Received: 20 September 2017 | Revised: 26 October 2017 | Accepted: 27 October 2017

DOI: 10.1002/wics.1421

WILEY WIREs COMPUTATIONAL STATISTICS

ADVANCED REVIEW

Minimum covariance determinant and extensions

Mia Hubert[1] | Michiel Debruyne[2] | Peter J. Rousseeuw[1]

[1]Department of Mathematics, KU Leuven, Leuven, Belgium

[2]Dexia Bank, Brussels, Belgium

Correspondence
Mia Hubert, Department of Mathematics, KU Leuven, Celestijnenlaan 200B, BE-3001 Leuven, Belgium.
Email: mia.hubert@kuleuven.be

The minimum covariance determinant (MCD) method is a highly robust estimator of multivariate location and scatter, for which a fast algorithm is available. Since estimating the covariance matrix is the cornerstone of many multivariate statistical methods, the MCD is an important building block when developing robust multivariate techniques. It also serves as a convenient and efficient tool for outlier detection. The MCD estimator is reviewed, along with its main properties such as affine equivariance, breakdown value, and influence function. We discuss its computation, and list applications and extensions of the MCD in applied and methodological multivariate statistics. Two recent extensions of the MCD are described. The first one is a fast deterministic algorithm which inherits the robustness of the MCD while being almost affine equivariant. The second is tailored to high-dimensional data, possibly with more dimensions than cases, and incorporates regularization to prevent singular matrices.

This article is categorized under:
Statistical and Graphical Methods of Data Analysis > Multivariate Analysis
Statistical and Graphical Methods of Data Analysis > Robust Methods
Statistical Learning and Exploratory Methods of the Data Sciences > Knowledge Discovery

KEYWORDS

algorithms, covariance matrix, multivariate statistics, outlier detection, robust estimation

1 | INTRODUCTION

The minimum covariance determinant (MCD) estimator is one of the first affine equivariant and highly robust estimators of multivariate location and scatter (Rousseeuw, 1984, 1985). Being resistant to outlying observations makes the MCD very useful for outlier detection. Although already introduced in 1984, its main use has only started since the construction of the computationally efficient FastMCD algorithm of Rousseeuw & Van Driessen (1999). Since then, the MCD has been applied in numerous fields such as medicine, finance, image analysis, and chemistry. Moreover, the MCD has also been used to develop many robust multivariate techniques, which includes robust principal component analysis, factor analysis, and multiple regression. Recent modifications of the MCD include a deterministic algorithm and a regularized version for high-dimensional data.

2 | DESCRIPTION OF THE MCD ESTIMATOR

2.1 | Motivation

In the multivariate location and scatter setting the data are stored in an $n \times p$ data matrix $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ with $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$ the ith observation, so n stands for the number of objects and p for the number of variables. We assume that the observations are

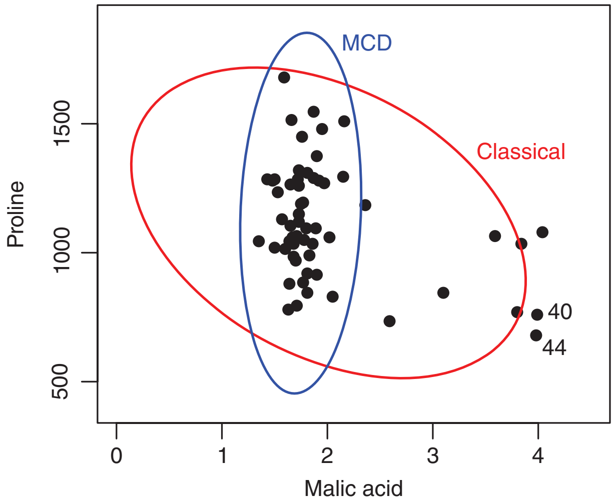

FIGURE 1 Bivariate wine data: tolerance ellipse of the classical mean and covariance matrix (red), and that of the robust location and scatter matrix (blue)

sampled from an elliptically symmetric unimodal distribution with unknown parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, where $\boldsymbol{\mu}$ is a vector with p components and $\boldsymbol{\Sigma}$ is a positive definite $p \times p$ matrix. To be precise, a multivariate distribution is called elliptically symmetric and unimodal if there exists a strictly decreasing real function g such that the density can be written in the form:

$$f(\mathbf{x}) = \frac{1}{\sqrt{|\boldsymbol{\Sigma}|}} g\left(d^2(\boldsymbol{x},\boldsymbol{\mu},\boldsymbol{\Sigma})\right) \tag{1}$$

in which the *statistical distance* $d(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by:

$$d(\boldsymbol{x},\boldsymbol{\mu},\boldsymbol{\Sigma}) = \sqrt{(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}. \tag{2}$$

To illustrate the MCD, we first consider the wine data set available in Hettich and Bay (1999) and also analyzed in Maronna, Martin, and Yohai (2006). This data set contains the quantities of 13 constituents found in three types of Italian wines. We consider the first group containing 59 wines, and focus on the constituents "Malic acid" and "Proline." This yields a bivariate data set, that is, $p = 2$. A scatter plot of the data is shown in Figure 1, in which we see that the points on the lower right-hand side of the plot are outlying relative to the majority of the data.

In the figure we see two ellipses. The classical tolerance ellipse is defined as the set of p-dimensional points $\boldsymbol{x}$ whose *Mahalanobis distance* is:

$$\mathrm{MD}(\boldsymbol{x}) = d(\boldsymbol{x},\overline{\boldsymbol{x}},\mathrm{Cov}(\boldsymbol{X})) = \sqrt{(\boldsymbol{x}-\overline{\boldsymbol{x}})'\mathrm{Cov}(\boldsymbol{X})^{-1}(\boldsymbol{x}-\overline{\boldsymbol{x}})} \tag{3}$$

equals $\sqrt{\chi^2_{p,0.975}}$. Here $\overline{\boldsymbol{x}}$ is the sample mean and $\mathrm{Cov}(\boldsymbol{X})$ the sample covariance matrix. The Mahalanobis distance $\mathrm{MD}(\boldsymbol{x}_i)$ should tell us how far away $\boldsymbol{x}_i$ is from the center of the data cloud, relative to its size and shape. In Figure 1 we see that the red tolerance

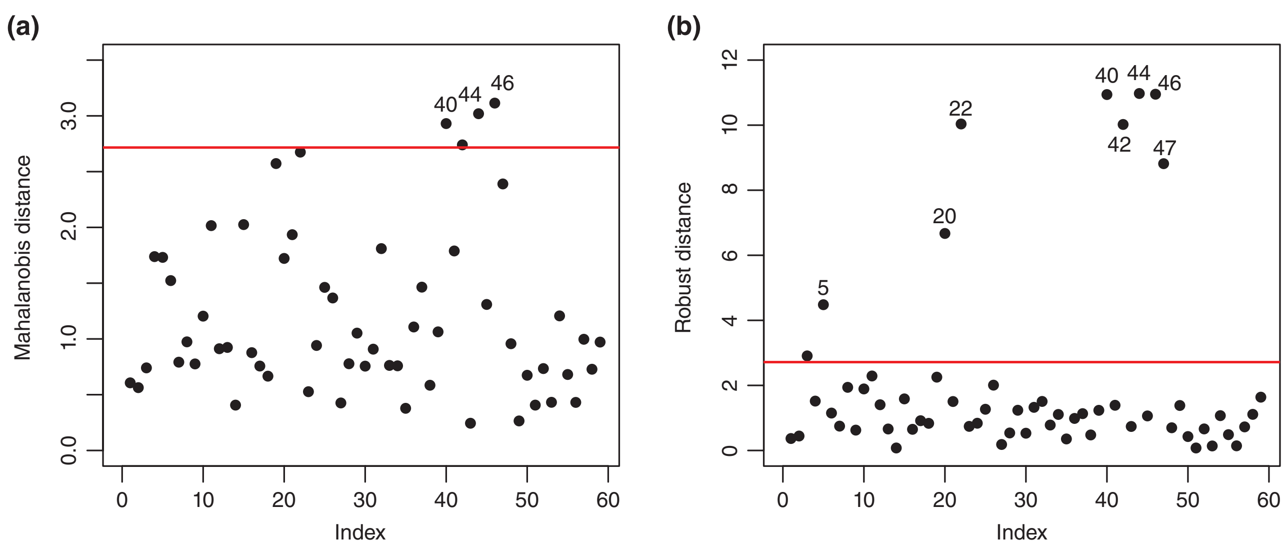

FIGURE 2 (a) Mahalanobis distances and (b) robust distances for the bivariate wine data

ellipse tries to encompass all observations. Therefore, none of the Mahalanobis distances is exceptionally large, as we can see in Figure 2a. Based on Figure 2a alone we would say there are only three mild outliers in the data (we ignore borderline cases).

On the other hand, the robust tolerance ellipse is based on the robust distances:

$$\mathrm{RD}(\boldsymbol{x}) = d\left(\boldsymbol{x}, \hat{\boldsymbol{\mu}}_{\mathrm{MCD}}, \hat{\boldsymbol{\Sigma}}_{\mathrm{MCD}}\right), \tag{4}$$

where $\hat{\boldsymbol{\mu}}_{\mathrm{MCD}}$ is the MCD estimate of location and $\hat{\boldsymbol{\Sigma}}_{\mathrm{MCD}}$ is the MCD covariance estimate, which we will explain in Section 2.2. In Figure 1 we see that the robust ellipse (in blue) is much smaller and only encloses the regular data points. The robust distances shown in Figure 2b now clearly expose eight outliers.

This illustrates the *masking effect*: the classical estimates can be so strongly affected by contamination that diagnostic tools such as the Mahalanobis distances become unable to detect the outliers. To avoid masking we instead need reliable estimators that can resist outliers when they occur. The MCD is such a robust estimator.

2.2 | Definition

The raw MCD estimator with tuning constant $n/2 \le h \le n$ is $\left(\hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0\right)$, where

1. the location estimate $\hat{\boldsymbol{\mu}}_0$ is the mean of the h observations for which the determinant of the sample covariance matrix is as small as possible;
2. the scatter matrix estimate $\hat{\boldsymbol{\Sigma}}_0$ is the corresponding covariance matrix multiplied by a consistency factor c_0.

Note that the MCD estimator can only be computed when $h > p$, otherwise the covariance matrix of any h-subset has determinant 0, so we need at least $n > 2p$. To avoid excessive noise it is, however, recommended that $n > 5p$, so that we have at least five observations per dimension. (When this condition is not satisfied one can instead use the *Minimum Regularized Covariance Determinant* (MRCD) method (Equation (11)) described in Section 7.2.) To obtain consistency at the normal distribution, the consistency factor, c_0, equals $\alpha / F_{\chi^2_{p+2}}(q_\alpha)$ with $\alpha = \lim_{n\to\infty} h(n)/n$, and q_α the α-quantile of the χ^2_p distribution (Croux & Haesbroeck, 1999). Also a finite-sample correction factor can be incorporated (Pison, Van Aelst, & Willems, 2002).

Consistency of the raw MCD estimator of location and scatter at elliptical models, as well as asymptotic normality of the MCD location estimator has been proved in Butler, Davies, and Jhun (1993). Consistency and asymptotic normality of the MCD covariance matrix at a broader class of distributions is derived in Cator & Lopuhaä, (2010, 2012).

The MCD estimator is the most robust when taking $h = [(n + p + 1)/2]$ where $[a]$ is the largest integer $\le a$. At the population level this corresponds to $\alpha = .5$. But unfortunately, the MCD then suffers from low efficiency at the normal model. For example, if $\alpha = .5$ the asymptotic relative efficiency of the diagonal elements of the MCD scatter matrix relative to the sample covariance matrix is only 6% when $p = 2$, and 20.5% when $p = 10$. This efficiency can be increased by considering a higher α such as $\alpha = .75$. This yields relative efficiencies of 26.2% for $p = 2$ and 45.9% for $p = 10$, see Croux and Haesbroeck (1999). On the other hand, this choice of α diminishes the robustness to possible outliers.

In order to increase the efficiency while retaining high robustness one can apply a weighting step (Lopuhaä, 1999; Lopuhaä & Rousseeuw, 1991). For the MCD this yields the estimates:

$$\hat{\boldsymbol{\mu}}_{\mathrm{MCD}} = \frac{\sum_{i=1}^{n} W\left(d_i^2\right)\boldsymbol{x}_i}{\sum_{i=1}^{n} W\left(d_i^2\right)}, \; \hat{\boldsymbol{\Sigma}}_{\mathrm{MCD}} = c_1 \frac{1}{n} \sum_{i=1}^{n} W\left(d_i^2\right)\left(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_{\mathrm{MCD}}\right)\left(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}_{\mathrm{MCD}}\right)' \tag{5}$$

with $d_i = d\left(\boldsymbol{x}, \hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0\right)$ and W an appropriate weight function. The constant c_1 is again a consistency factor. A simple yet effective choice for W is to set it to 1 when the robust distance is below the cutoff $\sqrt{\chi^2_{p,0.975}}$ and to 0 otherwise, that is, $W(d^2) = I\left(d^2 \le \chi^2_{p,0.975}\right)$. This is the default choice in the current implementations in R, SAS, Matlab and S-PLUS. If we take $\alpha = .5$ this weighting step increases the efficiency to 45.5% for $p = 2$ and to 82% for $p = 10$. In the example of the wine data (Figure 1) we applied the weighted MCD estimator with $\alpha = .75$, but the results were similar for smaller values of α.

Note that one can construct a robust correlation matrix from the MCD scatter matrix. The robust correlation between variables X_i and X_j is given by:

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}\, s_{jj}}}$$

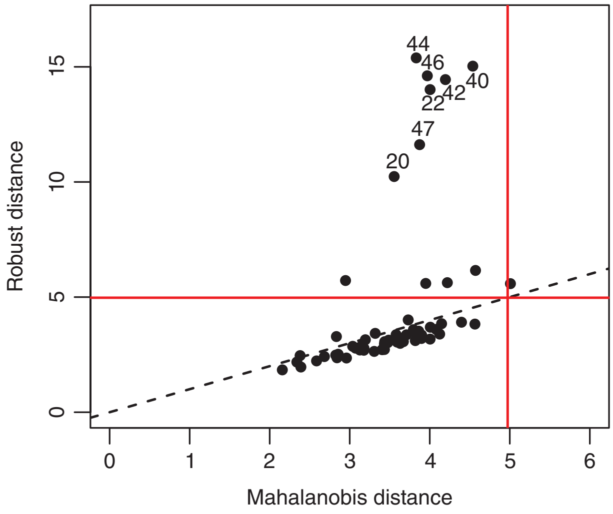

FIGURE 3 Distance-distance plot of the full 13-dimensional wine data set

with s_{ij} the (i, j)th element of the MCD scatter matrix. In Figure 1, the MCD-based robust correlation is $0.10 \approx 0$ because the majority of the data do not show a trend, whereas the classical correlation of -0.37 was caused by the outliers in the lower right part of the plot.

2.3 | Outlier detection

As already illustrated in Figure 2, the robust MCD estimator is very useful to detect outliers in multivariate data. As the robust distances (Equation (4)) are not sensitive to the masking effect, they can be used to flag the outliers (Cerioli, 2010; Rousseeuw & van Zomeren, 1990). This is crucial for data sets in more than three dimensions, which are difficult to visualize.

We illustrate the outlier detection potential of the MCD on the full wine data set, with all $p = 13$ variables. The *distance-distance plot* of Rousseeuw and Van Driessen (1999) in Figure 3 shows the robust distances based on the MCD versus the classical distances (Equation (3)). From the robust analysis we see that seven observations clearly stand out (plus some mild outliers), whereas the classical analysis does not flag any of them.

Note that the cutoff value $\sqrt{\chi^2_{p,0.975}}$ is based on the asymptotic distribution of the robust distances, and often flags too many observations as outlying. For relatively small n the true distribution of the robust distances can be better approximated by an F-distribution, see Hardin and Rocke (2005).

3 | PROPERTIES

3.1 | Affine equivariance

The MCD estimator of location and scatter is *affine equivariant*. This means that for any nonsingular $p \times p$ matrix $\boldsymbol{A}$ and any p-dimensional column vector $\boldsymbol{b}$ it holds that

$$\hat{\boldsymbol{\mu}}_{\text{MCD}}(\boldsymbol{X}\boldsymbol{A}' + \mathbf{1}_n\boldsymbol{b}') = \hat{\boldsymbol{\mu}}_{\text{MCD}}(\boldsymbol{X})\boldsymbol{A}' + \boldsymbol{b}, \tag{6}$$

$$\hat{\boldsymbol{\Sigma}}_{\text{MCD}}(\boldsymbol{X}\boldsymbol{A}' + \mathbf{1}_n\boldsymbol{b}') = = \boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\text{MCD}}(\boldsymbol{X})\boldsymbol{A}', \tag{7}$$

where the vector $\mathbf{1}_n$ is $(1, 1, \ldots, 1)'$ with n elements. This property follows from the fact that for each subset H of $\{1, 2, \ldots, n\}$ of size h and corresponding data set $\boldsymbol{X}_H$, the determinant of the covariance matrix of the transformed data equals:

$$|\boldsymbol{S}(\boldsymbol{X}_H\boldsymbol{A}')| = |\boldsymbol{A}\boldsymbol{S}(\boldsymbol{X}_H)\boldsymbol{A}'| = |\boldsymbol{A}|^2 |\boldsymbol{S}(\boldsymbol{X}_H)|.$$

Therefore, transforming an h-subset with lowest determinant yields an h-subset $\boldsymbol{X}_H\boldsymbol{A}'$ with lowest determinant among all h-subsets of the transformed data set $\boldsymbol{X}\boldsymbol{A}'$, and its covariance matrix is transformed appropriately. The affine equivariance of the raw MCD location estimator follows from the equivariance of the sample mean. Finally, we note that the robust distances $d_i = d\left(\boldsymbol{x}, \hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0\right)$ are affine *invariant*, meaning they stay the same after transforming the data, which implies that the weighted estimator is affine equivariant too.

Affine equivariance implies that the estimator transforms well under any nonsingular reparametrization of the space in which the $\boldsymbol{x}_i$ live. Consequently, the data might be rotated, translated or rescaled (e.g., through a change of measurement units) without affecting the outlier detection diagnostics.

The MCD is one of the first high-breakdown affine equivariant estimators of location and scatter, and was only preceded by the Stahel-Donoho estimator (Donoho & Gasko, 1992; Stahel, 1981). Together with the MCD also the Minimum Volume Ellipsoid estimator was introduced (Rousseeuw, 1984, 1985) which is equally robust but not asymptotically normal, and is harder to compute than the MCD.

3.2 | Breakdown value

The breakdown value of an estimator is the smallest fraction of observations that need to be replaced (by arbitrary values) to make the estimate useless. For a multivariate *location* estimator, T_n, the breakdown value is defined as:

$$\varepsilon_n^*(T_n;\boldsymbol{X}_n)=\frac{1}{n}\min\{m:\sup\|T_n(\boldsymbol{X}_{n,m})-T_n(\boldsymbol{X}_n)\|=+\infty\},$$

where $1 \le m \le n$ and the supremum is over all data sets $\boldsymbol{X}_{n,m}$ obtained by replacing any m data points $\boldsymbol{x}_{i_1},\ldots,\boldsymbol{x}_{i_m}$ of $\boldsymbol{X}_n$ by arbitrary points.

For a multivariate *scatter* estimator, C_n, we set

$$\varepsilon_n^*(C_n;\boldsymbol{X}_n)=\frac{1}{n}\min\left\{m:\sup\ \max_i|\log(\lambda_i(C_n(\boldsymbol{X}_{n,m})))-\log(\lambda_i(C_n(\boldsymbol{X}_n)))|=+\infty\right\}$$

with $\lambda_1(C_n) \ge \cdots \ge \lambda_p(C_n) > 0$ the eigenvalues of C_n. This means that we consider a scatter estimator to be broken when λ_1 can become arbitrarily large ("explosion") and/or λ_p can become arbitrary close to 0 ("implosion"). Implosion is a problem because it makes the scatter matrix singular whereas in many situations its inverse is required, for example, in Equation (4).

Let $k(\boldsymbol{X}_n)$ denote the highest number of observations in the data set that lie on an affine hyperplane in p-dimensional space, and assume $k(\boldsymbol{X}_n) < h$. Then the raw MCD estimator of location and scatter satisfies (Roelant, Van Aelst, & Willems, 2009):

$$\varepsilon_n^*(\hat{\boldsymbol{\mu}}_0;\boldsymbol{X}_n)=\varepsilon_n^*\left(\hat{\boldsymbol{\Sigma}}_0;\boldsymbol{X}_n\right)=\frac{\min(n-h+1,h-k(\boldsymbol{X}_n))}{n}. \tag{8}$$

If the data are sampled from a continuous distribution, then almost surely $k(\boldsymbol{X}_n) = p$ which is called *general position.* Then $\varepsilon_n^*(\hat{\boldsymbol{\mu}}_0;\boldsymbol{X}_n)=\varepsilon_n^*\left(\hat{\boldsymbol{\Sigma}}_0;\boldsymbol{X}_n\right)=\min(n-h+1,h-p)/n$, and consequently any $[(n+p)/2] \le h \le [(n+p+1)/2]$ gives the breakdown value $[(n-p+1)/2]$. This is the highest possible breakdown value for affine equivariant scatter estimators (Davies, 1987) at data sets in general position. Also for affine equivariant location estimators the upper bound on the breakdown value is $[(n-p+1)/2]$ under natural regularity conditions (Rousseeuw, 2005). Note that in the limit $\lim_{n\to\infty}\varepsilon_n^*=\min(1-\alpha,\alpha)$ which is maximal for $\alpha = .5$.

Finally, we note that the breakdown value of the weighted MCD estimators $\hat{\boldsymbol{\mu}}_{\text{MCD}}$ and $\hat{\boldsymbol{\Sigma}}_{\text{MCD}}$ is not lower than the breakdown value of the raw MCD estimator, as long as the weight function W used in Equation (5) is bounded and becomes 0 for large d_i, see Lopuhaä and Rousseeuw (1991).

3.3 | Influence function

The influence function (Hampel, Ronchetti, Rousseeuw, & Stahel, 1986) of an estimator measures the effect of a small (infinitesimal) fraction of outliers placed at a given point. It is defined at the population level hence it requires the functional form of the estimator T, which maps a distribution F to a value $T(F)$ in the parameter space. For multivariate location this parameter space is IR^p, whereas for multivariate scatter the parameter space is the set of all positive semidefinite $p \times p$ matrices. The influence function of the estimator T at the distribution F in a point $\boldsymbol{x}$ is then defined as:

$$IF(\boldsymbol{x},T,F)=\lim_{\varepsilon\to 0}\frac{T(F_\varepsilon)-T(F)}{\varepsilon} \tag{9}$$

with $F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_{\boldsymbol{x}}$ a contaminated distribution with point mass in $\boldsymbol{x}$.

The influence function of the raw and the weighted MCD has been computed in Croux and Haesbroeck (1999) and Cator and Lopuhaä (2012) and turns out to be bounded. This is a desirable property for robust estimators, as it limits the effect of a small fraction of outliers on the estimate. At the standard multivariate normal distribution, the influence function of the MCD location estimator becomes 0 for all $\boldsymbol{x}$ with $\|\boldsymbol{x}\|^2>\chi^2_{p,\alpha}$ hence far outliers do not influence the estimates at all. The same happens with the off-diagonal elements of the MCD scatter estimator. On the other hand, the influence function of the diagonal elements remains constant (different from 0) when $\|\boldsymbol{x}\|^2$ is sufficiently large. Therefore, the outliers still have a bounded influence on the estimator. All these influence functions are smooth, except at those $\boldsymbol{x}$ with $\|\boldsymbol{x}\|^2 = \chi^2_{p,\alpha}$. The weighted MCD estimator has an additional jump in $\|\boldsymbol{x}\|^2 = \chi^2_{p,0.975}$ due to the discontinuity of the weight function, but one could use a smooth weight function instead.

3.4 | Univariate MCD

For univariate data $x_1, \ldots, x_n$, the MCD estimates reduce to the mean and the standard deviation of the h-subset with smallest variance. They can be computed in $O(n \log n)$ time by sorting the observations and only considering contiguous h-subsets so that their means and variances can be calculated recursively (Rousseeuw & Leroy, 1987). For $h = [n/2] + 1$ the MCD location estimator has breakdown value $[(n + 1)/2]/n$ and the MCD scale estimator has $[n/2]/n$. These are the highest breakdown values that can be attained by univariate affine equivariant estimators (Croux & Rousseeuw, 1992). The univariate MCD estimators also have bounded influence functions, see Croux and Haesbroeck (1999) for details. Their maximal asymptotic bias is studied in Croux and Haesbroeck (2001, 2002) as a function of the contamination fraction.

Note that in the univariate case the MCD estimator corresponds to the least trimmed squares (LTS) regression estimator (Rousseeuw, 1984), which is defined as:

$$\hat{\beta}_{\mathrm{LTS}} = \underset{\mu}{\operatorname{argmin}} \sum_{i=1}^{h} \left(r_\beta^2\right)_{i:n}, \tag{10}$$

where $\left(r_\beta^2\right)_{1:n} \leq \left(r_\beta^2\right)_{2:n} \leq \cdots \leq \left(r_\beta^2\right)_{n:n}$ are the ordered squared residuals. For univariate data these residuals are simply $(r_\beta)_i = x_i - \beta$.

4 | COMPUTATION

The exact MCD estimator is very hard to compute, as it requires the evaluation of all $\binom{n}{h}$ subsets of size h. Therefore, one switches to an approximate algorithm such as the FastMCD algorithm of Rousseeuw and Van Driessen (1999) which is quite efficient. The key component of the algorithm is the C-step:

Theorem. *Take* $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ *and let* $H_1 \subset \{1, \ldots, n\}$ *be a subset of size h. Put* $\hat{\boldsymbol{\mu}}_1$ *and* $\hat{\boldsymbol{\Sigma}}_1$ *the empirical mean and covariance matrix of the data in* H_1*. If* $|\hat{\boldsymbol{\Sigma}}_1| \neq 0$ *define the relative distances* $d_1(i) := \mathrm{d}(\boldsymbol{x}_i, \hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\Sigma}}_1)$ *for* $i = 1, \ldots, n$*. Now take* H_2 *such that* $\{d_1(i); i \in H_2\} := \{(d_1)_{1:n}, \ldots, (d_1)_{h:n}\}$ *where* $(d_1)_{1:n} \leq (d_1)_{2:n} \leq \cdots \leq (d_1)_{n:n}$ *are the ordered distances, and compute* $\hat{\boldsymbol{\mu}}_2$ *and* $\hat{\boldsymbol{\Sigma}}_2$ *based on* H_2*. Then* $|\hat{\boldsymbol{\Sigma}}_2| \leq |\hat{\boldsymbol{\Sigma}}_1|$ *with equality if and only if* $\hat{\boldsymbol{\mu}}_2 = \hat{\boldsymbol{\mu}}_1$ *and* $\hat{\boldsymbol{\Sigma}}_2 = \hat{\boldsymbol{\Sigma}}_1$.

If $|\hat{\boldsymbol{\Sigma}}_1| > 0$, the C-step thus easily yields a new h-subset with lower covariance determinant. Note that the C stands for "concentration" since $\hat{\boldsymbol{\Sigma}}_2$ is more concentrated (has a lower determinant) than $\hat{\boldsymbol{\Sigma}}_1$. The condition $|\hat{\boldsymbol{\Sigma}}_1| \neq 0$ in the theorem is no real restriction because if $|\hat{\boldsymbol{\Sigma}}_1| = 0$ the minimal objective value is already attained (and in fact the h-subset H_1 lies on an affine hyperplane).

C-steps can be iterated until $|\hat{\boldsymbol{\Sigma}}_{\mathrm{new}}| = |\hat{\boldsymbol{\Sigma}}_{\mathrm{old}}|$. The sequence of determinants obtained in this way must converge in a finite number of steps because there are only finitely many h-subsets, and in practice converges quickly. However, there is no guarantee that the final value $|\hat{\boldsymbol{\Sigma}}_{\mathrm{new}}|$ of the iteration process is the global minimum of the MCD objective function. Therefore, an approximate MCD solution can be obtained by taking many initial choices of H_1 and applying C-steps to each, keeping the solution with lowest determinant.

To construct an initial subset H_1 one draws a random $(p + 1)$-subset J and computes its empirical mean $\hat{\boldsymbol{\mu}}_0$ and covariance matrix $\hat{\boldsymbol{\Sigma}}_0$. (If $|\hat{\boldsymbol{\Sigma}}_0| = 0$ then J can be extended by adding observations until $|\hat{\boldsymbol{\Sigma}}_0| > 0$.) Then the distances $d_0^2(i) := d^2\left(\boldsymbol{x}_i, \hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0\right)$ are computed for $i = 1, \ldots, n$ and sorted. The initial subset H_1 then consists of the h observations with smallest distance d_0. This method yields better initial subsets than drawing random h-subsets directly, because the probability of drawing an outlier-free $(p + 1)$-subset is much higher than that of drawing an outlier-free h-subset.

The FastMCD algorithm contains several computational improvements. Since each C-step involves the calculation of a covariance matrix, its determinant and the corresponding distances, using fewer C-steps considerably improves the speed of the algorithm. It turns out that after two C-steps, many runs that will lead to the global minimum already have a rather small determinant. Therefore, the number of C-steps is reduced by applying only two C-steps to each initial subset and selecting the 10 subsets with lowest determinants. Only for these 10 subsets further C-steps are taken until convergence.

This procedure is very fast for small sample sizes n, but when n grows the computation time increases due to the n distances that need to be calculated in each C-step. For large n FastMCD partitions the data set, which avoids doing all calculations on the entire data set.

Note that the FastMCD algorithm is itself affine equivariant. Implementations of the FastMCD algorithm are available in R (as part of the packages *rrcov*, *robust* and *robustbase*), in SAS/IML (SAS Institute, Cary, NC) Version 7 and SAS (SAS Institute, Cary, NC) Version 9 (in *PROC ROBUSTREG*), and in S-PLUS (TIBCO Software Inc, Palo Alto, CA) (as the built-in function *cov.mcd*). There is also a MATLAB (Mathworks, Natick, MA) version in LIBRA (LIBrary for Robust Analysis, Verboven & Hubert, 2005, 2010) which can be downloaded from http://wis.kuleuven.be/stat/robust. Moreover, it is available in the PLS toolbox of Eigenvector Research (http://www.eigenvector.com). Note that some MCD functions use $\alpha = .5$ by default, yielding a breakdown value of 50%, whereas other implementations use $\alpha = .75$. Of course, α can always be set by the user.

5 | APPLICATIONS

There are many applications of the MCD, for instance, in finance and econometrics (Gambacciani & Paolella, 2017; Welsh & Zhou, 2007; Zaman, Rousseeuw, & Orhan, 2001), medicine (Prastawa, Bullitt, Ho, & Gerig, 2004), quality control (Jensen, Birch, & Woodal, 2007), geophysics (Neykov, Neytchev, Van Gelder, & Todorov, 2007), geochemistry (Filzmoser, Garrett, & Reimann, 2005), image analysis (Lu, Wang, Kong, Zhang, & Zhang, 2006; Vogler, Goldenstein, Stolfi, Pavlovic, & Metaxas, 2007) and chemistry (van Helvoort, Filzmoser, & van Gaans, 2005), but this list is far from complete.

6 | MCD-BASED MULTIVARIATE METHODS

Many multivariate statistical methods rely on covariance estimation, hence, the MCD estimator is well-suited for constructing robust multivariate techniques. Moreover, the trimming idea of the MCD and the C-step have been generalized to many new estimators. Here we list some applications and extensions.

The MCD analog in regression is the LTS regression estimator (Rousseeuw, 1984) which minimizes the sum of the h-smallest-squared residuals (Equation (10)). Equivalently, the LTS estimate corresponds to the least squares fit of the h-subset with smallest sum of squared residuals. The FastLTS algorithm (Rousseeuw & Van Driessen, 2006) uses techniques similar to FastMCD. The outlier map introduced in Rousseeuw and van Zomeren (1990) plots the robust regression residuals versus the robust distances of the predictors, and is very useful for classifying outliers, see also Rousseeuw and Hubert (2017).

Moreover, MCD-based robust distances are also useful for robust linear regression (Coakley & Hettmansperger, 1993; Simpson, Ruppert, & Carroll, 1992), regression with continuous and categorical regressors (Hubert & Rousseeuw, 1996), and for logistic regression (Croux & Haesbroeck, 2003; Rousseeuw & Christmann, 2003). In the multivariate regression setting (i.e., with several response variables) the MCD can be used directly to obtain MCD-regression (Rousseeuw, Van Aelst, Van Driessen, & Agulló, 2004), whereas MCD applied to the residuals leads to multivariate LTS estimation (Agulló, Croux, & Van Aelst, 2008). Robust errors-in-variables regression is proposed in Fekri and Ruiz-Gazen (2004).

Covariance estimation is also important in principal component analysis and related methods. For low-dimensional data (with $n > 5p$) the principal components can be obtained as the eigenvectors of the MCD scatter matrix (Croux & Haesbroeck, 2000), and robust factor analysis based on the MCD has been studied in Pison, Rousseeuw, Filzmoser, and Croux (2003). The MCD was also used for invariant coordinate selection (Tyler, Critchley, Dümbgen, & Oja, 2009). Robust canonical correlation is proposed in Croux & Dehon (2002). For high-dimensional data, projection pursuit ideas combined with the MCD results in the ROBPCA method (Debruyne & Hubert, 2009; Hubert, Rousseeuw, & Vanden Branden, 2005) for robust PCA. In turn, ROBPCA has led to the construction of robust principal component regression (Hubert & Verboven, 2003) and robust partial least squares regression (Hubert & Vanden Branden, 2003; Vanden Branden & Hubert, 2004), together with appropriate outlier maps, see also Hubert, Rousseeuw, and Van Aelst (2008). Also methods for robust PARAFAC (Engelen & Hubert, 2011) and robust multilevel simultaneous component analysis (Ceulemans, Hubert, & Rousseeuw, 2013) are based on ROBPCA. The LTS subspace estimator (Maronna, 2005) generalizes LTS regression to subspace estimation and orthogonal regression.

An MCD-based alternative to the Hotelling test is provided in Willems, Pison, Rousseeuw, and Van Aelst (2002). A robust bootstrap for the MCD is proposed in Willems and Van Aelst (2004) and a fast cross-validation algorithm in Hubert and Engelen (2007). Computation of the MCD for data with missing values is explored in Cheng and Victoria-Feser (2002), Copt and Victoria-Feser (2004) and Serneels and Verdonck (2008). A robust Cronbach alpha is studied in Christmann and Van Aelst (2006). Classification (i.e., discriminant analysis) based on MCD is constructed in Hawkins and McLachlan (1997) and Hubert and Van Driessen (2004), whereas an alternative for high-dimensional data is developed in Vanden Branden and Hubert (2005). Robust clustering is handled in Rocke and Woodruff (1999), Hardin and Rocke (2004) and Gallegos and Ritter (2005).

The trimming procedure of the MCD has inspired the construction of maximum trimmed likelihood estimators (Čížek, 2008; Hadi & Luceño, 1997; Müller & Neykov, 2003; Vandev & Neykov, 1998), trimmed k-means (Cuesta-Albertos, Gordaliza, & Matrán, 1997; Cuesta-Albertos, Matrán, & Mayo-Iscar, 2008; García-Escudero, Gordaliza, San Martín, Van Aelst, & Zamar, 2009), least weighted squares regression (Víšek, 2002), and minimum weighted covariance determinant estimation (Roelant et al., 2009). The idea of the C-step in the FastMCD algorithm has also been extended to S-estimators (Hubert, Rousseeuw, Vanpaemel, & Verdonck, 2015; Salibian-Barrera & Yohai, 2006).

7 | RECENT EXTENSIONS

7.1 | Deterministic MCD

As the FastMCD algorithm starts by drawing random subsets, it does not necessarily give the same result at multiple runs of the algorithm. (To address this, most implementations fix the seed of the random selection.) Moreover, FastMCD needs to draw many initial subsets in order to obtain at least one that is outlier-free. To circumvent both problems, a deterministic algorithm for robust location and scatter has been developed, denoted as DetMCD (Hubert, Rousseeuw, & Verdonck, 2012). It uses the same iteration steps as FastMCD but does not start from random subsets. Unlike FastMCD it is permutation invariant, that is, the result does not depend on the order of the observations in the data set. Furthermore, DetMCD runs even faster than FastMCD, and is less sensitive to point contamination.

DetMCD computes a small number of deterministic initial estimates, followed by concentration steps. Let X_j denote the columns of the data matrix $\boldsymbol{X}$. First each variable X_j is standardized by subtracting its median and dividing by the Q_n scale estimator of Rousseeuw and Croux (1993). This standardization makes the algorithm location and scale equivariant, that is, Equation (6) holds for any nonsingular diagonal matrix $\boldsymbol{A}$. The standardized data set is denoted as the $n \times p$ matrix $\boldsymbol{Z}$ with rows z_i' ($i = 1, \ldots, n$) and columns Z_j ($j = 1, \ldots, p$).

Next, six preliminary estimates $\boldsymbol{S}_k$ are constructed ($k = 1, \ldots, 6$) for the scatter or correlation of $\boldsymbol{Z}$:

1. $\boldsymbol{S}_1 = \operatorname{corr}(\boldsymbol{Y})$ with $Y_j = \tanh(Z_j)$ for $j = 1, \ldots, p$.
2. Let R_j be the ranks of the column Z_j and put $\boldsymbol{S}_2 = \operatorname{corr}(\boldsymbol{R})$. This is the Spearman correlation matrix of $\boldsymbol{Z}$.
3. $\boldsymbol{S}_3 = \operatorname{corr}(\boldsymbol{T})$ with the normal scores $T_j = \Phi^{-1}((R_j - 1/3)/(n + 1/3))$.
4. The fourth scatter estimate is the spatial sign covariance matrix (Visuri, Koivunen, & Oja, 2000): define $\boldsymbol{k}_i = z_i/\|z_i\|$ for all i and let $\boldsymbol{S}_4 = (1/n)\sum_{i=1}^{n}\boldsymbol{k}_i\boldsymbol{k}_{i'}'$.
5. $\boldsymbol{S}_5$ is the covariance matrix of the $\lceil n/2 \rceil$ standardized observations z_i with smallest norm, which corresponds to the first step of the BACON algorithm (Billor, Hadi, & Velleman, 2000).
6. The sixth scatter estimate is the raw orthogonalized Gnanadesikan-Kettenring (OGK) estimator (Maronna & Zamar, 2002).

As these $\boldsymbol{S}_k$ may have very inaccurate eigenvalues, the following steps are applied to each of them:

1. Compute the matrix $\boldsymbol{E}$ of eigenvectors of $\boldsymbol{S}_k$ and put $\boldsymbol{V} = \boldsymbol{Z}\boldsymbol{E}$.
2. Estimate the scatter of $\boldsymbol{Z}$ by $\boldsymbol{S}_k(\boldsymbol{Z}) = \boldsymbol{E}\boldsymbol{\Lambda}\boldsymbol{E}'$, where $\boldsymbol{\Lambda} = \operatorname{diag}\left(Q_n^2(V_1),\ldots,Q_n^2(V_p)\right)$.
3. Estimate the center of $\boldsymbol{Z}$ by $\hat{\boldsymbol{\mu}}_k(\boldsymbol{Z}) = \boldsymbol{S}_k^{1/2}\left(\operatorname{comed}\left(\boldsymbol{Z}\boldsymbol{S}_k^{-1/2}\right)\right)$ where comed is the coordinatewise median.

For the six estimates $(\hat{\boldsymbol{\mu}}_k(\boldsymbol{Z}), \boldsymbol{S}_k(\boldsymbol{Z}))$ the statistical distances $d_{ik} = d(z_i, \hat{\boldsymbol{\mu}}_k(\boldsymbol{Z}), \boldsymbol{S}_k(\boldsymbol{Z}))$ of all points are computed as in Equation (2). For each initial estimate $k = 1, \ldots, 6$ the mean and covariance matrix of the $h_0 = \lfloor n/2 \rfloor$ observations with smallest d_{ik} are computed, and relative to those the statistical distances (denoted as d_{ik}^*) of all n points. For each $k = 1, \ldots, 6$ the h observations $\boldsymbol{x}_i$ with smallest d_{ik}^* are selected, and C-steps are applied to them until convergence. The solution with smallest determinant is called the raw DetMCD. Then a weighting step is applied as in Equation (5), yielding the final DetMCD.

DetMCD has the advantage that estimates can be quickly computed for a whole range of h values (and hence a whole range of breakdown values), as only the C-steps in the second part of the algorithm depend on h. Monitoring some diagnostics (such as the condition number of the scatter estimate) can give additional insights in the underlying data structure, as in the example in Hubert et al. (2015).

Note that even though DetMCD is not affine equivariant, it turns out that its deviation from affine equivariance is very small.

7.2 | Minimum regularized covariance determinant

In high dimensions we need a modification of MCD, since the existing MCD algorithms take long and are less robust in that case. For large p we can still make a rough estimate of the scatter as follows. First compute the first $q < p$ robust principal components of the data. For this we can use the MCD-based ROBPCA method (Hubert et al., 2005), which requires that the number of components q be set rather low. The robust PCA yields a center $\hat{\boldsymbol{\mu}}$ and q loading vectors. Then form the $p \times q$ matrix $\boldsymbol{L}$ with the loading vectors as columns. The principal component scores $\boldsymbol{t}_i$ are then given by $\boldsymbol{t}_i = \boldsymbol{L}'(\boldsymbol{x}_i - \boldsymbol{\mu})$. Now compute λ_j for $j = 1, \ldots, q$ as a robust variance estimate of the jth principal component, and gather all the λ_j in a diagonal matrix $\boldsymbol{\Lambda}$. Then we can robustly estimate the scatter matrix of the original data set $\boldsymbol{X}$ by $\hat{\boldsymbol{\Sigma}}(\boldsymbol{X}) = \boldsymbol{L}\boldsymbol{\Lambda}\boldsymbol{L}'$. Unfortunately, whenever $q < p$ the resulting matrix $\hat{\boldsymbol{\Sigma}}(\boldsymbol{X})$ will have $p - q$ eigenvalues equal to 0, hence $\hat{\boldsymbol{\Sigma}}(\boldsymbol{X})$ is singular.

If we require a nonsingular scatter matrix we need a different approach using regularization. The MRCD method (Boudt, Rousseeuw, Vanduffel, & Verdonck, 2017) was constructed for this purpose, and works when $n < p$ too. The MRCD minimizes

$$\det\{\rho \boldsymbol{T} + (1-\rho)\mathrm{Cov}(\boldsymbol{X}_H)\}, \tag{11}$$

where $\boldsymbol{T}$ is a positive definite "target" matrix and $\mathrm{Cov}(\boldsymbol{X}_H)$ is the usual covariance matrix of an h-subset $\boldsymbol{X}_H$ of $\boldsymbol{X}$. Even when $\mathrm{Cov}(\boldsymbol{X}_H)$ is singular by itself, the combined matrix is always positive definite hence invertible. The target matrix $\boldsymbol{T}$ depends on the application, and can for instance be the $p \times p$ identity matrix or be based on a rank correlation matrix. Perhaps surprisingly, it turns out that the C-step theorem can be extended to the MRCD. The MRCD algorithm is similar to the DetMCD described above, with deterministic starts followed by iterating these modified C-steps. The method simulates well even in 1000 dimensions.

Software for the DetMCD and MRCD methods is available from http://wis.kuleuven.be/stat/robust.

8 | CONCLUSIONS

In this study we have reviewed the MCD estimator of multivariate location and scatter. We have illustrated its resistance to outliers on a real data example. Its main properties concerning robustness, efficiency and equivariance were described, as well as computational aspects. We have provided a detailed reference list with applications and generalizations of the MCD in applied and methodological research. Finally, two recent modifications of the MCD make it possible to save computing time and to deal with high-dimensional data.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

RELATED WIREs ARTICLES

MATLAB library LIBRA
Anomaly detection by robust statistics
Minimum volume ellipsoid

ORCID

Mia Hubert http://orcid.org/0000-0001-6398-4850
Peter J. Rousseeuw http://orcid.org/0000-0002-3807-5353

REFERENCES

Agulló, J., Croux, C., & Van Aelst, S. (2008). The multivariate least trimmed squares estimator. *Journal of Multivariate Analysis*, *99*, 311–318.

Billor, N., Hadi, A., & Velleman, P. (2000). BACON: Blocked adaptive computationally efficient outlier nominators. *Computational Statistics & Data Analysis*, *34*, 279–298.

Boudt, K., Rousseeuw, P., Vanduffel, S., & Verdonck, T. (2017). The minimum regularized covariance determinant estimator. arXiv: 1701.07086.

Butler, R., Davies, P., & Jhun, M. (1993). Asymptotics for the minimum covariance determinant estimator. *The Annals of Statistics*, *21*(3), 1385–1400.

Cator, E., & Lopuhaä, H. (2010). Asymptotic expansion of the minimum covariance determinant estimators. *Journal of Multivariate Analysis*, *101*, 2372–2388.

Cator, E., & Lopuhaä, H. (2012). Central limit theorem and influence function for the MCD estimators at general multivariate distributions. *Bernouilli*, *18*, 520–551.

Cerioli, A. (2010). Multivariate outlier detection with high-breakdown estimators. *Journal of the American Statistical Association*, *105*(489), 147–156.

Ceulemans, E., Hubert, M., & Rousseeuw, P. (2013). Robust multilevel simultaneous component analysis. *Chemometrics and Intelligent Laboratory Systems*, *129*, 33–39.

Cheng, T.-C., & Victoria-Feser, M.-P. (2002). High breakdown estimation of multivariate location and scale with missing observations. *British Journal of Mathematical and Statistical Psychology*, *55*, 317–335.

Christmann, A., & Van Aelst, S. (2006). Robust estimation of Cronbach's alpha. *Journal of Multivariate Analysis*, *97*(7), 1660–1674.

Čížek, P. (2008). Robust and efficient adaptive estimation of binary-choice regression models. *Journal of the American Statistical Association*, *103*(482), 687–696.

Coakley, C., & Hettmansperger, T. (1993). A bounded influence, high breakdown, efficient regression estimator. *Journal of the American Statistical Association*, *88*, 872–880.

Copt, S., & Victoria-Feser, M.-P. (2004). Fast algorithms for computing high breakdown covariance matrices with missing data. In M. Hubert, G. Pison, A. Struyf, & S. Van Aelst (Eds.), Theory and applications of recent robust methods (pp. 71–82). Basel: Birkhäuser.

Croux, C., & Dehon, C. (2002). Analyse canonique basée sur des estimateurs robustes de la matrice de covariance. *La Revue de Statistique Appliquée*, *2*, 5–26.

Croux, C., & Haesbroeck, G. (1999). Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *Journal of Multivariate Analysis*, *71*, 161–190.

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

Croux, C., & Haesbroeck, G. (2001). Maxbias curves of robust scale estimators based on subranges. *Metrika*, *53*, 101–122.

Croux, C., & Haesbroeck, G. (2002). Maxbias curves of location estimators based on subranges. *Journal of Nonparametric Statistics*, *14*, 295–306.

Croux, C., & Haesbroeck, G. (2003). Implementing the Bianco and Yohai estimator for logistic regression. *Computational Statistics & Data Analysis*, *44*, 273–295.

Croux, C., & Rousseeuw, P. (1992). A class of high-breakdown scale estimators based on subranges. *Communications in Statistics*, *21*, 1935–1951.

Cuesta-Albertos, J., Gordaliza, A., & Matrán, C. (1997). Trimmed *k*-means: An attempt 19 to robustify quantizers. *The Annals of Statistics*, *25*, 553–576.

Cuesta-Albertos, J., Matrán, C., & Mayo-Iscar, A. (2008). Robust estimation in the normal mixture model based on robust clustering. *Journal of the Royal Statistical Society*, *70*, 779–802.

Davies, L. (1987). Asymptotic behavior of S-estimators of multivariate location parameters and dispersion matrices. *The Annals of Statistics*, *15*, 1269–1292.

Debruyne, M., & Hubert, M. (2009). The influence function of the Stahel-Donoho covariance estimator of smallest outlyingness. *Statistics & Probability Letters*, *79*, 275–282.

Donoho, D., & Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics*, *20*(4), 1803–1827.

Engelen, S., & Hubert, M. (2011). Detecting outlying samples in a parallel factor analysis model. *Analytica Chimica Acta*, *705*, 155–165.

Fekri, M., & Ruiz-Gazen, A. (2004). Robust weighted orthogonal regression in the errors-in-variables model. *Journal of Multivariate Analysis*, *88*(1), 89–108.

Filzmoser, P., Garrett, R., & Reimann, C. (2005). Multivariate outlier detection in exploration geochemistry. *Computers and Geosciences*, *31*, 579–587.

Gallegos, M., & Ritter, G. (2005). A robust method for cluster analysis. *The Annals of Statistics*, *33*, 347–380.

Gambacciani, M., & Paolella, M. S. (2017). Robust normal mixtures for financial portfolio allocation. *Econometrics and Statistics*, *3*, 91–111.

García-Escudero, L., Gordaliza, A., San Martín, R., Van Aelst, S., & Zamar, R. (2009). Robust linear clustering. *Journal of the Royal Statistical Society*, *71*, 1–18.

Hadi, A., & Luceño, A. (1997). Maximum trimmed likelihood estimators: A unified approach, examples and algorithms. *Computational Statistics & Data Analysis*, *25*, 251–272.

Hampel, F., Ronchetti, E., Rousseeuw, P., & Stahel, W. (1986). Robust statistics: The approach based on influence functions. New York, NY: Wiley.

Hardin, J., & Rocke, D. (2004). Outlier detection in the multiple cluster setting using the minimum covariance determinant estimator. *Computational Statistics & Data Analysis*, *44*, 625–638.

Hardin, J., & Rocke, D. M. (2005). The distribution of robust distances. *Journal of Computational and Graphical Statistics*, *14*(4), 928–946.

Hawkins, D., & McLachlan, G. (1997). High-breakdown linear discriminant analysis. *Journal of the American Statistical Association*, *92*, 136–143.

Hettich, S., & Bay, S. (1999). The UCI KDD archive [Computer software manual]. Irvine, CA: University of California, Department of Information and Computer Science. Retrieved from http://kdd.ics.uci.edu

Hubert, M., & Engelen, S. (2007). Fast cross-validation for high-breakdown resampling algorithms for PCA. *Computational Statistics & Data Analysis*, *51*, 5013–5024.

Hubert, M., & Rousseeuw, P. (1996). Robust regression with both continuous and binary regressors. *Journal of Statistical Planning and Inference*, *57*, 153–163.

Hubert, M., Rousseeuw, P., & Van Aelst, S. (2008). High breakdown robust multivariate methods. *Statistical Science*, *23*, 92–119.

Hubert, M., Rousseeuw, P., & Vanden Branden, K. (2005). ROBPCA: A new approach to robust principal components analysis. *Technometrics*, *47*, 64–79.

Hubert, M., Rousseeuw, P., Vanpaemel, D., & Verdonck, T. (2015). The DetS and DetMM estimators for multivariate location and scatter. *Computational Statistics & Data Analysis*, *81*, 64–75.

Hubert, M., Rousseeuw, P., & Verdonck, T. (2012). A deterministic algorithm for robust location and scatter. *Journal of Computational and Graphical Statistics*, *21*, 618–637.

Hubert, M., & Van Driessen, K. (2004). Fast and robust discriminant analysis. *Computational Statistics & Data Analysis*, *45*, 301–320.

Hubert, M., & Vanden Branden, K. (2003). Robust methods for partial least squares regression. *Journal of Chemometrics*, *17*, 537–549.

Hubert, M., & Verboven, S. (2003). A robust PCR method for high-dimensional regressors. *Journal of Chemometrics*, *17*, 438–452.

Jensen, W., Birch, J., & Woodal, W. (2007). High breakdown estimation methods for phase I multivariate control charts. *Quality and Reliability Engineering International*, *23*(5), 615–629.

Lopuhaä, H. (1999). Asymptotics of reweighted estimators of multivariate location and scatter. *The Annals of Statistics*, *27*, 1638–1665.

Lopuhaä, H., & Rousseeuw, P. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *The Annals of Statistics*, *19*, 229–248.

Lu, Y., Wang, J., Kong, J., Zhang, B., & Zhang, J. (2006). An integrated algorithm for MRI brain images segmentation. *Computer Vision Approaches to Medical Image Analysis*, *4241*, 132–1342.

Maronna, R. (2005). Principal components and orthogonal regression based on robust scales. *Technometrics*, *47*, 264–273.

Maronna, R., Martin, D., & Yohai, V. (2006). Robust statistics: Theory and methods. New York, NY: Wiley.

Maronna, R., & Zamar, R. (2002). Robust estimates of location and dispersion for high-dimensional data sets. *Technometrics*, *44*, 307–317.

Müller, C., & Neykov, N. (2003). Breakdown points of trimmed likelihood estimators and related estimators in generalized linear models. *Journal of Statistical Planning and Inference*, *116*, 503–519.

Neykov, N., Neytchev, P., Van Gelder, P., & Todorov, V. (2007). Robust detection of discordant sites in regional frequency analysis. *Water Resources Research*, *43*(6).

Pison, G., Rousseeuw, P., Filzmoser, P., & Croux, C. (2003). Robust factor analysis. *Journal of Multivariate Analysis*, *84*, 145–172.

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

Prastawa, M., Bullitt, E., Ho, S., & Gerig, G. (2004). A brain tumor segmentation framework based on outlier detection. *Medical Image Analysis*, *8*, 275–283.

Rocke, D., & Woodruff, D. (1999). A synthesis of outlier detection and cluster identification. Technical report.

Roelant, E., Van Aelst, S., & Willems, G. (2009). The minimum weighted covariance determinant estimator. *Metrika*, *70*, 177–204.

Rousseeuw, P. (1984). Least median of squares regression. *Journal of the American Statistical Association*, *79*, 871–880.

Rousseeuw, P. (1985). Multivariate estimation with high breakdown point. In W. Grossmann, G. Pflug, I. Vincze, & W. Wertz (Eds.), Mathematical statistics and applications (Vol. *B*, pp. 283–297). Dordrecht: Reidel.

Rousseeuw, P. (2005). Discussion on 'Breakdown and groups'. *Annals of Statistics*, *33*, 1004–1009.

Rousseeuw, P., & Christmann, A. (2003). Robustness against separation and outliers in logistic regression. *Computational Statistics & Data Analysis*, *43*, 315–332.

Rousseeuw, P., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, *88*, 1273–1283.

Rousseeuw, P., & Hubert, M. (2017). Anomaly detection by robust statistics. *WIREs Data Mining and Knowledge Discovery*. https://doi.org/10.1002/widm.1236

Rousseeuw, P., & Leroy, A. (1987). Robust regression and outlier detection. New York, NY: Wiley.

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

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

Rousseeuw, P., & Van Driessen, K. (2006). Computing LTS regression for large data sets. *Data Mining and Knowledge Discovery*, *12*, 29–45.

Rousseeuw, P., & van Zomeren, B. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, *85*, 633–651.

Salibian-Barrera, M., & Yohai, V. (2006). A fast algorithm for S-regression estimates. *Journal of Computational and Graphical Statistics*, *15*, 414–427.

Serneels, S., & Verdonck, T. (2008). Principal component analysis for data containing outliers and missing elements. *Computational Statistics & Data Analysis*, *52*, 1712–1727.

Simpson, D., Ruppert, D., & Carroll, R. (1992). On one-step GM-estimates and stability of inferences in linear regression. *Journal of the American Statistical Association*, *87*, 439–450.

Stahel, W. (1981). *Robuste Schätzungen: infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen.* (Unpublished doctoral dissertation). ETH Zürich, Zürich.

Tyler, D. E., Critchley, F., Dümbgen, L., & Oja, H. (2009). Invariant co-ordinate selection. *Journal of the Royal Statistical Society*, *71*(3), 549–592.

van Helvoort, P., Filzmoser, P., & van Gaans, P. (2005). Sequential factor analysis as a new approach to multivariate analysis of heterogeneous geochemical datasets: An application to a bulk chemical characterization of fluvial deposits (Rhine-Meuse delta, The Netherlands). *Applied Geochemistry*, *20*(12), 2233–2251.

Vanden Branden, K., & Hubert, M. (2004). Robustness properties of a robust PLS regression method. *Analytica Chimica Acta*, *515*, 229–241.

Vanden Branden, K., & Hubert, M. (2005). Robust classification in high dimensions based on the SIMCA method. *Chemometrics and Intelligent Laboratory Systems*, *79*, 10–21.

Vandev, D., & Neykov, N. (1998). About regression estimators with high breakdown point. *Statistics*, *32*, 111–129.

Verboven, S., & Hubert, M. (2005). LIBRA: A Matlab library for robust analysis. *Chemometrics and Intelligent Laboratory Systems*, *75*, 127–136.

Verboven, S., & Hubert, M. (2010). MATLAB library LIBRA. *WIREs Computational Statistics*, *2*, 509–515.

Víšek, J. (2002). The least weighted squares I. The asymptotic linearity of normal equations. *Bulletin of the Czech Econometric Society*, *9*, 31–58.

Visuri, S., Koivunen, V., & Oja, H. (2000). Sign and rank covariance matrices. *Journal of Statistical Planning and Inference*, *91*, 557–575.

Vogler, C., Goldenstein, S., Stolfi, J., Pavlovic, V., & Metaxas, D. (2007). Outlier rejection in high-dimensional deformable models. *Image and Vision Computing*, *25*(3), 274–284.

Welsh, R., & Zhou, X. (2007). Application of robust statistics to asset allocation models. *Revstat*, *5*, 97–114.

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

Willems, G., & Van Aelst, S. (2004). A fast bootstrap method for the MCD estimator. In J. Antoch (Ed.), Proceedings in computational statistics (pp. 1979–1986). Heidelberg: Springer.

Zaman, A., Rousseeuw, P., & Orhan, M. (2001). Econometric applications of high-breakdown robust regression techniques. *Economics Letters*, *71*, 1–8.

How to cite this article: Hubert M, Debruyne M, Rousseeuw PJ. Minimum covariance determinant and extensions. *WIREs Comput Stat*. 2018;10:e1421. https://doi.org/10.1002/wics.1421