

# Central limit theorem and influence function for the MCD estimators at general multivariate distributions

ERIC A. CATOR\* and HENDRIK P. LOPUHAÄ\*\*

*Delft University of Technology, DIAM, Mekelweg 4, 2628 CD Delft, The Netherlands.*  
*E-mail: \*e.a.cator@tudelft.nl; \*\*h.p.lopuhaa@tudelft.nl*

We define the minimum covariance determinant functionals for multivariate location and scatter through trimming functions and establish their existence at any multivariate distribution. We provide a precise characterization including a separating ellipsoid property and prove that the functionals are continuous. Moreover, we establish asymptotic normality for both the location and covariance estimator and derive the influence function. These results are obtained in a very general multivariate setting.

*Keywords:* asymptotic normality; influence function; minimum covariance determinant

## 1. Introduction

Consider the minimum covariance determinant (MCD) estimator introduced in [19], that is, for a sample  $X_1, X_2, \dots, X_n$  from a distribution  $P$  on  $\mathbb{R}^k$  and  $0 < \gamma \leq 1$ , consider subsamples  $S \subset \{X_1, \dots, X_n\}$  that contain  $h_n \geq \lceil n\gamma \rceil$  points. Define a corresponding trimmed sample mean and sample covariance matrix by

$$\begin{aligned}\widehat{T}_n(S) &= \frac{1}{h_n} \sum_{X_i \in S} X_i, \\ \widehat{C}_n(S) &= \frac{1}{h_n} \sum_{X_i \in S} (X_i - \widehat{T}_n(S))(X_i - \widehat{T}_n(S))'.\end{aligned}\tag{1.1}$$

Let  $S_n$  be a subsample that minimizes  $\det(\widehat{C}_n(S))$  over all subsamples of size  $h_n \geq \lceil n\gamma \rceil$ , where  $\lceil x \rceil$  denotes the smallest integer greater than or equal to  $x \in \mathbb{R}$ . Then the pair  $(\widehat{T}_n(S_n), \widehat{C}_n(S_n))$  is an MCD estimator. Today, the MCD estimator is one of the most popular robust methods to estimate multivariate location and scatter parameters. These estimators, in particular the covariance estimator, also serve as robust plug-ins in other multivariate statistical techniques, such as principal component analysis [5,21], multivariate linear regression [1,20], discriminant analysis [11], factor analysis [16], canonical correlations [22,25], error-in-variables models [8], invariant coordinate selection [24], among others (see also [12] for a more extensive overview). For this reason, the distributional and the robustness properties of the MCD estimators are essential for conducting inference and perform robust estimation in several statistical models.

The MCD estimators are known to have the same breakdown point as the minimum volume ellipsoid estimators [19], and for a suitable choice of  $\gamma$  they possess the maximal breakdown point possible for affine equivariant estimators (e.g., see [1,15]). However, their asymptotic properties, such as the rate of convergence, limit distribution and influence function, are not fully understood. Within the framework of unimodal elliptically contoured densities, Butler, Davies and Jhun [2] show that the MCD location estimator converges at  $\sqrt{n}$ -rate towards a normal distribution with mean equal to the MCD location functional. The rate of convergence and limit distribution of the covariance estimator still remains an open problem. Croux and Haesbroeck [4] give the expression for the influence function  $\text{IF}(x; C, P)$  of the MCD covariance functional  $C(P)$  at distributions  $P$  with a unimodal elliptically contoured density and use this to compute limiting variances of the MCD covariance estimator. However, existence, continuity and differentiability of the MCD functionals at perturbed distributions is implicitly assumed, but not proven. Moreover, the computation of the limiting variances via the influence function relies on the von Mises expansion, that is,

$$\widehat{C}_n(S_n) - C(P) = \frac{1}{n} \sum_{i=1}^n \text{IF}(X_i; C, P) + o_{\mathbb{P}}(n^{-1/2}), \quad (1.2)$$

which has not been established. The distribution and robustness properties of robust multivariate techniques that make use of the MCD, depend on the distribution and robustness properties of the MCD estimator, in particular those of the MCD covariance estimator. Despite the incomplete asymptotic theory for the MCD, at several places in the literature one prematurely assumes either a  $\sqrt{n}$  rate of convergence or asymptotic normality of the MCD covariance estimator, or uses the influence function of the covariance MCD functional to investigate the robustness of the specific multivariate method and to determine limiting variances based on the heuristic (1.2).

This paper is meant to settle these open problems and extend the asymptotic theory for the MCD estimator in a very general setting that allows a wide range of multivariate distributions. We will define the MCD functional by means of trimming functions which are in a wide class of measurable functions. Minimization of the determinant of our trimmed covariance functional has similarities with minimizing the variation functional corresponding to trimmed  $k$ -means considered in [6,7,9,10]. As opposed to our determinant functional, their variation functional contains the trimming function in a linear way. However, once we have established the existence of our MCD functionals, the characterization of our minimizing trimming functions is very similar to the characterization found in [6,10]. In fact, part of our characterization follows directly from results in [10]. These authors also recognized the advantage of employing a flexible class of trimming functions, which allows a uniform treatment at general probability measures, including empirical measures and perturbed measures needed for our purposes. We believe that obtaining our results for general multivariate distributions is an important contribution of this paper. To justify this claim, we will give several important examples of models where it is essential to study the MCD estimator for a class of distributions that is wider than the elliptically contoured distributions.

We prove existence of the MCD functional for any multivariate distribution  $P$  and provide a separating ellipsoid property for the functional. Furthermore, we prove continuity of the functional, which also yields strong consistency of the MCD estimators. Finally, we derive an asymptotic expansion of the functional, from which we rigorously derive the influence function, and

establish a central limit theorem for both MCD-estimators. We would like to emphasize that all results are obtained under very mild conditions on  $P$  and that essentially all conditions are satisfied for distributions with a density. For distributions with an elliptically contoured density that is unimodal we do not need any extra condition and recover the results in [2] and [4] as a special case (see [3]).

The paper is organized as follows. In Section 2, we define the MCD functional for general underlying distributions, discuss some of its basic properties and provide examples of models where it is essential to study behavior of the MCD estimator for underlying distributions that are beyond elliptically contoured distributions. In Section 3, we prove existence of the MCD functional and establish a separating ellipsoid property. Section 4 deals with continuity of the MCD functionals and consistency of the MCD estimators. Finally, in Section 5 we obtain an asymptotic expansion of the MCD estimators and MCD functional, from which we prove asymptotic normality and determine the influence function. In order to keep things readable, all proofs and technical lemmas have been postponed to an [Appendix](#) at the end of the paper.

## 2. Definition

Let  $P$  be a probability measure on  $\mathbb{R}^k$ . To define an MCD functional at  $P$ , we start by defining a trimmed mean and trimmed covariance functional in the following way. For a measurable function  $\phi : \mathbb{R}^k \rightarrow [0, 1]$ , define

$$\begin{aligned} T_P(\phi) &= \frac{1}{\int \phi \, dP} \int x \phi(x) P(dx), \\ C_P(\phi) &= \frac{1}{\int \phi \, dP} \int (x - T_P(\phi))(x - T_P(\phi))' \phi(x) P(dx). \end{aligned} \tag{2.1}$$

The function  $\phi$  determines the trimming of the mean and covariance matrix. For  $\phi = 1$ , the above functionals are the ordinary mean and covariance matrix corresponding to  $P$ . When  $P = P_n$ , the empirical measure, and  $\phi = \mathbb{1}_S$  for a subsample  $S$ , we recover (1.1). Next, we fix a proportion  $0 < \gamma \leq 1$  and require  $\phi$  to have at least mass  $\gamma$ , that is,

$$\int \phi \, dP \geq \gamma.$$

To ensure that the functionals in (2.1) are well defined, we take  $\phi$  in the class

$$K_P(\gamma) = \left\{ \phi : \mathbb{R}^k \rightarrow [0, 1]: \phi \text{ measurable, } \int \phi \, dP \geq \gamma, \int \|x\|^2 \phi(x) P(dx) < \infty \right\}.$$

If there exists  $\phi_P \in K_P(\gamma)$  which minimizes  $\det(C_P(\phi))$  over all  $\phi \in K_P(\gamma)$ , then the corresponding pair

$$(T_P(\phi_P), C_P(\phi_P))$$

is called an MCD functional at  $P$ . Note that, although for  $\phi \in K_P(\gamma)$  the functionals in (2.1) are well defined, the existence of a minimizing  $\phi$  is not guaranteed. Furthermore, if a minimizing  $\phi$  exists, it need not be unique.

To complete our definitions, note that each trimming function  $\phi$  determines an ellipsoid  $E(T_P(\phi), C_P(\phi), r_P(\phi))$ , where for each  $\mu \in \mathbb{R}^k$ ,  $\Sigma$  symmetric positive definite, and  $\rho > 0$ ,

$$E(\mu, \Sigma, \rho) = \{x \in \mathbb{R}^k: (x - \mu)' \Sigma^{-1} (x - \mu) \leq \rho^2\}, \tag{2.2}$$

and

$$r_P(\phi) = \inf\{s > 0: P(E(T_P(\phi), C_P(\phi), s)) \geq \gamma\}. \tag{2.3}$$

If a minimizing trimming function  $\phi_P$  exists, then  $E(T_P(\phi_P), C_P(\phi_P), r_P(\phi_P))$  is referred to as a “minimizing” ellipsoid.

Note that the functionals in (2.1) are affine equivariant in the following sense. Fix a non-singular  $k \times k$  matrix  $A$  and  $b \in \mathbb{R}^k$  and let  $h(x) = Ax + b$ , for  $x \in \mathbb{R}^k$ . If  $X \sim P$ , then  $AX + b \sim Q = P \circ h^{-1}$ . It is straightforward to see that  $\phi \in K_Q(\gamma)$  if and only if  $\phi \circ h \in K_P(\gamma)$ , which yields

$$T_Q(\phi) = AT_P(\phi \circ h) + b \quad \text{and} \quad C_Q(\phi) = AC_P(\phi \circ h)A',$$

as well as  $r_Q(\phi) = r_P(\phi \circ h)$ . Furthermore,  $\phi_Q$  minimizes  $\det(C_Q(\phi))$  over  $K_Q(\gamma)$  if and only if  $\phi_P = \phi_Q \circ h$  minimizes  $\det(C_P(\phi))$  over  $K_P(\gamma)$ . This means that if an MCD functional exists, it is affine equivariant, that is,  $T_Q(\phi_Q) = AT_P(\phi_P) + b$  and  $C_Q(\phi_Q) = AC_P(\phi_P)A'$ .

Butler *et al.* [2] define the MCD functional by minimizing over all indicator functions  $\mathbb{1}_B$  of measurable bounded Borel sets  $B \subset \mathbb{R}^k$  with  $P(B) = \gamma$ . These indicator functions form a subclass of  $K_P(\gamma)$ , that is sufficiently rich when one considers unimodal elliptically contoured densities. However, at perturbed distributions  $P_{\varepsilon,x} = (1 - \varepsilon)P + \varepsilon\delta_x$ , where  $\delta_x$  denotes the Dirac measure at  $x \in \mathbb{R}^k$ , their MCD functional may not exist. Croux and Haesbroeck [4] solve this problem by minimizing over all functions  $\mathbb{1}_B + \delta\mathbb{1}_{\{x\}}$ , with  $x \notin B$  and  $P(B) + \delta P(\{x\}) = \gamma$ . These functions form a subclass of  $K_P(\gamma)$ , that is sufficiently rich when one considers single-point perturbations of unimodal elliptically contoured densities, but the class  $K_P(\gamma)$  allows for functions other than  $\mathbb{1}_B + \delta\mathbb{1}_{\{x\}}$  for which the determinant of the covariance functional is strictly smaller. Moreover, minimization over the more flexible class  $K_P(\gamma)$  allows a uniform treatment of the functionals in (2.1) at general probability measures, including measures with atoms. Important examples are the empirical measure  $P_n$  corresponding to a sample from  $P$ , in which case the functionals relate to the MCD estimators, and perturbed measures  $P_{\varepsilon,x}$ , for which the functionals need to be investigated in order to determine the influence function. It should be noted that our Theorem 3.2 does show that a minimizer in the Croux–Haesbroeck sense does exist for all distributions  $P$ , but this is not at all obvious before hand.

Definition (2.1) might suggest that minimization of  $\det(C_P(\phi))$  is hindered by the fact that the denominator depends on  $\phi$ . However, the following property shows that if a minimum exists, it can always be achieved with a denominator in (2.1) equal to  $\gamma$ . Its proof is straightforward from definition (2.1).

**Lemma 2.1.** *For any  $0 < \lambda \leq 1$  and  $\phi \in K_P(\gamma)$ , such that  $\lambda\phi \in K_P(\gamma)$ , we have*

$$T_P(\lambda\phi) = T_P(\phi), \quad C_P(\lambda\phi) = C_P(\phi) \quad \text{and} \quad r_P(\lambda\phi) = r_P(\phi).$$

Since we can always construct a minimizing  $\phi$  in such a way that  $\int \phi dP = \gamma$ , it is tempting to replace the term  $\int \phi dP$  in (2.1) by  $\gamma$ . However, we will not do so, in order to keep enough flexibility for the functionals at probability measures  $P$  and trimming functions of the type  $\phi = \mathbb{1}_B$ , for measurable  $B \subset \mathbb{R}^k$  with  $P(B) > \gamma$ . An important example is the situation where  $P$  is the empirical measure.

## 2.1. Examples of non-elliptical models where the MCD is relevant

We will prove (see Theorem 4.2) that the MCD estimators converge (under mild conditions) to the MCD functionals at  $P$ . These functionals might not be related in any way to the expectation of  $P$  or the covariance matrix (in fact, our conditions allow for  $P$  whose expectation does not even exist) and one might question the relevance of the MCD-functional for general  $P$ .

First of all, we believe that it is not unreasonable to consider the MCD as a measure of location and scale on its own right, just like the median and the MAD. Our results then show how the natural estimator of this functional behaves. Especially in cases where the distribution has a heavy tail, the MCD functional might provide more useful quantitative information than the mean and covariance structure, for example for confidence sets of future realizations of  $P$ . In addition, we will give some explicit examples in which it is very relevant to extend the behavior of the MCD functional to general distributions.

### *Independent component analysis*

Consider a random vector  $Z \in \mathbb{R}^k$  with a density  $f$  that has the property that for each coordinate, the mapping  $y \mapsto f(z_1, \dots, y, \dots, z_k)$  is a univariate, symmetric unimodal function of  $y$  for each fixed  $z_1, \dots, z_k$ , and that  $f$  is invariant under coordinate-permutations. For example, this would be the case if all the marginals of  $f$  are independent and identically distributed according to a univariate symmetric and unimodal distribution. It is clear that if the MCD functional for  $f$  is unique, then from the symmetries it follows that the location functional is zero, and the covariance functional is a constant times the identity matrix. If we observe an affinely transformed sample from  $f$ , that is,  $X_1, \dots, X_n$  where  $X_i = AZ_i + \mu$  and  $Z_i$  has density  $f$ , then the MCD estimator would be a robust estimator of  $\mu$  and  $AA'$ . Note that the density of  $X_1, \dots, X_n$  is in general not elliptically contoured. The uniqueness of the MCD functional for an  $f$  of this kind would be similar to the results in [23] for  $S$ - and  $M$ -functionals. However, proving this is beyond the scope of this paper, and might in fact be quite hard, given the depths of the results in [23]. The above example has close connections with independent component analysis (ICA), a highly popular method within many applied areas, which routinely encounter multivariate data. For a good overview see [13]. The most common ICA model considers  $X$  arising as a convolution of  $k$  independent components, that is,  $X = AZ$ , where  $A$  is nonsingular, and the components of  $Z$  are independent. The main objective of ICA is to recover the mixing matrix  $A$  so that one can ‘unmix’  $X$  to obtain independent components.

### *Invariant coordinate selection*

Invariant coordinate selection (ICS), recently proposed in [24], compares two covariance estimators through so-called ICS roots to reveal departures from an elliptically contoured distribution.

The authors suggest one of the covariance estimators to be a class III scatter matrix, of which the MCD estimator is an example. Determining whether ICS roots differ significantly, or what power such a test would have, remains an open problem. This is precisely where the distribution of the MCD estimator at elliptical and non-elliptical distributions is essential.

*Contaminated distributions*

An important property for any robust estimator for location and scatter is that it is able to recover to some extent the mean and covariance matrix of the underlying distribution when this distribution is contaminated. For instance, when the contamination has small total mass or is very far away from the center of the underlying distribution, it should not affect the corresponding functional too much. For our MCD functional, this is precisely the content of the following theorem, whose proof can be found in the [Appendix](#). These results rely heavily on the methods used in this paper for general distributions, even if the uncontaminated distribution  $P$  is elliptically contoured.

**Theorem 2.1.** *Let  $P$  and  $Q$  be two probability measures on  $\mathbb{R}^k$  and define for  $x, r \in \mathbb{R}^k$  the translation  $\tau_r(x) = x + r$ . Consider, for  $\varepsilon < 1/2$ , the mixture*

$$P_{r,\varepsilon} = (1 - \varepsilon)P + \varepsilon Q \circ \tau_r^{-1}.$$

*Denote by  $\text{MCD}_\gamma(\cdot)$  the MCD functional of level  $\gamma$ . Choose  $\gamma$  such that  $\varepsilon < \gamma < 1 - \varepsilon$ , and suppose that*

$$P(H) < \frac{\gamma - \varepsilon}{1 - \varepsilon}$$

*for all hyperplanes  $H \subset \mathbb{R}^k$ .*

(i) *Then*

$$\lim_{\varepsilon \downarrow 0} \text{MCD}_\gamma(P_{r,\varepsilon}) = \text{MCD}_\gamma(P) \quad \text{and} \quad \lim_{\|r\| \rightarrow \infty} \text{MCD}_\gamma(P_{r,\varepsilon}) = \text{MCD}_{\gamma/(1-\varepsilon)}(P),$$

*where the first limit should be interpreted as: every limit point is an MCD functional at  $P$  of level  $\gamma$ , and the second limit similarly.*

(ii) *Furthermore, if in addition  $Q$  has a bounded support, then for all  $\gamma \in (\varepsilon, 1 - \varepsilon)$ , there exists  $r_0 \geq 0$  such that*

$$\text{MCD}_\gamma(P_{r,\varepsilon}) = \text{MCD}_{\gamma/(1-\varepsilon)}(P)$$

*for all  $r \in \mathbb{R}^k$  with  $\|r\| \geq r_0$ .*

As an illustration of Theorem 2.1, consider an elliptically contoured distribution  $P$  with parameter  $(\mu, \Sigma)$ . The second limit in (i) shows that if the contamination is far from zero, the MCD functionals of the contaminated distribution are close to  $\mu$  and a multiple of  $\Sigma$ . Part (ii) shows that for specific types of contamination, for example, single point contaminations, the MCD functionals at the contaminated distribution recovers these values exactly. The proof of Theorem 2.1 in principle provides a constructive (but elaborate) way to find  $r_0$  in terms of  $\varepsilon, \gamma, P$  and the support of  $Q$ .

### 3. Existence and characterization of an MCD-functional

By definition, the matrix  $C_P(\phi)$  is symmetric non-negative definite. Without imposing any assumptions on  $P$ , one cannot expect  $C_P(\phi)$  to be positive definite. We will assume that  $P$  satisfies:

$$P(H) < \gamma \quad \text{for every hyperplane } H \subset \mathbb{R}^k. \tag{3.1}$$

This is a reasonable assumption, since if  $P$  does not have this property, then there exists a  $\phi \in K_P(\gamma)$  with  $\det(C_P(\phi)) = 0$  (e.g.,  $\phi = 1_H$  with  $P(H) \geq \gamma$ ). This would prove the existence of a minimizing  $\phi$ , but obviously the corresponding MCD-functional is not very useful.

We first establish the existence of a minimizing  $\phi \in K_P(\gamma)$ . For later purposes, we do not only prove existence at  $P$ , but also at probability measures  $P_t$ , for which the sequence  $(P_t)$  converges weakly to  $P$ , as  $t \rightarrow \infty$ . For ease of notation, we continue to write  $P_0$  instead of  $P$  and for  $t \geq 0$  write

$$T_t = T_{P_t}, \quad C_t = C_{P_t}, \quad r_t = r_{P_t} \quad \text{and} \quad K_t(\gamma) = K_{P_t}(\gamma). \tag{3.2}$$

The next proposition shows that eventually the smallest eigenvalue of the covariance functional is bounded away from zero uniformly in  $\phi$  and  $t$ .

**Proposition 3.1.** *Suppose  $P_0$  satisfies (3.1) and let  $P_t \rightarrow P_0$  weakly. Then there exists  $\lambda_0 > 0$  and  $t_0 \geq 1$  such that for  $t = 0$ , all  $t \geq t_0$ , all  $\phi \in K_t(\gamma)$ , and all  $a \in \mathcal{S}^k$  (the sphere in  $\mathbb{R}^k$ ), we have*

$$\int (a'(x - T_t(\phi)))^2 \phi(x) P_t(dx) \geq \lambda_0.$$

*In particular, this means that the smallest eigenvalue of  $C_t(\phi)$  is at least  $\lambda_0$ .*

An immediate corollary is that if  $\det(C_t(\phi))$  is uniformly bounded, there exists a compact set that contains the location and covariance functionals for sufficiently large  $t$  (see Lemma A.1 in the Appendix). This will become very useful in establishing continuity of the functionals in Section 4. For the moment, we use this result to show that for minimizing  $\det(C_t(\phi))$ , one may restrict to functions  $\phi$  with bounded support.

For  $R > 0$ , define the ball  $B_R = \{x \in \mathbb{R}^k: \|x\| \leq R\}$  and for  $t \geq 0$  define the class

$$K_t^R(\gamma) = \{\phi \in K_t(\gamma): \{\phi \neq 0\} \subset B_R\}.$$

Clearly,  $K_t^R(\gamma) \subset K_t(\gamma)$ . The next proposition shows that for any  $\phi \in K_t(\gamma)$  we can always find a  $\psi$  with bounded support in  $K_t^R(\gamma)$  that has a smaller determinant.

**Proposition 3.2.** *Suppose that  $P_0$  satisfies (3.1) and  $P_t \rightarrow P_0$  weakly. There exists  $R > 0$  and  $t_0 \geq 1$  such that for  $t = 0$ , all  $t \geq t_0$  and all  $\phi \in K_t(\gamma)$ , there exists  $\psi \in K_t^R(\gamma)$  with*

$$\det(C_t(\psi)) \leq \det(C_t(\phi)).$$

Proposition 3.2 illustrates the general heuristic that if  $\phi$  has  $P_t$ -mass far away from  $T_t(\phi)$ , then moving this mass closer towards  $T_t(\phi)$  will decrease the determinant of the covariance matrix. Together with Proposition 3.1 this establishes the existence of at least one MCD functional for the probability measure  $P_0$ . Moreover, if  $P_t \rightarrow P_0$  weakly, then at least one MCD functional exists for  $P_t$  for sufficiently large  $t$ .

**Theorem 3.1.** *Suppose  $P_0$  satisfies (3.1) and let  $P_t \rightarrow P_0$  weakly. Then there exists  $R > 0$  and  $t_0 \geq 1$ , such that for  $t = 0$  and  $t \geq t_0$ , there exists  $\phi_t \in K_t^R(\gamma)$ , which minimizes  $\det(C_t(\phi))$  over  $K_t(\gamma)$ .*

In the remainder of this section, we provide a characterization of a minimizing  $\phi$ , which includes a separating ellipsoid property for the MCD functional. A similar result has been obtained in [2] for the empirical measure and in [4] for single-point perturbations of distributions with a unimodal elliptically contoured density. Our characterization is for general distributions and is very similar to the characterizations obtained for trimmed means in [10] and [6]. In fact, the first part of our result follows directly from [10]. We will denote the interior of a set  $E$  by  $E^\circ$ , and the (topological) boundary by  $\partial E$ .

**Theorem 3.2.** *Let  $\phi \in K_P(\gamma)$  be such that  $(T_P(\phi), C_P(\phi))$  is an MCD functional at  $P$  and let  $E_P(\phi) = E(T_P(\phi), C_P(\phi), r_P(\phi))$  be the corresponding minimizing ellipsoid. Then*

$$\int \phi \, dP = \gamma \quad \text{and} \quad \mathbb{1}_{E_P(\phi)^\circ} \leq \phi \leq \mathbb{1}_{E_P(\phi)}, \quad P\text{-a.e.}$$

*Furthermore, either  $\phi = 0$  on  $\partial E_P(\phi)$  ( $P$ -a.e.), or  $\phi = 1$  on  $\partial E_P(\phi)$  ( $P$ -a.e.), or there exists  $x \in \partial E_P(\phi)$  such that  $P(\partial E_P(\phi)) = P(\{x\})$ .*

The theorem shows that a minimizing trimming function  $\phi$  is almost the indicator function of an ellipsoid with center  $T_P(\phi)$  and covariance structure  $C_P(\phi)$ . When  $P$  has no mass on the boundary of the ellipsoid, then  $\phi$  is equal to the indicator function of this ellipsoid. If the interior of the ellipsoid  $E(T_P(\phi), C_P(\phi), r_P(\phi))$  has mass strictly smaller than  $\gamma$ , then either  $\phi$  equals 1 on the entire boundary of the ellipsoid, in which case the (closed) ellipsoid has  $P$  mass exactly  $\gamma$ , or  $P$  only has mass in exactly one point on the boundary, and  $\phi$  adapts its value in that point such that it has total  $P$  mass  $\gamma$ .

Theorem 3.2 holds for any probability measure, in particular for the empirical measure  $P_n$  and for perturbed measures  $P_{\varepsilon,x}$ , in which case we obtain results analogous to Theorem 2 in [2] and Proposition 1 in [4], respectively. However, note that the characterization in Theorem 3.2 is more precise and is such that the center and covariance structure of the separating ellipsoid are exactly the MCD functionals themselves.

### 4. Continuity of the MCD functional

Consider a sequence  $(T_t(\phi_t), C_t(\phi_t))$  of MCD functionals corresponding to a sequence of probability measures  $P_t \rightarrow P_0$  weakly. We investigate under what conditions  $(T_t(\phi_t), C_t(\phi_t))$  con-

verges and whether each limit point will be an MCD functional corresponding to  $P_0$ . Our approach requires  $\int \phi \, dP_t \rightarrow \int \phi \, dP_0$  uniformly in minimizing  $\phi$ . The following condition on  $P_0$  suffices:

$$\sup_{E \in \mathcal{E}} |P_t(E) - P_0(E)| \rightarrow 0, \quad \text{as } t \rightarrow \infty, \tag{4.1}$$

where  $\mathcal{E}$  denotes the class of all ellipsoids. This may seem restrictive, but it is either automatically fulfilled for sequences that are important for our purposes or a mild condition on  $P_0$  suffices. For instance, when  $P_t$  is a sequence of empirical measures, then (4.1) holds automatically by standard results from empirical process theory (e.g., see Theorem II.14 in [17]) because the ellipsoids form a class with polynomial discrimination or a Vapnik–Cervonenkis class. Condition (4.1) also holds for sequences of perturbed measures  $P_{\varepsilon,x}$ , as  $\varepsilon \downarrow 0$ . In general, if  $P_0(\partial C) = 0$  for all measurable convex  $C \subset \mathbb{R}^k$ , then condition (4.1) holds for any sequence  $P_t \rightarrow P_0$  weakly (see Theorem 4.2 in [18]). Note that this is always trivially true if  $P_0$  has a density.

For later purposes, we prove continuity not only for MCD functional minimizing functions  $\phi_t$ , but for any sequence of functions  $\psi_t$  with uniformly bounded support that satisfy the same characteristics as  $\phi_t$  and for which  $\det(C_t(\psi_t))$  is close to  $\det(C_t(\phi_t))$ .

**Theorem 4.1.** *Suppose  $P_0$  satisfies (3.1). Let  $P_t \rightarrow P_0$  weakly and suppose that (4.1) holds. For  $t \geq 1$ , let  $\psi_t \in K_t(\gamma)$  such that  $\psi_t \leq \mathbb{1}_{E_t}$ , where  $E_t = E(T_t(\psi_t), C_t(\psi_t), r_t(\psi_t))$ , and suppose there exist  $R > 0$  such that  $\{\psi_t \neq 0\} \subset B_R$ , for  $t$  sufficiently large. Suppose that*

$$\det(C_t(\psi_t)) - \det(C_t(\phi_t)) \rightarrow 0, \quad \text{as } t \rightarrow \infty,$$

where  $\phi_t$  minimizes  $\det(C_t(\phi))$  over  $K_t(\gamma)$ . Then

- (i) *there exist a convergent subsequence  $(T_{t_m}(\psi_{t_m}), C_{t_m}(\psi_{t_m}))$ ;*
- (ii) *the limit point of any convergent subsequence is an MCD functional at  $P_0$ .*

An immediate corollary is that in case the MCD functional at  $P_0$  is uniquely defined, all possible MCD functionals at  $P_t$  are consistent. For later purposes, we also need that  $r_t(\phi_t)$  converges. This may not be the case if  $P_0$  has no mass directly outside the boundary of its minimizing ellipsoid. For this reason, we also require that

$$P_0(E(T_0(\phi_0), C_0(\phi_0), r_0(\phi_0) + \varepsilon)) > \gamma \quad \text{for all } \varepsilon > 0. \tag{4.2}$$

Note that this condition is trivially true if  $P_0$  has a positive density in a neighborhood of the boundary of  $E(T_0(\phi_0), C_0(\phi_0), r_0(\phi_0))$ .

**Corollary 4.1.** *Suppose  $P_0$  satisfies (3.1) and that the MCD functional  $(T_0(\phi_0), C_0(\phi_0))$  is uniquely defined at  $P_0$ . Let  $P_t \rightarrow P_0$  weakly and suppose that (4.1) holds. For  $t \geq 1$ , let  $\psi_t \in K_t(\gamma)$  such that  $\psi_t \leq \mathbb{1}_{E_t}$ , where  $E_t = E(T_t(\psi_t), C_t(\psi_t), r_t(\psi_t))$ , and suppose there exist  $R > 0$  such that  $\{\psi_t \neq 0\} \subset B_R$ , for  $t$  sufficiently large. Suppose that*

$$\det(C_t(\psi_t)) - \det(C_t(\phi_t)) \rightarrow 0,$$

where  $\phi_t$  minimizes  $\det(C_t(\phi))$  over  $K_t(\gamma)$ . Then,

$$(i) (T_t(\psi_t), C_t(\psi_t)) \rightarrow (T_0(\phi_0), C_0(\phi_0)).$$

If in addition  $P_0$  satisfies (4.2), then

$$(ii) r_t(\psi_t) \rightarrow r_0(\phi_0).$$

Uniqueness of the MCD functional has been proven in [2] for distributions  $P_0$  that have a unimodal elliptically contoured density. For general distributions, one cannot expect such a general result. For instance, for certain bimodal distributions or for a spherically symmetric uniform distribution which is positive on a large enough disc, the MCD functional is no longer unique.

### 4.1. Consistency of the MCD estimators

For  $n = 1, 2, \dots$ , let  $P_n$  denote the empirical measure corresponding to a sample from  $P_0$ . From definitions (1.1) and (2.1), it is easy to see that the MCD estimators can be written in terms of the MCD functional as follows

$$\begin{aligned} \widehat{T}_n(S_n) &= T_n(\mathbb{1}_{S_n}), \\ \widehat{C}_n(S_n) &= C_n(\mathbb{1}_{S_n}), \end{aligned} \tag{4.3}$$

where we use the notation introduced in (3.2). Moreover, define  $\widehat{r}_n(S_n) = r_n(\mathbb{1}_{S_n})$ . We should emphasize that  $\widehat{T}_n(S_n)$  and  $\widehat{C}_n(S_n)$  may differ from the actual MCD functionals  $T_n(\phi_n)$  and  $C_n(\phi_n)$ , where  $\phi_n$  minimizes the MCD functional for  $P_n$ . Obviously, if these differences tend to zero, then consistency of the MCD estimators would follow immediately from Theorem 4.1, but unfortunately we have not been able to find an easy argument for this. However, we can show that the determinants of the covariance matrices are close with probability one, which suffices for our purposes.

**Proposition 4.1.** *Suppose  $P_0$  satisfies (3.1). Then for each MCD estimator minimizing subsample  $S_n$  and each MCD functional minimizing function  $\phi_n$ , we have*

$$\det(\widehat{C}_n(S_n)) - \det(C_n(\phi_n)) = O(n^{-1})$$

with probability one.

This does not necessarily mean that  $\widehat{T}_n(S_n) - T_n(\phi_n)$  and  $\widehat{C}_n(S_n) - C_n(\phi_n)$  are also of the order  $O(n^{-1})$ . But in view of Corollary 4.1, it suffices to establish a separating ellipsoid property and uniform bounded support for the minimizing subsample. The latter result can be found in the Appendix, whereas the separating ellipsoid property is stated in the next proposition.

**Proposition 4.2.** *Let  $S_n$  be a minimizing subsample for the MCD estimator and define corresponding ellipsoid  $\widehat{E}_n = E(\widehat{T}_n(S_n), \widehat{C}_n(S_n), \widehat{r}_n(S_n))$ . Then  $S_n$  has exactly  $\lceil n\gamma \rceil$  points,  $S_n \subset \widehat{E}_n$  and  $\widehat{E}_n$  only contains points of  $S_n$ .*

This separating ellipsoid property is somewhat different from the one in Theorem 3.2 (for the empirical measure) and from the one in [2]. The ellipsoid  $\widehat{E}_n$  has the MCD estimators as center and covariance structure instead of the trimmed sample mean and covariance corresponding to the minimizing subsample excluding a point that is most outlying (see [2]). The advantage of the characterization given in Proposition 4.2 is that integrating over  $S_n$  or  $\widehat{E}_n$  with respect to  $P_n$  is the same, which will become very useful later on. We now have the following theorem.

**Theorem 4.2.** *Suppose  $P_0$  satisfies (3.1) and that the MCD functional  $(T_0(\phi_0), C_0(\phi_0))$  is uniquely defined at  $P_0$ . For  $n \geq 1$ , let  $S_n$  be a minimizing subsample for the MCD estimator. Then*

$$(i) \quad (\widehat{T}_n(S_n), \widehat{C}_n(S_n)) \rightarrow (T_0(\phi_0), C_0(\phi_0)), \text{ with probability one.}$$

If, in addition  $P_0$  satisfies (4.2), then

$$(ii) \quad \widehat{r}_n(S_n) \rightarrow r_0(\phi_0), \text{ with probability one.}$$

As a special case, where  $P_0$  has a unimodal elliptically contoured density, we recover Theorem 3 in [2]. With Theorems 4.1 and 4.2, it turns out that the difference between the MCD estimator  $(\widehat{T}_n(S_n), \widehat{C}_n(S_n))$  and the MCD functional  $(T_n(\phi_n), C_n(\psi_n))$  indeed tends to zero with probability one. However, we were not able to find an easier, direct argument.

### 5. Asymptotic normality and influence function

For  $n = 1, 2, \dots$ , let  $S_n$  be a minimizing subsample for the MCD estimator and for ease of notation, write

$$\widehat{\mu}_n = \widehat{T}_n(S_n), \quad \widehat{\Sigma}_n = \widehat{C}_n(S_n) = \widehat{\Gamma}_n^2, \quad \widehat{\rho}_n = \widehat{r}_n(S_n) \quad \text{and} \quad \widehat{E}_n = E(\widehat{\mu}_n, \widehat{\Sigma}_n, \widehat{\rho}_n),$$

and define  $\widehat{\theta}_n = (\widehat{\mu}_n, \widehat{\Gamma}_n, \widehat{\rho}_n)$  in  $\mathbb{R}^k \times \text{PDS}(k) \times \mathbb{R}$ , where  $\text{PDS}(k)$  denotes the class of all positive definite symmetric matrices of order  $k$ . Note that  $\widehat{\Gamma}_n$  is uniquely defined in  $\text{PDS}(k)$ . Similarly, let  $P_n$  denote the empirical measure corresponding to a sample from  $P_0$ , and for  $n = 0, 1, 2, \dots$ , let  $\phi_n$  be a minimizing trimming function for the MCD functional and write

$$\mu_n = T_n(\phi_n), \quad \Sigma_n = C_n(\phi_n) = \Gamma_n^2, \quad \rho_n = r_n(\phi_n) \quad \text{and} \quad E_n = E(\mu_n, \Sigma_n, \rho_n), \quad (5.1)$$

where  $T_n, C_n$  and  $r_n$  are defined in (3.2), and write  $\theta_n = (\mu_n, \Gamma_n, \rho_n)$ . According to Corollary 4.1 and Theorem 4.2, under very mild conditions on  $P_0$ , we have  $\widehat{\theta}_n \rightarrow \theta_0$  and  $\theta_n \rightarrow \theta_0$  with probability one, where  $\theta_0 = (\mu_0, \Gamma_0, \rho_0)$  corresponds to  $P_0$  as defined in (5.1). The limit distribution of  $\widehat{\theta}_n - \theta_0$  and  $\theta_n - \theta_0$  are equal and can be obtained by the same argument. We briefly sketch the main steps for the MCD estimator.

Consider the estimator matrix equation in (1.1),

$$\widehat{\Sigma}_n = \frac{1}{P_n(S_n)} \int_{S_n} (x - \widehat{\mu}_n)(x - \widehat{\mu}_n)' P_n(dx).$$

After multiplying from the left and the right by  $\widehat{\Gamma}_n^{-1}$ , rearranging terms and replacing  $S_n$  by  $\widehat{E}_n$  (which leaves the integral unchanged according to Proposition 4.2), we obtain a covariance valued  $M$ -estimator type score equation:

$$0 = \int_{\widehat{E}_n} (\widehat{\Gamma}_n^{-1}(x - \widehat{\mu}_n)(x - \widehat{\mu}_n)' \widehat{\Gamma}_n^{-1} - I_k) P_n(dx).$$

Similarly, one can obtain a vector valued  $M$ -estimator type score equation from the location equation in (1.1) and the equality  $P_n(\widehat{E}_n) = \lceil n\gamma \rceil / n = \gamma + O(n^{-1})$  can be put into a real valued score equation. Putting everything together, we conclude that  $\widehat{\theta}_n$  satisfies

$$0 = \int \Psi(y, \widehat{\theta}_n) P_n(dy) + O(n^{-1}), \tag{5.2}$$

where  $\Psi = (\Psi_1, \Psi_2, \Psi_3)$ , defined as

$$\begin{aligned} \Psi_1(y, \theta) &= \mathbb{1}_{\{\|G^{-1}(y-m)\| \leq r\}} G^{-1}(y-m), \\ \Psi_2(y, \theta) &= \mathbb{1}_{\{\|G^{-1}(y-m)\| \leq r\}} (G^{-1}(y-m)(y-m)' G^{-1} - I_k), \\ \Psi_3(y, \theta) &= \mathbb{1}_{\{\|G^{-1}(y-m)\| \leq r\}} - \gamma, \end{aligned} \tag{5.3}$$

where  $\theta = (m, G, r)$ , with  $y, t \in \mathbb{R}^k$ ,  $r > 0$ , and  $G \in \text{PDS}(k)$ . Rewrite equation (5.2) as

$$\begin{aligned} 0 &= \Lambda(\widehat{\theta}_n) + \int \Psi(y, \theta_0)(P_n - P_0)(dy) \\ &\quad + \int (\Psi(y, \widehat{\theta}_n) - \Psi(y, \theta_0))(P_n - P_0)(dy) + O(n^{-1}), \end{aligned} \tag{5.4}$$

where

$$\Lambda(\theta) = \int \Psi(y, \theta) P_0(dy). \tag{5.5}$$

In order to determine the limiting distribution of  $\widehat{\theta}_n$ , we proceed as follows. The first term on the right-hand side of (5.4) can be approximated by a first order Taylor expansion that is linear in  $\widehat{\theta}_n - \theta_0$  and the second term can be treated by the central limit theorem. Most of the difficulty is contained in the third term, which must be shown to be of the order  $o_{\mathbb{P}}(n^{-1/2})$ . We apply empirical process theory, for which we need  $\int (\Psi(y, \widehat{\theta}_n) - \Psi(y, \theta_0)) P_0(dy) \rightarrow 0$ . For this, it suffices to impose

$$P_0(\partial E_0) = 0. \tag{5.6}$$

For the MCD functional  $\theta_n$  the argument is the same, apart from the fact that replacing  $\phi_n$  by  $\mathbb{1}_{E_n}$  requires an additional condition on  $P_0$ , that is,

$$P_0 \text{ has no atoms.} \tag{5.7}$$

Note that (5.6) and (5.7) are trivially true if  $P_0$  has a density. By representing elements of  $\mathbb{R}^k \times \text{PDS}(k) \times \mathbb{R}$  as vectors, we then have the following central limit theorem for the MCD estimators and the MCD functional at  $P_n$ .

**Theorem 5.1.** *Let  $P_0$  satisfy (3.1), (4.2) and (5.6). Suppose that  $(\mu_0, \Sigma_0)$  is uniquely defined at  $P_0$ . If  $\Lambda$ , as defined in (5.5), has a non-singular derivative at  $\theta_0$ , then*

$$\widehat{\theta}_n - \theta_0 = -\Lambda'(\theta_0)^{-1} \frac{1}{n} \sum_{i=1}^n (\Psi(X_i, \theta_0) - \mathbb{E}\Psi(X_i, \theta_0)) + o_{\mathbb{P}}(n^{-1/2}),$$

where  $\Psi$  is defined in (5.3). If in addition  $P_0$  satisfies (5.7), then

$$\theta_n - \theta_0 = -\Lambda'(\theta_0)^{-1} \frac{1}{n} \sum_{i=1}^n (\Psi(X_i, \theta_0) - \mathbb{E}\Psi(X_i, \theta_0)) + o_{\mathbb{P}}(n^{-1/2}).$$

In particular, this means that  $\sqrt{n}(\widehat{\theta}_n - \theta_0)$  and  $\sqrt{n}(\theta_n - \theta_0)$  are asymptotically normal with mean zero and covariance matrix

$$\Lambda'(\theta_0)^{-1} M \Lambda'(\theta_0)^{-1},$$

where  $M$  is the covariance matrix of  $\Psi(X_1, \theta_0)$ .

Now Theorem 5.1 has been established, it turns out that the MCD estimator and MCD functional (at  $P_n$ ) are asymptotically equivalent, that is,  $\widehat{\theta}_n - \theta_n = o_{\mathbb{P}}(n^{-1/2})$ . Although this seems natural, we have not been able to find an easier, direct argument for this, in which case we could have avoided establishing parallel results, such as the ones in Section 4.1. An immediate consequence of Theorem 5.1 is asymptotic normality of the MCD location estimator  $\sqrt{n}(\widehat{\mu}_n - \mu_0)$ . Furthermore, since

$$\widehat{\Sigma}_n - \Sigma_0 = (\widehat{\Gamma}_n + \Gamma_0)(\widehat{\Gamma}_n - \Gamma_0) = 2\Gamma_0(\widehat{\Gamma}_n - \Gamma_0) + o_{\mathbb{P}}(1),$$

Theorem 5.1 also yields asymptotic normality of the MCD covariance estimator  $\sqrt{n}(\widehat{\Sigma}_n - \Sigma_0)$  and of  $\sqrt{n}(\widehat{\rho}_n - \rho_0)$ . In [3], a precise expression is obtained for  $\Lambda'(\theta_0)$  for  $P_0$  with a density  $f$  and nonsingularity of  $\Lambda'(\theta_0)$  is proven if  $f$  has enough symmetry. This includes distributions with an elliptically contoured density, so that as a special case of Theorem 5.1, when  $P_0$  has a unimodal elliptically contoured density, one may recover Theorem 4 in [2] for the location MCD estimator.

To determine the influence function, let  $\phi_{\varepsilon,x}$  be the minimizing  $\phi$ -function for  $P_{\varepsilon,x}$  and let

$$\mu_{\varepsilon,x} = T_{P_{\varepsilon,x}}(\phi_{\varepsilon,x}), \quad \Sigma_{\varepsilon,x} = C_{P_{\varepsilon,x}}(\phi_{\varepsilon,x}) = \Gamma_{\varepsilon,x}^2, \quad \rho_{\varepsilon,x} = r_{P_{\varepsilon,x}}(\phi_{\varepsilon,x})$$

and  $E_{\varepsilon,x} = E(\mu_{\varepsilon,x}, \Sigma_{\varepsilon,x}, \rho_{\varepsilon,x})$  be an MCD functional at  $P_{\varepsilon,x}$  with corresponding minimizing ellipsoid. To determine the influence function, we follow the same kind of argument to obtain  $M$ -type score equations, by rewriting equations (2.1) at  $P_{\varepsilon,x}$  and replacing  $\phi_{\varepsilon,x}$  by  $\mathbb{1}_{E_{\varepsilon,x}}$ . Note

however, that from the characterization given in Theorem 3.2,

$$\int (\mathbb{1}_{E_{\varepsilon,x}} - \phi_{\varepsilon,x}) dP_0 = \begin{cases} P_0(\partial E_{\varepsilon,x}), & \text{if } \phi_{\varepsilon,x} = 0 \text{ on } \partial E_{\varepsilon,x}, \\ 0, & \text{if } \phi_{\varepsilon,x} = 1 \text{ on } \partial E_{\varepsilon,x}, \\ P_0(\{z\}), & \text{otherwise, for some } z \in \partial E_{\varepsilon,x}. \end{cases}$$

This means that in order to replace integrals over  $\phi_{\varepsilon,x}$  by integrals over  $E_{\varepsilon,x}$ , we need a stronger condition on  $P_0$ , that is,

$$P_0(\partial E) = 0 \quad \text{for any ellipsoid } E. \tag{5.8}$$

Now, denote  $\theta_{\varepsilon,x} = (\mu_{\varepsilon,x}, \Gamma_{\varepsilon,x}, \rho_{\varepsilon,x})$  then similar to (5.2), we obtain

$$0 = (1 - \varepsilon)\Lambda(\theta_{\varepsilon,x}) + \varepsilon\Phi_{\varepsilon}(x), \tag{5.9}$$

where  $\Lambda$  is defined in (5.5) and  $\Phi_{\varepsilon} = (\Phi_{1,\varepsilon}, \Phi_{2,\varepsilon}, \Phi_{3,\varepsilon})$ , with

$$\begin{aligned} \Phi_{1,\varepsilon}(x) &= \phi_{\varepsilon,x}(x)\Gamma_0^{-1}(x - \mu_0), \\ \Phi_{2,\varepsilon}(x) &= \phi_{\varepsilon,x}(x)[\Gamma_0^{-1}(x - \mu_0)(x - \mu_0)' \Gamma_0^{-1} - I_k], \\ \Phi_{3,\varepsilon}(x) &= \phi_{\varepsilon,x}(x) - \gamma. \end{aligned}$$

Define  $\Theta(P) = (\mu(P), \Gamma(P), \rho(P))$ , where  $\mu(P) = T_P(\phi_P)$ ,  $\Gamma(P)^2 = C_P(\phi_P)$ ,  $\rho(P) = r_P(\phi_P)$ , and  $\phi_P$  denotes a minimizing trimming function. The influence function of  $\Theta(P)$  at  $P_0$  is defined as

$$\text{IF}(x, \Theta, P_0) = \lim_{\varepsilon \downarrow 0} \frac{\Theta((1 - \varepsilon)P_0 + \varepsilon\delta_x) - \Theta(P_0)}{\varepsilon},$$

if this limit exists, where  $\delta_x$  is the Dirac measure at  $x \in \mathbb{R}^k$ . The following theorem shows that this limit exists and provides its expression.

**Theorem 5.2.** *Suppose  $P_0$  satisfies (3.1), (4.2), and (5.8). Suppose that  $(\mu_0, \Sigma_0)$  is uniquely defined at  $P_0$ . Suppose that  $x \notin \partial E(\mu_0, \Sigma_0, \rho_0)$ . If  $\Lambda$  has a non-singular derivative at  $\theta_0$ , then the influence function of  $\Theta$  at  $P_0$  is given by*

$$\text{IF}(x, \Theta, P_0) = -\Lambda'(\theta_0)^{-1}\Psi(x, \theta_0),$$

where  $\Psi$  is defined in (5.3).

From definition (5.3), we see that  $\text{IF}(x, \Theta, P_0)$  is bounded uniformly for  $x \notin \partial E(\mu_0, \Sigma_0, \rho_0)$ . When  $x \in \partial E(\mu_0, \Sigma_0, \rho_0)$ , then it is not clear what happens with  $\phi_{\varepsilon,x}(x)$ , as  $\varepsilon \downarrow 0$ . However, recall that there exist  $R > 0$  such that  $\{\phi_{\varepsilon,x} \neq 0\} \subset B_R$ , for  $\varepsilon > 0$  sufficiently small. This still implies that if  $\phi_{\varepsilon,x}(x)$  has a limit, as  $\varepsilon \downarrow 0$ , then  $\text{IF}(x; \Theta, P_0)$  exists and is bounded. In the case that  $\phi_{\varepsilon,x}(x)$  does not have a limit, as  $\varepsilon \downarrow 0$ , then we can still conclude that  $\theta_{\varepsilon,x} - \theta_0 = O(\varepsilon)$ , uniformly for  $x \in \partial E(\mu_0, \Sigma_0, \rho_0)$ .

Because  $\Sigma_{\varepsilon,x} - \Sigma_0 = 2\Gamma_0(\Gamma_{\varepsilon,x} - \Gamma_0) + o(1)$ , as  $\varepsilon \downarrow 0$ , it follows that the influence function of the covariance functional  $\Sigma(P) = C_P(\phi_P)$  is given by

$$\text{IF}(x; \Sigma, P_0) = 2\Gamma_0 \cdot \text{IF}(x; \Gamma, P_0).$$

As a special case of Theorem 5.2, when  $P_0$  has an elliptically contoured density, Theorem 1 in [4] may be recovered (see [3]). Finally, note that together with Theorem 5.1 it turns out that the von Mises expansion indeed holds, that is,

$$\widehat{\theta}_n - \theta_0 = \frac{1}{n} \sum_{i=1}^n \text{IF}(X_i; \Theta, P_0) + o_{\mathbb{P}}(n^{-1/2}),$$

which includes the heuristic (1.2).

On the basis of Theorems 5.1 and 5.2, one could compute asymptotic variances and robustness performance measures for the MCD estimators and compare them with other robust competitors. Assuming the influence function to exist and the expansion (1.2) to be valid, Croux and Haesbroeck [4] provide an extensive account of asymptotic and finite sample relative efficiencies for the components of the MCD covariance estimator separately at the multivariate standard normal, a contaminated multivariate normal and at several multivariate Student distributions, for a variety of dimensions  $k = 2, 3, 5, 10, 30$  and  $\gamma = 0.5, 0.75$ , as well as a comparison with  $S$ -estimators and reweighted versions. Of particular interest would be a comparison with the Stahel–Donoho (SD) estimator. Its asymptotic properties have been established by Zuo *et al.* [26–28], who also report an asymptotic and finite sample efficiency index for the SD location estimator and for the full SD covariance estimator at the multivariate normal and contaminated normal as well as a gross error sensitivity index and maximum bias curve for the SD covariance estimator. The first impression is that overall, apart from computational issues, the SD estimator performs better than the MCD. However, a honest comparison would require comparison of the same measure of efficiency and of the maximum bias curves. To determine the latter seems far from trivial for the MCD and we delay such a comparison to future research.

## Appendix

Because the proof of Theorem 2.1 relies heavily on the proof of the results in Sections 3 and 4, this proof is postponed to Section A.2.

### A.1. Proofs of existence and characterization (Section 3)

**Proof of Proposition 3.1.** Suppose that the statement is not true. This means that the inequality either doesn't hold for  $t = 0$  or for a sequence  $t_n \rightarrow \infty$ . Let us start with the last case: we would have that there exist  $a_n \in S^k$  and  $\phi_n \in K_{t_n}(\gamma)$  such that

$$\int (a'_n(x - T_{t_n}(\phi_n)))^2 \phi_n(x) P_n(dx) \rightarrow 0, \tag{A.1}$$

where  $P_n = P_{t_n} \rightarrow P_0$  weakly. Define the probability measures

$$Q_n = \frac{\phi_n \cdot P_n}{\int \phi_n dP_n}.$$

Since  $\{P_n\}$  is a tight (converging) sequence of probability measures and since  $\phi_n \leq 1$  is such that  $\int \phi_n dP_n \geq \gamma$  for all  $n$ , it is easy to see that  $\{Q_n\}$  is a tight sequence of probability measures. By going to subsequences, we may assume that  $Q_n \rightarrow Q_0$  weakly, for some probability measure  $Q_0$ . Furthermore, again by going to subsequences, we may assume that  $a_n \rightarrow a_0 \in S^k$  and that  $\int \phi_n dP_n \rightarrow \gamma_0 \geq \gamma$ . Finally, using the tightness of  $\{Q_n\}$ , we can see that  $a'_n T_{t_n}(\phi_n)$  is a bounded sequence, since otherwise the limsup of the integral would go to  $+\infty$ . This means that we can also assume that  $a'_n T_{t_n}(\phi_n) \rightarrow r \in \mathbb{R}$ . Since  $Q_n$  converges weakly to  $Q_0$ , we can use the Skorokhod representation theorem to define random variables  $X_n$  and  $X$  on a probability space  $(\Omega, \mu, \Sigma)$ , such that  $X_n \sim Q_n$ ,  $X \sim Q_0$  and  $X_n \rightarrow X$  almost surely. It follows that

$$(a'_n(X_n - T_{t_n}(\phi_n)))^2 \rightarrow (a'_0(X - r))^2, \quad \text{a.s.},$$

and from (A.1), together with Fatou's lemma we would conclude that

$$\begin{aligned} \int (a'_0(x - r))^2 Q_0(dx) &= \mathbb{E}(a'_0(X - r))^2 \\ &\leq \liminf_{n \rightarrow \infty} \mathbb{E}(a'_n(X_n - T_{t_n}(\phi_n)))^2 \\ &= \liminf_{n \rightarrow \infty} \frac{1}{\gamma_0} \int (a'_n(x - T_{t_n}(\phi_n)))^2 \phi_n(x) P_n(dx) = 0. \end{aligned}$$

This would mean that  $Q_0$  has to be concentrated on some closed hyperplane  $H$ . However, we know that for all positive bounded continuous functions  $g$ ,

$$\int g dQ_0 = \lim_{n \rightarrow \infty} \int g dQ_n \leq \lim_{n \rightarrow \infty} \frac{1}{\gamma} \int g dP_n = \frac{1}{\gamma} \int g dP_0.$$

This proves for the closed hyperplane  $H$  that  $1 = Q_0(H) \leq P_0(H)/\gamma$ , which contradicts assumption (3.1). The case  $t = 0$  can be proved using the same methods, just by defining  $Q_n = \phi_n \cdot P_0 / \int \phi_n dP_0$ . □

The proof of Proposition 3.2 relies on two lemmas. The first one is a direct consequence of Proposition 3.1, and shows that if  $\det(C_t(\phi))$  is bounded uniformly in  $t$  and  $\phi$ , then there exists a fixed compact set that contains all  $(T_t(\phi), C_t(\phi))$  eventually. The second lemma is a useful property involving the determinants of two nonnegative symmetric matrices. Furthermore, for  $R > 0$  and  $\mu \in \mathbb{R}^k$ , define

$$B(\mu, R) = \{x \in \mathbb{R}^k: \|x - \mu\| \leq R\}, \tag{A.2}$$

and write  $B_R$  in case  $\mu = 0$ .

**Lemma A.1.** *Suppose  $P_0$  satisfies (3.1) and let  $P_t \rightarrow P_0$  weakly. Fix  $M > 0$ . Then there exist  $t_0 \geq 1, 0 < \lambda_0 \leq \lambda_1 < \infty$  and  $L, \rho > 0$ , such that for  $t = 0$ , all  $t \geq t_0$ , all  $\phi$  such that  $T_t(\phi), C_t(\phi)$  exist,  $\int \phi dP_t \geq \gamma$ , and*

$$\det(C_t(\phi)) \leq M,$$

*we have that all eigenvalues of  $C_t(\phi)$  are between  $\lambda_0$  and  $\lambda_1$ ,  $\|T_t(\phi)\| \leq L$ , and  $r_t(\phi) \leq \rho$ .*

**Proof.** The existence of  $\lambda_0$  follows directly from Proposition 3.1. This also implies that the largest eigenvalue  $\lambda_{\max}$  of  $C_t(\phi)$  is smaller than  $M/\lambda_0^{k-1}$ . Finally, choose  $R > 0$  such that for all  $t \geq 0, P_t(B_R) \geq 1 - \gamma/2$ , with  $B_R$  as defined in (A.2). Suppose  $\|T_t(\phi)\| \geq R$  and according to Lemma 2.1, assume without loss of generality that  $\int \phi dP_t = \gamma$ . Then, since

$$\int_{B_R} \phi dP_t = \int \phi dP_t - \int_{\mathbb{R}^k \setminus B_R} \phi dP_t \geq \gamma - (1 - P_t(B_R)) \geq \frac{\gamma}{2},$$

we find

$$\begin{aligned} \lambda_{\max} &\geq \frac{1}{\gamma} \int \left( \frac{T_t(\phi)'(T_t(\phi) - x)}{\|T_t(\phi)\|} \right)^2 \phi(x) P_t(dx) \\ &\geq \frac{1}{\gamma} \int_{B_R} \left( \frac{T_t(\phi)'(T_t(\phi) - x)}{\|T_t(\phi)\|} \right)^2 \phi(x) P_t(dx) \\ &\geq \frac{1}{\gamma} \int_{B_R} (\|T_t(\phi)\| - R)^2 \phi(x) P_t(dx) \geq \frac{1}{2} (\|T_t(\phi)\| - R)^2. \end{aligned}$$

This proves that there exists  $L > 0$ , depending on  $R, M$  and  $\lambda_0$ , such that  $\|T_t(\phi)\| \leq L$ . Finally, since

$$(x - T_t(\phi))' C_t(\phi)^{-1} (x - T_t(\phi)) \leq \frac{(\|x\| + L)^2}{\lambda_0}$$

for  $\rho > 0$  large enough, the ellipsoid  $E(T_t(\phi), C_t(\phi), \rho)$ , as defined in (2.2), contains the ball  $B(0, \rho\sqrt{\lambda_0} - L)$ . Choose  $\rho > 0$  large enough, such that

$$P_t(B(0, \rho\sqrt{\lambda_0} - L)) \geq \gamma$$

for all  $t \geq 1$ . Then  $P_t(E(T_t(\phi), C_t(\phi), \rho)) \geq \gamma$  and by definition we must have  $r_t(\phi) \leq \rho$ . □

**Lemma A.2.** *Let  $\Sigma_1$  and  $\Sigma_2$  be two symmetric matrices, nonnegative and positive definite, respectively, such that  $\text{Tr}(\Sigma_2^{-1}(\Sigma_1 - \Sigma_2)) < 0$ . Then  $\det(\Sigma_1) < \det(\Sigma_2)$ . A similar result holds with  $\leq$  instead of strict inequalities.*

**Proof.** Without loss of generality we may assume that  $\Sigma_2 = I_k$ . Suppose  $\text{Tr}(\Sigma_1) < \text{Tr}(I_k)$ . This means that the eigenvalues  $\lambda_1, \dots, \lambda_k$  of  $\Sigma_1$  satisfy  $(\lambda_1 + \dots + \lambda_k)/k < 1$ . By means of the

inequality between the arithmetic mean and the geometric mean of nonnegative numbers, we find  $\det(\Sigma_1) = \lambda_1 \cdots \lambda_k < 1$ . □

We also need the following well known result.

**Lemma A.3.** *Suppose  $Q$  is a probability measure on  $\mathbb{R}^k$  such that  $\int \|x\|^2 Q(dx) < +\infty$  and  $Q$  is not supported by a hyperplane. Define  $\mu = \int x Q(dx)$ . Then for all  $a \in \mathbb{R}^k$ ,  $a \neq \mu$ , we have*

$$\det\left(\int (x - \mu)(x - \mu)' Q(dx)\right) < \det\left(\int (x - a)(x - a)' Q(dx)\right).$$

**Proof.** First note that

$$\int (x - a)(x - a)' Q(dx) = \int (x - \mu)(x - \mu)' Q(dx) + (a - \mu)(a - \mu)',$$

then apply Lemma A.2, remembering that  $\int (x - a)(x - a)' Q(dx)$  is invertible and therefore strictly positive definite. □

**Proof of Proposition 3.2.** Choose  $t_0 \geq 1$  and  $\lambda_0 > 0$  according to Proposition 3.1. Choose  $R' > 0$  such that  $P_t(B_{R'}) \geq \gamma$ , for all  $t \geq 0$ . Let  $\psi_0$  be a continuous bounded function such that  $\mathbb{1}_{B_{R'}} \leq \psi_0 \leq \mathbb{1}_{B_{R'+1}}$  and define  $D_0 = 2 \det(C_0(\psi_0))$ . Because  $P_t \rightarrow P_0$  weakly, and  $\psi_0$  has bounded support, we have

$$\begin{aligned} \int \psi_0 dP_t &\rightarrow \int \psi_0 dP_0, & \int \psi_0(x)x P_t(dx) &\rightarrow \int \psi_0(x)x P_0(dx) & \text{and} \\ & & \int \psi_0(x)xx' P_t(dx) &\rightarrow \int \psi_0(x)xx' P_0(dx), \end{aligned}$$

and hence  $C_t(\psi_0) \rightarrow C_0(\psi_0)$ , so that for  $t$  large enough,  $\det(C_t(\psi_0)) \leq D_0$ .

Now, consider  $\phi \in K_t(\gamma)$ . If  $\det(C_t(\phi)) \geq D_0 \geq \det(C_t(\psi_0))$ , then we are done, because  $\psi_0 \in K_t^{R'+1}(\gamma)$ . Therefore, suppose that  $\det(C_t(\phi)) < D_0$ . According to Lemma A.1, this implies there exist  $\lambda_1 \geq \lambda_0 > 0$  and  $L > 0$  such that

$$\lambda_0 \leq \lambda_{\min}(C_t(\phi)) \leq \lambda_{\max}(C_t(\phi)) \leq \lambda_1 \quad \text{and} \quad \|T_t(\phi)\| \leq L, \tag{A.3}$$

uniformly in  $t$  and  $\phi$ . According to Lemma 2.1, we may assume that  $\int \phi dP_t = \gamma$ . Choose any  $R > R' + 1$  and suppose that  $\phi > \mathbb{1}_{B_R} \cdot \phi$ . Because  $\int \phi dP_t = \gamma$  and  $P_t(B_{R'}) \geq \gamma$ , we know that

$$\int_{B_{R'}} (1 - \phi) dP_t \geq \int_{\mathbb{R}^k \setminus B_R} \phi dP_t.$$

Define  $h_1 = \mathbb{1}_{\mathbb{R}^k \setminus B_R} \cdot \phi$  and  $h_2 = \tau \mathbb{1}_{B_{R'}}(1 - \phi)$ , where we choose  $0 \leq \tau \leq 1$  such that

$$\int h_2 dP_t = \int_{B_{R'}} \tau(1 - \phi) dP_t = \int h_1 dP_t. \tag{A.4}$$

Furthermore, define  $\psi = \phi - h_1 + h_2$  and note that  $\psi \in K_t^R(\gamma)$ . Because according to (A.4),  $\int \psi \, dP_t = \int \phi \, dP_t$ , we can write

$$\begin{aligned} \det(C_t(\psi)) &= \det\left(\frac{1}{\int \psi \, dP_t} \int (x - T_t(\psi))(x - T_t(\psi))' \psi(x) P_t(dx)\right) \\ &\leq \det\left(\frac{1}{\int \psi \, dP_t} \int (x - T_t(\phi))(x - T_t(\phi))' \psi(x) P_t(dx)\right) \\ &= \det\left(C_t(\phi) + \frac{1}{\int \phi \, dP_t} \int (x - T_t(\phi))(x - T_t(\phi))' (\psi(x) - \phi(x)) P_t(dx)\right). \end{aligned} \tag{A.5}$$

For the inequality, we used Lemma A.3. So according to Lemma A.2, it suffices to show that

$$\frac{1}{\int \phi \, dP_t} \int (x - T_t(\phi))' C_t(\phi)^{-1} (x - T_t(\phi)) (h_2(x) - h_1(x)) P_t(dx) < 0. \tag{A.6}$$

To see that this is true, note that with (A.3) we get

$$\begin{aligned} \int (x - T_t(\phi))' C_t(\phi)^{-1} (x - T_t(\phi)) h_2(x) P_t(dx) &\leq \lambda_0^{-1} \int_{B_{R'}} \|x - T_t(\phi)\|^2 h_2(x) P_t(dx) \\ &\leq \lambda_0^{-1} (R' + L)^2 \int h_2 \, dP_t, \end{aligned}$$

and

$$\begin{aligned} \int (x - T_t(\phi))' C_t(\phi)^{-1} (x - T_t(\phi)) h_1(x) P_t(dx) &\geq \lambda_1^{-1} \int_{\mathbb{R}^k \setminus B_R} \|x - T_t(\phi)\|^2 h_1(x) P_t(dx) \\ &\geq \lambda_1^{-1} (R - L)^2 \int h_1 \, dP_t. \end{aligned}$$

So, together with (A.4), for  $R$  large enough (but independent of  $\phi$ !), this proves (A.6). □

**Proof of Theorem 3.1.** Choose  $R > 0$  and  $t_0 \geq 1$  according to Proposition 3.2. Then for  $t = 0$  and  $t \geq t_0$ , we may restrict minimization to  $\phi \in K_t^R(\gamma)$ . Since  $K_t^R(\gamma)$  is a weak\*-compact subset of  $L^\infty(P_t|_{B_R})$ , and since  $\phi \mapsto \det(C_t(\phi))$  is a weak\*-continuous function on this space, we conclude that there exists at least one minimum. □

**Proof of Theorem 3.2.** First, only consider minimizing functions with  $\int \phi \, dP = \gamma$ , which is always possible according to Lemma 2.1. Let  $\phi_\gamma \in K_P(\gamma)$  be such that  $\int \phi_\gamma \, dP = \gamma$  and  $\mathbb{1}_{E_P(\phi)^\circ} \leq \phi_\gamma \leq \mathbb{1}_{E_P(\phi)}$ . Define  $Y = \Gamma_P(\phi)^{-1}(X - T_P(\phi))$ , where  $X$  has distribution  $P$  and  $\Gamma_P(\phi)^2 = C_P(\phi)$ , and denote  $Q$  for the distribution of  $Y$ . Define

$$\begin{aligned} \tau_\gamma(y) &= \phi_\gamma(\Gamma_P(\phi)y + T_P(\phi)), \\ \tau_\phi(y) &= \phi(\Gamma_P(\phi)y + T_P(\phi)). \end{aligned} \tag{A.7}$$

Then  $\mathbb{1}_{B^\circ} \leq \tau_\gamma \leq \mathbb{1}_B$ , where  $B$  denotes the closed ball around the origin with radius  $r_P(\phi)$ . According to Lemma 1 in [10], with  $\Phi(t) = t^2$ , it then follows that

$$\int \tau_\gamma(y) \|y\|^2 Q(dy) \leq \int \tau_\phi(y) \|y\|^2 Q(dy)$$

and one has equality if and only if  $\mathbb{1}_{B^\circ} \leq \tau_\phi \leq \mathbb{1}_B$   $Q$ -a.s. This is equivalent with

$$\int (\phi_\gamma(x) - \phi(x))(x - T_P(\phi))' C_P(\phi)^{-1} (x - T_P(\phi)) P(dx) \leq 0,$$

where equality holds if and only if  $\mathbb{1}_{E_P(\phi)^\circ} \leq \phi \leq \mathbb{1}_{E_P(\phi)}$ ,  $P$ -a.e. Then with Lemma A.2, similar to arguments given in (A.5) and (A.6), it follows that  $\det(C_P(\phi_\gamma)) < \det(C_P(\phi))$  unless  $\mathbb{1}_{E_P^\circ} \leq \phi \leq \mathbb{1}_{E_P}$ ,  $P$ -a.e.

Now, suppose that  $\int \phi dP > \gamma$ . Then, according to Lemma 2.1, for some  $0 < \lambda < 1$ , the function  $0 \leq \lambda\phi < 1$  would also be minimizing and satisfies  $\int (\lambda\phi) dP = \gamma$ . But then, the argument above shows that  $\lambda\phi = 1$  on the interior of its own ellipsoid  $E(T_P(\lambda\phi), C_P(\lambda\phi), r_P(\lambda\phi))$ , which is a contradiction. We conclude that we must have  $\int \phi dP = \gamma$ .

The last statement of the theorem is a little bit more subtle. Suppose  $P(\partial E_P) > 0$ , since otherwise the statement is trivially true. Consider the following two functions on  $[0, 1]$ :

$$f_1(t) = \frac{P(\partial E_P \cap \{\phi \leq t\})}{P(\partial E_P)} \quad \text{and} \quad f_2(t) = \frac{P(\partial E_P \cap \{\phi \geq t\})}{P(\partial E_P)}.$$

If one realizes that  $f_1 + f_2 \geq 1$ ,  $f_1$  is nondecreasing and continuous from the right, whereas  $f_2$  is nonincreasing and continuous from the left, it is not hard to see that either  $f_1 = 1$  on  $[0, 1]$ , in which case  $\phi = 0$ ,  $P$ -a.e. on  $\partial E_P$ , or  $f_2 = 1$  on  $[0, 1]$ , in which case  $\phi = 1$ ,  $P$ -a.e. on  $\partial E_P$ , or there exists  $t \in (0, 1)$  such that  $f_1(t), f_2(t) > 0$ . For this  $t \in (0, 1)$ , define

$$A = \partial E_P \cap \{\phi \leq t\} \quad \text{and} \quad B = \partial E_P \cap \{\phi \geq t\}.$$

Either  $P(A \cup B) = P(\{x\})$  for some  $x \in \partial E_P$ , in which case  $P(\partial E_P) = P(\{x\})$ , or there exists  $x \in \text{supp}(P|_A)$  and  $y \in \text{supp}(P|_B)$  with  $x \neq y$ . We will show that this last assumption will lead to a contradiction, thereby finishing the proof. Choose  $\varepsilon > 0$  such that  $\varepsilon < \|x - y\|/3$ . Define  $A_\varepsilon = A \cap B_\varepsilon(x)$  and  $B_\varepsilon = B \cap B_\varepsilon(y)$ . By the choice of  $x$  and  $y$ , we know that  $P(A_\varepsilon), P(B_\varepsilon) > 0$ . Choose  $\eta < \min(tP(B_\varepsilon), (1-t)P(A_\varepsilon))$  and define

$$h_1(z) = \frac{\eta}{P(A_\varepsilon)} \mathbb{1}_{A_\varepsilon}(z) \quad \text{and} \quad h_2(z) = \frac{\eta}{P(B_\varepsilon)} \mathbb{1}_{B_\varepsilon}(z).$$

Since  $\phi \leq t$  on  $A_\varepsilon$  and  $\phi \geq t$  on  $B_\varepsilon$ , we get that

$$\psi = \phi + h_1 - h_2 \in K_P(\gamma).$$

Furthermore,  $\int \psi dP = \int \phi dP = \gamma$ . Since  $\varepsilon < \|x - y\|/3$ , we can see that

$$T_P(\psi) = T_P(\phi) + \frac{1}{\gamma P(A_\varepsilon)} \int_{A_\varepsilon} z P(dz) - \frac{1}{\gamma P(B_\varepsilon)} \int_{B_\varepsilon} z P(dz) \neq T_P(\phi).$$

Since  $C_P(\psi)$  is invertible, this means that (with strict inequality due to Lemma A.3)

$$\begin{aligned} \det(C_P(\psi)) &= \det\left(\frac{1}{\gamma} \int (z - T_P(\psi))(z - T_P(\psi))' \psi(z) P(dz)\right) \\ &< \det\left(\frac{1}{\gamma} \int (z - T_P(\phi))(z - T_P(\phi))' \psi(z) P(dz)\right) \\ &= \det\left(C_P(\phi) + \frac{1}{\gamma} \int (z - T_P(\phi))(z - T_P(\phi))' (h_1(z) - h_2(z)) P(dz)\right). \end{aligned} \tag{A.8}$$

Since  $A_\varepsilon \cup B_\varepsilon \subset \partial E_P$ , and we know that for  $z \in \partial E_P$ ,  $\text{Tr}((z - T_P(\phi))' C_P(\phi)^{-1} (z - T_P(\phi)))$  is constant, we can use Lemma A.2 to conclude that

$$\det(C_P(\psi)) < \det(C_P(\phi)),$$

which contradicts the minimizing property of  $\phi$ . □

### A.2. Proofs of continuity (Section 4)

The proof of Theorem 4.1 uses the following two lemmas.

**Lemma A.4.** *Suppose  $P_0$  satisfies (3.1). Let  $P_t \rightarrow P_0$  weakly and suppose that (4.1) holds. For  $t \geq 1$ , let  $\psi_t \in K_t(\gamma)$  such that  $\psi_t \leq \mathbb{1}_{E_t}$ , where  $E_t = E(T_t(\psi_t), C_t(\psi_t), r_t(\psi_t))$ , and suppose there exists  $R > 0$ , such that  $\{\psi_t \neq 0\} \subset B_R$ , for  $t$  sufficiently large. Then*

$$\int \psi_t dP_t - \int \psi_t dP_0 \rightarrow 0, \quad T_t(\psi_t) - T_0(\psi_t) \rightarrow 0 \quad \text{and} \quad C_t(\psi_t) - C_0(\psi_t) \rightarrow 0,$$

uniformly in  $\psi_t \in K_t(\gamma)$  with  $\{\psi_t \neq 0\} \subset B_R$ .

**Proof.** Because  $\{\psi_t \neq 0\} \subset B_R$  eventually, we can write

$$\int \psi_t dP_t - \int \psi_t dP_0 = \int_{B_R} \psi_t d(P_t - P_0)$$

for  $t$  sufficiently large. For the signed measure  $Q_t = P_t - P_0$  write  $Q_t = Q_t^+ - Q_t^-$ , where  $Q_t^+$  and  $Q_t^-$  are positive measures on  $\mathbb{R}^k$ . According to (4.1),

$$\sup_{E \in \mathcal{E}} Q_t^+(E) + \sup_{E \in \mathcal{E}} Q_t^-(E) \leq 2 \sup_{E \in \mathcal{E}} |P_t(E) - P_0(E)| \rightarrow 0.$$

This implies  $\sup_{E \in \mathcal{E}} Q_t^+(E) \rightarrow 0$  and  $\sup_{E \in \mathcal{E}} Q_t^-(E) \rightarrow 0$ . Because  $0 \leq \psi_t \leq \mathbb{1}_{E_t}$ , we find

$$\begin{aligned} 0 &\leq \int_{B_R} \psi_t(x) Q_t^+(dx) \leq Q_t^+(E_t \cap B_R) \leq Q_t^+(E_t) \rightarrow 0, \\ 0 &\leq \int_{B_R} \psi_t(x) Q_t^-(dx) \leq Q_t^-(E_t \cap B_R) \leq Q_t^-(E_t) \rightarrow 0, \end{aligned}$$

which implies that

$$\int \psi_t \, dP_t - \int \psi_t \, dP_0 = \int_{B_R} \psi_t(x) Q_t^+(dx) - \int_{B_R} \psi_t(x) Q_t^-(dx) \rightarrow 0. \quad (\text{A.9})$$

Now, write

$$\begin{aligned} T_t(\psi_t) - T_0(\psi_t) &= \frac{1}{\int \psi_t \, dP_t} \int_{B_R} \psi_t(x)x P_t(dx) - \frac{1}{\int \psi_t \, dP_0} \int_{B_R} \psi_t(x)x P_0(dx) \\ &= \frac{1}{\int \psi_t \, dP_t} \int_{B_R} \psi_t(x)x (P_t - P_0)(dx) + \left( \frac{1}{\int \psi_t \, dP_t} - \frac{1}{\int \psi_t \, dP_0} \right) \int_{B_R} \psi_t(x)x P_0(dx). \end{aligned} \quad (\text{A.10})$$

The first term in (A.10) tends to zero, because  $\gamma \leq \int \psi_t \, dP_t \leq 1$  and

$$\begin{aligned} 0 &\leq \int_{B_R} \psi_t(x) \|x\| Q_t^+(dx) \leq R Q_t^+(E_t \cap B_R) \leq R Q_t^+(E_t) \rightarrow 0, \\ 0 &\leq \int_{B_R} \psi_t(x) \|x\| Q_t^-(dx) \leq R Q_t^-(E_t \cap B_R) \leq R Q_t^-(E_t) \rightarrow 0, \end{aligned}$$

which implies that

$$\int_{B_R} \psi_t(x)x (P_t - P_0)(dx) = \int_{B_R} \psi_t(x)x Q_t^+(dx) - \int_{B_R} \psi_t(x)x Q_t^-(dx) \rightarrow 0.$$

The second term in (A.10) also tends to zero, because of (A.9) and the fact that

$$\left\| \int_{B_R} \psi_t(x)x P_0(dx) \right\| \leq R.$$

It follows that  $T_t(\psi_t) - T_0(\psi_t) \rightarrow 0$ . Similarly, one proves  $C_t(\psi_t) - C_0(\psi_t) \rightarrow 0$ .  $\square$

**Lemma A.5.** *Suppose  $P_0$  satisfies (3.1). Let  $P_t \rightarrow P_0$  weakly and suppose that (4.1) holds. For  $t \geq 1$ , let  $\psi_t \in K_t(\gamma)$  such that  $\psi_t \leq \mathbb{1}_{E_t}$ , where  $E_t = E(T_t(\psi_t), C_t(\psi_t), r_t(\psi_t))$ , and suppose there exist  $R > 0$  such that  $\{\psi_t \neq 0\} \subset B_R$ , for  $t$  sufficiently large. Then there exist a subsequence  $t_m \rightarrow \infty$  and  $\psi^* \in K_0^R(\gamma)$ , such that*

$$\lim_{m \rightarrow \infty} (T_0(\psi_{t_m}), C_0(\psi_{t_m})) = (T_0(\psi^*), C_0(\psi^*)).$$

**Proof.** Since  $0 \leq \psi_t \leq 1$  and  $\{\psi_t \neq 0\} \subset B_R$ , the  $\psi_t$  can be viewed as elements of the class

$$\mathfrak{L}_0^R = \{\psi \in L^\infty(P_0): 0 \leq \psi \leq 1, \{\psi \neq 0\} \subset B_R\},$$

which is a weak\*-compact subset of  $L^\infty(P_0|_{B_R})$ . Hence, there exist a subsequence  $(\psi_{t_m})$  that has a weak\* limit in  $\mathcal{L}_0^R$ , say  $\psi^*$ . This means that for any  $g \in L^1(P_0)$ ,

$$\lim_{m \rightarrow \infty} \int \psi_{t_m} g \, dP_0 = \int \psi^* g \, dP_0. \tag{A.11}$$

In particular,  $\int \psi_{t_m} \, dP_0 \rightarrow \int \psi^* \, dP_0$ . Because  $\int \psi_t \, dP_t \geq \gamma$ , together with Lemma A.4 this implies

$$\int \psi^* \, dP_0 = \lim_{m \rightarrow \infty} \int \psi_{t_m} \, dP_0 \geq \gamma - \lim_{m \rightarrow \infty} \int \psi_{t_m} \, d(P_{t_m} - P_0) = \gamma,$$

so that  $\psi^* \in K_0^R(\gamma)$ . Finally, since the support of both  $\psi_{t_m}$  and  $\psi^*$  lies in  $B_R$ , it follows from (A.11) that

$$\begin{aligned} T_0(\psi_{t_m}) &= \frac{1}{\int \psi_{t_m} \, dP_0} \int_{B_R} \psi_{t_m}(x) x P_0(dx) \\ &= \frac{1}{\int \psi^* \, dP_0} \int_{B_R} \psi_{t_m}(x) x P_0(dx) + \left( \frac{1}{\int \psi_{t_m} \, dP_0} - \frac{1}{\int \psi^* \, dP_0} \right) \int_{B_R} \psi_{t_m}(x) x P_0(dx) \\ &\rightarrow \frac{1}{\int \psi^* \, dP_0} \int_{B_R} \psi^*(x) x P_0(dx) = T_0(\psi^*), \end{aligned}$$

and similarly  $C_0(\psi_{t_m}) \rightarrow C_0(\psi^*)$ . □

**Proof of Theorem 4.1.** Consider the sequence  $(T_t(\phi_t), C_t(\phi_t))$ . According to Proposition 3.1 there exist  $\lambda_0 > 0$ , such that  $\lambda_{\min}(C_t(\phi_t)) \geq \lambda_0$  for  $t$  sufficiently large. Similar to the beginning of the proof of Proposition 3.2, we obtain  $\lambda_{\max}(C_t(\phi_t)) \leq \lambda_1$  (see (A.3)). Because  $\psi_t \in K_t(\gamma)$ , again according to Proposition 3.1,  $\lambda_{\min}(C_t(\psi_t)) \geq \lambda_0$ . Since  $\det(C_t(\phi_t)) \leq \lambda_1^k$ , for  $t$  sufficiently large, and  $\det(C_t(\psi_t)) - \det(C_t(\phi_t))$  tends to zero, it follows that  $\det(C_t(\psi_t)) \leq 2\lambda_1^k$  eventually, so that according to Lemma A.1, there exists a compact set which contains  $(T_t(\psi_t), C_t(\psi_t))$  for  $t$  sufficiently large. This means there exist a convergent subsequence.

Now, consider a subsequence, which we continue to denote by  $(T_t(\psi_t), C_t(\psi_t))$ , for which  $(T_t(\psi_t), C_t(\psi_t)) \rightarrow (T_0, C_0)$ . From Lemmas A.4 and A.5, we conclude that there exists a further subsequence  $(t_m)$ , such that

$$\begin{aligned} T_0 &= \lim_{m \rightarrow \infty} T_{t_m}(\psi_{t_m}) = T_0(\psi^*), \\ C_0 &= \lim_{m \rightarrow \infty} C_{t_m}(\psi_{t_m}) = C_0(\psi^*) \end{aligned}$$

for some  $\psi^* \in K_0^R(\gamma)$ . It remains to show that  $(T_0, C_0)$  is an MCD-functional, that is,  $\det(C_0(\psi^*))$  minimizes  $\det(C_0(\phi))$  over  $K_0^R(\gamma)$ . To this end, suppose there exists  $\delta > 0$  and  $\phi \in K_0^R(\gamma)$ , such that

$$\det(C_0(\phi)) \leq \det(C_0(\psi^*)) - \delta.$$

Since the set of bounded continuous functions is dense within  $K_0^R(\gamma)$ , we can construct a bounded continuous function  $\psi \in K_0^R(\gamma)$ , such that for all  $i, j = 1, 2, \dots, k$ :

$$\int |\psi - \phi| dP_0, \quad \int |\psi(x)x_i - \phi(x)x_i| P_0(x) \quad \text{and} \quad \int |\psi(x)x_i x_j - \phi(x)x_i x_j| P_0(x),$$

can be made arbitrarily small. Hence, we can construct a bounded continuous function  $\psi \in K_0^R(\gamma)$ , such that

$$\det(C_0(\psi)) \leq \det(C_0(\psi^*)) - \delta/2.$$

Now, since  $\psi(x)x$  is bounded and continuous on  $B_R$ , we have

$$T_{t_m}(\psi) = \frac{1}{\int \psi dP_{t_m}} \int \psi(x)x P_{t_m}(dx) \rightarrow \frac{1}{\int \psi dP_0} \int \psi(x)x P_0(dx) = T_0(\psi),$$

and similarly  $C_{t_m}(\psi) \rightarrow C_0(\psi)$ . Since also  $\det(C_{t_m}(\psi_{t_m})) - \det(C_{t_m}(\phi_{t_m})) \rightarrow 0$ , it would follow that

$$\begin{aligned} \lim_{m \rightarrow \infty} \det(C_{t_m}(\psi)) &= \det(C_0(\psi)) \leq \det(C_0(\psi^*)) - \frac{\delta}{2} \\ &= \lim_{m \rightarrow \infty} \det(C_{t_m}(\psi_{t_m})) - \frac{\delta}{2} = \lim_{m \rightarrow \infty} \det(C_{t_m}(\phi_{t_m})) - \frac{\delta}{2}. \end{aligned}$$

This would mean that for  $m$  sufficiently large,  $\det(C_{t_m}(\psi)) \leq \det(C_{t_m}(\phi_{t_m})) - \delta/4$ , which contradicts the minimizing property of  $\phi_{t_m}$ .  $\square$

**Proof of Theorem 2.1.** First note that when  $\varepsilon \downarrow 0$ , then  $P_{r,\varepsilon} \rightarrow P$  weakly. Condition (4.1) automatically holds, and because  $P(H) < (\gamma - \varepsilon)/(1 - \varepsilon) < \gamma$ , also condition (3.1) holds. According to Theorem 3.1 this means that  $\text{MCD}_\gamma(P_{r,\varepsilon})$  exists, for  $\varepsilon > 0$  sufficiently small, and the minimizing  $\phi_{r,\varepsilon} \in K_{P_{r,\varepsilon}}^R(\gamma)$ . Hence, together with Theorem 3.2, all conditions of Theorem 4.1 are satisfied, which yields the first limit in (i). The proof the second limit in (i) mimics the proof of Theorem 4.1. Note that although we are not dealing with a weakly convergent sequence of measures satisfying condition (4.1), we do have that for all continuous functions  $f$  with bounded support,

$$\lim_{\|r\| \rightarrow \infty} \int f dP_{r,\varepsilon} = \int f d(1 - \varepsilon)P \tag{A.12}$$

and for every fixed  $R > 0$ ,

$$\lim_{\|r\| \rightarrow \infty} \sup_{E \in \mathcal{E}, E \subset B_R} |P_{r,\varepsilon}(E) - (1 - \varepsilon)P(E)| = 0. \tag{A.13}$$

We first show the analogue of Proposition 3.2, that is, there exists  $R > 0$  such that for all  $\|r\|$  sufficiently large, the support of all minimizing  $\phi$  for  $P_{r,\varepsilon}$  lies in  $B_R$ .

Choose  $R' > 0$  large enough, such that  $P(B_{R'}) > \gamma/(1 - \varepsilon)$ . This shows that there exist  $\psi \in K_{\gamma/(1-\varepsilon)}(P)$  with support contained in  $B_{R'}$ , such that  $\int \psi dP_{r,\varepsilon} \geq \gamma$ , for all  $r \in \mathbb{R}^k$ , and from

(A.12) we find

$$\lim_{\|r\| \rightarrow \infty} (T_{P_{r,\varepsilon}}(\psi), C_{P_{r,\varepsilon}}(\psi)) = (T_{(1-\varepsilon)P}(\psi), C_{(1-\varepsilon)P}(\psi)).$$

When we take  $M = 2 \det(C_{(1-\varepsilon)P}(\psi))$ , there exists  $r_0 > 0$ , such that for all  $r$  with  $\|r\| > r_0$ ,  $\det(C_{P_{r,\varepsilon}}(\psi)) < M$ . It follows, that if  $\phi$  is a minimizing function for  $P_{r,\varepsilon}$  at level  $\gamma$ , we can conclude that  $\det(C_{P_{r,\varepsilon}}(\phi)) < M$ , for  $\|r\| > r_0$ . Also, since  $\int \phi \, dP \geq (\gamma - \varepsilon)/(1 - \varepsilon)$ , Proposition 3.1 yields that there exists  $\lambda_0 > 0$ , not depending on  $\phi$ , such that for all  $a \in \mathbb{R}^k$  with  $\|a\| = 1$

$$\int (a'(x - T_P(\phi)))^2 \phi \, dP \geq \lambda_0.$$

This implies that

$$\begin{aligned} \int (a'(x - T_{P_{r,\varepsilon}}(\phi)))^2 \phi \, dP_{r,\varepsilon} &\geq (1 - \varepsilon) \int (a'(x - T_{P_{r,\varepsilon}}(\phi)))^2 \phi \, dP \\ &\geq (1 - \varepsilon) \int (a'(x - T_P(\phi)))^2 \phi \, dP \geq (1 - \varepsilon)\lambda_0. \end{aligned}$$

This means that for all minimizing  $\phi$  for  $P_{r,\varepsilon}$ , we have a uniform lower bound on the smallest eigenvalue. From here on, we copy the proof of Lemma A.1. We choose  $R > 0$  and  $\delta > 0$  (independent of  $\phi$ ) such that  $P_{r,\varepsilon}(B_R) \geq (1 - \varepsilon)P(B_R) \geq 1 - \gamma + \delta$  and  $\int_{B_R} \phi \, dP_{r,\varepsilon} \geq \delta$ . This then shows that there exists  $\lambda_{\max}$  and  $L > 0$  such that for all  $r$  with  $\|r\| > r_0$  and for all minimizing  $\phi$ , we have  $\|T_{P_{r,\varepsilon}}(\phi)\| < L$  and the largest eigenvalue of  $C_{P_{r,\varepsilon}}(\phi)$  is smaller than  $\lambda_{\max}$ . Now we can follow the proof of Proposition 3.2, starting from (A.3), to conclude that there exists  $R > 0$  and  $r_0 > 0$ , such that for all  $r$  with  $\|r\| > r_0$ , the support of all minimizing  $\phi$  for  $P_{r,\varepsilon}$  lies within  $B_R$ .

Note that, because according to Proposition 3.2 all minimizing  $\phi$  for  $(1 - \varepsilon)P$  also have a fixed bounded support, this immediately yields statement (ii). Indeed, if  $Q$  has bounded support, then for  $\|r\|$  sufficiently large,  $\int \phi \, dQ \circ \tau_r^{-1} = 0$  for all  $\phi$  with a fixed bounded support. Hence, for  $\|r\|$  sufficiently large,  $\phi$  is minimizing for  $P_{r,\varepsilon}$  at level  $\gamma$  if and only if  $\phi$  is minimizing for  $(1 - \varepsilon)P$  at level  $\gamma$ , which means that  $\phi$  is minimizing for  $P$  at level  $\gamma/(1 - \varepsilon)$ .

To finish the proof of (i), we follow the proof of Theorem 4.1, from the point of considering a convergent subsequence of  $\text{MCD}_\gamma(P_{r,\varepsilon})$ . The conclusions of Lemmas A.4 and A.5 are still valid if we replace the condition of weak convergence by (A.12) and replace condition (4.1) by (A.13). This means that the proof of the second limit in (i) is completely similar to the remainder of the proof of Theorem 4.1, which proves (i). □

**Proof of Corollary 4.1.** Since the MCD functional at  $P_0$  is unique, it follows immediately from Theorem 4.1 that each convergent subsequence has the same limit point  $(T_0(\phi_0), C_0(\phi_0))$ , which proves part (i).

For  $t = 1, 2, \dots$ , write  $E_t(s) = E(T_t(\psi_t), C_t(\psi_t), s)$  and  $\rho_t = r_t(\psi_t)$ , as defined by (2.2) and (2.3), and write  $E_0(s)$  and  $\rho_0$  for the ellipsoid and radius corresponding to  $\phi_0$ . Let  $\varepsilon > 0$  and

suppose that  $\liminf \rho_t \leq \rho_0 - 3\varepsilon$ . Then there exists a subsequence  $t_n \rightarrow \infty$  such that

$$E_{t_n}(\rho_{t_n}) \subset E_{t_n}(\rho_0 - 2\varepsilon) \subset E_0(\rho_0 - \varepsilon),$$

and from (4.2) it would follow that  $P_{t_n}(E_{t_n}(\rho_{t_n})) \leq P_{t_n}(E_0(\rho_0 - \varepsilon)) \rightarrow P_0(E_0(\rho_0 - \varepsilon)) < \gamma$ , which is in contradiction with the definition of  $\rho_{t_n}$ . Similarly, if  $\limsup \rho_t \geq \rho_0 + 3\varepsilon$ , then there exists a subsequence  $t_n \rightarrow \infty$  such that

$$E_{t_n}(\rho_{t_n} - \varepsilon) \supset E_{t_n}(\rho_0 + \varepsilon) \supset E_0(\rho_0),$$

from which it would follow that  $P_{t_n}(E_{t_n}(\rho_{t_n} - \varepsilon)) \geq P_{t_n}(E_0(\rho_0)) \rightarrow P_0(E_0(\rho_0)) \geq \gamma$ , which again is in contradiction with the definition of  $\rho_{t_n}$ . Since  $\varepsilon > 0$  was arbitrary, this finishes the proof of part (ii).  $\square$

**Proof of Proposition 4.1.** Let  $\phi_n$  be a minimizing function for the MCD functional corresponding to  $P_n$ . Then by definition

$$\det(C_n(\phi_n)) \leq \det(C_n(\mathbb{1}_{S_n})) = \det(\widehat{C}_n(S_n)).$$

First note that  $\phi_n$  cannot be zero on the boundary of  $E_n = E(T_n(\phi_n), C_n(\phi_n), r_n(\phi_n))$ . Hence according to Theorem 3.2, we either have  $\phi_n = 1$  on  $\partial E_n$  or there exists a point  $x \in \partial E_n$  such that  $P_n(\{x\}) = P_n(\partial E_n)$ . In the first case

$$\det(\widehat{C}_n(S_n)) \leq \det(\widehat{C}_n(\widetilde{S}_n)) = \det(C_n(\phi_n))$$

for the subsample  $\widetilde{S}_n = \{X_i: \phi_n(X_i) = 1\}$ , which means  $\det(\widehat{C}_n(S_n)) = \det(C_n(\phi_n))$ .

Consider the other case. Suppose  $\phi_n = 1$  in  $k$  points other than  $x$  and suppose there are  $m$  sample points  $X_i = x$ . Then we must have  $\gamma > k/n$  and  $\phi_n(x) = \varepsilon_n$  for some  $0 < \varepsilon_n < 1$ , where

$$\gamma = \int \phi_n dP_n = \frac{k}{n} + \frac{m\varepsilon_n}{n}.$$

Now, let  $\widetilde{S}_n$  be the subsample consisting of the  $k$  points where  $\phi_n = 1$  and  $\lceil m\varepsilon_n \rceil$  points  $X_i = x$ . Then  $\widetilde{S}_n$  has  $nP_n(\widetilde{S}_n) = k + \lceil m\varepsilon_n \rceil = \lceil n\gamma \rceil$  points. According to Proposition 3.2, with probability one, there exists  $R > 0$  such that  $\widetilde{S}_n$  and  $\{\phi_n \neq 0\}$  are contained in  $B_R$ . This implies

$$\|T_n(\mathbb{1}_{\widetilde{S}_n}) - T_n(\phi_n)\| = \frac{nR}{\lceil n\gamma \rceil} \left( \frac{\lceil m\varepsilon_n \rceil}{n} - \frac{m\varepsilon_n}{n} \right) \leq \frac{R}{\lceil n\gamma \rceil},$$

and similarly  $C_n(\mathbb{1}_{\widetilde{S}_n}) - C_n(\phi_n) = O(n^{-1})$  and  $\det(C_n(\mathbb{1}_{\widetilde{S}_n})) - \det(C_n(\phi_n)) = O(n^{-1})$ , with probability one. This means

$$\det(\widehat{C}_n(S_n)) \leq \det(\widehat{C}_n(\widetilde{S}_n)) = \det(C_n(\mathbb{1}_{\widetilde{S}_n})) = \det(C_n(\phi_n)) + O(n^{-1})$$

with probability one.  $\square$

The proof of Proposition 4.2 relies partly on the following property.

**Lemma A.6.** Let  $S_m$  be a subsample of size  $m \geq 2$  and let  $X^* \in S_m$  have maximal Mahalanobis distance with respect to the corresponding trimmed sample mean  $T_m = \widehat{T}_n(S_m)$  and trimmed sample covariance  $C_m = \widehat{C}_n(S_m)$ , that is,

$$X^* = \operatorname{argmax}_{X_i \in S_m} (X_i - T_m)' C_m^{-1} (X_i - T_m).$$

Define subsample  $S_{m-1} = S_m \setminus \{X^*\}$  with trimmed sample covariance  $C_{m-1} = \widehat{C}_n(S_{m-1})$ . Then

$$\det(C_{m-1}) \leq \det(C_m).$$

**Proof.** We can write

$$\begin{aligned} \det(C_{m-1}) &= \det\left(\frac{1}{m-1} \sum_{X_i \in S_{m-1}} (X_i - T_{m-1})(X_i - T_{m-1})'\right) \\ &< \det\left(\frac{1}{m-1} \sum_{X_i \in S_{m-1}} (X_i - T_m)(X_i - T_m)'\right) \\ &= \det\left(\frac{m}{m-1} C_m - \frac{1}{m-1} (X^* - T_m)(X^* - T_m)'\right) \\ &= \det\left(C_m + \frac{1}{m-1} [C_m - (X^* - T_m)(X^* - T_m)']\right). \end{aligned}$$

From the definition of  $C_m$ , after multiplication with  $C_m^{-1}$  and taking traces, we find

$$k = \frac{1}{m} \sum_{X_i \in S_m} (X_i - T_m)' C_m^{-1} (X_i - T_m).$$

Therefore, since  $X^*$  has the largest value for  $(X_i - T_m)' C_m^{-1} (X_i - T_m)$ , it follows that

$$\operatorname{Tr}[I_k - C_m^{-1} (X^* - T_m)(X^* - T_m)'] = k - (X^* - T_m)' C_m^{-1} (X^* - T_m) \leq 0.$$

The lemma now follows from Lemma A.2. □

**Proof of Proposition 4.2.** Suppose that there is a point  $X_\ell \in \widehat{E}_n$  that is not in  $S_n$ . Because  $S_n$  must always have at least one point on the boundary of  $\widehat{E}_n$ , we can then interchange a point  $X_j \in S_n$  that lies on the boundary of  $\widehat{E}_n$  with  $X_\ell$ . We will show that this will always decrease  $\det(\widehat{C}_n(S_n))$ . Let  $S_n^* = (S_n \setminus \{X_j\}) \cup \{X_\ell\}$ . Then

$$\widehat{T}_n(S_n^*) = \frac{1}{[n\gamma]} \sum_{X_i \in S_n^*} X_i = \widehat{T}_n(S_n) + \frac{1}{[n\gamma]} (X_j - X_\ell) \neq \widehat{T}_n(S_n).$$

Therefore (with a strict inequality), similar to (A.8), we have

$$\det(\widehat{C}_n(S_n^*)) < \det\left(\widehat{C}_n(S_n) - \frac{1}{\lceil n\gamma \rceil} (X_j - \widehat{T}_n(S_n))(X_j - \widehat{T}_n(S_n))' + \frac{1}{\lceil n\gamma \rceil} (X_\ell - \widehat{T}_n(S_n))(X_\ell - \widehat{T}_n(S_n))'\right).$$

Because  $X_j$  is on the boundary of  $\widehat{E}_n$  and  $X_\ell$  inside  $\widehat{E}_n$ , we have

$$(X_\ell - \widehat{T}_n(S_n))' \widehat{C}_n(S_n)^{-1} (X_\ell - \widehat{T}_n(S_n)) - (X_j - \widehat{T}_n(S_n))' \widehat{C}_n(S_n)^{-1} (X_j - \widehat{T}_n(S_n)) \leq 0.$$

Therefore, it follows from Lemma A.2 that  $\det(\widehat{C}_n(S_n^*)) < \det(\widehat{C}_n(S_n))$ , which contradicts the minimizing property of  $S_n$ . We conclude that  $\{X_1, \dots, X_n\} \cap \widehat{E}_n \subset S_n$ . Since according to Lemma A.6 the subsample  $S_n$  has exactly  $\lceil n\gamma \rceil$  points, and by definition  $\widehat{E}_n$  contains at least  $\lceil n\gamma \rceil$  points, we conclude that  $\{X_1, \dots, X_n\} \cap \widehat{E}_n = S_n$ .  $\square$

**Lemma A.7.** *Suppose  $P_0$  satisfies (3.1). With probability one, there exists  $R > 0$  and  $n_0 \geq 1$ , such that for all  $n \geq n_0$  and all subsamples  $S_n$  with at least  $n\gamma$  points, there exists a subsample  $S_n^*$  with exactly  $\lceil n\gamma \rceil$  points contained in  $B_R$  such that*

$$\det(\widehat{C}_n(S_n^*)) \leq \det(\widehat{C}_n(S_n)).$$

**Proof.** The proof is along the lines of the proof of Proposition 3.2. We first choose  $R' > 0$  and construct a subsample  $S_{n0} \subset B_{R'}$  with at least  $n\gamma$  points, for which  $\det(\widehat{C}_n(S_{n0}))$  is uniformly bounded for  $n$  sufficiently large. By the law of large numbers,  $P_n \rightarrow P_0$  weakly with probability one. Hence, we can choose  $R' > 0$  such that for all  $n \geq 1$ ,

$$P_n(B_{R'}) \geq \max\{1 - \gamma/2, \gamma + (1 - \gamma)/2\},$$

with probability one, and define subsample  $S_{n0} = \{X_i : X_i \in B_{R'}\}$ . Then,  $S_{n0} \subset B_{R'}$  and because  $P_n(B_{R'}) \geq \gamma$ , we also have  $\mathbb{1}_{S_{n0}} \in K_n(\gamma)$ . Then, according to Proposition 3.1, with probability one, there exist a  $\lambda_0 > 0$  such that

$$\lambda_{\min}(C_n(\mathbb{1}_{S_{n0}})) \geq \lambda_0$$

for  $n$  sufficiently large. Define  $D_0 = 2 \det(C_0(\mathbb{1}_{B_{R'}}))$ . From (4.1), we have  $P_n(B_{R'}) \rightarrow P_0(B_{R'})$ , with probability one, and since the functions  $x$  and  $xx'$  bounded and continuous on  $B_{R'}$ , we also have

$$\int_{B_{R'}} x dP_n \rightarrow \int_{B_{R'}} x dP_0, \quad \text{and} \quad \int_{B_{R'}} xx' dP_n \rightarrow \int_{B_{R'}} xx' dP_0,$$

with probability one. Hence, together with (4.3), it follows that for  $n$  sufficiently large,

$$\det(\widehat{C}_n(S_{n0})) = \det(C_n(\mathbb{1}_{S_{n0}})) = \det(C_n(\mathbb{1}_{B_{R'}})) \leq D_0,$$

with probability one. Now, let  $S_n$  be a subsample with  $h_n \geq n\gamma$  points. According to Lemma A.6, without loss of generality, we may assume that it has exactly  $\lceil n\gamma \rceil$  points. When  $\det(\widehat{C}_n(S_n)) > D_0$ , then we are done because the subsample  $S_{n_0}$  has a smaller determinant, it is contained in  $B_{R'}$ , and according to Lemma A.6 we can reduce  $S_{n_0}$  if necessary to have exactly  $\lceil n\gamma \rceil$  points, without increasing the determinant. So suppose that  $S_n$  has  $\lceil n\gamma \rceil$  points and  $\det(\widehat{C}_n(S_n)) \leq D_0$ . From here on the proof is identical to that of Proposition 3.2 and is left to the reader.  $\square$

**Proof of Theorem 4.2.** With probability one  $P_n \rightarrow P_0$  weakly and (4.1) holds, since the class of ellipsoids has polynomial discrimination or forms a Vapnik–Cervonenkis class. According to (4.3) the MCD estimators can be written as MCD functionals with trimming function  $\psi_n = \mathbb{1}_{S_n}$ . From Propositions 4.1 and 4.2 together with Lemma A.7, it follows that  $\psi_n$  satisfies the conditions of Theorem 4.1 with probability one, which proves the theorem.  $\square$

### A.3. Proofs of asymptotic normality and IF (Section 5)

The proof of Theorem 5.1 relies on the following result from [17], which we state for easy reference.

**Theorem A.3 (Pollard, 1984).** *Let  $\mathcal{F}$  be a permissible class of real valued functions with envelope  $H \geq |\phi|$ ,  $\phi \in \mathcal{F}$ , and suppose that  $0 < \mathbb{E}[H(X)^2] < \infty$ . If the class of graphs of functions in  $\mathcal{F}$  has polynomial discrimination, then for each  $\eta > 0$  and  $\varepsilon > 0$  there exists  $\delta > 0$  such that*

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left\{ \sup_{(\phi_1, \phi_2) \in [\delta]} n^{1/2} \left| \int (\phi_1 - \phi_2) d(P_n - P_0) \right| > \eta \right\} < \varepsilon,$$

where  $[\delta] = \{(\phi_1, \phi_2): \phi_1, \phi_2 \in \mathcal{F} \text{ and } \int (\phi_1 - \phi_2)^2 dP_0 < \delta^2\}$ .

The theorem is not stated as such in [17], but it is a combination of the Approximation lemma (page 27), Lemma II.36 (page 34) and the Equicontinuity lemma (page 150). The polynomial discrimination of  $\mathcal{F}$  provides a suitable bound on the entropy of  $\mathcal{F}$  (Approximation lemma together with Lemma II.36). The stochastic equicontinuity stated in Theorem A.3 is then a consequence of the fact that the entropy of  $\mathcal{F}$  is small enough (Equicontinuity lemma). The classes of functions we will encounter in this way can always be indexed by the parameter set  $\mathbb{R}^k \times \text{PDS}(k) \times \mathbb{R}_+$ , and are easily seen to be permissible in the sense of Pollard [17].

**Proof of Theorem 5.1.** Consider equation (5.4) and define

$$\mathcal{F} = \{ \mathbb{1}_{\{\|G^{-1}(x-m)\| \leq r\}}: m \in \mathbb{R}^k, G \in \text{PDS}(k), r > 0 \}.$$

As subclass of the class of indicator functions of all ellipsoids, the class of graphs  $\mathcal{G}$  of functions in  $\mathcal{F}$  has polynomial discrimination and obviously  $\mathcal{F}$  has envelope  $H = 1$ . Hence, Theorem A.3 applies to  $\Psi_3$ . For the real valued components of  $\Psi_1$  and  $\Psi_2$ , use that there exists  $R > 0$ , such that for  $n = 0$  and  $n$  sufficiently large

$$\{x \in \mathbb{R}^k: \|\Gamma_n^{-1}(x - \mu_n)\| \leq \rho_n\} \subset B_R.$$

This means that for all  $i, j = 1, 2, \dots, k$ , the classes

$$\begin{aligned} \mathcal{F}_i &= \{x_i \mathbb{1}_{\{\|G^{-1}(x-m)\| \leq r\} \cap B_R} : m \in \mathbb{R}^k, G \in \text{PDS}(k), r > 0\}, \\ \mathcal{F}_{ij} &= \{x_i x_j \mathbb{1}_{\{\|G^{-1}(x-m)\| \leq r\} \cap B_R} : m \in \mathbb{R}^k, G \in \text{PDS}(k), r > 0\}, \end{aligned}$$

have uniformly bounded envelopes. According to Lemma 3 in [14], the corresponding classes of graphs have polynomial discrimination. Therefore, Theorem A.3 also applies to the components of  $\Psi_1$  and  $\Psi_2$ . It follows that

$$0 = \Lambda(\widehat{\theta}_n) + \int \Psi(y, \theta_0)(P_n - P_0)(dy) + o_{\mathbb{P}}(n^{-1/2}).$$

Now,  $\Lambda(\theta_0) = 0$  and since  $\Psi(y, \theta_0)$  has bounded support, the term

$$\int \Psi(y, \theta_0)(P_n - P_0)(dy) = \frac{1}{n} \sum_{i=1}^n (\Psi(X_i, \theta_0) - \mathbb{E}\Psi(X_i, \theta_0)),$$

behaves according to the central limit theorem and is therefore of the order  $O_{\mathbb{P}}(n^{-1/2})$ . Because  $\widehat{\theta}_n \rightarrow \theta_0$  with probability one, according to Theorem 4.2, we find

$$0 = \Lambda'(\theta_0)(\widehat{\theta}_n - \theta_0) + O_{\mathbb{P}}(n^{-1/2}) + o_{\mathbb{P}}(\|\widehat{\theta}_n - \theta_0\|).$$

Because  $\Lambda'(\theta_0)$  is non-singular, this gives  $\|\widehat{\theta}_n - \theta_0\| = O_{\mathbb{P}}(n^{-1/2})$  and when inserting this, we conclude that

$$\Lambda'(\theta_0)(\widehat{\theta}_n - \theta_0) = -\frac{1}{n} \sum_{i=1}^n (\Psi(X_i, \theta_0) - \mathbb{E}\Psi(X_i, \theta_0)) + o_{\mathbb{P}}(n^{-1/2}),$$

which proves the first statement. For the second statement note that

$$\int (\mathbb{1}_{E_n}(y) - \phi_n(y)) \|y\|^2 P_n(y) = O(n^{-1}),$$

with probability one. This follows from the characterization given in Theorem 3.2 and the fact that  $P_0$  satisfies (5.7). This means that the MCD functional  $\theta_n$  also satisfies equation (5.4). From here on the argument is the same as before, which proves the theorem.  $\square$

**Proof of Theorem 5.2.** Consider expansion (5.9) and write  $E_0 = E(\mu_0, \Sigma_0, \rho_0)$ . Because, according to Theorem 4.1,  $(\mu_{\varepsilon,x}, \Gamma_{\varepsilon,x}, \rho_{\varepsilon,x}) \rightarrow (\mu_0, \Gamma_0, \rho_0)$ , as  $\varepsilon \downarrow 0$ , for  $x \notin \partial E_0$  we get  $\phi_{\varepsilon,x}(x) \rightarrow \mathbb{1}_{E_0}(x)$  and hence

$$\lim_{\varepsilon \downarrow 0} \Phi_{\varepsilon}(x) = \Psi(x, \theta_0),$$

with  $\Psi$  defined in (5.3). Because  $\mathbb{1}_{E_0^{\circ}} \leq \phi_0 \leq \mathbb{1}_{E_0}$ , it follows from (5.8) that  $\Lambda(\theta_0) = 0$ . As  $\Lambda$  has a non-singular derivative at  $\theta_0$ , we find from (5.9),

$$0 = (1 - \varepsilon)\Lambda'(\theta_0)(\theta_{\varepsilon,x} - \theta_0) + \varepsilon\Phi_{\varepsilon}(x) + o(\|\theta_{\varepsilon,x} - \theta_0\|),$$

from which we first deduce that  $\theta_{\varepsilon,x} - \theta_0 = O(\varepsilon)$ , and then obtain the influence function:

$$\text{IF}(x; \Theta, P_0) = \lim_{\varepsilon \downarrow 0} \frac{\theta_{\varepsilon,x} - \theta_0}{\varepsilon} = -\Lambda'(\theta_0)^{-1} \Psi(x). \quad \square$$

## Acknowledgements

We like to thank two anonymous referees for their comments and suggestions, which were helpful for improving the manuscript. Special thanks goes to referee 1 for correcting a mistake in the proof of Corollary 4.1 and for pointing out the relation with other relevant work [6,7,9,10], which has led to shorter proofs of Proposition 3.1 and the first part of Theorem 3.2.

## References

- [1] Agulló, J., Croux, C. and Van Aelst, S. (2008). The multivariate least-trimmed squares estimator. *J. Multivariate Anal.* **99** 311–338. [MR2396966](#)
- [2] Butler, R.W., Davies, P.L. and Jhun, M. (1993). Asymptotics for the minimum covariance determinant estimator. *Ann. Statist.* **21** 1385–1400. [MR1241271](#)
- [3] Cator, E.A. and Lopuhaä, H.P. (2010). Asymptotic expansion of the minimum covariance determinant estimators. *J. Multivariate Anal.* **101** 2372–2388. [MR2719868](#)
- [4] Croux, C. and Haesbroeck, G. (1999). Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *J. Multivariate Anal.* **71** 161–190. [MR1735108](#)
- [5] Croux, C. and Haesbroeck, G. (2000). Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies. *Biometrika* **87** 603–618. [MR1789812](#)
- [6] Cuesta-Albertos, J.A., Gordaliza, A. and Matrán, C. (1995). Trimmed  $k$ -means and the Cauchy mean-value property. In *New Trends in Probability and Statistics, Vol. 3 (Tartu/Pühajärve, 1994)* 247–265. Utrecht: VSP. [MR1655012](#)
- [7] Cuesta-Albertos, J.A., Gordaliza, A. and Matrán, C. (1997). Trimmed  $k$ -means: An attempt to robustify quantizers. *Ann. Statist.* **25** 553–576. [MR1439314](#)
- [8] Fekri, M. and Ruiz-Gazen, A. (2004). Robust weighted orthogonal regression in the errors-in-variables model. *J. Multivariate Anal.* **88** 89–108. [MR2021862](#)
- [9] García-Escudero, L.A., Gordaliza, A. and Matrán, C. (1999). A central limit theorem for multivariate generalized trimmed  $k$ -means. *Ann. Statist.* **27** 1061–1079. [MR1724041](#)
- [10] Gordaliza, A. (1991). Best approximations to random variables based on trimming procedures. *J. Approx. Theory* **64** 162–180. [MR1091467](#)
- [11] Hawkins, D.M. and McLachlan, G.J. (1997). High-breakdown linear discriminant analysis. *J. Amer. Statist. Assoc.* **92** 136–143. [MR1436102](#)
- [12] Hubert, M., Rousseeuw, P.J. and Van Aelst, S. (2008). High-breakdown robust multivariate methods. *Statist. Sci.* **23** 92–119. [MR2431867](#)
- [13] Hyvärinen, A., Karhunen, J. and Oja, E. (2001). *Independent Component Analysis*. New York: Wiley.
- [14] Lopuhaä, H.P. (1997). Asymptotic expansion of  $S$ -estimators of location and covariance. *Statist. Neerlandica* **51** 220–237. [MR1466427](#)
- [15] Lopuhaä, H.P. and Rousseeuw, P.J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Statist.* **19** 229–248. [MR1091847](#)

- [16] Pison, G., Rousseeuw, P.J., Filzmoser, P. and Croux, C. (2003). Robust factor analysis. *J. Multivariate Anal.* **84** 145–172. [MR1965827](#)
- [17] Pollard, D. (1984). *Convergence of Stochastic Processes. Springer Series in Statistics.* New York: Springer. [MR0762984](#)
- [18] Ranga Rao, R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* **33** 659–680. [MR0137809](#)
- [19] Rousseeuw, P. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications, Vol. B (Bad Tatzmannsdorf, 1983)* 283–297. Dordrecht: Reidel. [MR0851060](#)
- [20] Rousseeuw, P.J., Van Driessen, K., Van Aelst, S. and Agulló, J. (2004). Robust multivariate regression. *Technometrics* **46** 293–305. [MR2082499](#)
- [21] Serneels, S. and Verdonck, T. (2008). Principal component analysis for data containing outliers and missing elements. *Comput. Statist. Data Anal.* **52** 1712–1727. [MR2422765](#)
- [22] Taskinen, S., Croux, C., Kankainen, A., Ollila, E. and Oja, H. (2006). Influence functions and efficiencies of the canonical correlation and vector estimates based on scatter and shape matrices. *J. Multivariate Anal.* **97** 359–384. [MR2234028](#)
- [23] Tatsuoka, K.S. and Tyler, D.E. (2000). On the uniqueness of  $S$ -functionals and  $M$ -functionals under nonelliptical distributions. *Ann. Statist.* **28** 1219–1243. [MR1811326](#)
- [24] Tyler, D.E., Critchley, F., Dümbgen, L. and Oja, H. (2009). Invariant co-ordinate selection. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 549–592. [MR2749907](#)
- [25] Zhou, J. (2009). Robust dimension reduction based on canonical correlation. *J. Multivariate Anal.* **100** 195–209. [MR2460487](#)
- [26] Zuo, Y. and Cui, H. (2005). Depth weighted scatter estimators. *Ann. Statist.* **33** 381–413. [MR2157807](#)
- [27] Zuo, Y., Cui, H. and He, X. (2004). On the Stahel–Donoho estimator and depth-weighted means of multivariate data. *Ann. Statist.* **32** 167–188. [MR2051003](#)
- [28] Zuo, Y., Cui, H. and Young, D. (2004). Influence function and maximum bias of projection depth based estimators. *Ann. Statist.* **32** 189–218. [MR2051004](#)

Received June 2010 and revised December 2010