different data sizes n on which the null hypothesis of normality was rejected at the 5% significance level. The samples are drawn from a normal distribution with varying percentage of central contamination

Central	$n=100$				$n=1000$			
	0%	1%	2%	4%	0%	1%	2%	4%
SW	0.052	0.053	0.050	0.061	0.052	0.057	0.079	0.156
b_2	0.059	0.066	0.071	0.089	0.058	0.087	0.117	0.197
P	0.035	0.040	0.050	0.071	0.040	0.068	0.101	0.196
LQW(0.125)	0.029	0.034	0.031	0.039	0.042	0.041	0.046	0.055
RQW(0.875)	0.028	0.030	0.031	0.041	0.042	0.046	0.048	0.058
LQW(0.25)	0.034	0.033	0.034	0.033	0.047	0.043	0.046	0.044
RQW(0.75)	0.034	0.031	0.035	0.035	0.047	0.044	0.044	0.042
LMC	0.033	0.033	0.031	0.036	0.046	0.043	0.055	0.056
RMC	0.029	0.028	0.033	0.035	0.048	0.044	0.051	0.055

Table 7
Fraction of 10,000 samples of different data sizes n on which the null hypothesis of normality was rejected at the 5% significance level. The samples are drawn from a normal distribution with varying percentage of right contamination

Right	$n=100$				$n=1000$			
	0%	1%	2%	4%	0%	1%	2%	4%
SW	0.052	1.000	1.000	1.000	0.052	1.000	1.000	1.000
b_2	0.059	1.000	1.000	1.000	0.058	1.000	1.000	1.000
P	0.035	0.041	0.046	0.060	0.040	0.058	0.077	0.160
LQW(0.125)	0.029	0.029	0.030	0.029	0.042	0.039	0.038	0.030
RQW(0.875)	0.028	0.042	0.054	0.127	0.042	0.106	0.236	0.816
LQW(0.25)	0.034	0.033	0.033	0.035	0.047	0.044	0.038	0.042
RQW(0.75)	0.034	0.043	0.044	0.057	0.047	0.062	0.089	0.187
LMC	0.033	0.031	0.030	0.030	0.046	0.039	0.038	0.034
RMC	0.029	0.043	0.048	0.085	0.048	0.078	0.141	0.393

5. Examples

Example 1. The stars data set (Rousseeuw and Leroy, 1987) contains the light intensity and the surface temperature of 47 stars in the direction of Cygnus. A scatter plot of the data and the robust LTS regression line (Rousseeuw, 1984) are shown in Fig. 7(a). In regression, it is important to check normality of the residuals. When a robust regression method is applied, it is sufficient that all residuals except those from the outlying observations are normally distributed. Figs. 7(b) and (c) contain the normal QQ-plot and the boxplot of the LTS residuals, from which five clear outliers are visible. It is known that the four largest residuals correspond with giant stars. The sixth observation that seems to deviate from the linear trend in the normal quantile plot is rather a borderline case with a standardized LTS residual of 3.47. Table 8 shows that the SW test and the b_2 test lead to very different conclusions whether or not these five outliers are included in the data. The same conclusion holds for RQW(0.875) and LQW(0.125) which is due to their low breakdown point of 12.5%. All the other robust tests, including P, do not reject the normality assumption, even in the presence of several outliers. We should be careful in interpreting these results as this data set is very small and consequently the robust tests are known to be very conservative. But still, this example shows again the nonrobustness of the SW and the b_2 test.

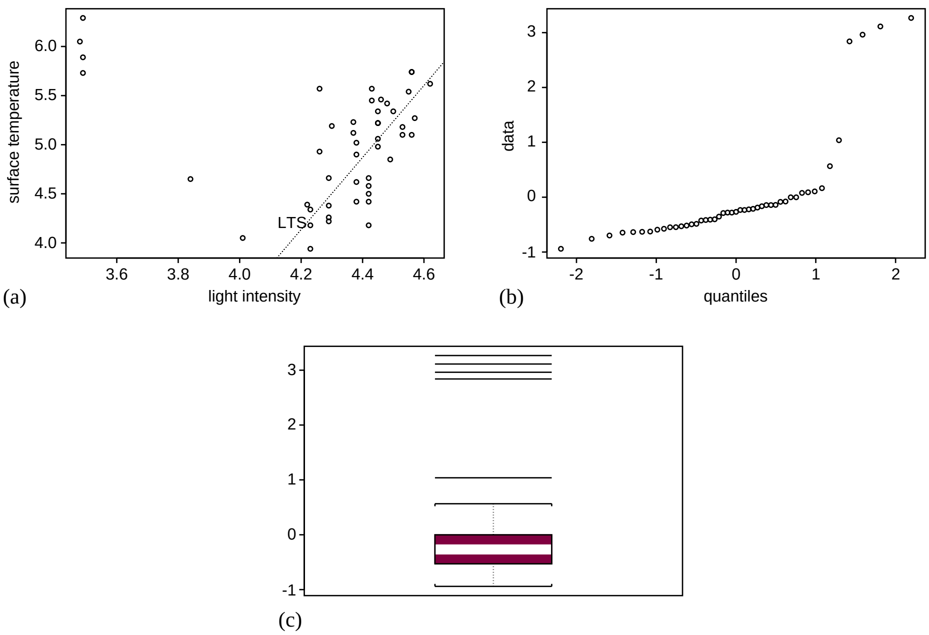

Fig. 7. The Stars data: (a) scatter plot with LTS regression line; (b) normal QQ-plot of the residuals; (c) boxplot of the residuals.

Table 8
Significance of the two-sided goodness-of-fit tests to the normal distribution (stars), to the χ^2_2 distribution (baseball) and to the Student(3) distribution (Procter) with (+) and without (–) outliers

	SW	b_2	P	LQW(0.125)	RQW(0.875)	LQW(0.25)	RQW(0.75)	LMC	RMC
Stars +	0.000	0.000	0.571	0.013	0.004	0.490	0.366	0.271	0.199
Stars –	0.622	0.375	0.622	0.193	0.734	0.308	0.672	0.228	0.669
Baseball +	—	0.000	0.074	0.892	0.165	0.697	0.107	0.695	0.134
Baseball –	—	0.848	0.427	0.981	0.480	0.699	0.189	0.706	0.206
Procter +	—	—	0.765	0.468	0.551	0.355	0.634	0.944	0.739
Procter –	—	—	0.722	0.404	0.578	0.457	0.659	0.987	0.805

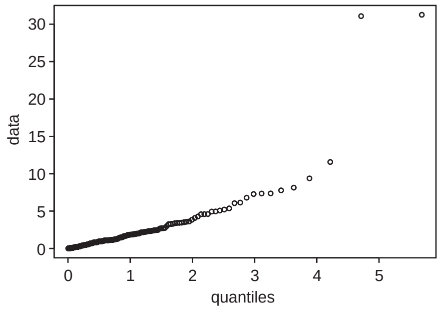

Fig. 8. The baseball data: χ^2_2 based QQ-plot of the robust distances.

Example 2. The baseball data (Reichler, 1991) consists of 162 major league baseball players who achieved true free agency. This means that the player could sell his services to the highest bidding team. A player is expected to handle in two possible directions. Or he plays badly in the year of his free agency, because he is unhappy with his current team and he will play much better in the next year. Or he pushes his performance in his free agency year in order to get to a better team, but then he will play less well the next year. Here, we wanted to test whether the batting average (hits per at bat) at the free agency year and at the next year is bivariate normally distributed. Therefore we computed the robust distances given by

$$(x-\hat{\mu})^{\mathrm{t}}\hat{\Sigma}^{-1}(x-\hat{\mu})$$

in which $\hat{\mu}$ and $\hat{\Sigma}$ are the Minimum Covariance Estimator (MCD) estimates of location and scatter (Rousseeuw, 1984) and $(.)^{\mathrm{t}}$ stands for matrix transpose. If the data follow a bivariate normal distribution, these robust distances are approximately χ^2_2 distributed. On the χ^2_2 based QQ-plot of Fig. 8 we notice two prominent outliers. With these outliers included, the b_2 test rejects the null hypothesis, but it clearly supports the null hypothesis when they are removed. The robust tests based on our tail weight measures are barely influenced by the

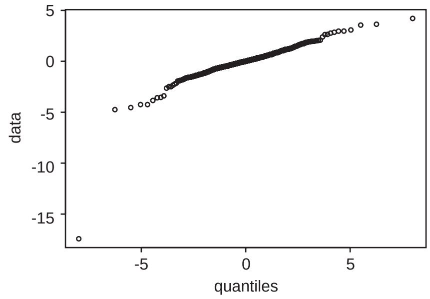

Fig. 9. The Procter & Gamble data: Student(3) based QQ-plot.

outliers and always accept H_0. Only the p-value of RQW(0.875) changes considerably, but the conclusion remains the same. Also P is rather sensitive to the outliers. At the 7.5% significance level, it even rejects the bivariate normality. Note that we cannot consider SW here, as it can only be used to test normality.

Example 3. From Datastream we collected the daily logarithmic returns of the Procter & Gamble stock from January 2000 to December 2003, leading to a univariate data set consisting of 1004 values. From the Student(3) based QQ-plot of Fig. 9, we could believe these data to be likewise distributed, apart from one very abnormal observation. This observation is noted on 7 March 2000, the day that Procter & Gamble has lost 40 billion of US Dollars due to a profit warning. All the robust tests do not reject the null hypothesis, even with the extreme value included. Moreover, they have the advantage that they can be performed to test the goodness-of-fit of distributions without finite fourth moments, such as the Student(3). On the contrary, the b_2 test cannot be applied here.

6. Conclusion

In this paper we have proposed several tail weight measures, based on robust measures of skewness. We considered left and right tail weight measures to make them applicable on asymmetric distributions. All of them follow the anti-skewness ordering of MacGillivray and Balanda (1988), making them intuitively correct tail weight measures. Moreover they do not depend on moments of the data.

We have shown that the measures are robust against outlying values. They all have a positive breakdown value and a bounded influence function. Except in the median, the influence function of the MC alternatives is continuous, while the QW alternatives have a stepwise influence function. Small perturbations may then lead to larger differences. Regarding the breakdown value, the LQW(0.25), RQW(0.75), LMC and RMC measures are preferable because of their breakdown value of 12.5%. This was confirmed with finite-sample simulations.

At a bunch of symmetric distributions, we found that the robust tests are more adequate to detect nonnormality. Regarding some goodness-of-fit tests, LQW(0.125) and RQW(0.875) appeared to be preferable, followed by LMC and RMC. In comparison with the commonly used Shapiro–Wilk test of normality or a test based on the classical kurtosis, the proposed measures show less power, but they are much more capable to handle some outlying values. In practice, we therefore recommend to perform both a robust and a nonrobust test. If they lead to contradictory conclusions, this can be due to the sensitivity of the nonrobust test towards outliers, or due to the conservative behavior of the robust test. In that case, a further investigation of the data is required.

Finally, the goodness-of-fit test proposed by Schmid and Trede (2003) gives results which are comparable with the QW and MC alternatives. But this test assumes inherently a symmetric distribution and it is more sensitive to central contamination. Moreover by considering a left and a right tail measures separately, we are able to perform separate tests on each tail of the distribution. This can provide additional insight in the shape of the data.

When we compare the QW and MC tail weight measures, we observe that the MC alternatives make a good compromise between robustness towards outlying values and adequately measuring tail weight. Moreover because of their low $O(n \log(n))$ computation time and their lack of any parameter p or q, we recommend LMC and RMC to use in practice. In our future work we will investigate how the goodness-of-fit test can be improved by considering the joint distribution of a robust tail weight measure and a robust skewness estimator.

Software. Source code to calculate all the mentioned measures in Matlab or S-plus and their asymptotic variances in Mathematica can be downloaded from `http://www.agoras.ua.ac.be/` and `http://www.wis.kuleuven.ac.be/stat/robust.html`.

Appendix

Proof of Theorem 1. For simplicity, we will assume that n is divisible by 4, for other values of n the proof is similar.

First, we prove that $\varepsilon_n^*(\mathrm{RMC}_n; X_n) \leqslant 1/n(\lceil n/8 \rceil + 1)$. By location invariance, assume w.l.o.g. that the third quartile $q_3(X_n) = 0$. Moreover, let $\mathrm{RMC}(X_n) \geqslant 0$ by symmetry. We will construct a contaminated sample X'_n by replacing $\lceil n/8 \rceil + 1$ data points from X_n such that the RMC becomes arbitrarily large, thus $\mathrm{RMC}(X'_n) > B$ for any $\mathrm{RMC}(X_n) < B < 1$. To contaminate our sample, we shift the $\lceil n/8 \rceil + 1$ $(= n - [7n/8] + 1)$ largest values of X_n by a constant $k > 2 \max |x_i| / (1 - B)$. (The notation $[x]$ stands for the largest integer smaller or equal to x.) Now, $q_3(X'_n) = q_3(X_n)$, the sample median $m_n(X'_n) = m_n(X_n)$ and for all $m_n \leqslant x_i \leqslant q_n$ we have that

$$h(x_i, x'_j) = \begin{cases} h(x_i, x_j) & \text{for } j = \dfrac{n}{2} + 1, \ldots, \left[\dfrac{7n}{8}\right] - 1, \\ \dfrac{x_j + x_i + k}{x_j - x_i + k} & \text{for } j = \left[\dfrac{7n}{8}\right], \ldots, n. \end{cases}$$

By definition of k, we obtain that $h(x_i, x'_j) > B$ for each $j \geqslant [7n/8]$. Since i runs over $n/4$ values, there are at least $n/4(\lceil n/8 \rceil + 1)$ kernels larger than B. We assumed that X_n is in general position and that n is divisible by 4, hence the RMC is computed as the median over $n^2/16$ values. It follows that $\mathrm{RMC}_n(X'_n) > B$.

It remains to show that $\varepsilon^*_n(\mathrm{RMC}_n; X_n) \geqslant 1/n(\lceil n/8 \rceil - 1)$. Replace $k < \lceil n/8 \rceil - 1$ data points by arbitrary values x'_i. We need to show that $\mathrm{RMC}(X'_n)$ does not depend on the contaminated data, and thus its absolute value should remain smaller than 1. The $\mathrm{RMC}(X'_n)$ is based upon the $n/2$ of the x'_i to the right of the median $m'_n = m_n(X'_n)$. Denote the third quartile of X'_n by q'_3, then there are a original data points lying in $(m'_n, q'_3]$ and b original data points larger than or equal to q'_3 with

$$\left[\frac{n}{8}\right] + 2 \leqslant \min\{a, b\} \quad \text{and} \quad \max\{a, b\} \leqslant \frac{n}{4}.$$

Also, it is clear that $a + b \geqslant n/2 - [n/8] + 2 = [3n/8] + 2$, such that the number of uncontaminated expressions $h(x_i, x_j)$ contributing to $\mathrm{RMC}(X'_n)$ is $ab \geqslant a([3n/8] + 2 - a)$. It is easy to verify that this lower bound is strictly larger than $[(n^2/16 + 1)/2]$, hence $\mathrm{RMC}(X'_n)$ is obtained as the median of one or two uncontaminated kernels. □

Proof of Theorem 2. Let the contaminated distribution of F be $F_\varepsilon = (1 - \varepsilon)F + \varepsilon\Delta_x$ and the corresponding quantile function $Q_\varepsilon = F_\varepsilon^{-1}$. Then, we can write

$$\mathrm{IF}(x, \mathrm{LQW}(p), F) = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left[\frac{Q_\varepsilon((1-p)/2) + Q_\varepsilon(p/2) - 2Q_\varepsilon(0.25)}{Q_\varepsilon((1-p)/2) + Q_\varepsilon(p/2)}\right]\Bigg|_{(\varepsilon=0)}.$$

Because $\mathrm{IF}(x, Q(p), F) = (\mathrm{d}/\mathrm{d}\varepsilon)Q_\varepsilon(p)|_{(\varepsilon=0)}$, simple calculus yields the given formula. □

Proof of Theorem 3. Similar to the proof of Theorem 2. □

Proof of Theorem 4. First, we rewrite (2.1) for a contaminated distribution $F_\varepsilon = (1 - \varepsilon)F + \varepsilon\Delta_x$. Let $\mathrm{LMC}_\varepsilon = \mathrm{LMC}(F_\varepsilon)$, $m_\varepsilon = F_\varepsilon^{-1}(0.5)$ and $q_\varepsilon = F_\varepsilon^{-1}(0.25)$, then the following equation holds:

$$\frac{1}{32} = \int_{q_\varepsilon}^{m_\varepsilon}\int_{-\infty}^{q_\varepsilon} I\left(-\frac{x_2 + x_1 - 2q_\varepsilon}{x_2 - x_1} \leqslant \mathrm{LMC}_\varepsilon\right) \mathrm{d}F_\varepsilon(x_1)\,\mathrm{d}F_\varepsilon(x_2).$$

Note that the conditions

$$-\frac{x_1 + x_2 - 2q_\varepsilon}{x_2 - x_1} \leqslant \mathrm{LMC}_\varepsilon, \quad x_1 \leqslant q_\varepsilon,\ q_\varepsilon \leqslant x_2 \leqslant m_\varepsilon,\ -1 \leqslant \mathrm{LMC}_\varepsilon \leqslant 1$$

are equivalent to

$$x_1 \geqslant \frac{x_2(-\mathrm{LMC}_\varepsilon - 1) + 2q_\varepsilon}{1 - \mathrm{LMC}_\varepsilon}, \quad q_\varepsilon \leqslant x_2 \leqslant m_\varepsilon,\ -1 \leqslant \mathrm{LMC}_\varepsilon \leqslant 1.$$

We now introduce the functions

$$g_1(v, \varepsilon) = \frac{v(-\mathrm{LMC}_\varepsilon - 1) + 2q_\varepsilon}{1 - \mathrm{LMC}_\varepsilon},$$
$$g_2(v, \varepsilon) = \frac{v(-\mathrm{LMC}_\varepsilon + 1) - 2q_\varepsilon}{-1 - \mathrm{LMC}_\varepsilon}$$

which for $\varepsilon = 0$ collapse with $g_{1,l}$ and $g_{2,l}$ defined in (3.1) and (3.2). With these notations, we obtain

$$\begin{aligned}\frac{1}{32} &= \int_{q_\varepsilon}^{m_\varepsilon} F_\varepsilon(g_1(x_2,\varepsilon))\,\mathrm{d}F_\varepsilon(x_2)\\ &= \int_{q_\varepsilon}^{m_\varepsilon} [(1-\varepsilon)F + \varepsilon\Delta_x](g_1(x_2,\varepsilon))d[(1-\varepsilon)F + \varepsilon\Delta_x](x_2)\\ &= (1-2\varepsilon)\int_{q_\varepsilon}^{m_\varepsilon} F(g_1(x_2,\varepsilon))\,\mathrm{d}F(x_2) + \varepsilon\int_{q_\varepsilon}^{m_\varepsilon} F(g_1(x_2,\varepsilon))\,\mathrm{d}\Delta_x(x_2)\\ &\quad + \varepsilon\int_{q_\varepsilon}^{m_\varepsilon} \Delta_x(g_1(x_2,\varepsilon))\,\mathrm{d}F(x_2) + \mathrm{O}(\varepsilon^2). \qquad (6.1)\end{aligned}$$

To compute $\mathrm{IF}(x, \mathrm{LMC}, F) = (\partial/\partial\varepsilon)\mathrm{LMC}(F_\varepsilon)|_{(\varepsilon=0)}$ we derive equality (6.1) with respect to ε, and let $\varepsilon \to 0$. Since the terms in ε^2 vanish, we have to derive the first three terms only, denoted by $T_{1,\varepsilon}$, $T_{2,\varepsilon}$ and $T_{3,\varepsilon}$.

$$\begin{aligned}\frac{\partial}{\partial\varepsilon} T_{1,\varepsilon}\bigg|_{(\varepsilon=0)} &= \frac{\partial}{\partial\varepsilon}\left[(1-2\varepsilon)\int_{q_\varepsilon}^{m_\varepsilon} F(g_1(x_2,\varepsilon))\,\mathrm{d}F(x_2)\right]\bigg|_{(\varepsilon=0)}\\ &= -2\int_{q_F}^{m_F} F(g_1(x_2))\,\mathrm{d}F(x_2) + \frac{\partial}{\partial\varepsilon}\int_{q_\varepsilon}^{m_\varepsilon} F(g_1(x_2,\varepsilon))\,\mathrm{d}F(x_2)\bigg|_{(\varepsilon=0)} \qquad (6.2)\end{aligned}$$

By definition of LMC_F, the first term in (6.2) equals $-1/32$, whereas Leibnitz' rule yields

$$\begin{aligned}&\frac{\partial}{\partial\varepsilon}\int_{q_\varepsilon}^{m_\varepsilon} F(g_1(x_2,\varepsilon))\,\mathrm{d}F(x_2)\bigg|_{(\varepsilon=0)}\\ &= \int_{q_F}^{m_F} F'(g_1(x_2,0))\frac{\partial}{\partial\varepsilon}g_1(x_2,\varepsilon)\bigg|_{(\varepsilon=0)}\mathrm{d}F(x_2) + F(g_1(m_F,0))F'(m_F)\frac{\partial}{\partial\varepsilon}m_\varepsilon\bigg|_{(\varepsilon=0)}\\ &\quad + F(g_1(q_F,0))F'(q_F)\frac{\partial}{\partial\varepsilon}q_\varepsilon\bigg|_{(\varepsilon=0)}.\end{aligned}$$

Calculus yields

$$\frac{\partial}{\partial\varepsilon}g_1(x_2,\varepsilon)\bigg|_{(\varepsilon=0)} = \frac{2(q_1 - x)\mathrm{IF}(x, \mathrm{LMC}_F, F) + 2\mathrm{IF}(x, q_1, F)(-\mathrm{LMC}_F + 1)}{(-\mathrm{LMC}_F + 1)^2}$$

hence

$$\begin{aligned}\frac{\partial}{\partial\varepsilon} T_{1,\varepsilon}\bigg|_{(\varepsilon=0)} &= -\frac{1}{16} + \mathrm{IF}(x, \mathrm{LMC}_F, F)\int_{q_1}^{m_F} \frac{2(q_1 - x_2)}{(-\mathrm{LMC}_F + 1)^2} f(g_1(x_2))\,\mathrm{d}F(x_2)\\ &\quad + 2\mathrm{IF}(x, q_1, F)\int_{q_1}^{m_F} \frac{f(g_{1,l}(x_2))}{-\mathrm{LMC}_F + 1}\,\mathrm{d}F(x_2)\\ &\quad - \frac{1}{4} f(q_1)\mathrm{IF}(x, q_1, F) + F(g_{1,l}(m_F))f(m_F)\mathrm{IF}(x, m_F, F). \qquad (6.3)\end{aligned}$$

The second term $T_{2,\varepsilon}$ in Eq. (6.1) has partial derivative

$$\begin{aligned}\frac{\partial}{\partial\varepsilon}T_{2,\varepsilon}\Big|_{(\varepsilon=0)} &= \frac{\partial}{\partial\varepsilon}\left[\varepsilon\int_{q_\varepsilon}^{m_\varepsilon} F(g_1(x_2,\varepsilon))\,\mathrm{d}\Delta_x(x_2)\right]\Big|_{(\varepsilon=0)}\\ &= \int_{q_\varepsilon}^{m_\varepsilon} F(g_1(x_2,\varepsilon))\,\mathrm{d}\Delta_x(x_2)|_{(\varepsilon=0)}\\ &= F(g_{1,l}(x))[I(x>q_1)-I(x>m_F)] \end{aligned}\tag{6.4}$$

whereas for the third term $T_{3,\varepsilon}$ we obtain

$$\begin{aligned}\frac{\partial}{\partial\varepsilon}T_{3,\varepsilon}\Big|_{(\varepsilon=0)} &= \int_{q_\varepsilon}^{m_\varepsilon} \Delta_x(g_1(x_2,\varepsilon))\,\mathrm{d}F(x_2)|_{(\varepsilon=0)}\\ &= \int_{q_\varepsilon}^{m_\varepsilon} I(x<g_1(x_2,\varepsilon))\,\mathrm{d}F(x_2)|_{(\varepsilon=0)}\\ &= \int_{q_\varepsilon}^{m_\varepsilon} I(x_2<g_2(x,\varepsilon))\,\mathrm{d}F(x_2)|_{(\varepsilon=0)}\\ &= \int_{q_\varepsilon}^{g_2(x,\varepsilon)} I(m_\varepsilon>g_2(x,\varepsilon))I(q_\varepsilon<g_2(x,\varepsilon))\,\mathrm{d}F(x_2)|_{(\varepsilon=0)}\\ &\quad+\int_{q_\varepsilon}^{m_\varepsilon} I(g_2(x,\varepsilon)>m_\varepsilon)\,\mathrm{d}F(x_2)|_{(\varepsilon=0)}\\ &= I(g_{2,l}(x)<m_F)I(x<q_1)\left[F(g_{2,l}(x))-\frac{1}{4}\right]\\ &\quad+I(g_{2,l}(x)>m_F)\frac{1}{4}\\ &= I(x>g_{1,l}(m_F))I(x<q_1)\left[F(g_{2,l}(x))-\frac{1}{4}\right]\\ &\quad+I(x<g_{1,l}(m_F))\frac{1}{4}. \end{aligned}\tag{6.5}$$

Combining Eqs. (6.1), (6.3), (6.4), and (6.5) and using the fact that

$$H'_F(\mathrm{LMC}_F)=16\int_{q_1}^{m_F} 2f(g_{1,l}(x_2))\left(\frac{q_1-x_2}{(-\mathrm{LMC}_F+1)^2}\right)\mathrm{d}F(x_2)$$

and

$$\mathrm{IF}(x,q_1,F)=\frac{0.25-I(x<q_1)}{f(q_1)}$$

finally leads to the influence function given in Theorem 4. □

Proof of Theorem 5. Similar to the Proof of Theorem 4. □

Proof of Theorem 6. The proof is similar as in Schmid and Trede (2003). For $0<p_1<\cdots<p_k<1$ and $F=\Phi$ the sample quantiles are asymptotically normal, i.e. $\mathrm{N}(0,\Sigma)$ with typical element in the covariance matrix $\Sigma = (\sigma_{ij})_{i,j=1,\ldots,k}$ given by

$$\sigma_{ij} = \frac{p_i(1-p_j)}{\phi(x_{p_i})\phi(x_{p_j})}$$

for $i \leqslant j$ and ϕ denoting the density of the standard normal distribution (Serfling, 1980). Let $k=4$ and $h : \mathbb{R}^3 \to \mathbb{R} : (x_1, x_2, x_3) \to (x_3+x_1-2x_2)(x_3-x_1)^{-1}$ with partial derivatives $h'(.) = (h_1(.), h_2(.), h_3(.))$ where $h_1(.) = 2(x_3-x_2)(x_3-x_1)^{-2}$, $h_2(.) = -2(x_3-x_1)^{-1}$ and $h_3(.) = 2(x_2-x_1)(x_3-x_1)^{-2}$. Using the delta method we arrive at

$$\sqrt{n}(\mathrm{QW}_n - \mathrm{QW}) \to_{\mathscr{D}} \mathrm{N}(0, h'\Sigma h),$$

where h has to be evaluated at the p_i quantiles of the corresponding quantiles of the QW alternative. Numerical calculations of $h'\Sigma h$ give the results in the theorem. □

References

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. Statistics of Extremes, Theory and Applications, Wiley, Chichester.

Bera, A., Jarque, C., 1981. Efficient tests for normality, heteroskedasticity and serial independence of regression residuals: Monte Carlo evidence. Econom. Lett. 7, 313–318.

Bickel, P.J., Lehmann, E.L., 1975. Descriptive statistics for nonparametric models, 1: Introduction. Ann. Statist. 3, 1038–1044.

Bonett, D.G., Seier, E., 2002. A test of normality with high uniform power. Comput. Statist. Data Anal. 40, 435–445.

Brys, G., Hubert, M., Struyf, A., 2003. A comparison of some new measures of skewness. In: Dutter, R., Filzmoser, P., Gather, U., Rousseeuw, P.J. (Eds.), Developments in Robust Statistics, ICORS 2001. Springer, Heidelberg, pp. 98–113.

Brys, G., Hubert, M., Struyf, A., 2004. A robust measure of skewness. J. Comput. Graphical Statist. (JCGS). 13 (No. 4) , 1–22.

Gleason, J.R., 1993. Understanding elongation: the scale contaminated normal family. J. Amer. Statist. Assoc. 88, 327–337.

Groeneveld, R.A., 1998. A class of quantile measures for kurtosis. Amer. Statist. 51, 325–329.

Groeneveld, R.A., Meeden, G., 1984. Measuring skewness and kurtosis. The Statistician 33, 391–399.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. Robust Statistics: the Approach Based on Influence Functions, Wiley-Interscience, New York.

Hastings, C., Mosteller, F., Tukey, J.W., Winsor, C.P., 1947. Low moments for small samples: a comparative study of order statistics. Ann. Math. Statist. 18, 413–426.

Hinkley, D.V., 1975. On power transformations to symmetry. Biometrika 62, 101–111.

Hoaglin, D.C., Mosteller, F., Tukey, J.W., 1985. Exploring Data Tables, Trends and Shapes, Wiley, New York.

Hsieh, P.-H., 1999. Robustness of tail index estimation. J. Comput. Graphical Statist. 8, 318–332.

Huber, P.J., 1981. Robust Statistics, Wiley, New York.

Johnson, N.L., 1949. Systems of frequency curves generated by methods of translations. Biometrika 36, 149–176.

MacGillivray, H.L., Balanda, K.P., 1988. The relationships between skewness and kurtosis. Austral. J. Statist. 30, 319–337.

Moors, J.J.A., Wagemakers, R.T.A., Coenen, V.M.J., Heuts, R.M.J., Janssens, M.J.B.T., 1996. Characterizing systems of distributions by quantile measures. Statist. Neerlandica 50, 417–430.

Oja, H., 1981. On location, scale, skewness and kurtosis of univariate distributions. Scand. J. Statist. 8, 154–168.

Rayner, G.D., MacGillivray, H.L., 2002. Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions. Statist. Comput. 12, 57–75.

Reichler, J.L., 1991. The Baseball Encyclopedia, Macmillan, New York.

Rousseeuw, P.J., 1984. Least median of squares regression. J. Amer. Statist. Assoc. 79, 871–881.
Rousseeuw, P.J., Leroy, A., 1987. Robust Regression and Outlier Detection, Wiley, New York.
Ruppert, D., 1987. What is kurtosis? An influence function approach. Amer. Statist. 41, 1–5.
Schmid, F., Trede, M., 2003. Simple tests for peakedness, fat tails and leptokurtosis based on quantiles. Comput. Statist. Data Anal. 43, 1–12.
Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). Biometrika 52, 591–611.
Vandewalle, B., Beirlant, J., Hubert, M., 2004a. A robust estimator of the tail index based on an exponential regression model. In: Hubert, M., Pison, G., Struyf, A., Van Aelst, S. (Eds.), Theory and Applications of Recent Robust Methods, Series: Statistics for Industry and Technology. Birkhauser, Basel, pp. 367–376.
Vandewalle, B., Hubert, M., Beirlant, J., 2004b. A robust estimator for the tail index of Pareto-type distributions, submitted. Available at http://www.wis.kuleuven.ac.be/stat/robust.html.
van Zwet, W.R., 1964. Convex Transformations of Random Variables, Mathematisch Centrum, Amsterdam.