TEST
DOI 10.1007/s11749-014-0357-7

ORIGINAL PAPER

On consistency factors and efficiency of robust *S*-estimators

Marco Riani · Andrea Cerioli · Francesca Torti

Received: 5 May 2013 / Accepted: 16 January 2014

Abstract We tackle the problem of obtaining the consistency factors of robust *S*-estimators of location and scale both in regression and multivariate analysis. We provide theoretical results, proving new formulae for their calculation and shedding light on the relationship between these factors and important statistical properties of *S*-estimators. We also give computational advances, suggesting efficient procedures so that hardly any time is lost for their calculation when computing *S*-estimates. In addition, when the purpose is to fix the efficiency of the scale estimator, we are able to quantify to what extent the approximate algorithms which are currently available provide an acceptable solution, and when it is necessary to resort to the exact formulae. Finally, even if this paper concentrates on *S*-estimates and Tukey's Biweight and optimal loss functions, the main results can be easily extended to calculate the tuning consistency factors for other popular loss function and other robust estimators.

Keywords Asymptotic variance · Biweight function · Breakdown point · Elliptical truncation · Mahalanobis distance · Robust regression

Mathematics Subject Classification 62G35 · 62E15 · 62E17 · 62J05

1 Introduction

The main goal of robust statistics is to obtain methods which are insensitive to deviations from the postulated model for the "good" part of the data. This model typically involves assumptions about normality or (elliptical) symmetry of the distribution

M. Riani (✉) · A. Cerioli
Dipartimento di Economia, Università di Parma, Parma, Italy
e-mail: mriani@unipr.it

F. Torti
European Commission, Joint Research Centre, Ispra, Italy

which is supposed to generate the data, either in the univariate, regression or multivariate case. As a consequence, great effort has been devoted to develop estimators with a high-breakdown point, which are not affected by the presence of a fraction of unknown outliers as large as 50 %. See, e.g., Hubert et al. (2008) for a recent overview.

In order to be successfully applied in practice, such high-breakdown estimators must possess sensible statistical properties also when the data are not contaminated. For instance, in the multivariate setting that we consider in Sect. 4, let $y_1, \ldots, y_n$ denote a sample of v-variate observations from the elliptical density $f(y, \mu, \Sigma) = \frac{1}{\sqrt{|\Sigma|}} c_g g(d(y, \mu, \Sigma))$, where μ is the mean vector and Σ is proportional to the covariance matrix (if the first two moments exist), g is the so called radial density, $c_g > 0$ is a normalizing factor which ensures that f integrates to one and $d(y, \mu, \Sigma) = (y - \mu)' \Sigma^{-1} (y - \mu)$ (Paindaveine 2012). Let $\hat{\Sigma} = \hat{\Sigma}(y_1, \ldots, y_n)$ be any affine equivariant estimator of Σ, such that $\hat{\Sigma} \to \hat{\Sigma}_\infty$ in probability as $n \to \infty$. It can be shown (Maronna et al. 2006, p. 217) that $\hat{\Sigma}_\infty$, is proportional to Σ, i.e.

$$\hat{\Sigma}_\infty = \kappa \Sigma, \tag{1}$$

where κ is a constant. This constant plays a crucial role in inference based on $\hat{\Sigma}$, because $\hat{\Sigma}/\kappa$ is a consistent estimator of Σ when all the data come from the postulated density $f(y, \mu, \Sigma)$.

Explicit and easily computable formulae for the scaling factor in (1) are available when $\hat{\Sigma}$ is an affine equivariant estimator whose high-breakdown properties derive from trimming and $y_i \sim N_v(\mu, \Sigma)$, with $i = 1, 2, \ldots, n$. In this case,

$$g(d(y, \mu, \Sigma)) = e^{-\frac{d(y,\mu,\Sigma)}{2}} \quad c_g = (2\pi)^{-\frac{v}{2}}. \tag{2}$$

For example, if $\hat{\Sigma}$ is the minimum covariance determinant (MCD) with asymptotic breakdown point equal to α and $h = [(2[\frac{n+v+1}{2}] - n + 2(n - [\frac{n+v+1}{2}](1-\alpha))]$ is the corresponding coverage, Croux and Haesbroeck (1999) show that at the multivariate normal model

$$\kappa = \kappa_h = \frac{\Pr\left(\chi^2_{v+2} < \chi^2_{v,h/n}\right)}{h/n}, \tag{3}$$

where $\chi^2_{v,h/n}$ is the h/n quantile of the χ^2_v distribution. Related results are obtained by Liu et al. (1999) in the case of data-depth based estimation of Σ, by Garcìa-Escudero and Gordaliza (2005) for a generalized radius process, by Cuesta-Albertos et al. (2008) under a likelihood approach to trimming, by Riani et al. (2009) for the forward search estimator of Σ, by Cerioli (2010) for the reweighted version of the MCD and by Cizek (2013) for the reweighted least trimmed squares estimator.

In this paper, we deal with robust S-estimators based on a sufficiently smooth weight function. Their asymptotic behavior is established by Davies (1987) and Lopuhaä (1989) for multivariate data, and by Rousseeuw and Yohai (1984) and Davies (1990) in regression. One crucial issue with these estimators is that smoothness of the weight function prevents from obtaining simple analytic expressions like (3) for their consistency factors. Instead, the asymptotic behavior of S-estimators is implicitly ruled by the features of the chosen weight function, as well as their robustness and efficiency properties (see, e.g., Rousseeuw and Leroy 1987; Croux et al. 2006).

More explicitly, in the "hard trimming procedures" like MCD, LTS or the forward search, where each observation receives weight 0 or 1, the number h of observations having weight equal to 1 determines the breakdown point (bdp), the consistency factor $\kappa = \kappa_h$, the variance of the estimator (asvar), and finally its efficiency (given that eff = asvar^{-1}; see for example Maronna et al. p. 182). Once one of the four factors h, bdp, asvar or eff is fixed, the values of the other three follow consequently. This easily explains why it is not possible, in a single step, to achieve a small asymptotic variance (high efficiency) and 50 % breakdown point simultaneously. In the context of S-estimation, the pivotal role of h is played by c, the parameter which appears in the typical ρ, ψ and weight functions of robust statistics, which we define in the next section [see, e.g., Eqs. (11), (36)]. Sometimes c is called a tuning constant and more often, as we do in this article, it is directly called the consistency factor for S-estimators.

Parameter c determines univocally both the breakdown point and the asymptotic variance of location and covariance estimators. Thus if we decide, for example, to have a fixed level of efficiency for the S-estimator of the covariance, we automatically determine the efficiency of the location estimator, the consistency factor c, the breakdown point of the location and covariance estimator.

The implicit nature of the consistency factors of S-estimators has two major consequences. The first one is that it can contribute to hide important statistical relationships, like the one between robustness and efficiency explored by Croux et al. (2011). The second consequence is that computation of the required scaling routinely involves iterative procedures which are considerably more complex and time consuming than expression (3) for trimmed estimators. The computed constant is then plugged into further iterative algorithms, such as those of Salibian-Barrera and Yohai (2006) and Salibian-Barrera et al. (2006), which provide an approximate solution to the nonlinear optimization problem of S-estimation. Finding simpler formulae for the consistency factors of S-estimators is thus an important task both from a conceptual and a practical point of view. Although we restrict to the hypothesis of normally distributed data like in (2), our approach could be potentially extended to observations from elliptical distributions.

Our goal in this paper is twofold. First, we prove new formulae for the calculation of the consistency factors of S-estimators. Our new method of proof is general and could be applied to any definition of the weight function based on polynomials, although we restrict ourselves to the case of Tukey's Biweight and optimal loss function, which are by far the most common. Our new formulae take advantage of the use of semifactorials, do not require the evaluation of the Gamma function and are conceptually simpler.

They can be interpreted as a natural extension of (3) to the case of S-estimators. They also open the door to efficient computational procedures, which ensure savings in computing time over the currently available algorithms. Similar improvements could be used by alternative popular routines available for robust data analysis.

Our second aim is to use the new formulae to provide a fresh look at the theory of consistency factors for S-estimators, by shedding light on the relationship between these factors and other important statistical properties of the estimators. Specifically, we provide theoretical results and graphical details relating consistency to robustness and asymptotic efficiency. Even if these plots could have been obtained using the traditional old formulae, we argue that so far they have not received sufficient attention. Finally, when the purpose is to fix the efficiency of the scale (covariance) estimator, we are also able to quantify to what extent the approximate algorithms which are currently available (e.g. Todorov and Filzmoser 2009) provide an acceptable solution, and when it is necessary to resort to the exact formulae. In addition, and perhaps unexpectedly, we uncover a few interesting relationships between consistency factors and some old (Tallis 1963) and new (Riani et al. 2009) results.

The outline of the paper is as follows. In Sect. 2 we recall *S*-estimation in a regression framework and we define the consistency factor for Tukey's Biweight function. In Sect. 3 we find easy expressions to compute this factor in regression and we illustrate its relationships with other statistical properties of the estimators. The multivariate case is dealt within Sects. 4 and 5. In Sect. 6 we compare the computational time it takes to compute the consistency factors using the routines currently downloadable from Internet with our new implementation. In Sect. 7 we extend the results previously obtained for the Tukey Biweight loss function to the optimal loss function. In Sect. 8 we introduce a numerical example which shows the need of considering different levels of breakdown point or different levels of nominal efficiency to understand the real structure of the data. Some concluding remarks are in Sect. 9.

2 *S*-estimation in a regression setup

As usual in a regression framework, we define y_i to be the response variable, which is related to the values of a set of $p-1$ explanatory variables $x_{i1},\ldots,x_{ip-1}$ by the relationship

$$y_i = \beta' x_i + \epsilon_i \quad i = 1, \ldots, n, \tag{4}$$

where, including an intercept, $\beta' = (\beta_0, \beta_1, \ldots, \beta_{p-1})$ and $x_i = (1, x_{i1}, \ldots, x_{ip-1})'$. Let $\sigma^2 = \text{var}(\epsilon_i)$, which is assumed to be constant for all $i = 1, \ldots, n$. We also take the quantities in x_i to be fixed and assume that $x_1, \ldots, x_n$ are not collinear. The case $p = 1$ corresponds to that of a univariate response without predictors. We call σ the scale of the distribution of the error term ϵ_i, when its density takes the form

$$\sigma^{-1} f\left(\frac{\epsilon}{\sigma}\right).$$

The *M*-estimator of the regression parameters which is scale equivariant (i.e. independent of the system of the units) is defined by

$$\hat{\beta}_M = \min_{\beta \in \Re^p} \frac{1}{n} \sum_{i=1}^{n} \rho\left(\frac{r_i}{s}\right) \tag{5}$$

where $r_i = y_i - \beta' x_i$ is the ith residual and ρ is a function with suitable properties and s is an estimate of σ. We also write $r_i(\beta)$ to emphasize the dependence of r_i

on β. Similarly, we write $s = s(r_1(\beta), \ldots, r_n(\beta))$ for the dispersion measure of the residuals, which is called the "scale statistic" (see, e.g., Van Aelst and Willems 2011).

Clearly, if we want to keep the *M*-estimate robust, s should be a robust estimate too. An *M*-estimator of scale in (4), say s, is defined to be the solution to the following equation

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{r_i}{s}\right) = \frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{y_i - \beta' x_i}{s}\right) = K, \tag{6}$$

Equation (6) is solved among all $(\beta, \sigma) \in \Re^p \times (0, \infty)$, where $0 < K < \sup \rho$. Rousseeuw and Yohai (1984) defined *S*-estimators by minimization of the dispersion s of the residuals

$$\hat{\beta}_S = \min_{\beta \in \Re^p} s(r_1(\beta), \ldots, r_n(\beta)) \tag{7}$$

with final scale estimate

$$\hat{\sigma}_S = s\left(r_1(\hat{\beta}_S), \ldots, r_n(\hat{\beta}_S)\right).$$

The dispersion s defined as the solution of Eq. (6). The *S*-estimates, therefore, can be thought as self-scaled *M*-estimates whose scale is estimated simultaneously with the regression parameters (Yohai 2006). Note, in fact, that when the scale and the regression estimates are simultaneously estimated, *S*-estimators for regression also satisfy (see for example Maronna et al. 2006; p. 131)

$$\hat{\beta}_S = \min_{\beta \in \Re^p} \sum_{i=1}^{n} \rho\left(\frac{r_i}{s}\right). \tag{8}$$

The estimator of β in (7) is called an *S*-estimator because it is derived from a scale statistic in an implicit way. *S*- and *M*-estimators for regression can also be extended to the case of heteroskedastic and time series data; see for example Croux et al. (2006), Maronna et al. (2006, Ch. 8), Boudt and Croux (2010), and Rieder (2012).

Rousseeuw and Leroy (1987, p. 139) show that if function ρ satisfies the following conditions:

1. it is symmetric and continuously differentiable, and $\rho(0) = 0$;
2. there exists a $c > 0$ such that ρ is strictly increasing on $[0, c]$ and constant on $[c, \infty)$;
3. it is such that

$$K/\rho(c) = \alpha \quad \text{with} \quad 0 < \alpha \leq 0.5. \tag{9}$$

then the asymptotic breakdown point of the *S*-estimator tends to α when $n \to \infty$. Note that if $\rho(c)$ is normalized in such a way that $\sup \rho(c) = 1$, constant K becomes exactly equal to the breakdown point of the *S*-estimator.

It is further assumed that

$$E_{\Phi_{0,1}}\left[\rho\left(\frac{r_i}{s}\right)\right] = K, \tag{10}$$

where $\Phi_{0,1}$ is the cdf of the standard normal distribution, to have a consistent scale estimate for normally distributed observations.

The most popular choice for the ρ function in (5) and (6) is Tukey's Biweight function

$$\rho(x) = \begin{cases} \frac{x^2}{2} - \frac{x^4}{2c^2} + \frac{x^6}{6c^4} & \text{if } |x| \le c \\ \frac{c^2}{6} & \text{if } |x| > c, \end{cases} \tag{11}$$

with associated ψ function:

$$\psi(x) = \rho'(x) = \begin{cases} x - \frac{2x^3}{c^2} + \frac{x^5}{c^4} = x\left\{1 - \left(\frac{x}{c}\right)^2\right\}^2 & \text{if } |x| \le c \\ 0 & \text{if } |x| > c. \end{cases} \tag{12}$$

Input x of ρ will be the scaled residuals (in a regression context) or the Mahalanobis distances (in multivariate analysis).

3 Consistency factors in regression

3.1 Consistency factors given a fixed breakdown point

We consider the case of finding c, once the breakdown point has been fixed. We work under the *assumption of normally distributed observations* which are a special case of the elliptical family with g given by (2). This assumption is the standard null model for the bulk of the data in the literature on robust estimation.

Conditions (9) and (10) imply that c for Tukey's Biweight must satisfy the following equation:

$$\int_{-c}^{c} \left(\frac{x^2}{2} - \frac{x^4}{2c^2} + \frac{x^6}{6c^4}\right) d\Phi_{0,1}(x) + \frac{c^2}{6}\Pr(|X| > c) = \alpha\frac{c^2}{6}. \tag{13}$$

If we put $v_c(k) = \int_{-c}^{c} x^k d\Phi_{0,1}(x)$, then Eq. (13) can be rewritten as follows

$$\frac{v_c(2)}{2} - \frac{v_c(4)}{2c^2} + \frac{v_c(6)}{6c^4} + \frac{c^2}{6}2(1 - \Phi_{0,1}(c)) = \alpha\frac{c^2}{6}. \tag{14}$$

At present [see, e.g., Appendix A of Croux et al. (2006) or the code in the R library `robustbase`], $v_c(k)$ is computed using the following expression

$$v_c(k) = \frac{2^{k/2}}{\sqrt{\pi}}\Gamma\left(\frac{k+1}{2}\right) P\left(\frac{c^2}{2}, \frac{k+1}{2}\right),$$

where Γ and P are the gamma and the incomplete gamma functions, respectively. In what follows we give a simpler expression which does not need the gamma function. We start with a simple lemma.

Lemma 3.1 *If X follows a standard normal distribution, then the kth order truncated central moment (where k is an even number) is given by*

$$v_c(k) = \int_{-c}^{c} x^k d\Phi(x) = (k-1)!!F_{\chi^2_{k+1}}(c^2), \tag{15}$$

where the double factorial $j!!$ *is defined as follows*

$$j!! = \begin{cases} j \cdot (j-2) \cdot \ldots \cdot 5 \cdot 3 \cdot 1, & j > 0, odd \\ j \cdot (j-2) \cdot \ldots \cdot 6 \cdot 4 \cdot 2, & j > 0, even \\ 1, & j = -1, 0 \end{cases}$$

Proof Using the well-known relationship between the standard normal and the χ^2, we start from the following equality

$$v_c(k) = \int_{-c}^{c} x^k d\Phi(x) = \int_{0}^{c^2} x^{k/2} dF_{\chi^2_1}(x).$$

Using repeatedly the recursive relationship for the χ^2 random variable

$$f_{\chi^2_{v+2}}(x) = \frac{x}{v} f_{\chi^2_v}(x), \tag{16}$$

where with symbol $f_{\chi^2_v}(x)$ we denote the density of a χ^2 random variable with v degrees of freedom, we obtain

$$\begin{aligned} v_c(k) &= \int_{0}^{c^2} x^{k/2-1} dF_{\chi^2_3}(x) \\ &= 3 \int_{0}^{c^2} x^{k/2-2} dF_{\chi^2_5}(x) \\ &= 5 \times 3 \int_{0}^{c^2} x^{k/2-3} dF_{\chi^2_7}(x) \\ &= \ldots \\ &= (k-1)(k-3)\ldots 1 \int_{0}^{c^2} dF_{\chi^2_{k+1}}(x) \\ &= (k-1)!!F_{\chi^2_{k+1}}(c^2). \end{aligned}$$

□

Remark 3.1 The result given in Appendix B of Riani et al. (2009), linking in an unexpected way the second moment of the truncated standard normal distribution with the χ^2_3, is a special case of Lemma 3.1.

Remark 3.2 If $c \to \infty$, then we recover in a simple way the well known relationship $E(X^k) = (k-1)!!$.

Remark 3.3 Given that

$$\begin{aligned}
v_c(2) &= F_{\chi^2_3}(c^2) = E(X^2)F_{\chi^2_3}(c^2)\\
v_c(4) &= 3F_{\chi^2_5}(c^2) = E(X^4)F_{\chi^2_5}(c^2)\\
v_c(6) &= 15F_{\chi^2_7}(c^2) = E(X^6)F_{\chi^2_7}(c^2)\\
\ldots &= \ldots\\
v_c(k) &= (k-1)!!F_{\chi^2_{k+1}}(c^2) = E(X^k)F_{\chi^2_{k+1}}(c^2),
\end{aligned}$$

Lemma 3.1 highlights the connection between consistency factors for robust estimation and the result anticipated by Tallis (1963): the moments of order k, for k even, when $X \sim N(\mu, \sigma^2)$ is truncated outside the interval $a < \frac{(x-\mu)^2}{\sigma^2} < b$ are those of the original distribution multiplied by the correction factor

$$\frac{F_{\chi^2_{k+1}}(b) - F_{\chi^2_{k+1}}(a)}{F_{\chi^2_1}(b) - F_{\chi^2_1}(a)}.$$

The quantity $v_c(k)$ is the numerator of the moment of the truncated distribution, i.e. the moment of the joint distribution of x and w, where $w = 1$ if $0 < \frac{(x-\mu)^2}{\sigma^2} < c^2$ and $w = 0$ otherwise.

Using the result of Lemma 3.1 it is immediate to obtain the following theorem.

Theorem 3.1 *The constant c linked to Tukey's Biweight ρ function which produces a given breakdown point (bdp), must satisfy the following constraint*

$$\text{bdp} = \frac{3}{c^2}\left\{{}_c\Psi_3 - 3\frac{{}_c\Psi_5}{c^2} + 5\frac{{}_c\Psi_7}{c^4} + \frac{c^2}{3}(1 - {}_c\Psi_1)\right\}, \tag{17}$$

where ${}_{[a,b]}\Psi_k = \Pr(\chi^2_k < b^2) - \Pr(\chi^2_k < a^2)$ *and* ${}_{[0,b]}\Psi_k = {}_b\Psi_k = \Pr(\chi^2_k < b^2)$.

Proof Using the expressions just found for $v_c(k)$ (Lemma 3.1), Eq. (14) can be rewritten as

$$6\left\{\frac{\Pr\left(\chi^2_3 < c^2\right)}{2} - 3\frac{\Pr\left(\chi^2_5 < c^2\right)}{2c^2} + 15\frac{\Pr\left(\chi^2_7 < c^2\right)}{6c^4} + \frac{c^2}{3}(1 - \Phi(c))\right\} = \alpha c^2. \tag{18}$$

Simplifying, the result follows immediately. □

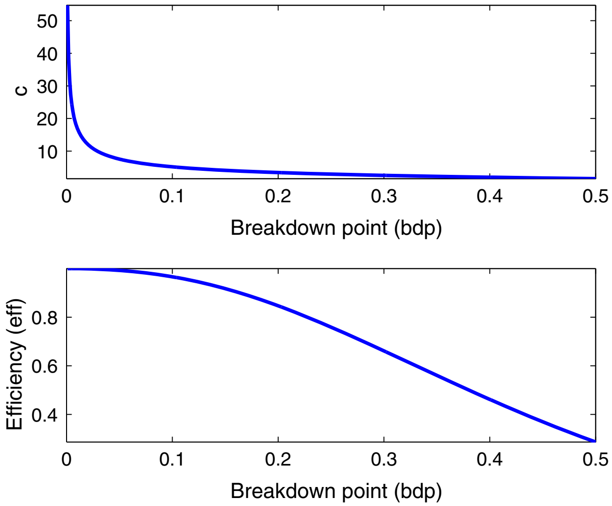

Fig. 1 Consistency factor c (*top panel*) and efficiency (*bottom panel*) as a function of the breakdown point (bdp) for Tukey's Biweight

Figure 1 shows the consistency factor c (top panel) and the efficiency (bottom panel), computed using Eq. (19), of the *S*-estimator as a function of the breakdown point, for Tukey's Biweight. Breakdown values close to 0.5 correspond to an efficiency smaller than 0.3. These results are in agreement with Hössjer (1992), who proved that an *S*-estimate with break down point equal to 0.5 has an asymptotic efficiency under normally distributed errors that is not larger than 0.33. In our opinion Fig. 1, and the others given in this paper, although feasible for a long time using the old formulae, have not received enough attention in the literature so far. For example, as is well known, results equal to traditional least squares are obtained when c tends to infinity which correspond to a 0 breakdown point (see lower panel). The slope of the curve at the bottom panel clearly shows that the transition from 0 to 0.1 in terms of breakdown point does not cause a significant decrease of nominal efficiency (which always remains above the 95 % threshold). On the other hand, an increase of breakdown point from, for example, 0.2–0.3 makes the efficiency decrease from 0.8467 to 0.6613. The maximum absolute loss of efficiency associated to an increase of 0.1 in breakdown point seems to be in the interval 0.3–0.4 (from 0.6613 to 0.4619 see Table 1). The new insight which comes from this figure is that it easily shows for which intervals, the change of breakdown point causes the most significant decrease in efficiency.

3.2 Consistency factors given a fixed efficiency

In this section, we give a simple expression to find c, if the user wants to fix the asymptotic efficiency of the robust *S*-estimator of the regression coefficients.

Theorem 3.2 *In order to obtain a fixed efficiency , in the case of Tukey's Biweight ρ function, the consistency factor c necessary to obtain a fixed efficiency (eff) for the location estimator must satisfy the following equality*

Table 1 Breakdown point, consistency factor and asymptotic efficiency at the normal model for Tukey's Biweight loss function in regression

Breakdown point (bdp)	Consistency factor (c)	Asymptotic efficiency at the normal model (eff)
0.05	7.5453	0.9924
0.10	5.1824	0.9662
0.20	3.4207	0.8467
0.25	2.9370	0.7590
0.30	2.5608	0.6613
0.40	1.9880	0.4619
0.50	1.5476	0.2868

$$\text{eff} = \frac{\left\{15\frac{{}_c\Psi_5}{c^4} - 6\frac{{}_c\Psi_3}{c^2} + {}_c\Psi_1\right\}^2}{945\frac{{}_c\Psi_{11}}{c^8} - 420\frac{{}_c\Psi_9}{c^6} + 90\frac{{}_c\Psi_7}{c^4} - 12\frac{{}_c\Psi_5}{c^2} + {}_c\Psi_3}. \tag{19}$$

Proof The asymptotic efficiency (eff) of the S-estimator at the Gaussian model is

$$\text{eff} = \frac{\left\{\int_{-c}^{c} \psi'(x) d\Phi(x)\right\}^2}{\int_{-c}^{c} \psi^2(x) d\Phi(x)};$$

see, e.g., Rousseeuw and Leroy (1987, p. 142). In the case of the Tukey's Biweight ρ function, we easily obtain

$$\{\psi(x)\}^2 = \begin{cases} \frac{1}{c^8}x^{10} - \frac{4}{c^6}x^8 + \frac{6}{c^4}x^6 - \frac{4}{c^2}x^4 + x^2 & \text{if } |x| \le c \\ 0 & \text{if } |x| > c, \end{cases} \tag{20}$$

$$\psi'(x) = \begin{cases} \frac{5}{c^4}x^4 - \frac{6}{c^2}x^2 + 1 & \text{if } |x| \le c \\ 0 & \text{if } |x| > c. \end{cases} \tag{21}$$

Hence,

$$\int_{-c}^{c} \psi'(x) d\Phi(x) = 15\frac{\Pr\left(\chi_5^2 < c^2\right)}{c^4} - 6\frac{\Pr\left(\chi_3^2 < c^2\right)}{c^2} + \Pr\left(\chi_1^2 < c^2\right). \tag{22}$$

Similarly, we obtain for $\{\psi(x)\}^2$

$$\begin{aligned} \int_{-c}^{c} \{\psi(x)\}^2 d\Phi(x) &= 9!!\frac{\Pr\left(\chi_{11}^2 < c^2\right)}{c^8} - 4\times 7!!\frac{\Pr\left(\chi_9^2 < c^2\right)}{c^6} + 6\times 5!!\frac{\Pr\left(\chi_7^2 < c^2\right)}{c^4} \\ &\quad - 4 \times 3!!\frac{\Pr\left(\chi_5^2 < c^2\right)}{c^2} + \Pr\left(\chi_3^2 < c^2\right). \end{aligned} \tag{23}$$

Taking the square of Eq. (22) and dividing by Eq. (23), the result follows. $\square$

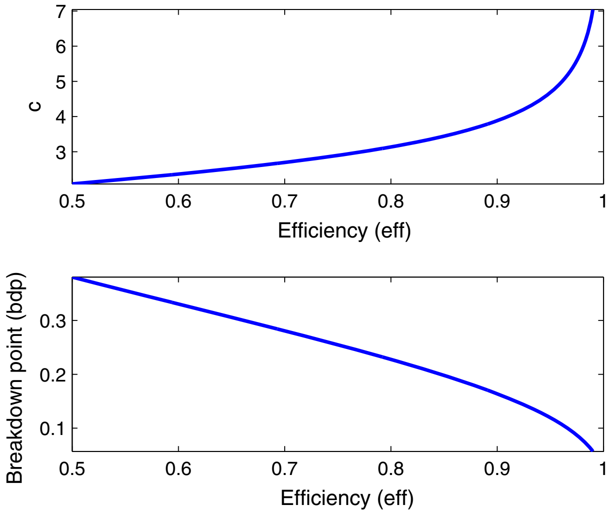

Fig. 2 Consistency factor c (*top panel*) and breakdown point (*bottom panel*) as a function of the efficiency (eff) for Tukey's Biweight

Table 2 Efficiency, consistency factor and breakdown point at the normal model for Tukey's Biweight loss function in regression

Asymptotic efficiency at the normal model (eff)	Consistency factor (c)	Breakdown point (bdp)
0.50	2.0871	0.3804
0.60	2.3666	0.3304
0.70	2.6972	0.2806
0.75	2.8972	0.2548
0.80	3.1369	0.2276
0.85	3.4437	0.1980
0.90	3.8827	0.1638
0.95	4.6851	0.1194
0.99	7.0414	0.0570

Theorem 3.2 gives a simple and efficient way to compute the efficiency values reported in Rousseeuw and Leroy (1987, p. 142). It also provides a straightforward way to generalize these results to ρ functions of powers other than Tukey's Biweight (such as the optimal loss function which will be considered in Sect. 7).

Figure 2 shows the consistency factor c (top panel) and the breakdown point (bottom panel) as a function of the efficiency. This plot and Table 2 clearly show that values of the efficiency greater than 0.9 lead to a considerable decrease in the breakdown point.

4 Consistency factors for multivariate estimators of location

In this section, we extend our results to multivariate analysis. We first introduce the notation, starting with M-estimators and then considering their S extension as in Sect. 2.

We assume to have an i.i.d. sample of observations $y_1, \ldots, y_n$ with each $y_i' = (y_{i1}, \ldots, y_{iv})$ having a v-variate normal distribution $N_v(\mu, \Sigma)$. The squared Mahalanobis distance between the vectors x and μ with respect to Σ is defined as

$$d^2(y; \mu, \Sigma) = (y - \mu)'\Sigma^{-1}(y - \mu). \tag{24}$$

Multivariate M-estimates are defined as the solution of the following system of equations (see, for example, Kent and Tyler 1991):

$$\sum_{i=1}^{n} W_1\left(d_i^2\right)\left(y_i - \hat{\mu}\right) = 0 \tag{25}$$

$$\frac{1}{n}\sum_{i=1}^{n} W_2\left(d_i^2\right)\left(y_i - \hat{\mu}\right)\left(y_i - \hat{\mu}\right)' = \hat{\Sigma}, \tag{26}$$

where W_1 and W_2 are real-valued functions on $[0, \infty)$. The above system of equations is known as the first-order M-estimator conditions.

Just as with regression estimates, where we aimed at making residuals small, it is possible to define estimates of location and dispersion that make the distances d_i, with $i = 1, \ldots, n$, small. To this end we look for estimates $\hat{\mu}$ and $\hat{\Sigma}$ which minimize some measure of largeness based on the vector of Mahalanobis distances. The constraint that $|\hat{\Sigma}| = 1$ is imposed to avoid trivial solutions. More formally, the problem can be reformulated in terms of minimizing a robust estimate of the scale of the distances (say $\hat{\sigma}$) as follows

$$\hat{\sigma}\left(d(y_1), d(y_2), \ldots, d(y_n)\right) = \min \quad \text{with} \quad \hat{\mu} \in \Re^v, \hat{\Sigma} \in S_v, |\hat{\Sigma}| = 1, \tag{27}$$

where S_v denotes the set of symmetric positive definite $v \times v$ matrices, and $\hat{\sigma}$ denotes a robust scale estimate. Call this minimization problem $\wp_n$. It is well known that if $\hat{\sigma}$ is the sample median we end up with the Minimum Volume Ellipsoid estimator (Rousseeuw and Leroy 1987). On the other hand, if $\hat{\sigma}$ is an M-estimate of scale that satisfies

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{d_i}{\hat{\sigma}}\right) = K, \tag{28}$$

we obtain the class of multivariate S-estimates (Davies 1987). It is possible to show (Lopuhaä 1989) that if ρ is differentiable the solution to the minimization problem $\wp_n$ also satisfies the first-order M-estimator conditions (25) and (26), and produces estimators which have the same asymptotic distribution of M-estimators.

As in the univariate case, the equations to compute the consistency factor for the location estimator $\hat{\mu}$, associated with a prefixed breakdown point or efficiency, are functions of the truncated normal moments. Thus, the first step consists in finding easy expressions to compute the multivariate truncated moments.

4.1 Consistency factors given a fixed breakdown point

In order to prove the main theorem of this section we need to recall the definition of central moments and we need a lemma.

Definition 1 The central moment of order k of the v dimensional random vector X is defined as (see for example Kotz et al. 2000, pp. 107–111)

$$\mu_{k_1,k_2,\ldots,k_v}(X) = E\left[(X_1-\mu_1)^{k_1}(X_2-\mu_2)^{k_2}\ldots(X_v-\mu_v)^{k_v}\right],$$

where $k_1+k_2+\cdots+k_v=k$ and $k_j \geq 0$. Simplifying the notation, we can rewrite this expression as

$$\mu_{k_1,k_2,\ldots,k_v}(X) = \mu_{1,2,\ldots,v}(X) = E\left[\prod_{j=1}^{v}(X_j-\mu_j)^{k_j}\right]$$

Lemma 4.1 *If $X = (X_1, \ldots, X_v)'$ follows a v-variate normal distribution with mean μ and covariance matrix $\sigma^2 I_v$, then the kth order central moment $\mu_{k_1,k_2,\ldots,k_v}(X)$, where k is an even number, is given by*

$$\sigma^k \frac{(v+k-2)!!}{(v-2)!!} = \sigma^k \prod_{j=0}^{k/2-1}(v+2j), \tag{29}$$

where $s!! \equiv 1$ if $s \leq 0$.

Notice that when $v=1$ the above formula reduces to $\sigma^k(k-1)!!$, which is consistent with the expression seen in the previous section. If $k=4$ the fourth central moment reduces to $\sigma^4 v(v+2)$, while if $k=6$ we obtain $\sigma^6 v(v+2)(v+4)$. The expression in Eq. (29) is intrinsically linked to the expression of noncentral moments given by Holmquist (1988) and Triantafyllopoulos (2003). In what follows we provide a simple and direct proof.

Proof If X is multivariate normal, due to Isserlis theorem (Isserlis 1918), the kth order central moment of X is zero if k is odd, while if k is even we obtain

$$\mu_{1,2,\ldots,v}(X) = \sum(\sigma_{ij}\sigma_{kl}\ldots\sigma_{yz}),$$

where for example σ_{ij} is the covariance of X_i and X_j and the sum is taken over all allocations of the set $\{1, 2, \ldots, k\}$ into $k/2$ unordered pairs. For example, in the fourth moment case, given that we are dealing with a spherical normal distribution with uncorrelated components, the terms which are not zeros are the v terms involving

$$E\left(X_i^4\right) = 3\sigma^4$$

and the $v(v-1)$ terms involving

$$E\left(X_i^2 X_j^2\right) = \sigma^4.$$

The fourth-order moment is given by

$$v3\sigma^4 + v(v-1)\sigma^4 = v(v+2)\sigma^4.$$

Proceeding in a similar way for the sixth moment, we obtain that there are v terms involving

$$E\left(X_i^6\right) = 15\sigma^6,$$

$v(v-1)$ terms involving the 3 permutations of the fourth- and second-order terms

$$E\left(X_i^4 X_j^2\right) = 3\sigma^6,$$

and $v(v-1)(v-2)$ terms involving

$$E\left(X_i^2 X_j^2 X_k^2\right) = \sigma^6.$$

Summing up the various terms we obtain

$$\sigma^6 \left\{15v + 9v(v-1) + v(v-1)(v-2)\right\} = \sigma^6 v(v+2)(v+4).$$

Therefore, the result easily follows by induction. □

At present (see for example the source code in the R library `robustbase`) the truncated central moment of order k (with k even) of the multivariate standard normal distribution of order v is computed using the expression which follows

$$\mu_{k_1,k_2,\ldots,k_v}(X) = \frac{\Gamma\left(\frac{v+k}{2}\right)}{\Gamma(v/2)} 2^{k/2} P\left(\frac{c^2}{2}, \frac{v+k}{2}\right), \tag{30}$$

where P and Γ are defined in Sect. 3.1.

In the lemma below we give a simpler expression which avoids the use of the two Gamma functions in (30). However, we preliminary need to define the concept of elliptical truncation under the multivariate normal distribution.

Definition 2 If $X \sim N_v(\mu, \Sigma)$, we define the elliptical truncation in the interval $[a\ b]$ the set of points in the v space belonging to E where

$$E = \left\{x | a \leq (x-\mu)'\Sigma^{-1}(x-\mu) \leq b\right\}, \quad 0 \leq a < b$$

Lemma 4.2 *The central moment of order k under elliptical truncation in the interval* $[0\ c^2]$ *satisfies the following identity*

$$\mu_{k_1,k_2,\ldots,k_v}(X) = \left\{\prod_{j=0}^{k/2-1}(v+2j)\right\} F_{\chi^2_{v+k}}(c^2). \tag{31}$$

The right-hand side of Eq. (31) extends in a natural way the result previously obtained for the univariate case [see Eq. (15)]. This result, which is proven in the Appendix, shows that in the v-variate normal distribution $N_v(\mu, \Sigma)$ the central moments of order k under elliptical truncation in the interval $[a\ b]$ are equal to the original moments multiplied by $F_{\chi^2_{v+k}}(b)-F_{\chi^2_{v+k}}(a)$. In our case $a=0, b=c^2$ and the original moments are given in Eq. (29). This result provides a simple alternative proof to the one given by Tallis (1963), which was based on the moment generating function of the multivariate normal distribution.

Using the results of the two previous lemmas, the following theorem follows immediately.

Theorem 4.1 *The constant c in Tukey's Biweight* ρ *function which produces a given breakdown point (bdp), in presence of v variables, must satisfy the following constraint*

$$\begin{aligned}\mathrm{bdp} = (3v/c^2)\Bigg\{ {}_c\Psi_{v+2} - (v+2)\frac{{}_c\Psi_{v+4}}{c^2} \\ +(v+2)(v+4)\frac{{}_c\Psi_{v+6}}{3c^4} + \frac{c^2}{3}(1-{}_c\Psi_v)/v\Bigg\},\end{aligned} \tag{32}$$

where ${}_c\Psi_k$ *is defined in (17).*

A result similar to that given in Eq. (32) has been obtained by Campbell (1984) using the theory of transformations. Our proof is much simpler and direct.

4.2 Consistency factors given a fixed efficiency

We now give the details about how to compute the consistency factor if the user decides to fix the level of asymptotic efficiency.

Theorem 4.2 *In presence of v variables, the consistency factor c associated to a fixed value of efficiency (eff) for Tukey's Biweight* ρ *function must satisfy the following equality*

$$\mathrm{eff} = \frac{\left(v_{1,2}\frac{{}_c\Psi_{v+4}}{c^4} - 2v_{1,1}\frac{{}_c\Psi_{v+2}}{c^2} + {}_c\Psi_v\right)^2}{v_{1,4}\frac{{}_c\Psi_{v+10}}{c^8} - 4v_{1,3}\frac{{}_c\Psi_{v+8}}{c^6} + 6v_{1,2}\frac{{}_c\Psi_{v+6}}{c^4} - 4v_{1,1}\frac{{}_c\Psi_{v+4}}{c^2} + {}_c\Psi_{v+2}}, \tag{33}$$

where $v_{i,k} = \prod_{j=i}^{k}(v+2j)$.

Proof We start by recalling that when $\hat{\theta} = (\hat{\mu}, \hat{\Sigma}_n)$ is a solution of the minimization problem $\wp_n$ [given in Eq. (27)] and the data come from an elliptical distribution with parameter $\theta = (\mu\ \Sigma)$, $\sqrt{n}(\hat{\theta} - \theta)$ has a limiting normal distribution. The covariance matrix of the limiting distribution of $\sqrt{n}(\hat{\mu} - \mu)$ is given by $(w_1/w_2^2)\Sigma$, where

$$w_1 = \frac{1}{v} E\left\{\psi^2(d)\right\}$$

and

$$w_2 = E\left\{\left(1 - \frac{1}{v}\right)\frac{\psi(d)}{d} + \frac{1}{v}\psi'(d)\right\}$$

(see for example Lopuhaä 1989). In the case of Tukey's Biweight function, applying Lemma 4.2 and using expression (20) we find after some algebra that

$$\begin{aligned} w_1 = \Bigg[& \left\{\prod_{j=0}^{4}(v+2j)\right\}\frac{\Pr\left(\chi^2_{v+10} < c^2\right)}{c^8} - 4\left\{\prod_{j=0}^{3}(v+2j)\right\}\frac{\Pr\left(\chi^2_{v+8} < c^2\right)}{c^6} \\ & + 6\left\{\prod_{j=0}^{2}(v+2j)\right\}\frac{\Pr\left(\chi^2_{v+6} < c^2\right)}{c^4} - 4v(v+2)\frac{\Pr\left(\chi^2_{v+4} < c^2\right)}{c^2} \\ & + v\Pr\left(\chi^2_{v+2} < c^2\right)\Bigg]/v \end{aligned}$$

and

$$\begin{aligned} w_2 = \Bigg\{ & v(v+2)(v+4)\frac{\Pr\left(\chi^2_{v+4} < c^2\right)}{c^4} \\ & -2v(v+2)\frac{\Pr\left(\chi^2_{v+2} < c^2\right)}{c^2} + v\Pr\left(\chi^2_{v} < c^2\right)\Bigg\}/v. \end{aligned}$$

It is now immediately seen that the efficiency $\frac{w_2^2}{w_1}$ is as claimed. □

Figure 3 shows that the efficiency becomes virtually independent of the breakdown point as the number of variables increases. For instance, the maximal breakdown 0.5 is associated with an efficiency close to 0.85 when $v = 5$, but the corresponding efficiency becomes close to 1 when v gets larger. This behavior is well known in the literature (see for example Rocke 1996; Maronna et al. 2006, p. 190), but the advantage of our procedure is that such comparisons can be easily made for any value of bdp, eff and v.

The effect of v on the breakdown point is made explicit in Fig. 4. The plot starts at $v = 5$, for which we pick up the value of c giving a 50 % breakdown point, say $c_{0.5}$.

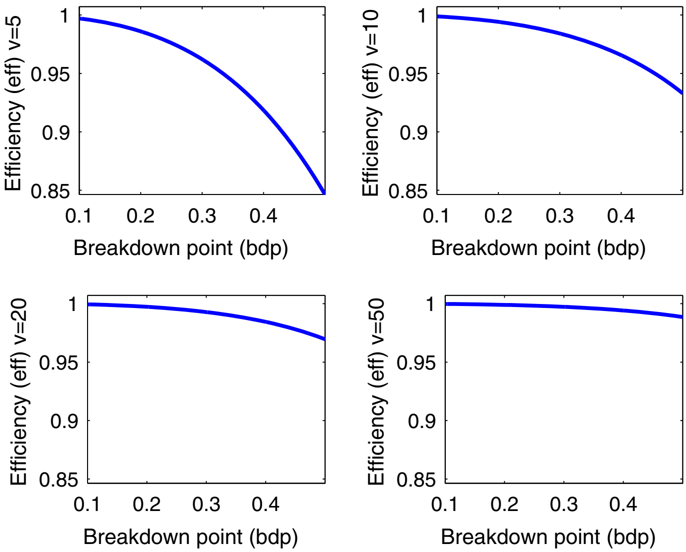

Fig. 3 Efficiency as a function of the breakdown point and of the number of variables for Tukey's Biweight: $v = 5$ (*top left panel*), $v = 10$ (*top right panel*), $v = 20$ (*bottom left panel*) and $v = 50$ (*bottom right panel*)

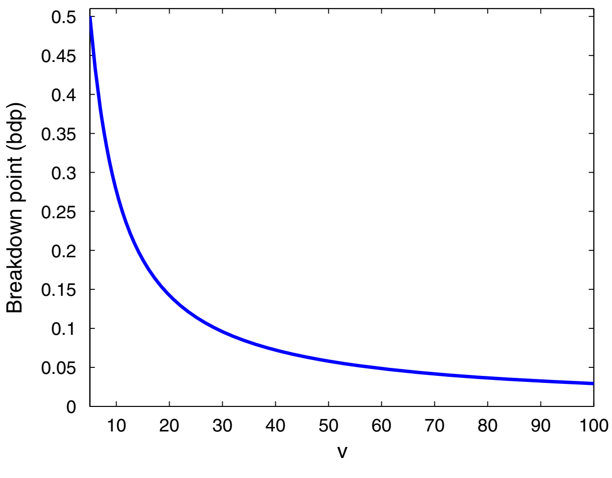

Fig. 4 Relationship between breakdown and v for Tukey's Biweight, when the false rejection probability is held constant

We then compute

$$\gamma = \Pr\left(\chi_5^2 > c_{0.5}\right),$$

which is the probability that a five-variate observation with squared Mahalanobis distance (24) is given null weight under Tukey's Biweight ρ function at the normal model. We see how c varies when v increases, by computing the quantile

$$c(v)_{0.5} = \chi^2_{v,1-\gamma}.$$

This quantile shows the effect of v on threshold c in (11), by keeping γ fixed. Figure 4 displays the values of the breakdown point corresponding to $c(v)_{0.5}$ for $v \geq 5$. The curve starts at 0.5 by construction, but then decreases rapidly. This means that the robustness properties of the S-estimator soon deteriorate in higher dimension, if the false rejection probability γ is kept constant.

5 Scale efficiency in multivariate analysis

Up to now we have dealt with location efficiency. The purpose of this section is to derive the expression of the constant c associated to a prefixed value of the scale efficiency.

Theorem 5.1 *If $X = (X_1, \ldots, X_v)'$ follows a v-variate normal distribution with mean μ and covariance matrix $\sigma^2 I_v$, the consistency factor c for Tukey's Biweight function associated to a fixed value of efficiency (eff_{sc}) of the estimator of the scale factor σ^2 must satisfy the following equality*

$$\text{eff}_{\text{sc}} = \left(\frac{v-1}{v} k_1 + IF2 \right)^{-1}, \tag{34}$$

where

$$k_1 = \frac{v_{2,5} \frac{{}_c\Psi_{v+12}}{c^8} - 4v_{2,4} \frac{{}_c\Psi_{v+10}}{c^6} + 6v_{2,3} \frac{{}_c\Psi_{v+8}}{c^4} - 4v_{2,2} \frac{{}_c\Psi_{v+6}}{c^2} + {}_c\Psi_{v+4}}{\left(v_{2,3} \frac{{}_c\Psi_{v+6}}{c^4} - 2v_{2,2} \frac{{}_c\Psi_{v+4}}{c^2} + \Psi_{v+2} \right)^2}, \tag{35}$$

$$IF2 = \frac{A - B^2}{\left(\frac{v_{0,0}}{2} {}_c\Psi_{v+2} - \frac{v_{0,1}}{2c^2} \Psi_{v+4} + \frac{v_{0,2}}{6c^4} \Psi_{v+6} \right)^2},$$

and

$$A = \frac{v_{0,1}}{4} {}_c\Psi_{v+4} - \frac{v_{0,2}}{2c^2} {}_c\Psi_{v+6} + \frac{5v_{0,3}}{12c^4} {}_c\Psi_{v+8} - \frac{v_{0,4}}{6c^6} {}_c\Psi_{v+10} + \frac{v_{0,5}}{36c^8} {}_c\Psi_{v+12},$$

$$B = \frac{v_{0,0}}{2} {}_c\Psi_{v+2} - \frac{v_{0,1}}{2c^2} {}_c\Psi_{v+4} + \frac{v_{0,3}}{6c^4} {}_c\Psi_{v+6} + \frac{c^2}{6} (1 - {}_c\Psi_v).$$

Proof Under the same assumptions of the proof of Theorem 4.2, the variance of the limiting distribution of $\sqrt{n}(\hat{\sigma}^2 - \sigma^2)$ where σ^2 is a generic element in the diagonal of the covariance matrix of the underlying data generation process (which due to the affine equivariance property we assume spherical), and $\hat{\sigma}^2$ is the associated S-estimator, is given by (see for example Lopuhaä 1989; eq. 5.4)

$$(2k_1 + k_2)\sigma^4,$$

where

$$k_1 = \frac{v(v+2)E[\psi^2(d)d^2]}{\left\{E\left[\psi'(d)d^2 + (v+1)\psi(d)d\right]\right\}^2}$$

and

$$k_2 = -\frac{2}{v}k_1 + \frac{4E\{\rho(d) - K\}^2}{E\{\psi(d)d\}}.$$

For Tukey's Biweight we find that

$$E\left\{\psi'(d)d^2 + (v+1)\psi(d)d\right\} = \left\{v(v+2)\Pr\left(\chi^2_{v+2} < c^2\right)\right.$$
$$\left. -2v(v+2)(v+4)\frac{\Pr\left(\chi^2_{v+4} < c^2\right)}{c^2} + v(v+2)(v+4)(v+6)\frac{\Pr\left(\chi^2_{v+6} < c^2\right)}{c^4}\right\},$$

while

$$E\left\{\psi^2(d)d^2\right\} = v(v+2)\Pr\left(\chi^2_{v+4} < c^2\right) - 4\left\{\prod_{j=0}^{2}(v+2j)\right\}\frac{\Pr\left(\chi^2_{v+6} < c^2\right)}{c^2}$$
$$+6\left\{\prod_{j=0}^{3}(v+2j)\right\}\frac{\Pr\left(\chi^2_{v+8} < c^2\right)}{c^4} - 4\left\{\prod_{j=0}^{4}(v+2j)\right\}\frac{\Pr\left(\chi^2_{v+10} < c^2\right)}{c^6}$$
$$+\left\{\prod_{j=0}^{5}(v+2j)\right\}\frac{\Pr\left(\chi^2_{v+12} < c^2\right)}{c^8},$$

$$E\left\{\rho^2(d)\right\} = v(v+2)\Pr\left(\chi^2_{v+4} < c^2\right) - \left\{\prod_{j=0}^{2}(v+2j)\right\}\frac{\Pr\left(\chi^2_{v+6} < c^2\right)}{2c^2}$$
$$+\frac{5}{12}\left\{\prod_{j=0}^{3}(v+2j)\right\}\frac{\Pr\left(\chi^2_{v+8} < c^2\right)}{c^4} - \left\{\prod_{j=0}^{4}(v+2j)\right\}\frac{\Pr\left(\chi^2_{v+10} < c^2\right)}{6c^6}$$
$$+\left\{\prod_{j=0}^{5}(v+2j)\right\}\frac{\Pr\left(\chi^2_{v+12} < c^2\right)}{36c^8}$$

and

$$E\{\rho(d)\} = K = \frac{v}{2}\Pr\left(\chi^2_{v+2} < c^2\right) - \frac{v(v+2)}{2c^2}\Pr\left(\chi^2_{v+4} < c^2\right)$$
$$+\frac{v(v+2)(v+4)}{6c^4}\Pr\left(\chi^2_{v+6} < c^2\right) + \frac{c^2}{6}\left\{1 - \Pr\left(\chi^2_{v} < c^2\right)\right\}.$$

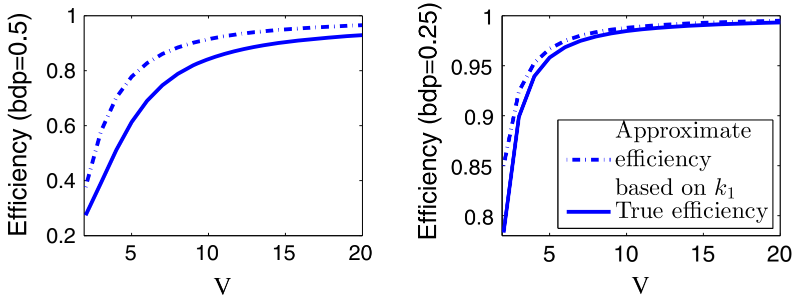

Fig. 5 True and approximate efficiency (based on k_1) as a function of v for a value of the consistency factor c associated to a breakdown point of 0.5 (*left panel*) and 0.25 (*right panel*)

Putting all pieces together, after some algebra we obtain the expression claimed in the theorem. □

In the case of $v \geq 2$, Tyler (1983) found after a simulation study that k_1 suffices as an index for the asymptotic variance based upon the robust covariance estimator. At present in the R library `robustbase` the efficiency of the scale estimator is computed as $1/k_1$. Our simple formulae enable us to immediately check under which conditions we can neglect the term $IF2$ (which is the expectation of the square influence function of the robust scale estimator) in Eq. (34). Figure 5 shows that if we fix the breakdown point to 0.5, the value of the curve associated to the true efficiency is much higher than the one based on k_1. As expected, the difference slowly decreases as v increases. The accuracy of the approximation increases if we decrease the value of the breakdown point (see the right panel).

6 Computational details

All the new formulae to compute the consistency factors, given a prefixed value of breakdown point or efficiency, have been included in the MATLAB toolbox named `FSDA` (see Riani et al. 2012). This toolbox can be freely downloaded from the Internet web address http://www.riani.it/MATLAB.

Given that in all cases the consistency factor has to be calculated through an iterative process, it is customary to use an iterative loop. In our implementation, we have used a searching algorithm which produces the required output in a much faster way. We also have removed, inside the loop, all redundant recalculations of factors which do not depend on c. It is worthwhile to notice that all terms $v_{i,k} = \prod_{j=i}^{k}(v + 2j)$ can be computed once and for all, and are fixed in the iterative procedure. Just to give an idea of the difference in speed between the old implementation and the new one, we show in Fig. 6 on the y axis the amount of seconds it takes to compute the consistency factors for the prefixed values of breakdown point in the sequence $0.01, 0.02, \ldots, 0.5$ as a function of the number of variables (v). The dashed line at the bottom is referred to the new implementation while the solid line is referred to the old procedure downloadable

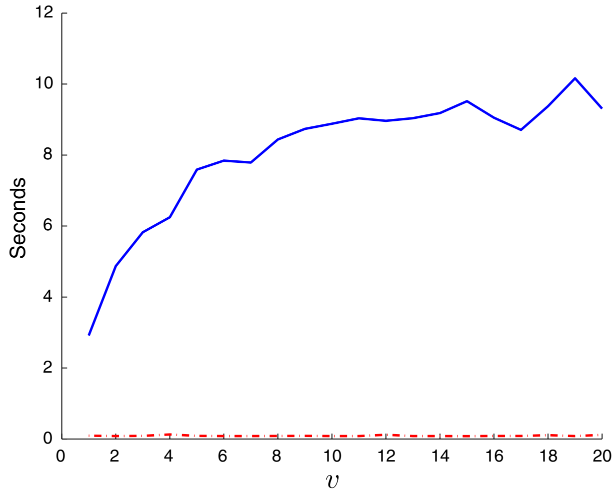

Fig. 6 Comparison of computational time (y axis) required to compute the consistency factors c for breakdown points in the sequence 0.01, 0.02, …, 0.5 for each value of v (x axis)

from Internet.[1] A referee showed that much of the difference of the computational time between the two implementations is due to the fact that the old procedure inefficiently uses calls to the gamma functions in each step of the loop and that considerable gain in speed could be obtained if these calls are brought out of it. The new formulae contained in this paper, however, completely avoid the call to the gamma function. This figure shows that the new computing time is always around 0.2 s, while the old one goes from 3 to 10 s as v reaches 20. It is also worthwhile to notice that the comparison between the computational times has been made just for the calculation of the consistency factor associated to the breakdown point (that is the easiest formulae). The computational advantage therefore is likely to be even more significant if we use the more complicated formulae which refer to the efficiency of location and scale given in Eqs. (19) and (34).

7 Extensions to the optimal loss function

So far we have dealt with the Tukey's bisquare ρ and ψ functions. However, it is important to notice that our method of proof is general and could be applied to any definition of the loss function based on polynomials. For example in case of the optimal

[1] The computing time of the old procedure is based on the MATLAB code which can be downloaded from http://www.econ.kuleuven.be/public/ndbae06/Programs/mme/MMrse.txt and uses subfunctions `Tbsc`, `Tbsb` and `gint`. The new code is based on function `TBbdp` which implements the new procedure and is contained in the FSDA toolbox. In the FSDA toolbox routine XXXbdp computes the value of c given as input the required breakdown point and the number of variables; routine XXXeff computes the value of c given as input the requested efficiency and the number of variables; routine XXXc returns the breakdown point and the efficiency associated to a particular value of c. Suffix XXX stands for the type of loss function.

loss function introduced by Yohai and Zamar (1997) we have that the (standardized) ρ function is given by:

$$\rho(x) = \begin{cases} 1.3846 \left(\frac{x}{c}\right)^2 & |x| \leq \frac{2}{3}c \\ 0.5514 - 2.6917 \left(\frac{x}{c}\right)^2 + 10.7668 \left(\frac{x}{c}\right)^4 - 11.6640 \left(\frac{x}{c}\right)^6 + 4.0375 \left(\frac{x}{c}\right)^8 & \frac{2}{3}c < |x| \leq c \\ 1 & |x| > c \end{cases} \tag{36}$$

with associated ψ function:

$$\psi(x) = \rho'(x) = \begin{cases} \frac{2.7692x}{c^2} & |x| \leq \frac{2}{3}c \\ -\frac{5.3834x}{c^2} + \frac{43.0672x^3}{c^4} - \frac{69.9840x^5}{c^6} + \frac{32.3x^7}{c^8} & \frac{2}{3}c < |x| \leq c \\ 0 & |x| > c. \end{cases} \tag{37}$$

Using the results in Lemmas 4.1 and 4.2 we can easily prove the following two theorems:

Theorem 7.1 *The constant c in the optimal ρ function which produces a given breakdown point (bdp), in presence of v variables, must satisfy the following constraint*

$$\begin{aligned} \text{bdp} = {} & 1.3846 \frac{{}_{\frac{2}{3}c}\Psi_{v+2}}{c^2} + 0.5514 \frac{{}_{\left[\frac{2}{3}c,c\right]}\Psi_v}{c^2} + -2.6917 v \frac{{}_{\left[\frac{2}{3}c,c\right]}\Psi_{v+2}}{c^2} \\ & + 10.7668 v_{0,1} \frac{{}_{\left[\frac{2}{3}c,c\right]}\Psi_{v+4}}{c^4} - 11.664 v_{0,2} \frac{{}_{\left[\frac{2}{3}c,c\right]}\Psi_{v+6}}{c^6} \\ & + 4.0375 v_{0,3} \frac{{}_{\left[\frac{2}{3}c,c\right]}\Psi_{v+8}}{c^8} + 1 - {}_c\Psi_v. \end{aligned}$$

Theorem 7.2 *In presence of v variables, the consistency factor c associated to a fixed value of efficiency (eff) for optimal ρ function must satisfy the following equality*

$$\text{eff} = \frac{w_1}{w_2}, \tag{38}$$

where

$$\begin{aligned} w_1 = {} & \frac{7.6685\, {}_{\frac{2}{3}c}\Psi_{v+2} + 28.9810\, {}_{\left[\frac{2}{3}c,c\right]}\Psi_{v+4}}{c^4} + 2608.2874(v_{1,2} + v_{1,1}) \frac{{}_{\left[\frac{2}{3}c,c\right]}\Psi_{v+6}}{c^8} \\ & + 7679.9013 v_{1,4} \frac{{}_{\left[\frac{2}{3}c,c\right]}\Psi_{v+10}}{c^{12}} + 1043.29 v_{1,6} \frac{{}_{\left[\frac{2}{3}c,c\right]}\Psi_{v+14}}{c^{16}} \end{aligned}$$

$$-463.6959 v_{1,1} \frac{\left[\frac{2}{3}c, c\right]^{\Psi_{v+4}}}{c^6} - 6375.7975 v_{1,3} \frac{\left[\frac{2}{3}c, c\right]^{\Psi_{v+6}}}{c^{10}}$$
$$-4520.9663 v_{1,5} \frac{\left[\frac{2}{3}c, c\right]^{\Psi_{v+6}}}{c^{14}}$$

and

$$w_2 = \frac{2.7692\,{}_{\frac{2}{3}c}\Psi_v - 5.3834_{\left[\frac{2}{3}c,c\right]}\Psi_v}{c^2} + 43.0672 v_{1,1} \frac{\left[\frac{2}{3}c, c\right]^{\Psi_{v+2}}}{c^4}$$
$$- 69.984 v_{1,2} \frac{\left[\frac{2}{3}c, c\right]^{\Psi_{v+4}}}{c^6} - 32.3 v_{1,3} \frac{\left[\frac{2}{3}c, c\right]^{\Psi_{v+6}}}{c^8}.$$

Yohai and Zamar (1997) showed that the ρ and ψ functions given above are optimal in the following highly desirable sense: the final *M*-estimate has a breakdown point of one-half, and minimizes the maximum bias under contamination distributions (locally for small fraction of contamination), subject to achieving a desired efficiency when the data are Gaussian. Figure 7 compares the efficiency of Tukey's Biweight loss function

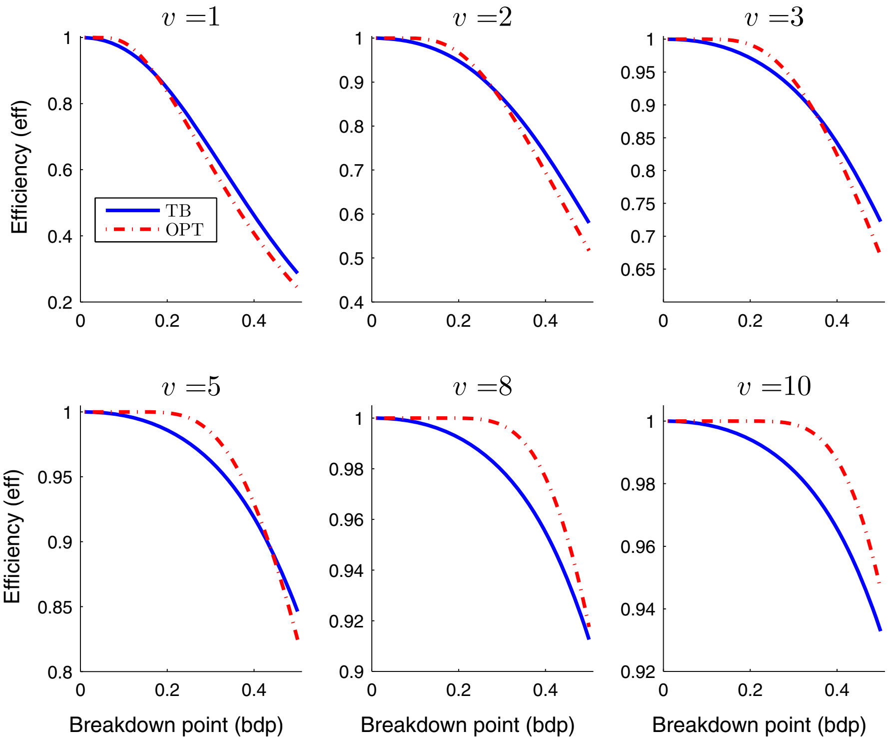

Fig. 7 Comparison of efficiency of Tukey's Biweight loss function (*solid line*) with optimal loss function (*dashed dotted line*) for six different values of v

(solid line) with that of the optimal loss function (dashed dotted line) for six different values of v. This plot shows that none of the two curves is uniformly above the other. In presence of large values of breakdown point and moderate values of v (say $v \leq 5$) the efficiency of Tukey's loss function is above that of the optimal loss function. However, for moderate values of breakdown point 0.1–0.3 and $v > 1$, the optimal loss function is above that of the Tukey's Biweight. Finally, it is important to notice that the optimal loss function when v is very large suffers from the same drawbacks of the Tukey's bisquare loss function discussed in Sect. 4.2, that is becomes virtually independent of the breakdown point. This effect seems even more pronounced for the optimal loss function.

8 A numerical example

The paper focuses on *S*-estimators, for which there is a trade-off between breakdown point and efficiency. However, the formulae to tune the efficiency of the *S*-estimators can easily be extended to *MM*-estimators, which are believed not to suffer from this trade-off. For example, *MM*-estimators for regression are obtained as

$$\hat{\beta}_{MM} = \arg \min_{\beta \in \Re^p} \sum_{i=1}^{n} \rho\left(\frac{r_i}{s}\right), \tag{39}$$

where s is precalculated and kept fixed in each iteration (Yohai 1987; Maronna et al. 2006). Generally s is the estimate of the scale which is obtained by *S*-estimators with bdp = 0.5, and the initial β to start the iteration is $\hat{\beta}_S$ obtained from Eq. (8). In the loop above the efficiency parameter eff associated to the ρ function is generally fixed to 0.90 or 0.95.

An important advantage of our new formulae is that they provide the user with an easy way to compute the consistency factor for different levels of breakdown point and efficiency, thus opening the way to much wider comparisons that would be problematic otherwise. The need to repeatedly compute robust estimators for different values of bdp and eff, to find the best solution, has become a prominent task in robust statistics and has lead to the development of new fast algorithms (Hubert et al. 2012; Riani et al. 2013) and to tests to balance the trade-off between efficiency and robustness using *MM*-estimators (Dehon et al. 2012). In this section we apply our results to a well-known data set to compare the findings that are obtained when we consider different values of bdp (for *S*-estimators), and different values of eff (for *MM*-estimators) combined with two different loss functions in a regression context.

Before dealing with the data it is interesting to compare the loss ($\rho(x)$) and weight ($\psi(x)/x$) functions associated to a nominal breakdown point of 50 % or a nominal efficiency of 95 % for Tukey Biweight and optimal. Figure 8 shows that fixing a value of breakdown point equal to 0.5 leads to a ρ function which has a strong upward slope as soon as we move away from 0 (see solid line in the left panels). Similarly, the corresponding weight functions starts a strong decrease when $|x|$ is greater than 1.5. On the other hand, if we fix the efficiency to a high value (and implicitly to a low break down point) we end up with a ρ function which increases at a much lower

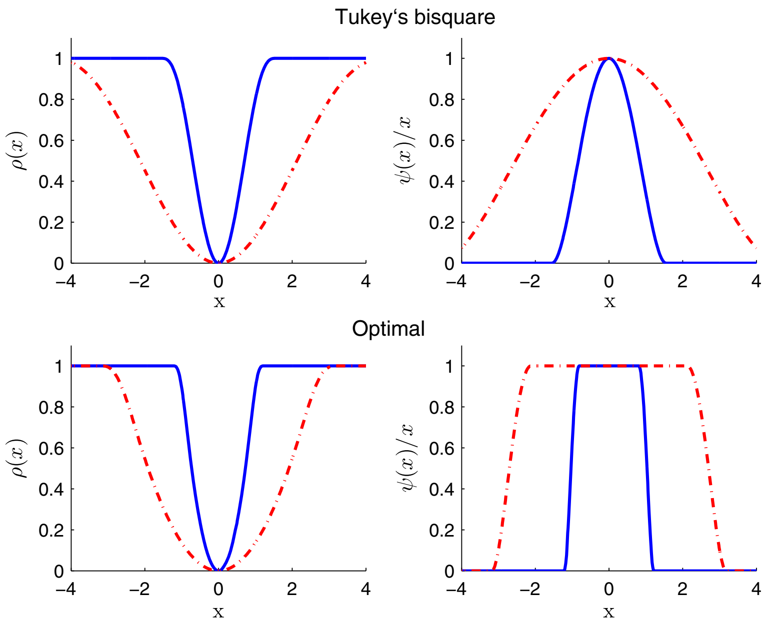

Fig. 8 Analysis of $\rho(x)$ and weight function $(\psi(x)/x)$ for Tukey Biweight (*top panels*) and optimal (*bottom panels*), for fixed levels of breakdown point and efficiency. The *solid line* is associated to a value of c linked to a breakdown point of 0.5, while the *dashed line* is associated to a value of c linked to a nominal efficiency of 0.95

rate and similarly to a weight function which assigns a high weight in correspondence of values of $|x|$ up to 2 in absolute value. Furthermore, it is interesting to observe that while the weight function of optimal ρ is similar to a hard trimming approach, in which units get 0 or 1 weight, the Tukey's Biweight seems to decrease to zero much more smoothly. Therefore, in the present observations very far from the bulk of the data, Tukey's Biweight and optimal are likely to produce equal results, but they may differ in the presence of masked outliers, which are are not extremely remote as we are going to see in the data described below.

Atkinson and Riani (2000), pp. 5–9, give an example of a regression dataset with 60 observations on three explanatory variables where there are 6 masked outliers (observations 9, 21, 30, 31, 38 and 47) and a swamping case (unit 43) that cannot be detected using standard analyses. The scatter plot of the response y against the three explanatory variables and the traditional plot of residuals against fitted values, as well as the qqplot of OLS residuals, (not given here for lack of space) do not reveal particular observations far from the bulk of the data.

Figure 9 shows the index plot of scaled robust S residuals using four different values of breakdown point and Tukey's Biweight loss function (the corresponding plot using the optimal loss function is very similar so it is not given). As the reader can see, the two panels on the left which are associated to bdp = 0.5 (top panel) and bdp = 0.3 (bottom panel), convey completely different information about outliers from the one that can be inferred from the panels on the right, which are associated respectively to

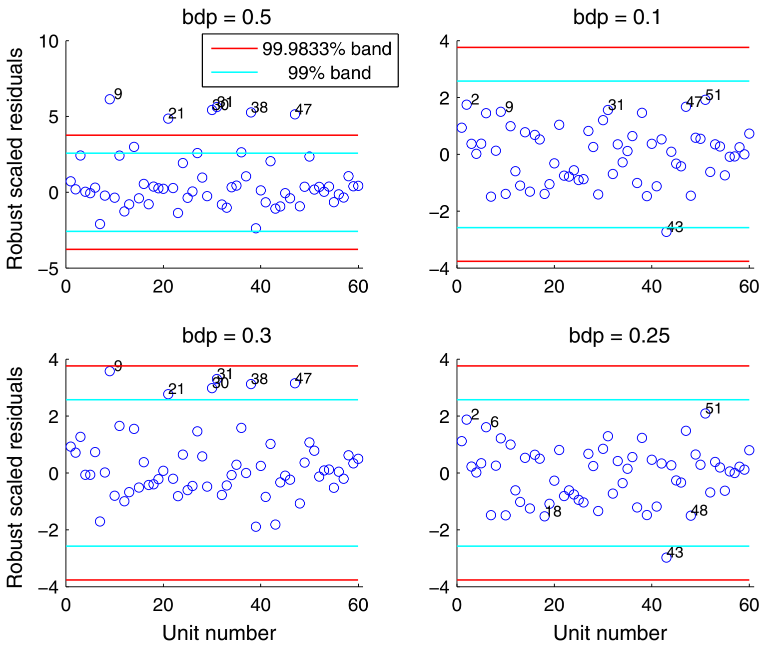

Fig. 9 Index plots of robust scale residuals obtained using $\hat{\beta}_S$, Tukey's Biweight loss function and four different levels of breakdown point. *Top left panel* bdp = 0.5, *top right panel* bdp = 0.1, *bottom left panel* bdp = 0.3, *bottom right panel* bdp = 0.25 The *horizontal lines* correspond to the 99 % individual and simultaneous bands using the standard normal

bdp = 0.1 (top right panel) and bdp = 0.25 (bottom right panel). It is surprising to notice that, changing bdp from 0.25 to 0.3, gives a picture of residuals which is drastically different. It is also interesting to point out that, while with a value of bdp = 0.5 the six masked outliers have a residual which is above the 99 % individual asymptotic confidence band, when bdp = 0.3 the residuals are in between the 99 % individual and simultaneous thresholds. Finally, when bdp = 0.1 and bdp = 0.25, the swamping case (unit 43) has a residual which is just above the 99 % individual confidence band.

Figure 10 shows the MM robust scaled residuals obtained using a Tukey's Biweight loss function when the efficiency parameter is set to 0.9 (left panel) and 0.95 (right panel). While the plot on the right (which is similar to the masked index plot of OLS residuals) highlights the presence of a unit (number 43) which is on the boundary of the simultaneous confidence band, the plot on the left (based on smaller efficiency) shows that the six outliers have a residual which is very close to the 99 % simultaneous threshold. In addition, there are two units (39 and 43) whose negative residual is very close to the 99 % individual threshold. Figure 11, on the other hand, shows what happens when we use the optimal loss function. In this case, there is not an appreciable difference between the two panels and the use of two levels of efficiency leads to the correct conclusion of identifying the six masked outliers.

This small example has, in our opinion, shown the added value of considering the output which comes from using different levels of breakdown point or different values

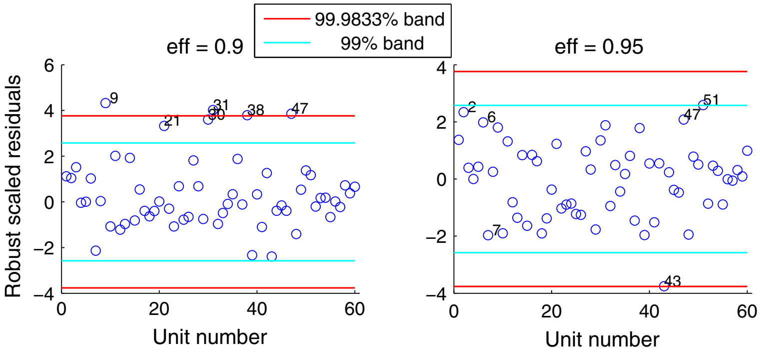

Fig. 10 Index plots of robust scaled residuals obtained using $\hat{\beta}_{MM}$ with a preliminary S-estimate of scale based on a 50 % breakdown point and Tukey Biweight loss function. *Left panel* 90 % nominal efficiency, *right panel* 95 % nominal efficiency. The *horizontal lines* correspond to the 99 % individual and simultaneous bands using the standard normal

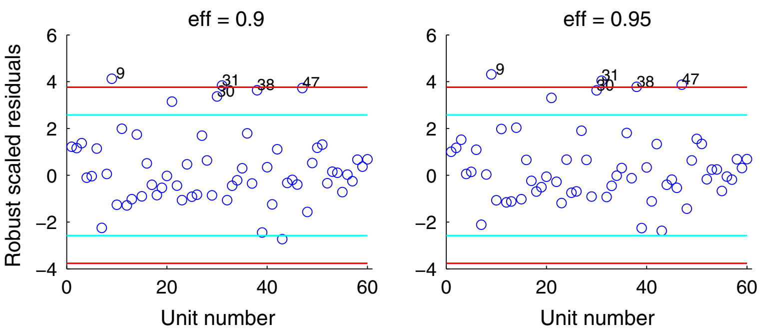

Fig. 11 Index plots of robust scaled residuals obtained using $\hat{\beta}_{MM}$ with a preliminary S-estimate of scale based on a 50 % breakdown point and optimal loss function. *Left panel* 90 % nominal efficiency, *right panel* 95 % nominal efficiency. The *horizontal lines* correspond to the 99 % individual and simultaneous bands using the standard normal

of nominal efficiency in *S*- and *MM*-estimators. At present, it is customary to use a value of bdp equal to 0.5 or 0.25 and values of the efficiency 0.85, 0.9 or 0.95. However, we think that the new computing power enables us to investigate for which values of efficiency (or robustness) there is a change to a nonrobust fit, hence enabling us to use the highest possible efficiency still maintaining a robust fit. This is the key idea of the paper by Riani et al. (2014). The traditional approach to robust statistics is to compare a robust with a nonrobust fit in a static way. However, the user is interested in understanding what is the interplay between the different units and how the model changes if a different level of bdp or efficiency is used. This is also the keystone idea of the forward search which investigates how the model changes when new units are introduced into the initial subset. In all cases where observations are sequentially included, the consistency factors must be continuously recalculated and it is necessary to consider values of breakdown point beyond those in the sequence $0.5, 0.49, \ldots, 0.01$ (same thing for the efficiency). Finally, it is also possible to envisage cases where in certain steps of the forward search S- or MM-estimators are used. In addition, also in the context of robust cluster analysis, as in Tclust (García-Escudero et al. 2008) extended to fuzzy clustering or mixture models to choose the optimal level of trimming or the best number of groups, a variety of solutions which consider several values of trimming are considered for a different number of groups. In this case, it could be also important to have efficient procedures for computing consistency factors which easily enable us to compare likelihoods based on different levels of trimming. Finally, even in a situation with just one prefixed level of overall trimming (α) given that the imposed value of α, due to the different variability of the groups, can spread in a different way over them, it could be interesting inside each step of the optimization procedure to have a likelihood function which explicitly incorporates the different trimming levels associated to the different groups. Therefore, it is important to have algorithms which enable us to obtain the corresponding consistency factors in an easily understandable and computationally efficient way, so that the user can easily try a large number of possibilities losing hardly any time for this calculation.

9 Conclusions

In this paper, we have tackled the problem of obtaining the consistency factors of robust S-estimators of location and scale both in regression and multivariate analysis for normally distributed observations. We have provided theoretical advances by proving new formulae for the calculation of these factors. Our new formulae extend to S-estimators much of the conceptual simplicity of the consistency factors already available for trimmed estimators. They are also helpful to shed new light on the relationship between consistency and other important statistical properties of the estimators. The relationship between the consistency factor c and all the other quantities studied in this paper, which are automatically determined, is summarized in Table 3, where a reference to the corresponding formula is given. Finally, we have also provided new efficient code implementations, which ensure considerable savings in computing time over the currently available algorithms.

Although in this paper we have only considered the Tukey's Biweight and optimal ρ functions for S-estimation, it is important to note that our method of proof is general and could be applied to any definition of the weight function based on polynomials (i.e. the Hampel weight function). This fact can become relevant when the number of variables increases and the robustness properties of Tukey's Biweight and optimal

Table 3 One to one relationship among the different quantities of S-estimation

Consistency factor	Breakdown point	Asymptotic variance of location estimator	Variance of (shape) covariance estimator	Asymptotic efficiency of location estimator	Asymptotic efficiency of (shape) covariance estimator
c	bdp = (32)	asvar = 1/(33)	$\text{asvar}_{\text{sc}} = 1/(33)$	eff = (33)	$\text{eff}_{\text{sc}} = (34)$

loss functions deteriorate. An effective alternative for weight computation in high dimension is provided by the Biflat function proposed by Rocke (1996). The extension of our approach to that and other ρ functions will be given elsewhere. Finally, it is necessary to point out that while in this paper we have mainly concentrated on *S*-estimators, the formulae given here can be also used to obtain the consistency factors associated to tune the efficiency of *MM*-estimators.

Acknowledgments This work was supported by the project MIUR PRIN *MISURA—multivariate models for risk assessment*. The authors are grateful to Prof. Pier Luigi Conti for useful comments on an earlier draft. We are most grateful to the referees and to the associate editor whose thoughtful and detailed comments led to improvement and clarification of our paper.

Appendix

Proof of Lemma 4.2

Proof We start from the definition of the central moment of order k of a v dimensional random vector $X = (X_1, X_2, \dots, X_v)'$

$$\mu_{k_1,k_2,\dots,k_v}(X) = E\left[(X_1-\mu_1)^{k_1}(X_2-\mu_2)^{k_2}\dots(X_v-\mu_v)^{k_v}\right].$$

If $X \sim N(0, \sigma^2 I_v)$ with the constraint that the region of integration is $0 < x'x < c^2$, we can rewrite the previous expression as:

$$\int\int\dots\int_{0<x'x<c^2} x_1^{k_1}\dots x_v^{k_v}\, d\Phi(X_1)\dots d\Phi(X_v).$$

This v-dimensional integral can be rewritten in the transformed space as a univariate integral as follows:

$$\begin{aligned}
\int_0^{c^2} y^{k/2} f_{\chi^2_v}(y)\mathrm{d}y &= \int_0^{c^2} \frac{y^{\frac{v+k}{2}-1}e^{-y/2}}{2^{v/2}\Gamma(v/2)}dy \\
&= \int_0^{c^2/2} \frac{(2t)^{\frac{v+k}{2}-1}e^{-t}dt}{2^{v/2}\Gamma(v/2)}dy \\
&= 2^{k/2}\frac{\Gamma\left(\frac{v+k}{2}\right)}{\Gamma(v/2)}\int_0^{c^2/2}\frac{t^{\frac{v+k}{2}-1}e^{-t}}{\Gamma\left(\frac{v+k}{2}\right)}dt \\
&= 2^{k/2}\frac{\Gamma\left(\frac{v+k}{2}\right)}{\Gamma(v/2)}P\left(\frac{c^2}{2}, \frac{v+k}{2}\right).
\end{aligned}$$

Now, it is easy to verify that

$$2^{k/2}\frac{\Gamma(\frac{v+k}{2})}{\Gamma(v/2)} = v(v+2)\dots(v+k-2),$$

because

$$\begin{aligned}\Gamma\left(\frac{v+k}{2}\right) &= \left(\frac{v+k}{2}-1\right)\times\left(\frac{v+k}{2}-2\right)\times\cdots\times\left(\frac{v+k}{2}-\frac{k}{2}\right)\Gamma\left(\frac{v}{2}\right)\\ &= \frac{v+k-2}{2}\times\frac{v+k-4}{2}\times\dots\frac{v}{2}\times\Gamma\left(\frac{v}{2}\right)\\ &= v(v+2)\cdots(v+k-2)\times 2^{-k/2}\times\Gamma\left(\frac{v}{2}\right).\end{aligned}$$

Similarly, it is easy to verify that

$$P\left(\frac{c^2}{2},\frac{v+k}{2}\right) = F_{\chi^2_{v+k}}(c^2),$$

because

$$\begin{aligned}F_{\chi^2_{v+k}}(c^2) &= \int_0^{c^2}\frac{t^{\frac{v+k}{2}-1}e^{-t/2}}{2^{(v+k)/2}\Gamma((v+k)/2)}dt\\ &= \frac{2}{2^{\frac{v+k}{2}}\Gamma((v+k)/2)}\int_0^{c^2}(2y)^{\frac{v+k}{2}-1}e^{-y}dy\\ &= \frac{\int_0^{c^2/2}y^{\frac{v+k}{2}-1}e^{-y}}{\Gamma((v+k)/2)}dy\\ &= P\left(\frac{c^2}{2},\frac{v+k}{2}\right).\end{aligned}$$

□

References

Atkinson AC, Riani M (2000) Robust diagnostic regression analysis. Springer, New-York

Boudt K, Croux C (2010) Robust M-estimation of multivariate GARCH models. Comput Stat Data Anal 54:2459–2469

Campbell NA (1984) Mixture models and atypical values. Math Geol 16:465–477

Cerioli A (2010) Multivariate outlier detection with high-breakdown estimators. J Am Stat Assoc 105:147–156

Cizek P (2013) Reweighted least trimmed squares: an alternative to one-step estimators. Test 22:514–533

Croux C, Haesbroeck G (1999) Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. J Multivar Anal 71:161–190

Croux C, Dhaene G, Hoorelbeke D (2006) Testing the information matrix equality with robust estimators. J Stat Plan Inference 136:3583–3613

Croux C, Dehon C, Yadine A (2011) On the optimality of multivariate *S*-estimators. Scand J Stat 38:332–341
Cuesta-Albertos JA, Matrán C, Mayo-Iscar A (2008) Trimming and likelihood: robust location and dispersion estimation in the elliptical model. Ann Stat 36:2284–2318
Davies PL (1987) Asymptotic behaviour of *S*-estimates of multivariate location parameters and disperion matrices. Ann Stat 15:1269–1292
Davies PL (1990) The asymptotics of *S*-estimators in the linear regression model. Ann Stat 18:1651–1675
Dehon C, Gassner M, Verardi V (2012) Extending the Hausman test to check for the presence of outliers. Adv Econom Essays Honor Jerry Hausman 29:435–453
Garcìa-Escudero LA, Gordaliza A (2005) Generalized radius processes for elliptically contoured distributions. J Am Stat Assoc 100:1036–1045
García-Escudero LA, Gordaliza A, Matrán C, Mayo-Iscar A (2008) A general trimming approach to robust cluster analysis. Ann Stat 36:1324–1345
Holmquist B (1988) Moments and cumulants of the multivariate normal distribution. Stoch Anal Appl 6:273–278
Hössjer O (1992) On the optimality of *S*-estimators. Stat Probab Lett 14:413–419
Hubert M, Rousseeuw PJ, Van Aelst S (2008) High-breakdown robust multivariate methods. Stat Sci 23:92–119
Hubert M, Rousseeuw PJ, Verdonck T (2012) A deterministic algorithm for robust location and scatter. J Comput Graph Stat 21:618–637
Isserlis L (1918) On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. Biometrika 12:34–39
Kent JT, Tyler DE (1991) Redescending *M*-estimates of multivariate location and scatter. Ann Stat 19:2102–2119
Kotz S, Balakrishnan N, Johnson NL (2000) Continuous multivariate distributions, vol 1, 2nd edn. Wiley, New York
Liu RY, Parelius JM, Singh K (1999) Multivariate analysis by data depth: descriptive statistics, graphics and inference. Ann Stat 27:783–840
Lopuhaä HP (1989) On the relation between *S*-estimators and *M*-estimators of multivariate location and covariance. Ann Stat 17:1662–1683
Maronna RA, Martin RD, Yohai VJ (2006) Robust statistics. Wiley, New York
Paindaveine D (2012) Elliptical symmetry. In: El-Shaarawi AH, Piegorsch W (eds) Encyclopedia of environmetrics, 2nd edn. Wiley, Chichester, pp 802–807
Riani M, Atkinson AC, Cerioli A (2009) Finding an unknown number of multivariate outliers. J R Stat Soc Ser B 71:447–466
Riani M, Perrotta D, Torti F (2012) FSDA: a MATLAB toolbox for robust analysis and interactive data exploration. Chemom Intell Lab Syst 116:17–32
Riani M, Perrotta D, Cerioli A (2013) The forward search for very large datasets (Submitted)
Riani M, Cerioli A, Atkinson AC, Perrotta D (2014) Monitoring robust regression (Submitted)
Rieder S (2012) Robust parameter estimation for the Ornstein–Uhlenbeck process. Stat Methods Appl 21:411–436
Rocke DM (1996) Properties of *S*-estimators of multivariate location and shape in high dimension. Ann Stat 24:1327–1345
Rousseeuw PJ, Leroy AM (1987) Robust regression and outlier detection. Wiley, New York
Rousseeuw PJ, Yohai VJ (1984) Robust regression by means of *S*-estimators. Robust and nonlinear time series analysis. Lecture notes in statistics, vol 26. Springer, New York, pp 256–272
Salibian-Barrera M, Yohai V (2006) A fast algorithm for *S*-regression estimates. J Comput Graph Stat 15:414–427
Salibian-Barrera M, Van Aelst S, Willems G (2006) Principal components analysis based on multivariate *MM* estimators with fast and robust bootstrap. J Am Stat Assoc 101:1198–1211
Tallis GM (1963) Elliptical and radial truncation in normal populations. Ann Math Stat 34:940–944
Todorov V, Filzmoser P (2009) An object-oriented framework for robust multivariate analysis. J Stat Softw 32:1–47
Triantafyllopoulos K (2003) On the central moments of the multidimensional Gaussian distribution. Math Sci 28:125–128
Tyler DE (1983) Robustness and efficiency properties of scatter matrices. Biometrika 70:411–420
Van Aelst S, Willems G (2011) Robust and efficient one-way MANOVA tests. J Am Stat Assoc 106:706–718

Yohai VJ (1987) High breakdown-point and high efficiency robust estimates for regression. Ann Stat 15(2):642–656

Yohai VJ (2006) The teaching of robust statistics for regression. In: Proceedings of ICOTS-7, Salvador. ISBN-13: 978–90-73592-24-7

Yohai VJ, Zamar RH (1997) Optimal locally robust *M*-estimates of regression. J Plan Stat Inference 64:309–323