Journal of Multivariate Analysis **78**, 11–36 (2001)
doi:10.1006/jmva.2000.1942, available online at http://www.idealibrary.com on IDEAL®

Highly Robust Estimation of Dispersion Matrices

Yanyuan Ma and Marc G. Genton

Massachusetts Institute of Technology
E-mail: yanyuan@math.mit.edu, genton@math.mit.edu

Received January 4, 1999; published online April 6, 2001

In this paper, we propose a new componentwise estimator of a dispersion matrix, based on a highly robust estimator of scale. The key idea is the elimination of a location estimator in the dispersion estimation procedure. The robustness properties are studied by means of the influence function and the breakdown point. Further characteristics such as asymptotic variance and efficiency are also analyzed. It is shown in the componentwise approach, for multivariate Gaussian distributions, that covariance matrix estimation is more difficult than correlation matrix estimation. The reason is that the asymptotic variance of the covariance estimator increases with increasing dependence, whereas it decreases with increasing dependence for correlation estimators. We also prove that the asymptotic variance of dispersion estimators for multivariate Gaussian distributions is proportional to the asymptotic variance of the underlying scale estimator. The proportionality value depends only on the underlying dependence. Therefore, the highly robust dispersion estimator is among the best robust choice at the present time in the componentwise approach, because it is location-free and combines small variability and robustness properties such as high breakdown point and bounded influence function. A simulation study is carried out in order to assess the behavior of the new estimator. First, a comparison with another robust componentwise estimator based on the median absolute deviation scale estimator is performed. The highly robust properties of the new estimator are confirmed. A second comparison with global estimators such as the method of moment estimator, the minimum volume ellipsoid, and the minimum covariance determinant estimator is also performed, with two types of outliers. In this case, the highly robust dispersion matrix estimator turns out to be an interesting compromise between the high efficiency of the method of moment estimator in noncontaminated situations and the highly robust properties of the minimum volume ellipsoid and minimum covariance determinant estimators in contaminated situations.

AMS subject classifications: 62H12; 62G35.

Key words and phrases: breakdown point; componentwise; influence function; robustness; scale estimator.

1. INTRODUCTION

Dispersion matrices, i.e. covariance and correlation matrices, play an important role in many methods of multivariate statistics. For instance,

0047-259X/01 $35.00

they are the cornerstones of principal component analysis, discriminant analysis, factor analysis, canonical correlation analysis, and many others (e.g. Mardia *et al.*, 1979). Moreover, dispersion matrices are themselves quantities of interest since they represent a measure of association or interdependence between several characteristics. They provide information about the shape of the ellipsoid of the data cloud in a multidimensional space. Therefore, reliable estimators of dispersion matrices are of prime importance. Unfortunately, classical sample dispersion matrices are known to be very sensitive to outlying values in the data, which can typically be hidden in the high dimensionality of the space of variables. As a consequence, eigenvalues and eigenvectors of the dispersion matrix inherit this sensitivity. A principal component analysis could thus reveal an artificial structure in the data, that does not really exist but is merely created by a few outliers.

In the past three decades, many attempts to overcome the poor resistance properties of the classical sample dispersion matrix have been made. The robust proposals can be classified in two main categories: robust componentwise estimation and robust global estimation of the dispersion matrix. The first one can be approached via location estimation, or scale estimation, as described in Section 3. It has the advantage of being able to deal with missing values in the data, but is not affine invariant and does not provide a positive definite matrix directly. The second category usually insures affine invariance and positive definiteness, but is less appropriate to deal with missing data.

In this paper, we propose the use of a highly robust estimator of scale, denoted by Q_n, in the componentwise approach. In fact, we show that it is among the best robust choice available at the present time in the componentwise approach. Of course, other robust and efficient scales estimators could be used, for example like the τ-scales proposed by Yohai and Zamar (1988). However, the highly robust estimator of scale Q_n possesses the location-free property and has already been successfully used in the context of regression (Hössjer, Croux, and Rousseeuw, 1994; Croux, Rousseeuw, and Hössjer, 1994), as well as for variogram estimation (Genton, 1998) in spatial statistics, and autocovariance estimation (Ma and Genton, 2000) in time series. In the next section, we start by recalling some of the dispersion matrix estimators that can be found in the literature. The third section describes the highly robust estimator of dispersion matrices. Robustness properties are discussed in Section 4. The influence function for covariance and correlation estimators are studied, as well as their breakdown point. The asymptotic variance and efficiency are derived for the new estimator in the case of multivariate Gaussian distributions. In the end, we compare the suggested method with some other methods (componentwise and global) and carry out some simulations.

In the sequel of the paper, we use the following notations. In the case of two random variables, we typically use X and Y to represent them and use $\mathbf{x} = (x_1, x_2, ..., x_n)^T$, $\mathbf{y} = (y_1, y_2, ..., y_n)^T$, to represent the observation vectors. In the case of p random variables, we use X_i, $i = 1, 2, ..., p$ to denote the random variables. The n observations of each random variable X_i are represented by $x_{1i}, x_{2i}, ..., x_{ni}$, and they are gathered into a vector $\mathbf{x}_{(i)}$. The n realizations of the random vector $(X_1, X_2, ..., X_p)$ is represented by $\mathbf{x}_j = (x_{j1}, x_{j2}, ..., x_{jp})$, $j = 1, 2, ..., n$. Therefore, the data matrix $\mathbf{X}$ can be represented in the following format:

$$\begin{array}{cc} & \begin{array}{cccc} X_1 & X_2 & \cdots & X_p \end{array} \\ \mathbf{X} = & \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \end{array}$$

or $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)^T = (\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, ..., \mathbf{x}_{(p)})$.

2. DISPERSION MATRIX ESTIMATORS

In this section, we describe some commonly used estimators for the dispersion matrix, as well as some recent robust proposals. We focus on the estimation of covariance matrices, since estimation of correlation matrices can be derived in the same way.

Suppose that the sample $\mathbf{x}_1, ..., \mathbf{x}_n$, with $\mathbf{x}_i \in \mathbb{R}^p$, $i = 1, ..., n$, is independently and identically distributed according to a multivariate distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. Note that estimation of the correlation matrix R can always be derived from the relation $R = D\Sigma D$, where $D = \operatorname{diag}(1/\sqrt{\Sigma_{11}}, ..., 1/\sqrt{\Sigma_{pp}})$. The method of moment estimator (MME) of the covariance matrix Σ is

$$\hat{\Sigma}_{MME} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T, \tag{1}$$

where $\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$.

The breakdown point is an important feature of reliability of an estimator. It indicates, roughly speaking, the largest proportion of data that can be replaced by arbitrary values to bring the estimator to the boundaries of the parameter space. More details can be found for instance in Donoho and Huber (1982), Huber (1981, 1984), Hampel *et al.* (1986). The breakdown point of the method of moment estimator (1) is zero, indicating its very poor resistance.

Affine equivariant M-estimators for dispersion matrices were first suggested by Hampel (1973), and studied by Maronna (1976) and Huber (1977, 1981). Unfortunately, their breakdown point is at most $1/(p+1)$. This is not satisfactory, because it means that the breakdown point becomes smaller with increasing dimension, where there are more opportunities for outliers to occur. The performance of some M-estimators were studied by mean of a Monte Carlo study by Devlin *et al.* (1975, 1981).

Stahel (1981) and Donoho (1982) were first to independently propose robust affine equivariant estimators of multivariate location and dispersion having a high breakdown point (asymptotically 1/2) for any dimension. They are defined as weighted mean and weighted dispersion, where the weights are functions of a measure of "outlyingness" obtained by considering all univariate projections of the data. Subsequently, other high breakdown point equivariant multivariate estimators have been introduced. The most well known is probably the Minimum Volume Ellipsoid (MVE) estimator, introduced by Rousseeuw (1984, 1985), and discussed in Rousseeuw and Leroy (1987), Rousseeuw and van Zomeren (1990). The method seeks an ellipsoid of minimum volume, containing $m = \lfloor (n+p+1)/2 \rfloor$ points, where $\lfloor \cdot \rfloor$ denotes the integer part. More precisely, it consists in finding $\hat{\boldsymbol{\mu}}_{MVE}$ and $\hat{\Sigma}_{MVE}$ such that the determinant of Σ is minimized subject to

$$\#\{i \mid (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \leqslant a^2\} \geqslant m, \tag{2}$$

where a^2 is a fixed constant, for example from χ^2_p in the case of Gaussian data. The MVE has a finite sample breakdown point of m, i.e. 50% asymptotically. Two algorithms (resampling and projection) to compute an approximate solution of MVE can be found in Rousseeuw and van Zomeren (1990).

The MVE estimator has been generalized to multivariate S-estimators (Davies, 1987; Lopuhaä, 1989; Lopuhaä and Rousseeuw, 1991). Li and Chen (1985) proposed a dispersion matrix estimator based on robustifying principal components via projection pursuit techniques. A class of projection estimators for dispersion matrices were studied by Maronna, Stahel and Yohai (1992). Tyler (1994) discusses finite sample breakdown point of projection based estimators, in particular the Stahel-Donoho estimator. Maronna and Yohai (1995) studied asymptotic and finite-sample behaviors of the Stahel–Donoho robust multivariate estimators. From a simulation study, they concluded that they compare favorably with other proposals like multivariate M- or S-estimators, and Rousseeuw's MVE. However, the main drawback remains the lack of feasible methods to compute the estimators for dimensions larger than $p = 2$.

Recently, Rousseeuw and Van Driessen (1999) proposed a fast algorithm (FAST-MCD) for the Minimum Covariance Determinant (MCD) estimator. Originally proposed by Rousseeuw (1984, 1985), the use of this estimator was until now hampered by the high computation time of existing algorithms. The MCD objective is to find h observations out of n whose classical covariance matrix has the lowest determinant. The MCD estimator, $\hat{\Sigma}_{MCD}$, of the covariance matrix is then the method of moment estimator of these h observations. Rousseeuw and Van Driessen (1999) have proved that the finite sample breakdown point of MCD is m defined above, when $h=m$, i.e. 50% asymptotically. Moreover, Croux and Haesbroeck (1999) showed that MCD is more efficient than MVE in high dimensions, and therefore recommend the use of MCD.

3. THE HIGHLY ROBUST ESTIMATOR

3.1. *Dispersion between Two Random Variables*

Traditionally, covariance estimation between two random variables X and Y is based on a location approach, since $\mathrm{Cov}(X, Y) = \mathrm{E}[(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))]$, yielding for example the estimator (1) of Σ. However, covariance estimation can also be based on a scale approach, by means of the following identity (Huber, 1981; Gnanadesikan, 1997):

$$\mathrm{Cov}(X, Y) = \frac{\alpha\beta}{4}\left[\mathrm{Var}(X/\alpha + Y/\beta) - \mathrm{Var}(X/\alpha - Y/\beta)\right], \qquad \forall \alpha, \beta \in \mathbb{R}^*. \tag{3}$$

In general, X and Y may be measured in different units, and the choice $\alpha = \sigma_X$ and $\beta = \sigma_Y$ is recommended (Gnanadesikan and Kettenring, 1972), where $\sigma_X = \sqrt{\mathrm{Var}(X)}$ and $\sigma_Y = \sqrt{\mathrm{Var}(Y)}$. The choice of a robust estimator of the variance in (3) produces a robust estimator of the covariance between X and Y.

In the context of scale estimation, Rousseeuw and Croux (1992, 1993) proposed a simple, explicit and highly robust estimator of scale, Q_n,

$$Q_n(\mathbf{z}) = d\,\{|z_i - z_j|;\, i<j,\, i, j = 1, 2, \ldots, n\}_{(k)}, \tag{4}$$

where $\mathbf{z} = (z_1, \ldots, z_n)^T$ is a sample of a random variable Z, $k = \lfloor (\binom{n}{2} + 2)/4 \rfloor + 1$ and $\lfloor \cdot \rfloor$ denotes the integer part. The factor d is for consistency: for the Gaussian distribution, $d = 2.2191$. This means that we sort the set of all absolute differences $|z_i - z_j|$ in increasing order for $i<j$, $i, j = 1, 2, \ldots, n$ and then compute its kth order statistic (approximately the 1/4 quantile for large n). This value is multiplied by d, thus yielding Q_n. Note that this estimator computes the k-th order statistic of the $\binom{n}{2}$ interpoint distances.

It is of interest to remark that Q_n does not rely on any location knowledge and is therefore said to be location-free. This is in contrast to the classical sample covariance matrix (1), which can be obtained by inserting the classical sample variance estimator in Eq. (3). Therefore, the use of the highly robust scale estimator (4) in the identity (3) will produce a highly robust covariance estimator which is also location-free. At first sight, the estimator Q_n appears to need $O(n^2)$ computation time, which would be a disadvantage. However, it can be computed using no more than $O(n \log n)$ time and $O(n)$ storage, by means of the fast algorithm described in Croux and Rousseeuw (1992).

Using the identity (3) and the definition (4) of the scale estimator Q_n, we propose the following highly robust estimator to compute the covariance γ between two random variables X and Y. First, use Q_n to estimate the standard deviations σ_X and σ_Y of X and Y. Then, use Q_n again to estimate the standard deviations σ_+ and σ_- of $X/\sigma_X + Y/\sigma_Y$ and $X/\sigma_X - Y/\sigma_Y$. The covariance γ between X and Y is $\sigma_X \sigma_Y(\sigma_+^2 - \sigma_-^2)/4$. Therefore, the highly robust estimator $\hat{\gamma}_Q$ of the covariance γ is

$$\hat{\gamma}_Q(\mathbf{x}, \mathbf{y}) = \frac{\alpha\beta}{4}\left[Q_n^2(\mathbf{x}/\alpha + \mathbf{y}/\beta) - Q_n^2(\mathbf{x}/\alpha - \mathbf{y}/\beta)\right], \tag{5}$$

where $\alpha = Q_n(\mathbf{x})$, $\beta = Q_n(\mathbf{y})$. As will be shown in Section 4, it has a breakdown point of 50%, which is the same as the Q_n estimator. Here, 50% breakdown point means that among the n observation pairs $\{x_i, y_i\}$, $i = 1, ..., n$, half of them can contain contaminated (arbitrary) values and the estimation will not be totally destroyed. Note that the highly robust covariance estimator $\hat{\gamma}_Q$ can also be carried out with $O(n \log n)$ time and $O(n)$ storage.

In order to obtain a highly robust estimator of the correlation ρ between two random variables X and Y, we could divide the estimator $\hat{\gamma}_Q(\mathbf{x}, \mathbf{y})$ in Eq. (5) by $Q_n(\mathbf{x})$ and $Q_n(\mathbf{y})$, yielding

$$\tfrac{1}{4}\left[Q_n^2(\mathbf{x}/\alpha + \mathbf{y}/\beta) - Q_n^2(\mathbf{x}/\alpha - \mathbf{y}/\beta)\right], \tag{6}$$

where $\alpha = Q_n(\mathbf{x})$, $\beta = Q_n(\mathbf{y})$. However, this is not a natural correlation estimator because it is not bounded between -1 and 1. Therefore, we consider the following highly robust correlation estimator $\hat{\rho}_Q$ of ρ,

$$\hat{\rho}_Q(\mathbf{x}, \mathbf{y}) = \frac{Q_n^2(\mathbf{x}/\alpha + \mathbf{y}/\beta) - Q_n^2(\mathbf{x}/\alpha - \mathbf{y}/\beta)}{Q_n^2(\mathbf{x}/\alpha + \mathbf{y}/\beta) + Q_n^2(\mathbf{x}/\alpha - \mathbf{y}/\beta)}, \tag{7}$$

where the denominator is an estimator of the value 4 that insures $|\hat{\rho}_Q(\mathbf{x}, \mathbf{y})| \leqslant 1$. Note that $\hat{\gamma}_Q$ depends upon the choice of the constant d appearing in

Eq. (4), whereas $\hat{\rho}_Q$ is independent of the choice of d. Nevertheless, d can be computed for various distributions, although the Gaussian case is usually preferred.

3.2. *Dispersion between p Random Variables*

In the case of n observations of a p-dimensional random vector, we use the estimator $\hat{\gamma}_Q$ to estimate every covariance between X_i and X_j ($i, j = 1, ..., p$, $i \neq j$) to get the (i, j) entry of the covariance matrix Σ. The diagonal entries are estimated using Q_n^2 directly on the X_i's ($i = 1, ..., p$). This provides a highly robust componentwise estimator $\hat{\Sigma}_Q$ of the covariance matrix Σ.

Using $\hat{\rho}_Q$, we can estimate the entries of the correlation matrix R similarly as in the covariance matrix case, thus yielding a highly robust componentwise estimator $\hat{R}_Q$. We set all the diagonal entries of $\hat{R}_Q$ to 1's.

Note that since the method we propose is componentwise instead of global, there is no guarantee that we get a positive definite matrix at the end of the estimation. Rousseeuw and Molenberghs (1993) proposed three kinds of methods to transform the estimated matrix to a positive definite matrix. They are respectively the shrinking method, the eigenvalue method, and the scaling method. When the covariance itself is the quantity of interest, one should transform it to a positive definite matrix using one of these methods, while if some particular entries in the matrix are the values of interest, then the estimated values should provide a good estimation of the real values.

4. PROPERTIES OF THE ESTIMATOR

4.1. *Breakdown Point*

It is known that the breakdown point of Q_n is 50% (Rousseeuw and Croux, 1993). Inspecting $X/\alpha + Y/\beta$ (or $X/\alpha - Y/\beta$), we can see that as long as x_i (or y_i) is contaminated, then $x_i/\alpha + y_i/\beta$ (or $x_i/\alpha - y_i/\beta$) is contaminated. So in the pairs $(x_1, y_1), ..., (x_n, y_n)$, we can at most have half of the pairs containing contaminated data. If we look at one pair as one observation, then the estimators $\hat{\gamma}_Q$ and $\hat{\rho}_Q$ are robust against at most half of the contaminated observations. So, they have breakdown point of 50%. In estimating the covariance matrix Σ and the correlation matrix R, we form pairs of all the observations of X_i and X_j ($i, j = 1, ..., p$), and the estimator allows at most half of the pairs to be contaminated. Therefore, among the n observation vectors $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$, at most half of them can contain contaminated data. In other words, the breakdown point of the highly robust componentwise estimators $\hat{\Sigma}_Q$ and $\hat{R}_Q$ is 50%. Note that in the context of dispersion matrix estimation, another interesting type of breakdown point is when outliers cause the estimated matrix to become singular. However, this is not the case for our dispersion estimators $\hat{\Sigma}_Q$ and $\hat{R}_Q$. Even without outliers, they have to be transformed to positive definiteness by means of one of the three methods mentioned in Section 3.2.

4.2. *Influence Function*

The influence function (Hampel, 1974) is a tool to describe the robustness properties of an estimator. Its importance lies in its appealing heuristic interpretation: it measures the asymptotic bias caused by an infinitesimal contamination of the observations. Denote by γ_Q, ρ_Q, and Q the statistical functional (e.g. Huber, 1981; Hampel *et al.*, 1986) corresponding to the estimators $\hat{\gamma}_Q$, $\hat{\rho}_Q$, and Q_n respectively. The influence function of dispersion estimators has been derived by Genton and Ma (1999). It is based on the influence function of the underlying scale estimator. In our case, the influence function of Q_n is (Rousseeuw and Croux, 1993):

$$IF(u; Q, F) = d\,\frac{\frac{1}{4} - F(u + d^{-1}) + F(u - d^{-1})}{\int f(x + d^{-1})\, f(x)\, dx}, \tag{8}$$

where f is the density function of the distribution F and d is the same coefficient as in the Q_n estimator. Based on (5), the influence function of the covariance estimator $\hat{\gamma}_Q$ is:

$$\begin{aligned} IF((u, v); \gamma_Q, \mathbf{F}) = \frac{1}{2}\bigg[& \sigma_+ IF\left(\frac{u}{\sigma_X} + \frac{v}{\sigma_Y}; Q, F_+\right) \\ & - \sigma_- IF\left(\frac{u}{\sigma_X} - \frac{v}{\sigma_Y}; Q, F_-\right)\bigg] \sigma_X \sigma_Y. \end{aligned} \tag{9}$$

Here, F_+ is the distribution function of $\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}$, F_- is the distribution function of $\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}$, $\mathbf{F}$ is the bivariate distribution of X and Y, with marginal distributions F_X and F_Y. The influence functions $IF(\cdot; Q, F_+)$ and $IF(\cdot; Q, F_-)$ are given by Eq. (8).

More information on the justification and properties of Eq. (9) can be found in Genton and Ma (1999). One way to understand it intuitively is: σ_X and σ_Y in Eq. (9) can be replaced by any non-zero constants α and β. Then from Eq. (5), noticing the connection between the influence function and the first order derivative, we know Eq. (9) gives the influence function. In particular, for $\alpha = \sigma_X$ and $\beta = \sigma_Y$, Eq. (9) still remains valid. One may suspect that since σ_X and σ_Y themselves have to be estimated first, our influence function should take the perturbation of these two estimators into account too, hence should have a more complicated form than the one given in Eq. (9). Fortunately, this is not the case and we can understand it in this way: how far the estimated σ_X and σ_Y are from the true values does not have any direct effect on the estimation since even if we take arbitrary α and β, the estimator is still valid. The values of α and β only have an effect in carrying out the estimation of $\frac{X}{\sigma_X}+\frac{Y}{\sigma_Y}$ and $\frac{X}{\sigma_X}-\frac{Y}{\sigma_Y}$, and this is taken care of in the influence function of these two estimators.

Since the correlation estimator $\hat{\rho}_Q(\mathbf{x},\mathbf{y})$ can be written as in Eq. (7), we have:

$$
\begin{aligned}
IF((u,v);\rho_Q,\mathbf{F}) = \frac{2}{(\sigma_+^2+\sigma_-^2)^2}\Bigg[&(\sigma_+^2+\sigma_-^2)\left(\sigma_+ IF\left(\frac{u}{\sigma_X}+\frac{v}{\sigma_Y};Q,F_+\right)\right. \\
&\left.-\sigma_- IF\left(\frac{u}{\sigma_X}-\frac{v}{\sigma_Y};Q,F_-\right)\right) \\
&-(\sigma_+^2-\sigma_-^2)\left(\sigma_+ IF\left(\frac{u}{\sigma_X}+\frac{v}{\sigma_Y};Q,F_+\right)\right. \\
&\left.+\sigma_- IF\left(\frac{u}{\sigma_X}-\frac{v}{\sigma_Y};Q,F_-\right)\right)\Bigg].
\end{aligned} \tag{10}
$$

Thus, we obtain the following influence function for the correlation estimator $\hat{\rho}_Q$:

$$
\begin{aligned}
IF((u,v);\rho_Q,\mathbf{F}) = \frac{\sigma_+\sigma_-}{4}\Bigg[&\sigma_- IF\left(\frac{u}{\sigma_X}+\frac{v}{\sigma_Y};Q,F_+\right) \\
&-\sigma_+ IF\left(\frac{u}{\sigma_X}-\frac{v}{\sigma_Y};Q,F_-\right)\Bigg].
\end{aligned} \tag{11}
$$

It can be checked that the influence functions of both the covariance estimator and the correlation estimator satisfy $\int IF\, d\mathbf{F}=0$.

4.3. *Asymptotic Variance*

Under regularity conditions, both $\hat{\gamma}_Q$ and $\hat{\rho}_Q$ are consistent estimators, since Q_n is consistent (Rousseeuw and Croux, 1993). Moreover, they are asymptotically normal with asymptotic variance (of order $1/n$) given by:

$$
\begin{aligned}
V(\gamma_Q,\mathbf{F}) &= \int IF((u,v);\gamma_Q,\mathbf{F})^2\, d\mathbf{F}(u,v), \\
V(\rho_Q,\mathbf{F}) &= \int IF((u,v);\rho_Q,\mathbf{F})^2\, d\mathbf{F}(u,v), \\
V(Q,F) &= \int IF(u;Q,F)^2\, dF(u).
\end{aligned} \tag{12}
$$

Subsequently, we assume a bivariate Gaussian distribution $\mathbf{F}=\mathbf{\Phi}$ for $(X, Y)^T$, i.e.:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \gamma \\ \gamma & \sigma_Y^2 \end{pmatrix}\right) = N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right),$$

where γ is the covariance and ρ is the correlation between X and Y. We have:

PROPOSITION 1. *The asymptotic variance of the covariance estimator $\hat{\gamma}_Q$ is*

$$V(\gamma_Q, \mathbf{\Phi}) = 2V(Q, \Phi)(\sigma_X^2\sigma_Y^2 + \gamma^2) = 1.215(\sigma_X^2\sigma_Y^2 + \gamma^2), \tag{13}$$

and the asymptotic variance of the correlation estimator $\hat{\rho}_Q$ is

$$V(\rho_Q, \mathbf{\Phi}) = 2V(Q, \Phi)(1-\rho^2)^2 = 1.215(1-\rho^2)^2, \tag{14}$$

where Φ represents the standard Gaussian distribution function, i.e. with mean zero and variance one.

In Table I, we compute the variance of the covariance estimator and of the correlation estimator for various underlying variances and covariances. The results are presented in the fourth and fifth columns of Table I. Proposition 1 is in fact valid for a dispersion estimator based on any statistical functional of scale. For instance, we can replace the Q_n estimator in Proposition 1 with the maximum likelihood estimator of scale MLE, and calculate the closed form of the variance of the covariance estimator $\hat{\gamma}_{MLE}$ and of the correlation estimator $\hat{\rho}_{MLE}$:

TABLE I

Asymptotic Variance and Efficiency of the Dispersion Estimators $\hat{\gamma}_Q$ and $\hat{\rho}_Q$, in the Case of Gaussian Distributions

σ_X^2	σ_Y^2	γ	$V(\gamma_Q, \mathbf{\Phi})$	$V(\rho_Q, \mathbf{\Phi})$	$\mathrm{Eff}(\gamma_Q, \mathbf{\Phi})$	$\mathrm{Eff}(\rho_Q, \mathbf{\Phi})$
1	1	0	1.215	1.215	0.823	0.823
1	1	0.2	1.264	1.120	0.701	0.791
1	1	0.5	1.519	0.683	0.296	0.658
1	1	0.8	1.993	0.157	0.040	0.501
1	2	0.5	2.735	0.930	0.498	0.732
1	3	0.5	3.950	1.021	0.589	0.758
1	10	0.5	12.458	1.155	0.745	0.803

Note. The numerical values of the asymptotic variances were computed with Proposition 1 and the numerical values of the asymptotic efficiencies were computed with Proposition 3.

Corollary 1. *The asymptotic variance of the covariance estimator* $\hat{\gamma}_{MLE}$ *is*

$$V(\gamma_{MLE}, \mathbf{\Phi}) = \sigma_X^2\sigma_Y^2 + \gamma^2, \tag{15}$$

and the asymptotic variance of the correlation estimator $\hat{\rho}_{MLE}$ *is*

$$V(\rho_{MLE}, \mathbf{\Phi}) = (1-\rho^2)^2. \tag{16}$$

Thus, the asymptotic variance of covariance estimators increases with increasing dependence, whereas it decreases with increasing dependence for correlation estimators. In fact, we see that the asymptotic variance of dispersion estimators for multivariate Gaussian distributions is proportional to the asymptotic variance of the underlying scale estimator. The proportionality value depends only on the underlying dependence.

4.4. *Fisher Information*

For Gaussian distributions, a closed form of the Fisher information of both covariance and correlation can be obtained:

Proposition 2. *The Fisher information of the covariance* γ *is*

$$I(\gamma, \mathbf{\Phi}) = \frac{\sigma_X^2\sigma_Y^2 + \gamma^2}{(\sigma_X^2\sigma_Y^2 - \gamma^2)^2}, \tag{17}$$

and the Fisher information of the correlation ρ *is*

$$I(\rho, \mathbf{\Phi}) = \frac{1+\rho^2}{(1-\rho^2)^2}. \tag{18}$$

Note that from the Fisher information for the covariance γ, it is straightforward to get the Fisher information for the correlation, since the correlation ρ is simply $\frac{\gamma}{\sigma_X\sigma_Y}$.

4.5. *Efficiency*

Efficiency is defined as the inverse of the product of the Fisher information and the asymptotic variance of the estimator. Thus, for Gaussian distributions, we can calculate the asymptotic efficiency of $\hat{\gamma}_Q$ and $\hat{\rho}_Q$.

Proposition 3. *The asymptotic efficiency of the covariance estimator* $\hat{\gamma}_Q$ *is*

$$\mathrm{Eff}(\gamma_Q, \mathbf{\Phi}) = \frac{(\sigma_X^2\sigma_Y^2 - \gamma^2)^2}{2V(Q, \Phi)(\sigma_X^2\sigma_Y^2 + \gamma^2)^2} = 0.823\,\frac{(\sigma_X^2\sigma_Y^2 - \gamma^2)^2}{(\sigma_X^2\sigma_Y^2 + \gamma^2)^2}, \tag{19}$$

and the asymptotic efficiency of the correlation estimator $\hat{\rho}_Q$ *is*

$$\mathrm{Eff}(\rho_Q, \mathbf{\Phi}) = \frac{1}{2V(Q, \Phi)(1+\rho^2)} = 0.823\,\frac{1}{1+\rho^2}, \tag{20}$$

We present the efficiency of both the covariance and the correlation estimators in the sixth and seventh column of Table I, calculated by Proposition 3. In fact, Proposition 3 is valid for a dispersion estimator based on any statistical functional of scale. For instance, we can again replace the Q_n estimator in Proposition 3 with the maximum likelihood estimator of scale MLE, and calculate the closed form of the asymptotic efficiency of the covariance estimator $\hat{\gamma}_{MLE}$ and of the correlation estimator $\hat{\rho}_{MLE}$:

COROLLARY 2. *The asymptotic efficiency of the maximum likelihood estimator of the covariance* $\hat{\gamma}_{MLE}$ *is*

$$\mathrm{Eff}(\gamma_{MLE}, \mathbf{\Phi}) = \frac{(\sigma_X^2\sigma_Y^2 - \gamma^2)^2}{(\sigma_X^2\sigma_Y^2 + \gamma^2)^2}, \tag{21}$$

and the asymptotic efficiency of the maximum likelihood estimator of the correlation $\hat{\rho}_{MLE}$ *is*

$$\mathrm{Eff}(\rho_{MLE}, \mathbf{\Phi}) = \frac{1}{1+\rho^2}. \tag{22}$$

5. COMPARISONS

We first compare the estimator we proposed here, $\hat{\gamma}_Q$, with the maximum likelihood one, $\hat{\gamma}_{MLE}$, and another componentwise robust estimator, $\hat{\gamma}_{MAD}$, based on the median absolute deviation (e.g. Hampel *et al.*, 1986). Next we compare $\hat{\Sigma}_Q$ with the global estimators $\hat{\Sigma}_{MME}$, $\hat{\Sigma}_{MVE}$, and $\hat{\Sigma}_{MCD}$. We focus on covariance estimation here since as we will point out in Section 5.1, it is more difficult than correlation estimation.

5.1. *Comparison with MLE and MAD*

As we have pointed out, Proposition 1 is valid for any dispersion estimator based on an M-estimator of scale (Genton and Ma, 1999). In Fig. 1, we plot the asymptotic variance of the three covariance estimators $\hat{\gamma}_Q$, $\hat{\gamma}_{MLE}$ and $\hat{\gamma}_{MAD}$, for a standardized Gaussian distribution with correlation ρ. Similarly we also plot the asymptotic variance of the three corresponding correlation estimators in Fig. 2. The three curves in Fig. 1 and in Fig. 2 are computed with the formula in Proposition 1 and in Corollary 1. We can see that

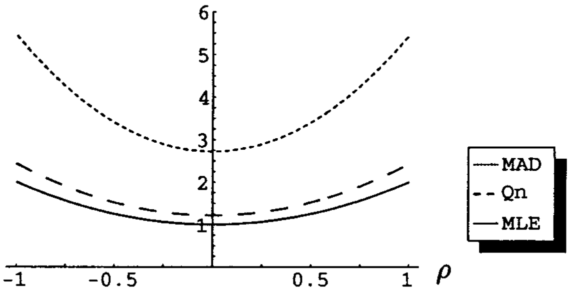

FIG. 1. The asymptotic variance of the covariance estimators based on Q_n, MLE, and MAD, for a standardized bivariate Gaussian distribution with correlation ρ. The $\hat{\gamma}_{MLE}$ estimator has the smallest asymptotic variance, the asymptotic variance of the $\hat{\gamma}_Q$ estimator is slightly larger, whereas $\hat{\gamma}_{MAD}$ has an asymptotic variance much larger than the other two. For all three estimators, the asymptotic variance increases when the covariance between the two random variables increases.

when the covariance (correlation) between two random variables increases, the asymptotic variance of the covariance estimator increases, while the asymptotic variance of the correlation estimator decreases. As a consequence, correlation estimation is easier than covariance estimation, in the sense that it has smaller variability. In the independent standard Gaussian distribution case, i.e. $\rho = 0$, the asymptotic variance of the covariance estimator and the correlation estimator have the same value.

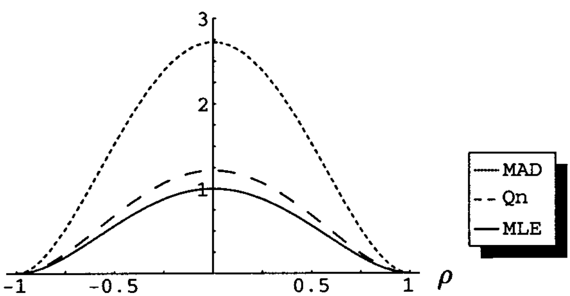

FIG. 2. The asymptotic variance of the correlation estimators based on Q_n, MLE, and MAD, for a standardized bivariate Gaussian distribution with correlation ρ. The $\hat{\rho}_{MLE}$ estimator has the smallest asymptotic variance, the asymptotic variance of the $\hat{\rho}_Q$ estimator is slightly larger, whereas $\hat{\rho}_{MAD}$ has an asymptotic variance much larger than the other two. For all three estimators, the asymptotic variance decreases when the covariance between the two random variables increases.

TABLE II

The Mean and Variance of the Covariance Estimators $\hat{\gamma}_Q$, $\hat{\gamma}_{MLE}$, and $\hat{\gamma}_{MAD}$

Sample size	Mean			Variance		
	$\hat{\gamma}_Q$	$\hat{\gamma}_{MLE}$	$\hat{\gamma}_{MAD}$	$\hat{\gamma}_Q$	$\hat{\gamma}_{MLE}$	$\hat{\gamma}_{MAD}$
20	−0.007	0.005	−0.002	1.630	0.966	2.684
100	−0.002	−0.002	−0.002	1.257	0.988	2.865
200	−0.003	−0.003	−0.003	1.320	1.057	2.794

Note. The data followed an independent standard Gaussian distribution, and we calculated the mean and variance after running 1000 samples. The three estimators are all unbiased, and the variance of the $\hat{\gamma}_{MAD}$ is significantly larger than the other two.

We carried out some simulations to test the mean and variance of the dispersion estimators based on the Q_n, MLE, and MAD estimators. The simulation was on two standardized Gaussian random variables with covariance 0 and 0.5, and based on 1000 samples. The sample sizes were 20, 100 and 200. The results are presented in Table II and III. We can see that the estimators are unbiased and the variance of the estimators increases as the variance between the two random variables increases.

5.2. *Comparison with MME, MVE, and MCD*

In order to compare the highly robust componentwise estimator $\hat{\Sigma}_Q$ with the global estimators $\hat{\Sigma}_{MME}$, $\hat{\Sigma}_{MVE}$, and $\hat{\Sigma}_{MCD}$, we carried out some

TABLE III

The Mean and Variance of the Covariance Estimators $\hat{\gamma}_Q$, $\hat{\gamma}_{MLE}$, and $\hat{\gamma}_{MAD}$

Sample size	Mean			Variance		
	$\hat{\gamma}_Q$	$\hat{\gamma}_{MLE}$	$\hat{\gamma}_{MAD}$	$\hat{\gamma}_Q$	$\hat{\gamma}_{MLE}$	$\hat{\gamma}_{MAD}$
20	0.526	0.477	0.506	2.018	1.163	3.497
100	0.499	0.493	0.496	1.715	1.302	3.477
200	0.504	0.500	0.500	1.649	1.258	3.254

Note. The data followed a Gaussian distribution with mean zero and variance one, and the covariance γ between the two random variables was 0.5. We calculated the mean and variance after running 1000 samples. The three estimators are all unbiased, and the variance of the $\hat{\gamma}_{MAD}$ is significantly larger than the other two.

simulations on three variables, i.e. Σ is a 3×3 matrix, from a multivariate Gaussian distribution. In Table IV,

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 1.0 & 0.9 & -0.5 \\ 0.9 & 1.0 & 0.2 \\ -0.5 & 0.2 & 3.0 \end{pmatrix}, \tag{23}$$

and in Table V,

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 1.0 & 0.8 & 0.5 \\ 0.8 & 1.0 & 0.8 \\ 0.5 & 0.8 & 1.0 \end{pmatrix}. \tag{24}$$

Both situations include some large correlations (0.9 in (23) and 0.8 in (24)). We generated 1000 sets of data, each with sample size 100 and we used the four estimators to calculate the covariance matrix Σ. In the statistical software S-PLUS, the $\hat{\Sigma}_{MME}$, $\hat{\Sigma}_{MVE}$, and $\hat{\Sigma}_{MCD}$ estimators are respectively implemented as `var`, `cov.mve`, and `cov.mcd$cov` (note that the last two functions yield one-step reweighted estimators based on MVE and MCD, see e.g. Rousseeuw and Van Driessen (1999)). We implemented $\hat{\Sigma}_Q$ in S-PLUS from a C-routine provided by Croux and Rousseeuw (1992). Based on the 1000 estimated covariance matrices, we computed the mean and the variance of the estimations. The results are presented in Table IV and V. In the first columns, the data do not contain any outliers, in the second column, 10% of the data have a covariance matrix 9Σ (explode type outliers), in the third column, 10% of the data have a covariance matrix $\Sigma/9$ (implode type outliers). In these examples, the matrices $\hat{\Sigma}_Q$ are positive definite. In case they are not positive definite, a transformation as described at the end of Section 3.2 must be applied. For convenience, we call the sum of the absolute values of all the entries of a matrix the 1-norm of the matrix, and denote it by $\|\cdot\|_1$. The smallest 1-norm in each column is emphasized by boldface font. From the tables, we can see that when there is no outliers, $\hat{\Sigma}_{MME}$ behaves the best, $\hat{\Sigma}_Q$ is slightly worse, while $\hat{\Sigma}_{MVE}$ and $\hat{\Sigma}_{MCD}$ behave the worst. When the outliers are of explode type (the observation tends to be much larger than the true value), $\hat{\Sigma}_{MVE}$ has the best estimation, followed by $\hat{\Sigma}_{MCD}$ and $\hat{\Sigma}_Q$, whereas $\hat{\Sigma}_{MME}$ gives the worst result. For outliers that are of implode type (the observation tends to be much smaller than the true value), $\hat{\Sigma}_{MME}$ and $\hat{\Sigma}_Q$ both give relatively good estimations, whereas $\hat{\Sigma}_{MVE}$ is worse and $\hat{\Sigma}_{MCD}$ gives the worst result. This can be understood if we notice that the estimators $\hat{\Sigma}_{MVE}$ and $\hat{\Sigma}_{MCD}$ only take into account half of the observations which are distributed nearest to an estimated center. Thus exploding outliers will not have much

TABLE IV

The Biases and Variances of the Estimators $\hat{\Sigma}_Q$, $\hat{\Sigma}_{MME}$, $\hat{\Sigma}_{MVE}$, and $\hat{\Sigma}_{MCD}$, with $\boldsymbol{\mu}$ and Σ Given in Eq. (23)

	No outliers	10% explode	10% implode
$\text{bias}(\hat{\Sigma}_Q)$	$\begin{pmatrix} 0.011 & 0.008 & -0.010 \\ 0.008 & 0.008 & -0.005 \\ -0.010 & -0.005 & 0.021 \end{pmatrix}$	$\begin{pmatrix} 0.268 & 0.239 & -0.146 \\ 0.239 & 0.262 & 0.040 \\ -0.146 & 0.040 & 0.794 \end{pmatrix}$	$\begin{pmatrix} -0.137 & -0.125 & 0.068 \\ -0.125 & -0.137 & -0.025 \\ 0.068 & -0.025 & -0.405 \end{pmatrix}$
$\\|\text{bias}(\hat{\Sigma}_Q)\\|_1$	0.086	2.174	1.115
$\text{variance}(\hat{\Sigma}_Q)$	$\begin{pmatrix} 0.026 & 0.023 & 0.043 \\ 0.023 & 0.025 & 0.041 \\ 0.043 & 0.041 & 0.259 \end{pmatrix}$	$\begin{pmatrix} 0.049 & 0.042 & 0.081 \\ 0.042 & 0.043 & 0.071 \\ 0.081 & 0.071 & 0.433 \end{pmatrix}$	$\begin{pmatrix} 0.022 & 0.020 & 0.038 \\ 0.020 & 0.023 & 0.036 \\ 0.038 & 0.036 & 0.202 \end{pmatrix}$
$\\|\text{variance}(\hat{\Sigma}_Q)\\|_1$	0.524	0.913	0.435
$\text{bias}(\hat{\Sigma}_{MME})$	$\begin{pmatrix} 0.004 & 0.003 & -0.004 \\ 0.003 & 0.001 & -0.004 \\ -0.004 & -0.004 & 0.003 \end{pmatrix}$	$\begin{pmatrix} 0.796 & 0.707 & -0.436 \\ 0.707 & 0.778 & 0.124 \\ -0.436 & 0.124 & 2.410 \end{pmatrix}$	$\begin{pmatrix} -0.091 & -0.083 & 0.041 \\ -0.083 & -0.092 & -0.020 \\ 0.041 & -0.020 & -0.260 \end{pmatrix}$
$\\|\text{bias}(\hat{\Sigma}_{MME})\\|_1$	**0.030**	6.518	**0.731**
$\text{variance}(\hat{\Sigma}_{MME})$	$\begin{pmatrix} 0.020 & 0.018 & 0.031 \\ 0.018 & 0.019 & 0.029 \\ 0.031 & 0.029 & 0.189 \end{pmatrix}$	$\begin{pmatrix} 0.183 & 0.164 & 0.294 \\ 0.164 & 0.176 & 0.267 \\ 0.294 & 0.267 & 1.712 \end{pmatrix}$	$\begin{pmatrix} 0.019 & 0.017 & 0.030 \\ 0.017 & 0.018 & 0.027 \\ 0.030 & 0.027 & 0.168 \end{pmatrix}$
$\\|\text{variance}(\hat{\Sigma}_{MME})\\|_1$	**0.384**	3.521	**0.353**

$\text{bias}(\hat{\Sigma}_{MVE})$	$\begin{pmatrix} -0.071 & -0.066 & 0.033 \\ -0.066 & -0.076 & -0.019 \\ 0.033 & -0.019 & -0.218 \end{pmatrix}$	$\begin{pmatrix} -0.004 & -0.006 & -0.006 \\ -0.006 & -0.010 & -0.011 \\ -0.006 & -0.011 & -0.022 \end{pmatrix}$	$\begin{pmatrix} -0.203 & -0.182 & 0.102 \\ -0.182 & -0.201 & -0.038 \\ 0.102 & -0.038 & -0.597 \end{pmatrix}$
$\Vert\text{bias}(\hat{\Sigma}_{MVE})\Vert_1$	0.601	**0.082**	1.645
$\text{variance}(\hat{\Sigma}_{MVE})$	$\begin{pmatrix} 0.024 & 0.022 & 0.039 \\ 0.022 & 0.024 & 0.037 \\ 0.039 & 0.037 & 0.226 \end{pmatrix}$	$\begin{pmatrix} 0.034 & 0.030 & 0.055 \\ 0.030 & 0.033 & 0.049 \\ 0.055 & 0.049 & 0.308 \end{pmatrix}$	$\begin{pmatrix} 0.024 & 0.021 & 0.038 \\ 0.021 & 0.024 & 0.034 \\ 0.038 & 0.034 & 0.215 \end{pmatrix}$
$\Vert\text{variance}(\hat{\Sigma}_{MVE})\Vert_1$	0.470	**0.643**	0.449
$\text{bias}(\hat{\Sigma}_{MCD})$	$\begin{pmatrix} -0.167 & -0.153 & 0.077 \\ -0.153 & -0.174 & -0.044 \\ 0.077 & -0.044 & -0.515 \end{pmatrix}$	$\begin{pmatrix} -0.111 & -0.101 & 0.049 \\ -0.101 & -0.114 & -0.032 \\ 0.049 & -0.032 & -0.344 \end{pmatrix}$	$\begin{pmatrix} -0.290 & -0.260 & 0.150 \\ -0.260 & -0.287 & -0.051 \\ 0.150 & -0.051 & -0.864 \end{pmatrix}$
$\Vert\text{bias}(\hat{\Sigma}_{MCD})\Vert_1$	1.404	0.933	2.363
$\text{variance}(\hat{\Sigma}_{MCD})$	$\begin{pmatrix} 0.031 & 0.029 & 0.057 \\ 0.029 & 0.032 & 0.053 \\ 0.057 & 0.053 & 0.299 \end{pmatrix}$	$\begin{pmatrix} 0.037 & 0.034 & 0.062 \\ 0.034 & 0.037 & 0.057 \\ 0.062 & 0.057 & 0.335 \end{pmatrix}$	$\begin{pmatrix} 0.026 & 0.024 & 0.044 \\ 0.024 & 0.027 & 0.042 \\ 0.044 & 0.042 & 0.248 \end{pmatrix}$
$\Vert\text{variance}(\hat{\Sigma}_{MCD})\Vert_1$	0.640	0.715	0.521

Note. In each column, the smallest 1-norm is emphasized by boldface font.

TABLE V

The Biases and Variances of the Estimators $\hat{\Sigma}_Q$, $\hat{\Sigma}_{MME}$, $\hat{\Sigma}_{MVE}$, and $\hat{\Sigma}_{MCD}$, with $\boldsymbol{\mu}$ and Σ Given in Eq. (24)

	No outliers	10% explode	10% implode
$\text{bias}(\hat{\Sigma}_Q)$	$\begin{pmatrix} 0.012 & 0.012 & 0.004 \\ 0.012 & 0.009 & 0.003 \\ 0.004 & 0.003 & 0.002 \end{pmatrix}$	$\begin{pmatrix} 0.261 & 0.203 & 0.132 \\ 0.203 & 0.254 & 0.211 \\ 0.132 & 0.211 & 0.263 \end{pmatrix}$	$\begin{pmatrix} -0.122 & -0.094 & -0.059 \\ -0.094 & -0.124 & -0.100 \\ -0.059 & -0.100 & -0.131 \end{pmatrix}$
$\Vert\text{bias}(\hat{\Sigma}_Q)\Vert_1$	0.061	1.870	0.883
$\text{variance}(\hat{\Sigma}_Q)$	$\begin{pmatrix} 0.025 & 0.021 & 0.016 \\ 0.021 & 0.027 & 0.022 \\ 0.016 & 0.022 & 0.027 \end{pmatrix}$	$\begin{pmatrix} 0.044 & 0.037 & 0.029 \\ 0.037 & 0.045 & 0.038 \\ 0.029 & 0.038 & 0.048 \end{pmatrix}$	$\begin{pmatrix} 0.021 & 0.019 & 0.015 \\ 0.019 & 0.024 & 0.020 \\ 0.015 & 0.020 & 0.023 \end{pmatrix}$
$\Vert\text{variance}(\hat{\Sigma}_Q)\Vert_1$	0.197	0.345	0.176
$\text{bias}(\hat{\Sigma}_{MME})$	$\begin{pmatrix} -0.009 & -0.006 & -0.005 \\ -0.006 & -0.005 & -0.005 \\ -0.005 & -0.005 & -0.004 \end{pmatrix}$	$\begin{pmatrix} 0.789 & 0.619 & 0.376 \\ 0.619 & 0.772 & 0.612 \\ 0.376 & 0.612 & 0.769 \end{pmatrix}$	$\begin{pmatrix} -0.081 & -0.064 & -0.040 \\ -0.064 & -0.084 & -0.069 \\ -0.040 & -0.069 & -0.089 \end{pmatrix}$
$\Vert\text{bias}(\hat{\Sigma}_{MME})\Vert_1$	**0.050**	5.544	**0.600**
$\text{variance}(\hat{\Sigma}_{MME})$	$\begin{pmatrix} 0.019 & 0.016 & 0.012 \\ 0.016 & 0.019 & 0.016 \\ 0.012 & 0.016 & 0.020 \end{pmatrix}$	$\begin{pmatrix} 0.175 & 0.143 & 0.107 \\ 0.143 & 0.171 & 0.140 \\ 0.107 & 0.140 & 0.173 \end{pmatrix}$	$\begin{pmatrix} 0.018 & 0.016 & 0.013 \\ 0.016 & 0.019 & 0.016 \\ 0.013 & 0.016 & 0.019 \end{pmatrix}$
$\Vert\text{variance}(\hat{\Sigma}_{MME})\Vert_1$	**0.146**	1.299	**0.146**

$\mathrm{bias}(\hat{\Sigma}_{MVE})$	$\begin{pmatrix} -0.084 & -0.066 & -0.042 \\ -0.066 & -0.080 & -0.067 \\ -0.042 & -0.067 & -0.083 \end{pmatrix}$	$\begin{pmatrix} -0.014 & -0.014 & -0.005 \\ -0.014 & -0.019 & -0.010 \\ -0.005 & -0.010 & -0.010 \end{pmatrix}$	$\begin{pmatrix} -0.185 & -0.147 & -0.091 \\ -0.147 & -0.190 & -0.154 \\ -0.091 & -0.154 & -0.196 \end{pmatrix}$
$\Vert\mathrm{bias}(\hat{\Sigma}_{MVE})\Vert_1$	0.597	**0.101**	1.355
$\mathrm{variance}(\hat{\Sigma}_{MVE})$	$\begin{pmatrix} 0.023 & 0.019 & 0.016 \\ 0.019 & 0.025 & 0.021 \\ 0.016 & 0.021 & 0.027 \end{pmatrix}$	$\begin{pmatrix} 0.032 & 0.026 & 0.020 \\ 0.026 & 0.032 & 0.027 \\ 0.020 & 0.027 & 0.033 \end{pmatrix}$	$\begin{pmatrix} 0.023 & 0.021 & 0.016 \\ 0.021 & 0.026 & 0.022 \\ 0.016 & 0.022 & 0.025 \end{pmatrix}$
$\Vert\mathrm{variance}(\hat{\Sigma}_{MVE})\Vert_1$	0.187	**0.243**	0.192
$\mathrm{bias}(\hat{\Sigma}_{MCD})$	$\begin{pmatrix} -0.177 & -0.143 & -0.091 \\ -0.143 & -0.179 & -0.147 \\ -0.091 & -0.147 & -0.181 \end{pmatrix}$	$\begin{pmatrix} -0.124 & -0.099 & -0.056 \\ -0.099 & -0.119 & -0.089 \\ -0.056 & -0.089 & -0.108 \end{pmatrix}$	$\begin{pmatrix} -0.281 & -0.224 & -0.139 \\ -0.224 & -0.285 & -0.229 \\ -0.139 & -0.229 & -0.288 \end{pmatrix}$
$\Vert\mathrm{bias}(\hat{\Sigma}_{MCD})\Vert_1$	1.299	0.951	2.038
$\mathrm{variance}(\hat{\Sigma}_{MCD})$	$\begin{pmatrix} 0.031 & 0.026 & 0.022 \\ 0.026 & 0.032 & 0.027 \\ 0.022 & 0.027 & 0.033 \end{pmatrix}$	$\begin{pmatrix} 0.037 & 0.030 & 0.025 \\ 0.030 & 0.035 & 0.030 \\ 0.025 & 0.030 & 0.036 \end{pmatrix}$	$\begin{pmatrix} 0.026 & 0.023 & 0.020 \\ 0.023 & 0.029 & 0.025 \\ 0.020 & 0.025 & 0.029 \end{pmatrix}$
$\Vert\mathrm{variance}(\hat{\Sigma}_{MCD})\Vert_1$	0.246	0.278	0.220

Note. In each column, the smallest 1-norm is emphasized by boldface font.

effect on the estimators, whereas imploding outliers can bring significant challenge to the estimators. In other words, $\hat{\Sigma}_{MVE}$ and $\hat{\Sigma}_{MCD}$ are robust only against exploding outliers, not imploding outliers. $\hat{\Sigma}_{MME}$ gives very good results in the imploding case because the implode values we tested are not extreme case and they only take 10% of the data, so under the averaging procedure, the effect of imploding is very small. $\hat{\Sigma}_Q$ is not the best in any of the three simulations, but it is relatively good in all three simulations. So, in practice when one does not really know what kind of outliers exist and how many percentage of the data are contaminated, $\hat{\Sigma}_Q$ is a suitable estimator to use. In particular with no outliers, the bias of $\hat{\Sigma}_Q$ is almost as small as the bias of $\hat{\Sigma}_{MME}$. Note that the simulation results for $\hat{\Sigma}_Q$ are valid only for (23) and (24) because our estimator is not affine invariant. However, the results for (23) and (24) are quite similar.

6. CONCLUSION

A new componentwise estimator of a dispersion matrix, based on a highly robust estimator of scale, has been proposed in this article. Its robustness properties were studied by means of the influence function and the breakdown point. Further characteristics such as asymptotic variance and efficiency were also analyzed. A major advantage of the novel estimator is that its behavior is close to the method of moment estimator in noncontaminated situations, whereas it is highly robust in contaminated ones. It was shown in the componentwise approach, for multivariate Gaussian distributions, that covariance matrix estimation is more difficult than correlation matrix estimation. The reason is that the asymptotic variance of the covariance estimator increases with increasing dependence, whereas it decreases with increasing dependence for correlation estimators. We also proved that the asymptotic variance of dispersion estimators for multivariate Gaussian distributions is proportional to the asymptotic variance of the underlying scale estimator. The proportionality value depends only on the underlying dependence. Therefore, the highly robust dispersion estimator is the best robust choice at the present time in the componentwise approach, because it combines small variability and robustness properties such as high breakdown point and bounded influence function. A simulation study was carried out in order to assess the behavior of the new estimator. First, a comparison with another robust componentwise estimator based on the median absolute deviation scale estimator, was performed. The highly robust properties of the new estimator were confirmed. Moreover, it has been shown that the behavior of the new estimator is better than the one based on the MAD, although the latter is the most B-robust componentwise dispersion estimator (Genton and Ma, 1999). A second comparison with global estimators like the method of moment estimator, the minimum volume ellipsoid estimator, and the minimum covariance determinant estimator, has also been performed, with two types of outliers. In this case, the highly robust dispersion matrix estimator turns out to be a compromise between the high efficiency of the method of moment estimator in noncontaminated situations and the highly robust properties of the minimum volume ellipsoid and minimum covariance determinant estimators in contaminated situations, with exploding type of outliers.

7. PROOFS

7.1. *Proof of Proposition* 1

The asymptotic variance of $\hat{\gamma}_Q$ at $\mathbf{\Phi}$ is

$$\begin{aligned} V(\gamma_Q, \mathbf{\Phi}) &= \iint IF^2((u,v); \gamma_Q, \mathbf{\Phi})\, d\mathbf{\Phi}(u,v) \\ &= \frac{\sigma_X^2 \sigma_Y^2}{4} \iint \left[\sigma_+ IF\left(\frac{u}{\sigma_X} + \frac{v}{\sigma_Y}; Q, \Phi_+\right) \right. \\ &\quad \left. - \sigma_- IF\left(\frac{u}{\sigma_X} - \frac{v}{\sigma_Y}; Q, \Phi_-\right)\right]^2 d\mathbf{\Phi}(u,v). \end{aligned}$$

The change of variables

$$\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} \dfrac{1}{\sigma_+ \sigma_X} & \dfrac{1}{\sigma_+ \sigma_Y} \\ \dfrac{1}{\sigma_- \sigma_X} & \dfrac{-1}{\sigma_- \sigma_Y} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

yields

$$ds\, dt = \frac{2}{\sigma_+ \sigma_- \sigma_X \sigma_Y}\, du\, dv$$

and corresponds to the random variables $\frac{X}{\sigma_+ \sigma_X} + \frac{Y}{\sigma_+ \sigma_Y}$ and $\frac{X}{\sigma_- \sigma_X} - \frac{Y}{\sigma_- \sigma_Y}$, each of which follows the standard normal distribution Φ and is independent of each other. Therefore

$$\begin{aligned} V(\gamma_Q, \mathbf{\Phi}) &= \frac{\sigma_X^2 \sigma_Y^2}{4} \left[\sigma_+^4 \iint IF^2(s; Q, \Phi)\, d\Phi(s)\, d\Phi(t) \right. \\ &\quad \left. + 0 + \sigma_-^4 \iint IF^2(t; Q, \Phi)\, d\Phi(s)\, d\Phi(t) \right]. \end{aligned}$$

Note that we use the linear property of the influence function (Hampel *et al.*, 1986): $IF(\alpha x; Q, \Phi_{\alpha X}) = \alpha IF(x; Q, \Phi_X)$, $\forall \alpha \in \mathbb{R}$. Thus:

$$\begin{aligned}
V(\gamma_Q, \mathbf{\Phi}) &= \frac{\sigma_X^2 \sigma_Y^2}{4} [\sigma_+^4 + 0 + \sigma_-^4]\, V(Q, \Phi) \\
&= \frac{\sigma_X^2 \sigma_Y^2}{4} \left[\left(2 + 2\frac{\gamma}{\sigma_X \sigma_Y}\right)^2 + \left(2 - 2\frac{\gamma}{\sigma_X \sigma_Y}\right)^2 \right] V(Q, \Phi) \\
&= 2V(Q, \Phi)(\sigma_X^2 \sigma_Y^2 + \gamma^2) \\
&= 1.215(\sigma_X^2 \sigma_Y^2 + \gamma^2).
\end{aligned} \tag{30}$$

Similarly, the asymptotic variance of $\hat{\rho}_Q$ at $\mathbf{\Phi}$ is

$$\begin{aligned}
V(\rho_Q, \mathbf{\Phi}) &= \iint IF^2((u, v); \rho_Q, \mathbf{\Phi})\, d\mathbf{\Phi}(u, v) \\
&= \frac{16\sigma_+^2 \sigma_-^2}{(\sigma_+^2 + \sigma_-^2)^4} \iint \left[\sigma_- IF\left(\frac{u}{\sigma_X} + \frac{v}{\sigma_Y}; Q, \Phi_+\right) \right. \\
&\quad \left. - \sigma_+ IF\left(\frac{u}{\sigma_X} - \frac{v}{\sigma_Y}; Q, \Phi_-\right) \right]^2 d\mathbf{\Phi}(u, v).
\end{aligned}$$

Using the same technique as above, we have:

$$\begin{aligned}
V(\rho_Q, \mathbf{\Phi}) &= \frac{16\sigma_+^2 \sigma_-^2}{(\sigma_+^2 + \sigma_-^2)^4} \left[\sigma_+^2 \sigma_-^2 \iint IF^2(s; Q, \Phi)\, d\Phi(s)\, d\Phi(t) \right. \\
&\quad \left. + 0 + \sigma_+^2 \sigma_-^2 \iint IF^2(t; Q, \Phi)\, d\Phi(s)\, d\Phi(t) \right]. \\
&= \frac{16\sigma_+^2 \sigma_-^2}{(\sigma_+^2 + \sigma_-^2)^4} [2\sigma_+^2 \sigma_-^2]\, V(Q, \Phi) \\
&= \frac{32\sigma_+^4 \sigma_-^4}{(\sigma_+^2 + \sigma_-^2)^4} V(Q, \Phi) \\
&= 2V(Q, \Phi)\left(1 - \frac{\gamma^2}{\sigma_X^2 \sigma_Y^2}\right) \\
&= 1.215(1 - \rho^2)^2.
\end{aligned}$$

7.2. *Proof of Proposition* 2

We write out the probability density function of the bivariate Gaussian distribution

$$\begin{aligned}\mathbf{\Phi}(u, v) &= \frac{1}{2\pi\sqrt{ab-\gamma^2}} \exp\left(-\frac{1}{2}(u \ v)\begin{pmatrix} a & \gamma \\ \gamma & b\end{pmatrix}^{-1}\begin{pmatrix} u \\ v\end{pmatrix}\right) \\ &= \frac{1}{2\pi\sqrt{ab-\gamma^2}} \exp\left(\frac{bu^2+av^2-2\gamma uv}{-2(ab-\gamma^2)}\right) \\ &= \frac{1}{\pi\sqrt{2B}} \exp\left(-\frac{A}{B}\right),\end{aligned}$$

where $A = bu^2 + av^2 - 2\gamma uv$ and $B = 2ab - 2\gamma^2$. Following the definition of the Fisher information, we have

$$\begin{aligned} I(\gamma, \mathbf{\Phi}) &= \iint \left(\frac{\partial}{\partial \gamma} \log \phi(u, v)\right)^2 \phi(u, v)\, du\, dv \\ &= \iint 2\left(\frac{\gamma e^{-A/B}}{2\pi(ab-\gamma^2)^{3/2}} + \frac{\left(\frac{2uv}{B} - \frac{4\gamma A}{B^2}\right)e^{-A/B}}{2\pi(ab-\gamma^2)^{1/2}}\right)^2 \frac{\pi(ab-\gamma^2)^{1/2}}{e^{-A/B}}\, du\, dv \\ &= \iint \frac{e^{-A/B}}{2\pi}(ab-\gamma^2)^{-9/2}\,[(ab-\gamma^2)\,\gamma - \gamma(bu^2+av^2) \\ &\quad + (ab+\gamma^2)\, uv]^2\, du\, dv. \end{aligned} \tag{25}$$

Let

$$\begin{cases} s = \sqrt{b}\, u + \sqrt{a}\, v, \\ t = \sqrt{b}\, u - \sqrt{a}\, v. \end{cases}$$

Then, we have

$$bu^2 + av^2 = \frac{s^2+t^2}{2},$$

$$uv = \frac{s^2-t^2}{2},$$

$$du\, dv = \frac{1}{2\sqrt{ab}}\, ds\, dt,$$

and Eq. (25) becomes

$$I(\gamma, \mathbf{\Phi}) = \iint \frac{e^{-A/B}}{2\pi} (ab-\gamma^2)^{-9/2} \left[(ab-\gamma^2)\,\gamma - \gamma\,\frac{s^2+t^2}{2} \right.$$

$$\left. + (ab+\gamma^2)\,\frac{s^2-t^2}{4\sqrt{ab}} \right]^2 \frac{1}{2\sqrt{ab}}\, ds\, dt$$

$$= \frac{(ab-\gamma^2)^{-9/2}}{4\pi\sqrt{ab}} \iint e^{-(\sqrt{ab}-\gamma)\,s^2 + (\sqrt{ab}+\gamma)\,t^2/[4(ab-\gamma^2)\sqrt{ab}]}$$

$$\times \left[(ab-\gamma^2)\,\gamma + \frac{(\sqrt{ab}-\gamma)^2\, s^2}{4\sqrt{ab}} - \frac{(\sqrt{ab}+\gamma)^2\, t^2}{4\sqrt{ab}} \right]^2 ds\, dt. \tag{26}$$

Let $p=\sqrt{s^2/[4\sqrt{ab}(\sqrt{ab}+\gamma)]}$ and $q=\sqrt{t^2/[4\sqrt{ab}(\sqrt{ab}-\gamma)]}$. Then Eq. (26) becomes

$$I(\gamma, \mathbf{\Phi}) = \frac{(ab-\gamma^2)^{-9/2}}{4\pi\sqrt{ab}} \iint e^{-p^2-q^2} \left[(ab-\gamma^2)\,\gamma + (\sqrt{ab}+\gamma)(\sqrt{ab}-\gamma)^2\, p^2 \right.$$

$$\left. -(\sqrt{ab}-\gamma)(\sqrt{ab}+\gamma)^2 q^2 \right]^2 \cdot 4\sqrt{ab}\,\sqrt{ab-\gamma^2}\, dp\, dq$$

$$= (ab-\gamma^2)^{-2}\,(ab+\gamma^2).$$

By the definition of the Fisher information, we know

$$I(\rho, \mathbf{\Phi}) = I(\gamma, \mathbf{\Phi}) \left(\frac{d\gamma}{d\rho} \right)^2,$$

where $\gamma=\sqrt{ab}\,\rho$ in this case. Using Eq. (17), we get

$$I(\rho, \mathbf{\Phi}) = (ab - ab\rho^2)^{-2}\,(ab + ab\rho^2)(ab)$$

$$= (1-\rho^2)^{-2}\,(1+\rho^2).$$

ACKNOWLEDGMENTS

The authors are grateful to Xavier de Luna for his comments on an earlier version of the paper. We also thank two anonymous referees for comments which helped improve the paper.

REFERENCES

1. C. Croux and P. J. Rousseeuw, Time-efficient algorithms for two highly robust estimators of scale, *Comput. Statist.* **2** (1992), 411–428.
2. C. Croux, P. J. Rousseeuw, and O. Hössjer, Generalized *S*-estimators, *J. Amer. Statist. Assoc.* **89** (1994), 1271–1281.

3. C. Croux and G. Haesbroeck, Influence function and efficiency of the minimum covariance determinant scatter matrix estimator, *J. Multivariate Anal.* **71** (1999), 161–190.
4. P. L. Davies, Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices, *Ann. of Statist.* **15** (1987), 1269–1292.
5. S. J. Devlin, R. Gnanadesikan, and J. R. Kettenring, Robust estimation and outlier detection with correlation coefficients, *Biometrika* **62** (1975), 531–545.
6. S. J. Devlin, R. Gnanadesikan, and J. R. Kettenring, Robust estimation of dispersion matrices and principal components, *J. Amer. Statist. Assoc.* **76** (1981), 354–362.
7. D. L. Donoho and P. J. Huber, The notion of breakdown point, *in* "A Festschrift for Erich L. Lehmann" (P. J. Bickel, K. A. Doksum, and J. L. Hodges Jr., Eds.), pp. 157–184, 1982.
8. M. G. Genton, Highly robust variogram estimation, *Math. Geol.* **30** (1998), 213–221.
9. M. G. Genton and Y. Ma, Robustness properties of dispersion estimators, *Statist. Probab. Lett.* **44** (1999), 343–350.
10. R. Gnanadesikan and J. R. Kettenring, Robust estimates, residuals, and outlier detection with multiresponse data, *Biometrics* **28** (1972), 81–124.
11. R. Gnanadesikan, "Methods for Statistical Data Analysis of Multivariate Observations," 2nd ed., Wiley, New York, 1997.
12. F. R. Hampel, Robust estimation: A condensed partial survey, *Z. Wahr. Verw. Gebiete* **27** (1973), 87–104.
13. F. R. Hampel, The influence curve and its role in robust estimation, *J. Amer. Statist. Assoc.* **69** (1974), 383–393.
14. F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, "Robust Statistics: The Approach Based on Influence Functions," Wiley, New York, 1986.
15. O. Hössjer, C. Croux, and P. J. Rousseeuw, Asymptotics of generalized S-estimators, *J. Multivariate Anal.* **51** (1994), 148–177.
16. P. J. Huber, Robust covariances, *in* "Statistical Decision Theory and Related Topics 2" (S. S. Gupta and D. S. Moore, Eds.), pp. 165–191, Academic Press, San Diego, 1977.
17. P. J. Huber, "Robust Statistics," Wiley, New York, 1981.
18. P. J. Huber, Finite sample breakdown point of M- and P-estimators, *Ann. of Statist.* **12** (1984), 119–126.
19. G. Li and Z. Chen, Projection-pursuit approach to robust dispersion matrices and principal components: primary theory and Monte Carlo, *J. Amer. Statist. Assoc.* **80** (1985), 759–766.
20. H. Lopuhaä, On the relation between S-estimators and M-estimators of multivariate location and covariance, *Ann. Statist.* **17** (1989), 1662–1683.
21. H. Lopuhaä and P. J. Rousseeuw, Breakdown point of affine equivariant estimators of multivariate location and covariance matrices, *Ann. of Statist.* **19** (1991), 229–248.
22. Y. Ma and M. G. Genton, Highly robust estimation of the autocovariance function, *J. Time Series Anal.* **21** (2000), 663–684.
23. C. L. Mallows, "On Some Topics in Robustness," Bell Labs Tech. Memo, Murray Hill, NJ, 1976.
24. K. V. Mardia, J. T. Kent, and J. M. Bibby, "Multivariate Analysis," Academic Press, San Diego, 1979.
25. R. A. Maronna, Robust M-estimators of multivariate location and scatter, *Ann. of Statist.* **4** (1976), 51–67.
26. R. A. Maronna, W. A. Stahel, and V. J. Yohai, Bias-robust estimation of multivariate scatter based on projections, *J. Multivariate Anal.* **42** (1992), 141–161.
27. R. A. Maronna and V. J. Yohai, The behavior of the Stahel–Donoho robust multivariate estimator, *J. Amer. Statist. Assoc.* **90** (1995), 330–341.

28. P. J. Rousseeuw, Least median of squares regression, *J. Amer. Statist. Assoc.* **79** (1984), 871–880.
29. P. J. Rousseeuw, Multivariate estimation with high breakdown point, *in* "Mathematical Statistics and Applications" (W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, Eds.), pp. 283–297, Reidel, Dordrecht, 1985.
30. P. J. Rousseeuw and A. M. Leroy, "Robust Regression and Outlier Detection," Wiley, New York, 1987.
31. P. J. Rousseeuw and G. Molenberghs, Transformation of non positive semidefinite correlation matrices, *Commun. Statist. Theory Meth.* **22** (1993), 965–984.
32. P. J. Rousseeuw and B. C. van Zomeren, Unmasking multivariate outliers and leverage points, *J. Amer. Statist. Assoc.* **85** (1990), 633–651.
33. P. J. Rousseeuw and C. Croux, Explicit scale estimators with high breakdown point, *in* "L_1 Statistical Analyses and Related Methods" (Y. Dodge, Ed.), pp. 77–92, Elsevier, Amsterdam, 1992.
34. P. J. Rousseeuw and C. Croux, Alternatives to the median absolute deviation, *J. Amer. Statist. Assoc.* **88** (1993), 1273–1283.
35. P. J. Rousseeuw and G. Molenberghs, Transformation of non positive semidefinite correlation matrices, *Commun. Statist. Theory Meth.* **22** (1993), 965–984.
36. P. J. Rousseeuw and K. Van Driessen, A fast algorithm for the minimum covariance determinant estimator, *Technometrics* **41** (1999), 212–223.
37. W. A. Stahel, "Breakdown of Covariance Estimators," Research Report 31, Fachgruppe für Statistik ETH, Zürich, 1981.
38. D. E. Tyler, Finite sample breakdown points of projection based multivariate location and scatter statistics, *Ann. of Statist.* **22** (1994), 1024–1044.
39. V. J. Yohai and R. Zamar, High breakdown-point estimates of regression by means of the minimization of an efficient scale, *J. Amer. Statist. Assoc.* **83** (1988), 406–413.