Stat Comput
DOI 10.1007/s11222-008-9096-5

Controlling the size of multivariate outlier tests with the MCD estimator of scatter

Andrea Cerioli · Marco Riani · Anthony C. Atkinson

Received: 19 February 2008 / Accepted: 1 September 2008

Abstract Multivariate outlier detection requires computation of robust distances to be compared with appropriate cut-off points. In this paper we propose a new calibration method for obtaining reliable cut-off points of distances derived from the MCD estimator of scatter. These cut-off points are based on a more accurate estimate of the extreme tail of the distribution of robust distances. We show that our procedure gives reliable tests of outlyingness in almost all situations of practical interest, provided that the sample size is not much smaller than 50. Therefore, it is a considerable improvement over all the available MCD procedures, which are unable to provide good control over the size of multiple outlier tests for the data structures considered in this paper.

Keywords Minimum covariance determinant estimator · Robust distances · Multiple outliers · Simultaneous testing · Calibration factor · Simulation

1 Introduction

Multivariate outlier detection is a fundamental and pervasive step of most statistical analyses. For data from location-scale families, its accomplishment requires estimation of location and scatter in v dimensions. However, it is well known that the estimates of the mean and covariance matrix using all the data are extremely sensitive to the presence of outliers. The circularity breaks down only if robust estimates are employed instead of the classical unbiased ones (Rousseeuw and van Zomeren 1990; Becker and Gather 1999). In particular, in this paper we deal with the highly-robust Minimum Covariance Determinant (MCD) estimator of Rousseeuw and Van Driessen (1999) and with its several refinements. This estimator has an intuitive appeal and benefits from the availability of software implementation in different languages, including R, S-Plus, Fortran and Matlab. It also has good asymptotic properties that compare favourably with those of other high-breakdown estimators (Butler et al. 1993; Croux and Haesbroeck 1999). For these reasons the MCD estimator has gained much popularity, not only for outlier identification but also as an ingredient of many robust multivariate techniques (Croux and Haesbroeck 2000; Willems et al. 2002; Pison and Van Aelst 2004; Rousseeuw et al. 2004; Todorov 2006; Todorov and Filzmoser 2008).

To identify multivariate outliers in a sample $y = (y_1, \ldots, y_n)'$ of n observations from a v-variate population with mean μ and dispersion matrix Σ, the MCD estimates of μ and Σ, say $\hat{\mu}_{(\mathrm{MCD})}$ and $\hat{\Sigma}_{(\mathrm{MCD})}$, are plugged into the Mahalanobis formula to obtain the n squared robust distances

$$d^2_{i(\mathrm{MCD})} = (y_i - \hat{\mu}_{(\mathrm{MCD})})' \hat{\Sigma}^{-1}_{(\mathrm{MCD})} (y_i - \hat{\mu}_{(\mathrm{MCD})}), \quad i = 1, \ldots, n, \tag{1}$$

which do not suffer from masking and swamping. Observations for which $d^2_{i(\mathrm{MCD})}$ exceeds a specified threshold are then labelled as outliers.

In order to achieve consistency under the normal population model, $\hat{\Sigma}_{(\mathrm{MCD})}$ must incorporate a correction factor that

A. Cerioli (✉) · M. Riani
Dipartimento di Economia, Università di Parma, Via Kennedy 6, 43100 Parma, Italy
e-mail: andrea.cerioli@unipr.it

M. Riani
e-mail: mriani@unipr.it

A.C. Atkinson
Department of Statistics, The London School of Economics, Houghton Street, London WC2A 2AE, UK
e-mail: a.c.atkinson@lse.ac.uk

allows for the fact that Σ is estimated from the 'central' part of the data only. An additional ingredient that is often, but not always, part of MCD estimation is a small sample bias-correction factor obtained by simulation (Pison et al. 2002). Sometimes a one-step reweighted version of $\hat{\Sigma}_{(\mathrm{MCD})}$ is used in definition (1) to gain efficiency while retaining the same breakdown point as the initial MCD estimator (Rousseeuw and Van Driessen 1999, p. 218; Lopuhaä 1999). However, there is no general agreement on whether or not the same correction factors should apply after reweighting. Despite the intuitive appeal of robust distances, the user must then choose among several different versions of (1) and needs guidance about the performance of each of them. We provide such guidance.

An unsolved problem is that the exact distribution of robust distances is unknown for finite sample sizes. The standard approach has been to compare the squared distances to the percentage points of their asymptotic χ^2_v distribution. However, the χ^2_v approximation is known to be liberal and leads to the nomination of too many outliers in small and moderate samples. Hardin and Rocke (2005) provided compelling evidence of this behaviour and showed that the χ^2_v approximation can be substantially inadequate even for relatively large values of n, depending on the actual value of v. They suggested an improved approximation based on the F distribution. The results of Hardin and Rocke refer to the basic MCD distances (1) and do not apply to the supposedly more efficient reweighted ones. Therefore, they are only partially useful for the purpose of comparing the available versions of such distances.

Another important issue is that of simultaneity of the n tests performed by means of the squared robust distances $d^2_{i(\mathrm{MCD})}, i = 1, \ldots, n$. The usual suggestion in the MCD literature has been to control the size of each individual test by comparing each $d^2_{i(\mathrm{MCD})}$ to the $\alpha\%$ cut-off point of the reference distribution. Most published results are based on $0.001 \leq \alpha \leq 0.05$, with $\alpha = 0.025$ a popular choice (e.g. Rousseeuw and Van Driessen 1999; Pison et al. 2002). However, the resulting family-wise error rate soon approaches 1 when n is larger than a few dozen and the user should be prepared to declare at least one outlier (and often many more) in *any data set* of realistic size, even when contaminated observations are not present.

The individual testing scenario is inappropriate in many situations, especially when one has repeatedly to check for outliers in several samples supposed to come from the same population. For instance, Arsenis et al. (2005) and Riani et al. (2009) analyzed bivariate trade data arising in the European Union market. Outliers are of paramount importance because some of them may correspond to fraudulent transactions. However, there are hundreds of transactions to be inspected over thousands of markets, corresponding to different traded commodities, and controlling the size of individual tests would lead to a plethora of false signals for anti-fraud services. Another striking example arises in the process of producing microarray data, where it is crucial to evaluate the quality of an array and to identify those with low quality. Cohen Freue et al. (2007) developed a multivariate technique based on robust distances similar to (1) computed on different subsets of variables. Several distances are obtained for each array and ignoring the multiplicity of the resulting tests increases the probability of labelling false outliers and of discarding potentially useful information. In both these examples more effective conclusions could be reached by controlling the *proportion of 'good' data sets* that are wrongly declared to contain outliers.

The goal of this paper is twofold. First, after a brief review of the MCD methodology, in Sect. 3 we examine the null performance of the outlier tests that can be obtained through alternative versions of (1). We investigate both the individual testing scenario, thus complementing the published simulation results of Hardin and Rocke (2005), and the more realistic situation where control is over the family-wise error rate of all the n tests, as suggested by Becker and Gather (1999). We will see that performance is often worse than expected and surprisingly bad for the purpose of simultaneous outlier identification. Some hints on the weakness of MCD-based outlier identification rules in the multiple testing framework were also given by Becker and Gather (2001), who provided evidence of the inadequacy of asymptotic cut-off values in a limited number of cases. Their results thus find extensive confirmation in our paper. We will also see that satisfactory behaviour of an individual outlier test does not necessarily reproduce in the multiple testing scenario.

Then, in Sect. 4 we extend our simulation results to derive improved correction factors for the robust distances. These factors are obtained by calibrating the tail of the distribution of the distances, not just their mean and variance. Our corrections turn out to be essential for controlling the size of outlier tests in the multiple testing framework, where the currently available MCD procedures are ineffective. They are computed for a number of specific values of n and v, but are then generalized through interpolation. Their impact on the power of outlier tests is also investigated. The paper ends with some concluding remarks in Sect. 5.

2 Options in MCD estimation

Suppose we have a sample y of n v-dimensional observations. The MCD subset $y_{(\mathrm{MCD})}$ is defined to be the subsample of h observations, with $n/2 \leq h < n$, whose covariance matrix has the smallest determinant. We are interested in outlier detection, as in Hardin and Rocke (2005). Hence we take the value of h yielding the maximum possible break-

down point, i.e.

$$h = \left\lfloor \frac{n+v+1}{2} \right\rfloor, \tag{2}$$

where $\lfloor \cdot \rfloor$ denotes the integer part. A larger value of h would result in more efficient estimates, but at the expense of a reduced breakdown value. The MCD estimate of location is the average of the MCD subset,

$$\hat{\mu}_{(\mathrm{MCD})} = \frac{1}{h} \sum_{i \in y_{(\mathrm{MCD})}} y_i, \tag{3}$$

whereas the MCD estimate of scatter is proportional to the dispersion matrix of this subset:

$$\hat{\Sigma}_{(\mathrm{MCD})} = \frac{c(h)s(h,n,v)}{h-1} \sum_{i \in y_{(\mathrm{MCD})}} (y_i - \hat{\mu}_{(\mathrm{MCD})})(y_i - \hat{\mu}_{(\mathrm{MCD})})'. \tag{4}$$

The proportionality constant $c(h)$ makes $\hat{\Sigma}_{(\mathrm{MCD})}$ Fisher-consistent when the distribution of y is elliptically symmetric and unimodal with mean μ and dispersion matrix Σ (Butler et al. 1993; Croux and Haesbroeck 1999). If $y \sim N(\mu, \Sigma)$

$$c(h) = \frac{h/n}{P(\chi^2_{v+2} < \chi^2_{v,1-h/n})}, \tag{5}$$

where $\chi^2_{v,\alpha}$ denotes the $\alpha\%$ cut-off point of the χ^2_v distribution which leaves $\alpha\%$ of the values at its right.

The second proportionality constant, $s(h,n,v)$, serves the purpose of reducing the small sample bias of $\hat{\Sigma}_{(\mathrm{MCD})}$. The actual value of this factor depends also on n and v. It was obtained by Pison et al. (2002) through a combination of Monte Carlo simulation and parametric interpolation, under the assumption that $s(h,n,v) \to 1$ as $n \to \infty$ for fixed v. It is worth noting that $s(h,n,v)$ is just a first-order correction factor for $\hat{\Sigma}_{(\mathrm{MCD})}$. Hence, it might be expected to work reasonably well for the purpose of adjusting the mean of the squared robust distances (1), but it is likely to be inadequate in the extreme tail of their distribution, the one of importance for outlier detection, especially in the simultaneous testing framework of Sect. 3.2. Improved first-order corrections for very small samples are given by Todorov (2008) but are not considered here.

Equations (3) and (4) define the raw MCD estimates of location and scatter. To increase efficiency, a one-step reweighted version of them is often used in practice. These estimators are computed by giving weight 0 to observations for which $d^2_{(\mathrm{MCD})i}$ exceeds a threshold value. Thus a first subset of h observations is used to select a second subset of m from which the parameters are estimated. The reweighted MCD estimates of location and scatter are then

$$\hat{\mu}_{\mathrm{RMCD}} = \frac{1}{m} \sum_{i=1}^{n} w_i y_i \tag{6}$$

and

$$\hat{\Sigma}_{\mathrm{RMCD}} = \frac{c^*(m)s^*(m,n,v)}{m-1} \times \sum_{i=1}^{n} w_i (y_i - \hat{\mu}_{(\mathrm{RMCD})})(y_i - \hat{\mu}_{(\mathrm{RMCD})})', \tag{7}$$

where $w_i = 0$ if $d^2_{(\mathrm{MCD})i} > d^2_{(\mathrm{MCD})*}$, $w_i = 1$ otherwise, and $m = \sum_{i=1}^{n} w_i$. The usual suggestion for the threshold (e.g. Rousseeuw and Leroy 1987, p. 260; Rousseeuw and Van Driessen 1999, p. 218; Pison and Van Aelst 2004, p. 312) is to take the 0.025% cut-off point of the χ^2_v distribution:

$$d^2_{(\mathrm{MCD})*} = \chi^2_{v,0.025}. \tag{8}$$

This threshold is the default implementation in most available software and is chosen independently of the size α of the outlier test. An alternative proposal is described in Sect. 3.2. The factors $c^*(m)$ and $s^*(m,n,v)$ guarantee consistency of the reweighted estimator and improve its small sample behaviour, as do the corresponding factors in (4). They were not supported in the initial MCD literature but were advocated later by Croux and Haesbroeck (1999) and by Pison et al. (2002). However, only a few software functions actually implement them (Todorov 2008). Most of the available functions simply take the empirical covariance matrix of the subset of m observations as the reweighted MCD estimate of scatter.

3 Null performance of the MCD estimators

In this section we investigate the null performance of multivariate outlier tests based on squared robust Mahalanobis distances. The reported values of n and v represent a subset of the finer grid to be considered in Sect. 4. We focus on the nominal test size $\alpha = .01$, although similar results were also obtained for $\alpha = 0.05$ and $\alpha = 0.025$. We choose among the available MCD options those that should guarantee the best control over the size of the resulting tests. Therefore, we include both the consistency correction and the small sample correction in all our estimates of scatter. We take $c(h)$ as in (5). Similarly,

$$c^*(m) = \frac{m/n}{P(\chi^2_{v+2} < \chi^2_{v,m/n})}.$$

Table 1 Size of the multivariate outlier tests MCD, MCD-HR and RMCD for testing the n individual hypotheses $H_{0i} : y_i \sim N(\mu, \Sigma)$, each at nominal level $\alpha = 0.01$. Size is computed on 50,000 simulations for each combination of n and v

		$n = 50$	$n = 75$	$n = 100$	$n = 150$	$n = 200$	$n = 300$	$n = 500$
	MCD	0.070	0.050	0.043	0.033	0.027	0.021	0.017
$v = 2$	MCD-HR	0.001	0.003	0.005	0.007	0.008	0.009	0.009
	RMCD	0.019	0.015	0.013	0.012	0.011	0.011	0.010
	MCD	0.137	0.090	0.070	0.046	0.035	0.024	0.018
$v = 4$	MCD-HR	0.009	0.012	0.012	0.012	0.011	0.010	0.010
	RMCD	0.047	0.025	0.019	0.014	0.013	0.011	0.011
	MCD	0.194	0.122	0.091	0.056	0.041	0.028	0.019
$v = 6$	MCD-HR	0.019	0.017	0.015	0.013	0.011	0.010	0.010
	RMCD	0.090	0.038	0.025	0.017	0.014	0.013	0.011
	MCD	0.237	0.152	0.111	0.066	0.046	0.031	0.020
$v = 8$	MCD-HR	0.026	0.020	0.016	0.013	0.011	0.010	0.009
	RMCD	0.150	0.058	0.033	0.020	0.016	0.013	0.012
	MCD	0.266	0.179	0.131	0.077	0.053	0.034	0.022
$v = 10$	MCD-HR	0.031	0.021	0.016	0.012	0.010	0.009	0.009
	RMCD	0.217	0.088	0.046	0.023	0.018	0.015	0.013
	MCD	0.283	0.203	0.152	0.088	0.060	0.037	0.024
$v = 12$	MCD-HR	0.034	0.021	0.016	0.011	0.010	0.008	0.008
	RMCD	0.270	0.131	0.065	0.029	0.021	0.016	0.013

The small sample factors $s(h, n, v)$ and $s^*(m, n, v)$ are those implemented in the function *covMcd*() of the R package *robustbase* available at http://cran.r-project.org/. The MCD subset $y_{(\mathrm{MCD})}$ is obtained through the Fortran algorithm *FAST-MCD* of Rousseeuw and Van Driessen (1999) available at http://www.agoras.ua.ac.be/. Simulation uses the pseudo-random number generator of Matsumoto and Nishimura (1998), the default choice also in the R function *RNG*.

For fixed n and v, the actual size of each test is estimated by simulating 50,000 independent n-dimensional samples y from the v-variate $N(0, I)$ distribution. The results are valid for any $y \sim N(\mu, \Sigma)$ of the same dimensions thanks to the affine invariance property of robust Mahalanobis distances. Size is estimated by the proportion of false rejections of the null hypothesis over the total number of tests performed.

3.1 Individual testing

In the first part of our simulation study we work under the inferential setting of Rousseeuw and Van Driessen (1999), Pison et al. (2002) and Hardin and Rocke (2005), among others. In this case each of the n *individual null hypotheses*

$$H_{0i} : y_i \sim N(\mu, \Sigma), \quad i = 1, \ldots, n, \tag{9}$$

is tested at nominal size $\alpha = 0.01$. The n hypotheses (9) must be checked on each data set. The total number of tests performed in the simulation study is thus $50{,}000n$.

We focus on three different outlier tests.

- MCD. This test compares the squared distances $d^2_{i(\mathrm{MCD})}$ to their asymptotic χ^2_v distribution, as in the standard approach of Pison et al. (2002). The raw MCD estimators (3) and (4) are used.
- MCD-HR. This test uses the same MCD options as above. However, the squared distances $d^2_{i(\mathrm{MCD})}$ are compared to the improved approximation of Hardin and Rocke (2005), based on a scaled F distribution.
- RMCD. In this test the efficient one-step reweighted MCD distances

$$d^2_{i(\mathrm{RMCD})} = (y_i - \hat{\mu}_{(\mathrm{RMCD})})' \hat{\Sigma}^{-1}_{(\mathrm{RMCD})} (y_i - \hat{\mu}_{(\mathrm{RMCD})})$$

are used, with $\hat{\mu}_{(\mathrm{RMCD})}$ and $\hat{\Sigma}_{(\mathrm{RMCD})}$ defined as in (6) and (7). The weights are computed from the raw MCD estimates (3) and (4), with the threshold $d^2_{(\mathrm{MCD})*}$ at its default value (8).

Table 1 displays the main findings obtained for the three outlier tests. If the distribution of the robust distances were well approximated by the reference distribution, we would expect to declare about $\lfloor \alpha n \rfloor$ false outliers in each simulated data set. Ideally each entry in the cells of Table 1 should be

Table 2 Estimated standard deviation of the proportion of outliers found by each test for some values of n and v

		$n=50$	$n=100$	$n=200$	$n=500$
	MCD	0.071	0.042	0.020	0.008
$v=4$	MCD-HR	0.020	0.017	0.011	0.006
	RMCD	0.045	0.016	0.008	0.005
	MCD	0.067	0.048	0.021	0.008
$v=8$	MCD-HR	0.037	0.020	0.010	0.005
	RMCD	0.086	0.024	0.009	0.005
	MCD	0.046	0.048	0.023	0.008
$v=12$	MCD-HR	0.042	0.020	0.009	0.005
	RMCD	0.063	0.037	0.011	0.005

close to 0.01. However, it is seen that the MCD test based on the χ^2_v distribution is largely unsatisfactory for all the reported values of n and v, a confirmation of the findings of Hardin and Rocke (2005). The improved test based on the F distribution, MCD-HR, is the only one that guarantees reasonable performance, although its true size can still be a bit different from the nominal value when $n < 100$, with departures in both directions. The increased efficiency of the reweighted test RMCD comes at a cost. Its size is generally larger than 0.01 and the test becomes considerably more liberal than MCD-HR for small sample sizes. In all instances it is apparent that the small sample corrections do not work properly when $n < 100$ and v increases. One explanation is that $s(h,n,v)$ and $s^*(m,n,v)$ are first-order correction factors that do not allow for the variability in the tail of the distribution of robust distances.

Table 2 provides a summary of the variability associated with our simulations. It shows the estimated standard deviation of the proportion of outliers found by each test for some values of n and v. The standard errors of the test sizes given in Table 1, which measure the Monte Carlo precision of our estimated sizes, are easily obtained dividing these standard deviations by $\sqrt{50{,}000}$.

3.2 Multiple testing

The situation is much less reassuring in the simultaneous testing framework introduced in Sect. 1. The null hypothesis of interest is now the *intersection hypothesis*

$$H_{0s} : \{y_1 \sim N(\mu, \Sigma)\} \cap \{y_2 \sim N(\mu, \Sigma)\} \cap \cdots \cap \{y_n \sim N(\mu, \Sigma)\} \tag{10}$$

that no outliers are present in the data. Given a cut-off d^2_γ for the squared robust distances, the size of this test is the probability that at least one outlier is erroneously found using d^2_γ as a cut-off. This probability is

$$P\{\max_{i=1}^n d^2_{(\mathrm{MCD})i} > d^2_\gamma \mid \mathrm{H_{0s}\ is\ true}\} \tag{11}$$

for the MCD-based distances, and

$$P\{\max_{i=1}^n d^2_{(\mathrm{RMCD})i} > d^2_\gamma \mid \mathrm{H_{0s}\ is\ true}\} \tag{12}$$

if the reweighted estimators are used. Size is then estimated as the proportion of simulated data sets for which the event in (11) is verified.

We control the multiplicity in (10) through a Bonferroni approach, similar to the simultaneous outlier identification rule of Becker and Gather (1999) and Becker and Gather (2001). A Bonferroni-type argument is appropriate because the observations not included in $y_{(\mathrm{MCD})}$ are approximately independent of $\hat{\Sigma}_{(\mathrm{MCD})}$ (Hardin and Rocke 2005). The correlation between the test statistics $d^2_{(\mathrm{MCD})i}$ and $d^2_{(\mathrm{MCD})j}$, and that between $d^2_{(\mathrm{RMCD})i}$ and $d^2_{(\mathrm{RMCD})j}$, should then be negligible if y_i and y_j do not belong to the MCD subset. Given the liberal behaviour exhibited by tests MCD and RMCD in Table 1, the potential conservativeness of the method should not be of great concern in the present context. A confirmation of the minor effect of correlation on test size is provided by the simulation results of Sect. 4.3. Use of the sharper Šidák inequality (Šidák 1967) gave a negligible advantage over the Bonferroni approach.

We set $\gamma = \alpha/n$ and define

$$d^2_\gamma = \chi^2_{v,\alpha/n} \tag{13}$$

for MCD and RMCD. For the MCD-HR test,

$$d^2_\gamma = \frac{vn^*}{n^* - v + 1} F_{v,n^*-v+1,\alpha/n}, \tag{14}$$

where $F_{\nu_1,\nu_2,\alpha}$ is the $\alpha\%$ cut-off point of the F_{ν_1,ν_2} distribution. The value n^* in the denominator degrees of freedom is computed using the adjusted asymptotic method of Hardin and Rocke (2005, Sect. 3.1.1). The functions *DCHIIN* and *DFIN* of the IMSL library are employed to obtain the numerical values of (13) and (14).

We note that, under the Bonferroni approach, there is a potential source of incoherence in the default RMCD procedure described in Sect. 3.1. In this procedure each observation is first tested for outlyingness at size 0.025 to compute its weight, see (8), and then at size $\gamma = \alpha/n$ to verify the corresponding component of the intersection hypothesis (10). Since typically $\gamma \ll 0.025$, one may control the multiplicity of the n tests also in the computation of the weights. We thus consider a modified version of the RMCD procedure with Bonferroni adjustment of the threshold value $d^2_{(\mathrm{MCD})*}$. This adjustment is introduced by Riani et al. (2009) and is in agreement with the general definition of reweighted MCD estimators provided by Croux and Haesbroeck (1999).

- RMCD-CH. This test uses the same reweighting scheme as RMCD, but with threshold

$$d^2_{(\mathrm{MCD})*} = \chi^2_{v,\gamma}. \tag{15}$$

Table 3 Size of the multivariate outlier tests MCD, MCD-HR, RMCD and RMCD-CH for testing the intersection hypothesis (10) at nominal level $\alpha = 0.01$. Size is computed on 50,000 simulations for each combination of n and v

		$n = 50$	$n = 75$	$n = 100$	$n = 150$	$n = 200$	$n = 300$	$n = 500$
	MCD	0.426	0.336	0.297	0.230	0.185	0.129	0.083
$v = 2$	MCD-HR	0.000	0.000	0.000	0.001	0.002	0.004	0.007
	RMCD	0.074	0.045	0.034	0.024	0.020	0.017	0.014
	RMCD-CH	0.026	0.018	0.014	0.012	0.011	0.010	0.011
	MCD	0.790	0.645	0.553	0.400	0.295	0.182	0.091
$v = 4$	MCD-HR	0.002	0.008	0.013	0.017	0.017	0.016	0.013
	RMCD	0.284	0.131	0.078	0.041	0.030	0.021	0.015
	RMCD-CH	0.085	0.038	0.026	0.017	0.016	0.013	0.011
	MCD	0.934	0.807	0.706	0.501	0.363	0.202	0.097
$v = 6$	MCD-HR	0.021	0.030	0.031	0.028	0.022	0.017	0.013
	RMCD	0.549	0.246	0.138	0.056	0.035	0.023	0.018
	RMCD-CH	0.176	0.063	0.039	0.022	0.018	0.014	0.013
	MCD	0.982	0.904	0.812	0.595	0.419	0.227	0.101
$v = 8$	MCD-HR	0.050	0.051	0.044	0.031	0.023	0.017	0.012
	RMCD	0.794	0.402	0.210	0.077	0.047	0.029	0.020
	RMCD-CH	0.333	0.102	0.053	0.030	0.023	0.017	0.014
	MCD	0.997	0.958	0.888	0.680	0.489	0.266	0.114
$v = 10$	MCD-HR	0.082	0.064	0.050	0.031	0.023	0.014	0.010
	RMCD	0.947	0.606	0.321	0.108	0.059	0.034	0.021
	RMCD-CH	0.567	0.171	0.078	0.039	0.027	0.020	0.015
	MCD	0.999	0.983	0.942	0.759	0.565	0.307	0.128
$v = 12$	MCD-HR	0.108	0.068	0.053	0.032	0.022	0.012	0.009
	RMCD	0.995	0.803	0.483	0.156	0.077	0.038	0.024
	RMCD-CH	0.814	0.283	0.124	0.053	0.036	0.024	0.016

Our default choice in (15) is $\gamma = 0.01/n$.

The performances of the four robust procedures for testing the intersection hypothesis of no outliers in the data are shown in Table 3. The results for MCD are exceptionally bad, especially if $n \leq 200$, with sizes up to almost 100%. As n increases the χ^2_v approximation improves, but even when $n = 500$, a value by which asymptotics could be expected to be a reasonable guide, the sizes fluctuate between 0.08 and 0.13. The MCD test is thus clearly unusable for the purpose of simultaneous outlier identification over the whole range of selected values of n and v.

Also the RMCD procedure of Sect. 3.1 is prone to strong liberality and is thus unusable even in relatively large samples, its sizes being close to the hoped-for value only when n approaches 500. The modified test RMCD-CH improves the situation but does not solve the problem: when $n = 100$ its size is around 0.03 for $v = 5$ and around 0.08 for $v = 10$. Therefore, the Bonferroni adjustment in threshold (15) is not sufficient to control the size of the reweighted test. There are at least two reasons for this, perhaps surprising, failure. The first one is the same that leads to the unsatisfactory performance of MCD. For small and moderate samples, the χ^2_v approximation is poor and threshold (15) discards more observations than the expected γn for the reweighting step. The second motivation is that the small sample corrections $s^*(m, n, v)$ were obtained by Pison et al. (2002) using the $\chi^2_{v,0.025}$ threshold and are not appropriate for RMCD-CH.

The MCD-HR test based on the F distribution was seen in Sect. 3.1 to be the only one with reasonable performance across all values of n and v. However, its performance worsens considerably in the multiple testing framework of Table 3. One disappointing feature of this test is that it can be either extremely conservative, with sizes < 0.001, when n or v are small, or moderately liberal, with sizes > 0.05, when v increases. Another surprising feature is that the test size does not necessarily improve as a monotone function of n for given v. For instance, when $v = 6$ the size at $n = 50$ is closer to the target than for n in the range [100, 200]. We conclude that the scaled-F approximation to the distribution of robust distances, which works well for quantiles in the

0.1–0.01 range, is substantially inaccurate for the quantiles needed to perform multiple comparisons.

In summary, none of the four robust distances computed from the MCD estimator of scatter is able to provide good control over the size of the test of the intersection hypothesis (10), even if the (potentially conservative) Bonferroni method is adopted for the definition of cut-off values. The reason is that the reference distributions defined by (13) and (14) provide poor approximations to the extreme tail of the distribution of robust distances, even after correction for consistency and small-sample bias, at least when $n < 500$. New correction factors that considerably improve this approximation are developed in the following section.

4 Improved calibration factors

4.1 Simulation-based calibration

The traditional way to get better control on the size of outlier tests based on squared robust distances has been to calibrate the dispersion matrix of the subset of units used for estimation. The basic corrections are the consistency factors $c(h)$ and $c^*(m)$, in (4) and (7) respectively, which account for trimming in a multivariate normal sample. The factors $s(h, n, v)$ and $s^*(m, n, v)$ developed by Pison et al. (2002) provide first-order corrections suitable for small samples. The F approximation suggested by Hardin and Rocke (2005) also takes into account the variance of the diagonal elements of the MCD dispersion matrix, but was seen in Sect. 3.2 to be rather inaccurate in the extreme tail of the distribution. Furthermore, it is not applicable to the robust distances computed from the reweighted estimator $\hat{\Sigma}_{\text{RMCD}}$.

We obtain improved calibration factors for the robust distances of Sect. 3 by Monte Carlo simulation. Our aim is to calibrate the cut-off values of the reference distribution of each squared distance, not just its expected value and its variance, in order to obtain good control over the size of the corresponding test. In what follows, we concentrate on the more problematic test of the intersection hypothesis of no outliers in the data. We consider the two basic and most widely adopted procedures MCD and RMCD, although the approach is quite general and could be applied to MCD-HR and RMCD-CH as well.

Let, without loss of generality, d_i^2 represent the squared robust distance actually selected for testing outlyingness of observation y_i. Instead of relying on distributional assumptions such as (13), in principle a Monte Carlo estimate of d_γ^2 could be obtained from K independent replicates $y^{(k)} = (y_1^{(k)}, \ldots, y_n^{(k)})'$, $k = 1, \ldots, K$, of y under the null hypothesis (10). If $d_{[i]}^{(k)}$ is the i-th ordered distance, $i = 1, \ldots, n$, in sample k, this estimate would be

$$\tilde{d}_\gamma^2 = \frac{1}{K} \sum_{k=1}^{K} \{d_{[l]}^{(k)}\}^2,$$

where $l = \lfloor (n+1)(1-\gamma) \rfloor$. However, with $\gamma = \alpha/n$ we obtain that $l = n$ if $\alpha < n/(n+1)$, so that $\tilde{d}_\gamma^2$ is not an appropriate estimate in virtually all situations of practical interest.

We resort to a different approximation of d_γ^2 based on the K replicates $y^{(k)}$. Let $Y = (y^{(1)}, \ldots, y^{(K)})'$ be the pooled sample of $N = Kn$ observations obtained by simulation from the v-variate $N(0, I)$ distribution. We rely on approximate independence of $\hat{\Sigma}_{(\text{MCD})}$ and the observations not included in $y_{(\text{MCD})}$. From this result the squared robust distances computed from the observations in Y that do not contribute to any MCD subset can be taken as an approximate random sample from the same distribution. Write Y^* for this sample of $N^* = K(n-h) \approx N/2$ approximately independent observations in Y. Let $d_{[i]}$ be the i-th ordered distance in the pooled sample Y, and $d_{[i]}^*$ be the i-th ordered distance in the approximately random sample Y^*. The vector $d^* = (d_{[1]}^*, d_{[2]}^*, \ldots, d_{[N^*]}^*)'$ contains the order statistics on which we base our estimate of d_γ^2.

To motivate the estimation procedure, write $y_{(\overline{\text{MCD}})}$ for the set of $n - h \approx n/2$ units that do not belong to the MCD subset for sample y. These are the units that contribute to the random (pooled) sample Y^* and to the vector of order statistics d^*. Hardin and Rocke (2005, p. 936) showed that for γ in the usual range of test size values

$$P\{i \in y_{(\overline{\text{MCD}})} \mid d_i^2 > d_\gamma^2\} \to 1 \quad \text{as } n \to \infty. \tag{16}$$

Therefore, the size of the intersection hypothesis (10) can be approximated by

$$P\{\max_* d_i^2 > d_\gamma^2\}, \tag{17}$$

where $\max_*$ denotes the maximum over the units in $y_{(\overline{\text{MCD}})}$. For K sufficiently large, convergence of the empirical quantiles of Y^* to their population values suggests estimating d_γ^2 as

$$\hat{d}_\gamma^2 = \{d_{[L^*]}^*\}^2, \tag{18}$$

with $L^* = \lfloor (N^*+1)(1-\alpha/(n-h)) \rfloor$. It is also straightforward to see that under (16) the samples Y and Y^* share the same extreme observations. Therefore

$$N - N^* + L^* = \lfloor (N+1)(1-\alpha/n) \rfloor = L,$$

say, so that $\hat{d}_\gamma^2 = \{d_{[L]}\}^2$ with probability tending to 1.

The calibration factors for tests MCD and RMCD at simultaneous size α are defined as

$$\hat{\kappa}_{\alpha,n,v} = \hat{d}_\gamma^2 / \chi_{v,\alpha/n}^2. \tag{19}$$

Table 4 Calibration factors for the multivariate outlier tests MCD and RMCD for testing the intersection hypothesis (10) at nominal level $\alpha = 0.01$. Calibration factors are computed on the same 50,000 simulations that give rise to Table 3

		$n = 50$	$n = 75$	$n = 100$	$n = 150$	$n = 200$	$n = 300$	$n = 500$
	MCD	3.200	2.426	2.164	1.836	1.686	1.489	1.336
$v = 2$	RMCD	1.556	1.277	1.188	1.113	1.084	1.055	1.043
	MCD	5.151	3.416	2.848	2.147	1.827	1.533	1.287
$v = 4$	RMCD	2.690	1.630	1.378	1.173	1.122	1.067	1.037
	MCD	7.073	4.016	3.092	2.194	1.794	1.478	1.253
$v = 6$	RMCD	4.493	1.987	1.488	1.204	1.130	1.076	1.046
	MCD	8.595	4.437	3.220	2.199	1.776	1.466	1.247
$v = 8$	RMCD	6.722	2.493	1.639	1.242	1.151	1.086	1.050
	MCD	10.407	4.916	3.374	2.225	1.801	1.444	1.235
$v = 10$	RMCD	9.330	3.113	1.841	1.282	1.170	1.095	1.059
	MCD	12.000	5.173	3.570	2.250	1.804	1.432	1.233
$v = 12$	RMCD	12.402	3.837	2.079	1.330	1.185	1.101	1.064

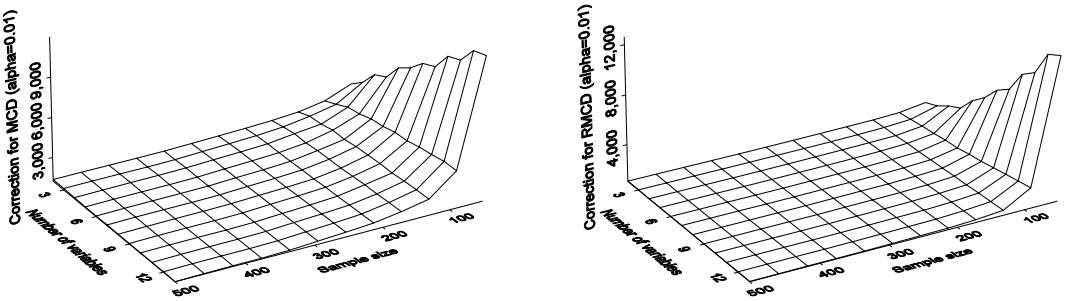

Fig. 1 Three-dimensional surfaces of simulated calibration factors (19) as a function of n and v, for MCD (*left-hand panel*) and RMCD (*right-hand panel*) and $\alpha = 0.01$

Table 4 provides these factors with $\hat{d}^2_\gamma$ computed from the same set of $K = 50{,}000$ simulations that give rise to Table 3, in the case $\alpha = 0.01$. As expected all the reported values of $\hat{\kappa}_{\alpha,n,v}$ are larger than 1, and considerably so when $n \leq 100$. They represent the extent to which the asymptotic cut-off $\chi^2_{v,0.01/n}$ is underestimating the true cut-off d^2_γ in (11) and (12).

4.2 Parametric calibration

It is important for the usability of our corrections to have simple interpolation formulas which could reproduce the required calibration factor for all n and v, without the need of performing any additional simulation. For this purpose, we now extend the simulation design of Sect. 3 by considering a finer grid of sample sizes and dimensions. We computed our calibration factors $\hat{\kappa}_{\alpha,n,v}$ for $n \in \{50, 55, 60, 75, 90, 100, 125, 150, 200, 300, 500\}$, $2 \leq v \leq 13$ and $\alpha \in \{0.01, 0.025, 0.05\}$. We performed 100,000 simulations in the case $n \leq 75$, 75,000 simulations for $n = 90$ and 50,000 simulations if $n \geq 100$, in order to have comparable pooled sample sizes N^* across different values of n.

Figure 1 displays the three-dimensional surfaces of $\hat{\kappa}_{\alpha,n,v}$ for MCD and RMCD as a function of n and v, when $\alpha = 0.01$. It is clear from these pictures that the simulated calibration factors (19) generally decrease with n and increase with v, as already suggested by Table 4, and that there is strong interaction between n and v. Both surfaces are rather flat in the region $n \geq 200$ but increase steeply when $n < 100$. The slope for the MCD test decays more slowly than that for RMCD, the value still being larger than 1.2 at $n = 500$. It is interesting to note that the eastern border of each surface is not smooth, an indication that the behaviour of $\hat{\kappa}_{\alpha,n,v}$ for v even is somewhat different from that for v odd when the sample size is small and v increases. A plausible explanation is the slight truncation effect induced by the definition of h in (18) when $n + v$ is even and n is small. Furthermore, failure of the *FAST-MCD* algorithm to provide the global minimizer of the covariance determinant may enhance this behaviour.

For fixed n, we smooth the relationship between $\hat{\kappa}_{0.01,n,v}$ and v through the exponential function

$$\tilde{\kappa}_{\alpha,n,v} = b_0 + b_1 I(v) + b_2 \exp(b_3 v), \tag{20}$$

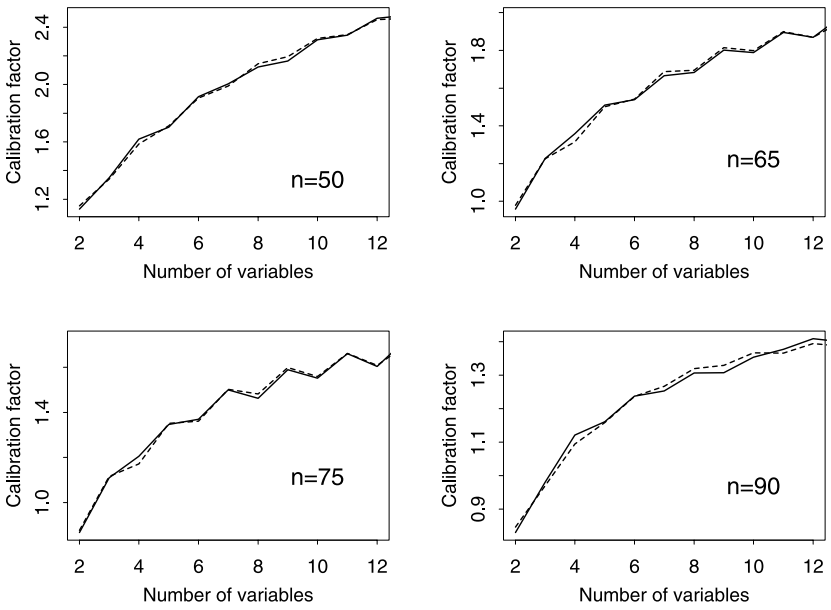

Fig. 2 MCD: relationship between $\ln(\hat{\kappa}_{0.01,n,v})$ and v for different values of n. *Solid lines*: simulated calibration factors; *dashed lines*: fit from (20)

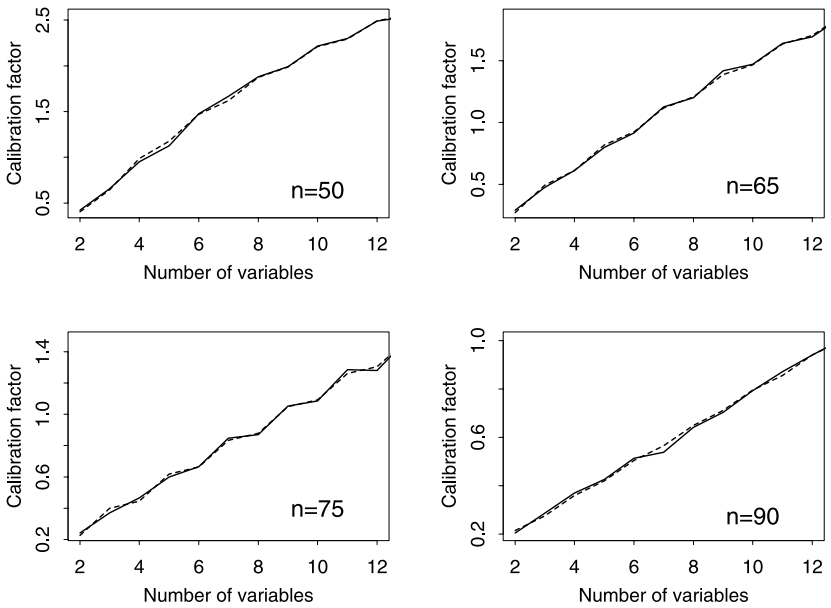

Fig. 3 RMCD: relationship between $\ln(\hat{\kappa}_{0.01,n,v})$ and v for different values of n. *Solid lines*: simulated calibration factors; *dashed lines*: fit from (20)

where $\tilde{\kappa}_{\alpha,n,v} = \ln(\hat{\kappa}_{\alpha,n,v})$ and the parameters b_0, b_1 and b_2 depend on the chosen value of n. The dummy $I(v)$ takes the value 1 if v is odd and 0 otherwise. It accounts for the different behaviour of calibration factors for v odd and n small, as already depicted in Fig. 1. The parameter b_1 is constrained to be 0 if $n > 100$. Function (20) is fitted to the simulated calibration factors by nonlinear least squares and a Gauss code for performing this task is available at http://www.riani.it/mcd. The web site also contains the interpolated calibration factors $\exp(\tilde{\kappa}_{\alpha,n,v})$, for $v = 2, \ldots, 15$, $\alpha \in \{0.01, 0.025, 0.05\}$ and n as in our simulation grid, together with the fitted coefficients b_0, b_1 and b_2. Figures 2 and 3 show the relationship between $\ln(\hat{\kappa}_{0.01,n,v})$ and v for selected values of n, together with the interpolated calibration factors from function (20). From these pictures it is apparent the fit is almost perfect for both MCD and RMCD. The plots for the other values of α are similar and are therefore omitted.

Equation (20) provides a way to compute the required calibration factor for any v, given a value of n and a test size α belonging to our simulation grid. The final step is linear interpolation of the smoothed calibration factors $\exp(\tilde{\kappa}_{\alpha,n,v})$ with respect to n and α, to obtain the desired correction, say $\kappa_{\alpha,n,v}$, for the value of n at hand and for the required size α.

Table 5 Estimated size of the calibrated tests MCD and RMCD under the intersection hypothesis (10) with nominal level $\alpha = 0.01$. Size is computed on 5,000 new simulations for each combination of n and v

		$n = 48$	$n = 52$	$n = 58$	$n = 70$	$n = 82$	$n = 95$	$n = 175$
	MCD	0.008	0.011	0.009	0.012	0.011	0.009	0.007
$v = 6$	RMCD	0.009	0.008	0.008	0.008	0.009	0.009	0.010
	MCD	0.009	0.009	0.009	0.012	0.008	0.007	0.009
$v = 8$	RMCD	0.008	0.008	0.008	0.011	0.007	0.009	0.008
	MCD	0.010	0.010	0.010	0.011	0.010	0.007	0.007
$v = 10$	RMCD	0.010	0.009	0.008	0.010	0.008	0.008	0.010
	MCD	0.008	0.008	0.006	0.005	0.006	0.008	0.007
$v = 11$	RMCD	0.009	0.008	0.007	0.006	0.004	0.009	0.007
	MCD	0.010	0.009	0.011	0.014	0.010	0.007	0.007
$v = 12$	RMCD	0.009	0.010	0.007	0.008	0.007	0.008	0.007
	MCD	0.014	0.014	0.015	0.012	0.008	0.011	0.008
$v = 14$	RMCD	0.013	0.011	0.011	0.009	0.007	0.011	0.010
	MCD	0.016	0.010	0.008	0.010	0.009	0.012	0.010
$v = 15$	RMCD	0.013	0.008	0.005	0.006	0.005	0.010	0.009

4.3 Size of calibrated tests

The adequacy of our parametric calibration procedure is checked through a new Monte Carlo experiment. In this experiment we generate 5,000 independent n-dimensional samples y from the v-variate $N(0, I)$ distribution for a number of values of n and v. As in Sect. 3.2, size is estimated as the proportion of simulated data sets for which hypothesis (10) is rejected. The cut-off squared distance for rejection is now the calibrated threshold

$$d_\gamma^2 = \kappa_{\alpha,n,v}\chi^2_{v,\alpha/n}, \tag{21}$$

with $\kappa_{\alpha,n,v}$ computed following the procedure of Sect. 4.2.

Table 5 summarizes the results for $\alpha = 0.01$ in the more critical situation where $n < 200$ and $v \geq 6$. To avoid the possible danger of overfitting we only focus on sample sizes not belonging to the simulation grid of Sect. 4.2. Indeed, all the reported values of n are intermediate between the grid nodes used in the calibration study. The results are thus intended to show the *worst-case* behaviour of our procedure. In spite of this the performance of the calibrated tests is generally excellent, with estimated sizes close to the nominal 0.01 and moderate conservativeness occurring occasionally when $v > 10$.

It is worth noting that performance is still satisfactory close to, and even beyond, the boundary of the interpolation range, as shown when $v \geq 12$ and $n \leq 52$. This is a remarkable result in view of the sparsity of multivariate space with so few observations per dimension. Due to the curse of dimensionality, Rousseeuw and van Zomeren (1990, p. 649) stated that "*any* outlier method can get into trouble" if n/v is relatively small and, as a rule of thumb, they recommended applying robust multivariate methods only when $n/v > 5$. Table 5 demonstrates that this is not the case for our technique. The null hypothesis of no outliers in the data can be safely tested via (21) in otherwise problematic situations with at most four observations per dimension.

These findings are complemented by Table 6, which shows the estimated sizes of the calibrated tests under the individual hypothesis (9) with nominal level $\alpha = 0.01/n$ for the smallest values of n. The agreement between nominal and actual sizes of individual tests is good, confirming that (18) yields reliable estimates of d_γ^2 even in relatively small samples. Furthermore, as anticipated in Sect. 3.2, the effect of Bonferronization when combining the n individual tests into a simultaneous one is modest and produces only a slight amount of conservativeness.

4.4 Hints on power of calibrated tests

A decrease in power is to be expected when strong control on the size of multiple outlier tests based on robust distances is achieved. Some evidence of this behaviour is given in Riani et al. (2009), who consider the tests MCD-HR and MCD-CH. Here we focus on the calibrated tests of Sect. 4.2, MCDCAL and RMCDCAL say, and compare them to their non-calibrated counterparts. For this purpose, we now simulate n-dimensional samples from a v-variate location-shift model with constant contamination on all variables, $y_i \sim N(\delta e, I)$, where δ is a positive scalar and e is a column-vector of ones. For a specified contamination rate $\omega < 0.5$,

Table 6 Size of the calibrated tests MCD and RMCD under the individual hypothesis (9) with nominal level $\alpha = 0.01/n$. Size is computed on the same simulations as Table 5

		$v = 6$	$v = 8$	$v = 10$	$v = 11$	$v = 12$
$n = 48$	MCD	0.00020	0.00024	0.00024	0.00020	0.00030
$\alpha = 0.00021$	RMCD	0.00033	0.00023	0.00025	0.00024	0.00025
$n = 52$	MCD	0.00023	0.00020	0.00024	0.00020	0.00020
$\alpha = 0.00019$	RMCD	0.00021	0.00020	0.00021	0.00020	0.00022
$n = 58$	MCD	0.00018	0.00018	0.00021	0.00012	0.00024
$\alpha = 0.00017$	RMCD	0.00020	0.00017	0.00020	0.00013	0.00017

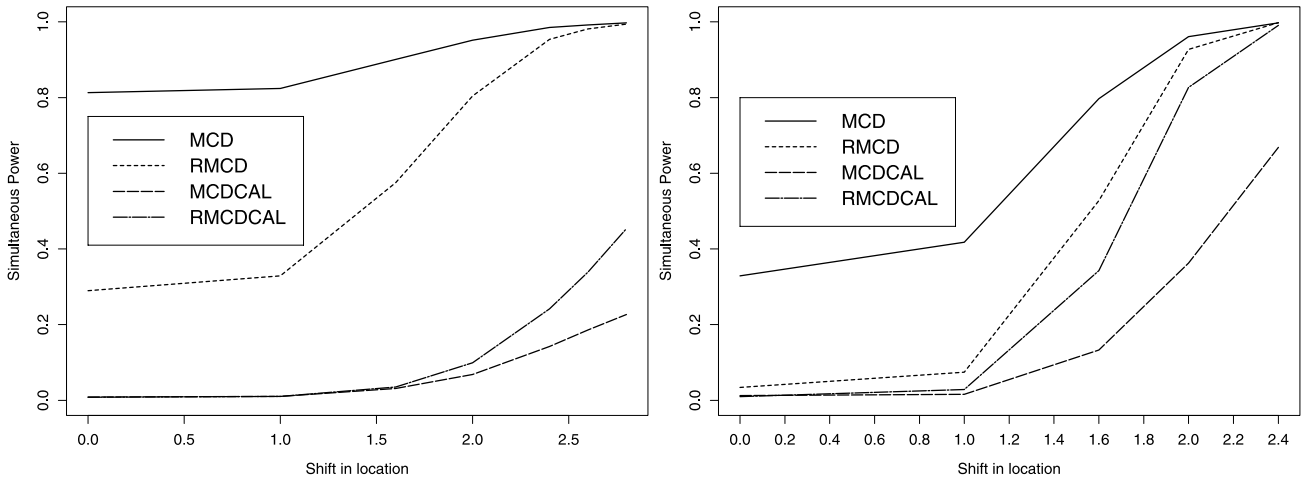

Fig. 4 Simultaneous power for MCD-based tests under a multivariate location-shift model with $v = 5$ and contamination rate $\omega = 0.05$: $n = 60$ (*left*) and $n = 200$ (*right*)

each simulated data set is composed of $100(1 - \omega)\%$ observations from $N(0, I)$ and $100\omega\%$ observations from the contaminated model.

The results on rejection of the simultaneous hypothesis (10) in the case $\omega = 0.05$ are shown in Fig. 4, for two values of n, $v = 5$ and several values of δ. Simultaneous power is defined as the probability of finding at least one outlier and is estimated using 10,000 simulated data sets. Obviously the results for $\delta = 0$ repeat the findings of Table 3, showing that the standard MCD tests have unacceptable error rates in small and moderate samples. The shape of the power function of the calibrated tests is similar to that of their liberal versions, but now, starting from approximately the same size, RMCDCAL clearly outperforms MCDCAL as δ increases. The power of the calibrated tests also increases with n and the gap between RMCDCAL and RMCD reduces as the contamination shift is more pronounced. This gap becomes negligible for $n = 200$ and $\delta \geq 2$.

When the intersection hypothesis (10) has been rejected and at least one outlier has been found, it remains to answer the important question of which of the observations actually come from the contaminated model. Figure 5 reports the average power of the different procedures, defined as the proportion of contaminated observations which are correctly named as outliers. Since these power results can only be obtained under the alternative, the plots now start at $\delta = 1.0$. The main findings remain unaltered, although it is seen that the average power of all the tests is appreciably smaller than simultaneous power for a given contamination shift.

In summary, our power simulations show that the calibrated RMCD test is the one to be recommended among the various MCD procedures considered in this paper. It achieves strong control on the size of the hypothesis of no outliers in the data and has reasonably good power properties for moderate to high contamination.

5 Conclusions

In this paper we have proposed a new calibration method for obtaining reliable cut-off points of robust distances derived from the MCD estimator of scatter, as implemented by Rousseeuw and Van Driessen (1999) and subsequently modified by Pison et al. (2002). We have shown that our proce-

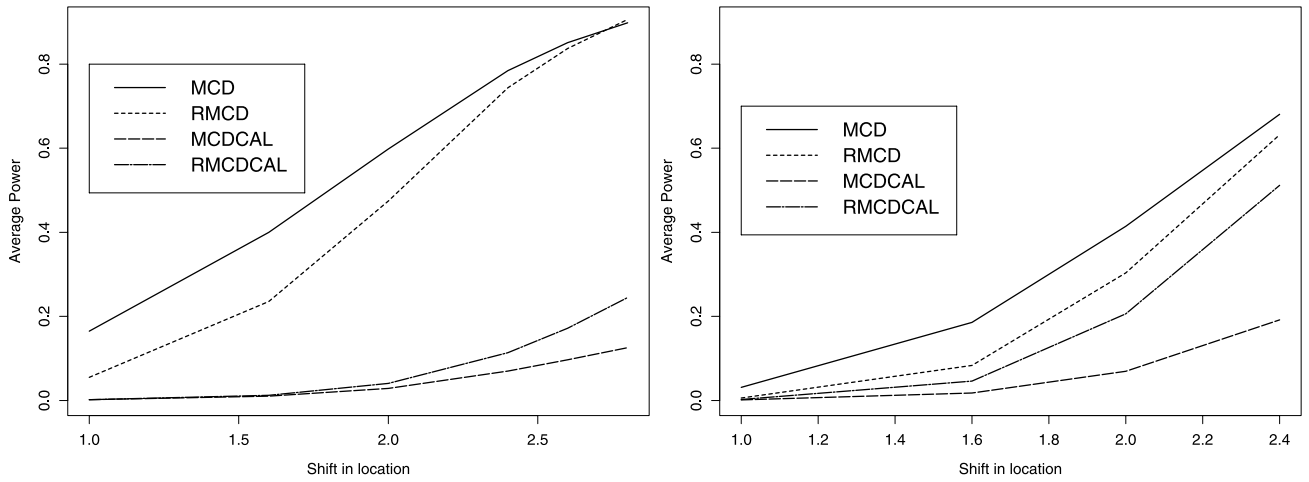

Fig. 5 Average power for MCD-based tests under a multivariate location-shift model with $v = 5$ and contamination rate $\omega = 0.05$: $n = 60$ (*left*) and $n = 200$ (*right*)

dure gives reliable tests of outlyingness in almost all situations of practical interest, provided that the sample size is not much smaller than 50. It is a considerable improvement over all the available MCD procedures, even the adjusted versions of Croux and Haesbroeck (1999) and Hardin and Rocke (2005), which were seen to be unable to provide good control over the size of multiple outlier tests for the data structures considered here. Furthermore, our corrections are also available for the more efficient reweighted MCD test not considered by Hardin and Rocke (2005). On the other hand, additional work should be carried out in order to obtain appropriate calibration factors for very small samples.

We have been mainly concerned with the problem of detecting multiple outliers in a sample. We have addressed the problem of multiplicity of tests through a Bonferroni approach. Although we have shown that conservativeness is not a matter of great concern due to approximate independence of individual tests and that our calibrated tests have reasonable power properties, there might be more efficient ways to address simultaneity in the present context. The development of more powerful procedures is the subject of ongoing research.

Acknowledgements We thank an Associate Editor and two anonymous referees for helpful comments. We are also grateful to Dr. Valentin Todorov for providing advance drafts of Todorov (2008) and Todorov and Filzmoser (2008). Our work was supported by the grants "Metodi statistici multivariati per la valutazione integrata della qualità dei servizi di pubblica utilità: efficacia-efficienza, rischio del fornitore, soddisfazione degli utenti" and "Metodologie statistiche per l'analisi di impatto e la valutazione della regolamentazione" of Ministero dell'Università e della Ricerca—PRIN 2006.

References

Arsenis, S., Perrotta, D., Torti, F.: Price outliers in EU external trade data. Internal working document on work presented at the "Enlargement and Integration Workshop 2005", Joint Research Centre of the European Commission, http://theseus.jrc.it/index.php?id=1298 (2005)

Becker, C., Gather, U.: The masking breakdown point of multivariate outlier identification rules. J. Am. Stat. Assoc. **94**, 947–955 (1999)

Becker, C., Gather, U.: The largest nonidentifiable outlier: a comparison of multivariate simultaneous outlier identification rules. Comput. Stat. Data Anal. **36**, 119–127 (2001)

Butler, R.W., Davies, P.L., Jhun, M.: Asymptotics for the minimum covariance determinant estimator. Ann. Stat. **21**, 1385–1400 (1993)

Cohen Freue, G.V., Hollander, Z., Shen, E., Zamar, R.H., Balshaw, R., Scherer, A., McManus, B., Keown, P., McMaster, W.R., Ng, R.T.: MDQC: A new quality assessment method for microarrays based on quality control reports. Bioinformatics **23**, 3162–3169 (2007)

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

Croux, H., Haesbroeck, G.: Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies. Biometrika **87**, 603–618 (2000)

Hardin, J., Rocke, D.M.: The distribution of robust distances. J. Comput. Graph. Stat. **14**, 910–927 (2005)

Lopuhaä, H.P.: Asymptotics of reweighted estimators of multivariate location and scatter. Ann. Stat. **27**, 1638–1665 (1999)

Matsumoto, M., Nishimura, T.: Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans. Model. Comput. Simul. **8**, 3–30 (1998)

Pison, G., Van Aelst, S.: Diagnostic plots for robust multivariate methods. J. Comput. Graph. Stat. **13**, 310–329 (2004)

Pison, G., Van Aelst, S., Willems, G.: Small sample corrections for LTS and MCD. Metrika **55**, 111–123 (2002)

Riani, M., Cerioli, A., Atkinson, A., Perrotta, D., Torti, F.: Fitting mixtures of regression lines with the Forward Search. In: Fogelman-Soulié, F., Perrotta, D., Piskorski, J., Steinberger, R. (eds.) Mining Massive Data Sets for Security. IOS Press, Amsterdam (2008)

Riani, M., Atkinson, A.C., Cerioli, A.: Finding an unknown number of multivariate outliers. J. R. Stat. Soc. Ser. B **71** (2009)

Rousseeuw, P.J., Leroy, A.M.: Robust Regression and Outlier Detection. Wiley, New York (1987)

Rousseeuw, P.J., Van Driessen, K.: A fast algorithm for the minimum covariance determinant estimator. Technometrics **41**, 212–223 (1999)

Rousseeuw, P.J., Van Zomeren, B.C.: Unmasking multivariate outliers and leverage points. J. Am. Stat. Assoc. **85**, 633–9 (1990)

Rousseeuw, P.J., Van Aelst, S., Van Driessen, K., Agulló, J.: Robust multivariate regression. Technometrics **46**, 293–305 (2004)

Šidák, Z.: Rectangular confidence regions for the means of multivariate normal distributions. J. Am. Stat. Assoc. **62**, 626–633 (1967)

Todorov, V.: Robust selection of variables in linear discriminant analysis. Stat. Methods Appl. **15**, 395–407 (2006)

Todorov, V.: A note on the MCD consistency and small sample correction factors. Unpublished manuscript (2008, in preparation)

Todorov, V., Filzmoser, P.: Robust statistics for the one-way MANOVA. Unpublished manuscript (2008, submitted for publication)

Willems, G., Pison, G., Rousseeuw, P.J., Van Aelst, S.: A robust Hotelling test. Metrika **55**, 125–138 (2002)