Journal of Multivariate Analysis 126 (2014) 167–183

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Strong consistency and robustness of the Forward Search estimator of multivariate location and scatter

Andrea Cerioli [a,*], Alessio Farcomeni [b], Marco Riani [a]

[a] University of Parma, Via Kennedy 6; 43100 Parma, Italy

[b] Sapienza - University of Rome, piazzale Aldo Moro, 5; 00186 Roma, Italy

ARTICLE INFO

Article history:
Received 16 May 2013
Available online 15 January 2014

AMS subject classifications:
62H10
62H12
62E20
62F12

Keywords:
Elliptical truncation
Forward Search
Generalized Mahalanobis distances
High-breakdown estimation
Multivariate trimming
Robust diagnostics

ABSTRACT

The Forward Search is a powerful general method for detecting anomalies in structured data, whose diagnostic power has been shown in many statistical contexts. However, despite the wealth of empirical evidence in favor of the method, only few theoretical properties have been established regarding the resulting estimators. We show that the Forward Search estimators are strongly consistent at the multivariate normal model. We also obtain their finite sample breakdown point. Our results put the Forward Search approach for multivariate data on a solid statistical ground, which formally motivates its use in robust applied statistics. Furthermore, they allow us to compare the Forward Search estimators with other well known multivariate high-breakdown techniques.

1. Introduction

The Forward Search (FS) is a powerful general method for detecting anomalies in structured data [2,6]. The idea behind the FS is simple and attractive. Given a sample of n observations and a generating model for them, the method starts from a subset of cardinality $m^* \ll n$ – often only few observations are required in practice, unless n is very large – which is robustly chosen to contain observations coming from the postulated model. This subset is used for fitting the model and the residuals, or other deviance measures, are computed. The subsequent fitting subset is then obtained by taking the $m^* + h$ observations with the smallest deviance measures. The algorithm iterates this fitting and updating scheme until all the observations are used in the fitting subset, thus yielding the classical statistical summary of the data. In practice h must be a finite number, possibly depending on n and on the postulated model. For instance, the typical choice with independent observations and moderate sample sizes is $h = 1$, while higher values are suitable with correlated data or very large samples. In the asymptotic framework of this work we have that $h \to \infty$ as $n \to \infty$, but we still assume a finite number of steps in the FS.

A major advantage of the FS is that it provides clear evidence of the impact that each unit, or block of units, exerts on the fitting process, with outliers and other peculiar observations entering in the last steps of the search. The presence of observations deviating from the null model can be displayed through pictures that monitor relevant quantities along the search, such as model residuals, distances, and their order statistics. For instance, if only $m < n$ units actually belong to the postulated population, we typically observe a peak in the monitoring plot of the minimum residual (distance) outside the fitting subset, when this subset only contains the m 'good' observations and the first outlier is about to enter. A further

* Corresponding author.
E-mail addresses: andrea.cerioli@unipr.it (A. Cerioli), alessio.farcomeni@uniroma1.it (A. Farcomeni), mriani@unipr.it (M. Riani).

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2013.12.010

bonus of the FS is that its main findings are usually insensitive to the specific choice of the initial subset, provided that it is outlier free, and virtually identical results are obtained through different criteria [7]. Typical methods for initializing the FS are Least Trimmed Squares in regression [21] and robust bivariate projections in multivariate analysis [36], but several alternative choices are also feasible.

The diagnostic power of the FS has been shown in many statistical contexts. For instance, in regression [3,4,8] the deletion residuals computed at each step of the FS can be monitored along the search, together with some of their relevant order statistics, for the detection of outliers and unsuspected structure in data, and so for building robust models. Such informative pictures are often called *forward plots*, because they are drawn by collecting several pieces of information, each of which is gathered from a different subset as the algorithm progresses. The full power of the FS thus stems from the combination of the different pieces, like in a "data movie" as opposed to a "data picture". Similar tools have also been developed for correlated observations [13], like in the case of spatial autoregressive models and in the kriging model of geostatistics. The FS for multivariate data replaces residuals with Mahalanobis distances, but keeps the general diagnostic approach unchanged. This leads to a (partial) ordering of multivariate data, and to robust and efficient diagnostic tools for the detection of multivariate outliers [5,30,27,18].

However, despite the wealth of empirical and simulation evidence in favor of the method, only few theoretical properties are available for the resulting estimators. The key ingredient for deriving such properties is the distribution of the basic quantities, i.e. residuals or distances, which are monitored along the search. These quantities are computed after a sequence of data driven steps. Therefore, obtaining their distribution is far from trivial, even in an asymptotic framework. Some approximate results that are useful in practice are available in the regression setting, based on the combination of the distribution theory of order statistics for residuals and truncation arguments under the normal distribution. Similar results are also available in the multivariate framework, with Mahalanobis distances in place of model residuals, and provide sound statistical thresholds for outlier nomination in finite samples, even of small and moderate sizes [30]. A detailed asymptotic analysis for the FS estimators has been developed only recently in [23,24], but for the univariate regression context only. Their study involves theory for a new class of weighted and marked empirical processes, quantile process theory, and a fixed point argument to describe the iterative nature of the FS algorithm. The main analytic results are an asymptotic representation of the FS residuals, scaled by the estimated variance, and convergence of the corresponding empirical process.

In this paper we deal with the estimators obtained through the single-population multivariate version of the FS, for which no asymptotic result is available yet. We do agree that the ultimate goal should be to study the weak convergence of the empirical process defined through the FS as the algorithm progresses, e.g. by extending the complex analytic machinery of [24] to the multivariate case. However, this difficult task is outside the scope of this paper and is left for further research. Our goal in the present work is slightly less ambitious. The multivariate FS estimators are usually assumed to be consistent and robust, following intuition and empirical experience, but formal proofs of such properties are still lacking. Our purpose is to fill the gap and to provide justification for such statements. Our asymptotic results thus put the FS approach for multivariate data on a solid statistical ground, which formally motivates its use in robust applied statistics and provides justification for its very good diagnostic properties. Furthermore, our proofs of consistency and robustness are important because they allow us to compare the FS estimators (3) and (4) with other well known multivariate high-breakdown estimators, for which similar properties have been established in the past. These include the Minimum Covariance Determinant and its reweighted version [14,25,15,10,11], S-estimators [17,20,35], Projection estimators [28,37,34] and Trimmed Likelihood estimators [16]. Some preliminary comparisons are provided in the paper, while more extensive theoretical and empirical results will be given elsewhere.

A problem related to the one that motivates our work was also considered by García-Escudero and Gordaliza [19], who derived the asymptotic distribution of the so-called radius process. This process is defined by trimmed Mahalanobis distances similar to (5) below, when the trimming level $1-\gamma$ varies in $(0, 1]$. However, an important difference is that in the radius process the multivariate estimators of location and scatter are computed only once and for all; then, they are kept fixed when the trimming level changes. This makes the radius process a valuable tool for the purpose of multivariate outlier detection, when sufficiently "good" robust parameter estimators already exist. On the other hand, the adaptive nature of the FS implies that the fitting subset changes with trimming level. New location and scatter estimators are thus defined at each step of the FS. These estimators, as well as the corresponding ellipsoids in the multivariate normal model, are dependent and the degree of dependence is unknown. We conjecture that, even with this additional degree of dependence at subsequent steps of the FS, the corresponding radius process is weakly convergent, but a formal proof of this statement is lacking. Therefore, in this paper we limit ourselves to analyze the pointwise asymptotic properties of the FS estimators, when they are computed for a finite sequence of steps. The strong consistency and asymptotic equicontinuity properties that we derive are clearly positive results towards our conjecture, and point to weak convergence results similar to those obtained by García-Escudero and Gordaliza [19]. A careful asymptotic analysis of the relationship between the radius process and the empirical process defined through the FS requires techniques that go well beyond the scope of this paper. Nevertheless, we note that our requirement of the existence of "good" robust parameter estimators is less stringent than in the radius process. For the consistency results derived in this paper, it is sufficient to start the FS with consistent estimators (see Assumption 1 in Section 3), while García-Escudero and Gordaliza [19] assume convergence with a rate of $n^{-1/2}$.

Our basic model for the data generating mechanism is the multivariate normal distribution

$$M_0 : y_i \sim N_v(\mu, \Sigma), \quad i = 1, \ldots, n, \tag{1}$$

where $y_i = (y_{i1}, \ldots, y_{iv})'$ is a v-variate observation, $\mu \in \mathbb{R}^v$, Σ is a positive-definite $v \times v$ matrix of constants, and $y_1, \ldots, y_n$ are independent. Let m be the sample size of the fitting subset in a generic intermediate step of the FS, with $m^* < m < n$. We do not consider the situations corresponding to classical statistical inference ($m = n$) and to the initial step ($m = m^*$). Under model (1) asymptotic results for the case $m = n$ are of course well known, while those for $m = m^*$ will depend on the specific properties of the initialization method chosen by the user.

Let $0 < \gamma < 1$ be such that

$$m = \lfloor n\gamma \rfloor, \tag{2}$$

where $\lfloor \cdot \rfloor$ denotes the floor function. The estimators of the parameters μ and Σ in (1) at step γ of the FS, based on a sample of size n, are defined as follows

$$\dot{\mu}_{\gamma,n} = \sum_{i=1}^{n} \frac{w_{i,\gamma,n}}{W_{\gamma,n}} y_i, \tag{3}$$

$$\dot{\Sigma}_{\gamma,n} = c_\gamma \sum_{i=1}^{n} \frac{w_{i,\gamma,n}}{W_{\gamma,n}} (y_i - \dot{\mu}_{\gamma,n})(y_i - \dot{\mu}_{\gamma,n})', \tag{4}$$

where either $w_{i,\gamma,n} = 0$ or $w_{i,\gamma,n} = 1$, $W_{\gamma,n} = \sum_{i=1}^{n} w_{i,\gamma,n}$, and c_γ is a scaling factor ensuring consistency of $\dot{\Sigma}_{\gamma,n}$ at the normal model (1). Both the binary weights $w_{i,\gamma,n}$, $i = 1, \ldots, n$, and the consistency factor c_γ will be specified in the following sections. Based on estimates (3) and (4), individual deviations from the null model M_0 are measured through the generalized Mahalanobis squared distances

$$d^2_{i,\gamma,n} = (y_i - \dot{\mu}_{\gamma,n})' \dot{\Sigma}^{-1}_{\gamma,n} (y_i - \dot{\mu}_{\gamma,n}), \quad i = 1, \ldots, n. \tag{5}$$

The goal of this paper is to show that the FS estimators (3) and (4) are strongly consistent when the multivariate normal model (1) holds for all the data. For this purpose, we start from the ideal case where the parameters μ and Σ are known. Then we consider the effect of parameter estimation. In our asymptotic scheme we take $m, n \to \infty$ in such a way that

$$\frac{m}{n} = \gamma + O\left(n^{-1}\right), \quad 0 < \gamma < 1. \tag{6}$$

Our results are thus equally valid both under (2), and under other practically sensible choices of m, such as $m = \lfloor n\gamma \rfloor + 1$ or $m = \lfloor (n + v + 1)\gamma \rfloor$. An important consequence of consistency is that the asymptotic distribution of the squared distances (5) is χ^2_v, like that of classical Mahalanobis distances. Furthermore, we show that the FS estimators, and thus also the squared distances (5), possess high-breakdown properties under contamination.

The structure of the paper is as follows. In Section 2 we obtain consistency of the FS estimators in the ideal situation where the parameters are known. The case of unknown parameters is dealt with in Section 3. Robustness is shown in Section 4, while the empirical performance of the FS estimators is investigated in Section 5. The paper ends with some concluding remarks and discussion of open issues in Section 6.

2. Case 1: fixed step with known parameters

We start our asymptotic analysis of the FS estimators (3) and (4) in the ideal case where the multivariate normal model M_0 holds with known μ and Σ. In this case, given a sample $y_1, \ldots, y_n$ of size n, progression in the FS would simply require to order the n squared (population) Mahalanobis distances

$$d^2_{i,n} = (y_i - \mu)' \Sigma^{-1} (y_i - \mu), \quad i = 1, \ldots, n, \tag{7}$$

which correspond to the deviance measures (5) when the population parameters are known. Under M_0 the squared distances (7) are i.i.d. according to a χ^2_v distribution for any sample size n.

In what follows, we denote by $F_{\chi^2_v}(x)$ the distribution function of a χ^2_v random variable. When dealing with vectors and matrices we assume convergence to be componentwise, which is equivalent to assuming convergence in the Euclidean (matrix) norm $\|\cdot\|$.

Fix a step $0 < \gamma < 1$. In the known parameter case the FS estimators at γ are given by expressions (3) and (4), with the general weights $w_{i,\gamma,n}$ taking the form

$$\tilde{w}_{i,\gamma,n} = I(d^2_{i,n} \le \delta^2_{\gamma,n}) \quad i = 1, \ldots, n, \tag{8}$$

where $I(\cdot)$ is the indicator function and $\delta^2_{\gamma,n}$ is the γ-th quantile among the n squared distances (7). Let $d^2_{(1),n}, d^2_{(2),n}, \ldots, d^2_{(n),n}$ be the order statistics of these squared distances. Note that, under M_0, $d^2_{(1),n} < d^2_{(2),n} < \cdots < d^2_{(n),n}$ with probability 1. We thus take

$$\delta^2_{\gamma,n} = d^2_{(m),n}, \tag{9}$$

where m is defined in (2). The scaling factor in the scatter estimate is given by

$$c_\gamma = \frac{\gamma}{F_{\chi^2_{v+2}}(\chi^2_{v,\gamma})}, \tag{10}$$

where $\chi^2_{v,\gamma}$ is the γ-th quantile of the χ^2_v distribution:

$$\chi^2_{v,\gamma} = F^{-1}_{\chi^2_v}(\gamma). \tag{11}$$

Let $\tilde{\mu}_{\gamma,n}$ and $\tilde{\Sigma}_{\gamma,n}$ be the FS estimators (3) and (4) when the weights are defined by (8). The purpose of this section is to show that $\tilde{\mu}_{\gamma,n}$ and $\tilde{\Sigma}_{\gamma,n}$ are strongly consistent estimators of μ and Σ, when $m, n \to \infty$ so that (6) holds. But, before then, we study the "idealized" estimators whose weights are based on the γ-th quantile of the χ^2_v distribution, i.e. the distribution of the squared distances (7) when M_0 holds. For the moment, we also use the known value of μ when estimating the scatter. These "idealized" estimators are thus defined as

$$\tilde{\mu}^*_{\gamma,n} = \sum_{i=1}^{n} \frac{\tilde{w}^*_{i,\gamma,n}}{\tilde{W}^*_{\gamma,n}} y_i \tag{12}$$

$$\tilde{\Sigma}^*_{\gamma,n} = c_\gamma \sum_{i=1}^{n} \frac{\tilde{w}^*_{i,\gamma,n}}{\tilde{W}^*_{\gamma,n}} (y_i - \mu)(y_i - \mu)', \tag{13}$$

where

$$\tilde{w}^*_{i,\gamma,n} = I(d^2_{i,n} \le \chi^2_{v,\gamma}) \tag{14}$$

and $\tilde{W}^*_{\gamma,n} = \sum_{i=1}^{n} \tilde{w}^*_{i,\gamma,n}$. In the "idealized" estimators $\tilde{\mu}^*_{\gamma,n}$ and $\tilde{\Sigma}^*_{\gamma,n}$, the sum $\tilde{W}^*_{\gamma,n}$ is a random variable, differently from what happens for $\tilde{\mu}_{\gamma,n}$ and $\tilde{\Sigma}_{\gamma,n}$.

We start with a preliminary Lemma and then provide a consistency result for $\tilde{\mu}^*_{\gamma,n}$ and $\tilde{\Sigma}^*_{\gamma,n}$. Our asymptotic properties are for $m, n \to \infty$ so that (6) holds.

Lemma 1. *Under M_0 with known μ and Σ,*

$$\frac{1}{n}\sum_{i=1}^{n} \tilde{w}^*_{i,\gamma,n} y_i \xrightarrow{a.s.} \gamma\mu$$

$$\frac{1}{n}\sum_{i=1}^{n} \tilde{w}^*_{i,\gamma,n} (y_i - \mu)(y_i - \mu)' \xrightarrow{a.s.} \gamma c_\gamma^{-1}\Sigma.$$

Proof. Under model M_0, the results on elliptical truncation [33,1,31] give

$$\mathrm{E}(y_i|\tilde{w}^*_{i,\gamma,n} = 1) = \mu \qquad \mathrm{Var}(y_i|\tilde{w}^*_{i,\gamma,n} = 1) = c_\gamma^{-1}\Sigma. \tag{15}$$

It then follows from the binary nature of $\tilde{w}^*_{i,\gamma,n}$ that

$$\mathrm{E}(\tilde{w}^*_{i,\gamma,n} y_i) = \gamma \mathrm{E}(y_i|\tilde{w}^*_{i,\gamma,n} = 1) = \gamma\mu, \tag{16}$$

and

$$\mathrm{E}\left\{\tilde{w}^*_{i,\gamma,n}(y_i - \mu)(y_i - \mu)'\right\} = \gamma \mathrm{Var}(y_i|\tilde{w}^*_{i,\gamma,n} = 1) = \gamma c_\gamma^{-1}\Sigma. \tag{17}$$

The random vectors $\tilde{w}^*_{i,\gamma,n} y_i$, $i = 1, \ldots, n$, are i.i.d. and satisfy the Kolmogorov condition for the SLLN [32, p. 27]. In fact, according to (17), the variance of $\tilde{w}^*_{i,\gamma,n} y_{ij}$ is finite for any $i = 1, \ldots, n$, $j = 1, \ldots, v$. Let C be such that $\max_j \mathrm{Var}\left(\tilde{w}^*_{i,\gamma,n} y_{ij}\right) \le C < \infty$. Then, we have

$$\lim_{n\to\infty} \sum_{i=1}^{n} \frac{\mathrm{Var}\left(\tilde{w}^*_{i,\gamma,n} y_{ij}\right)}{i^2} \le C \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty,$$

from which the first statement of this lemma follows. To prove the second statement, we only need to verify the Kolmogorov condition for the elements of the i.i.d. vectors $\tilde{w}^*_{i,\gamma,n} y_i y_i'$, $i = 1, \ldots, n$. This condition holds because $\mathrm{E}[\tilde{w}^*_{i,\gamma,n} y^2_{ij_1} y^2_{ij_2}]$ is finite for any $j_1, j_2 = 1, \ldots, v$, again as a consequence of results on elliptical truncation under model M_0 [33]. □

Theorem 1. *Under M_0 with known μ and Σ, the estimators $\tilde{\mu}^*_{\gamma,n}$ and $\tilde{\Sigma}^*_{\gamma,n}$ are strongly consistent, i.e.*

$$\tilde{\mu}^*_{\gamma,n} \xrightarrow{a.s.} \mu$$
$$\tilde{\Sigma}^*_{\gamma,n} \xrightarrow{a.s.} \Sigma,$$

when $m, n \to \infty$ so that (6) holds.

Proof. Consider the location estimator. First, note that

$$\tilde{\mu}^*_{\gamma,n} = \left(\frac{n}{\tilde{W}^*_{\gamma,n}}\right)\left(\frac{1}{n}\sum_{i=1}^{n} \tilde{w}^*_{i,\gamma,n} y_i\right). \tag{18}$$

Given that the distribution of $\tilde{W}^*_{\gamma,n}$ is Binomial, $\tilde{W}^*_{\gamma,n} \sim \mathcal{B}(n,\gamma)$, we have

$$\frac{n}{\tilde{W}^*_{\gamma,n}} \xrightarrow{a.s.} \frac{1}{\gamma}.$$

It then follows from Lemma 1 and the continuous mapping theorem [32, p. 24] that

$$\tilde{\mu}^*_{\gamma,n} \xrightarrow{a.s.} \mu. \tag{19}$$

A similar reasoning can be used to show that

$$\tilde{\Sigma}^*_{\gamma,n} = \frac{nc_\gamma}{\tilde{W}^*_{\gamma,n}}\frac{1}{n}\sum_{i=1}^{n} \tilde{w}^*_{i,\gamma,n}(y_i-\mu)(y_i-\mu)' \xrightarrow{a.s.} \Sigma. \quad \square$$

We now derive a result of independent interest, which will also be useful in Lemma 2 that follows. We show that the FS weight of each unit at step γ, i.e. $\tilde{w}_{i,\gamma,n}$ in (8), converges to the weight $\tilde{w}^*_{i,\gamma,n}$ defined through the γ-th quantile of the χ^2_v distribution. In this way the standard fitting subset of the FS, which is bounded by an empirical quantile of the squared Mahalanobis distances, is shown to be asymptotically equivalent to a fitting subset bounded by an ellipsoid.

Theorem 2. *Under M_0 with known μ and Σ, for any $i = 1, \ldots, n$, the sequence of weights for observation y_i converges a.s. to the weight (14):*

$$|\tilde{w}_{i,\gamma,n} - \tilde{w}^*_{i,\gamma,n}| \xrightarrow{a.s.} 0$$

when $m, n \to \infty$ so that (6) holds.

Proof. Recall that

$$\tilde{w}_{i,\gamma,n} = I(d^2_{i,n} \le \delta^2_{\gamma,n}) \quad i = 1, \ldots, n,$$

where $\delta^2_{\gamma,n}$ is the empirical quantile (9). Since the squared distances are an i.i.d. sample from the χ^2_v distribution, it is then a standard result [32, p. 75] that

$$\delta^2_{\gamma,n} \xrightarrow{a.s.} \chi^2_{v,\gamma} \tag{20}$$

when $m, n \to \infty$ so that (6) holds. Given (20), by definition

$$P(\lim|\delta^2_{\gamma,n} - \chi^2_{v,\gamma}| > 0) = 0. \tag{21}$$

We can now study

$$P(\lim|\tilde{w}_{i,\gamma,n} - \tilde{w}^*_{i,\gamma,n}| > 0).$$

The latter corresponds, for any $i = 1, \ldots, n$, to

$$P(\lim I(\delta^2_{\gamma,n} \ne \chi^2_{v,\gamma})),$$

which is equal to (21). Therefore, for any $i = 1, \ldots, n$,

$$P(\lim|\tilde{w}_{i,\gamma,n} - \tilde{w}^*_{i,\gamma,n}| = 0) = 1. \quad \square$$

The following lemma extends the results on elliptical truncation of [33].

Lemma 2. *Under M_0 with known μ and Σ, when $m, n \to \infty$ so that (6) holds,*

$$\mathrm{E}(y_i|\tilde{w}_{i,\gamma,n} = 1) = \mu_n \to \mu \tag{22}$$

$$\mathrm{Var}(y_i|\tilde{w}_{i,\gamma,n} = 1) = \Sigma_n \to c_\gamma^{-1}\Sigma. \tag{23}$$

Furthermore, all moments of $y_i\tilde{w}_{i,\gamma,n}$ are finite.

Proof. Statements (22) and (23) follow from (20) and the continuous mapping theorem, since

$$\mathrm{E}(y_i|d^2_{i,n} \le \delta^2_{\gamma,n}) \to \mathrm{E}(y_i|d^2_{i,n} \le \chi^2_{v,\gamma}) = \mu$$

and

$$\mathrm{Var}(y_i|d^2_{i,n} \le \delta^2_{\gamma,n}) \to \mathrm{Var}(y_i|d^2_{i,n} \le \chi^2_{v,\gamma}) = c_\gamma^{-1}\Sigma.$$

The discontinuity in the conditional moments of y_i, as a function of $\delta^2_{\gamma,n}$, has zero probability measure at the limit, and thus does not violate the hypotheses of the continuous mapping theorem.

To see that all moments of $y_i\tilde{w}_{i,\gamma,n}$ are finite, note that under M_0 all moments of y_i are finite. We have that for any $\kappa_j \in \mathbb{N}$, $j = 1, \ldots, v$,

$$\mathrm{E}\left\{\prod_{j=1}^{v}(y_{ij}\tilde{w}_{i,\gamma,n})^{\kappa_j}\right\} = \mathrm{E}\left(\prod_{j=1}^{v} y_{ij}^{\kappa_j}\tilde{w}_{i,\gamma,n}\right) = \mathrm{E}\left(\prod_{j} y_{ij}^{\kappa_j}|\tilde{w}_{i,\gamma,n} = 1\right)P(\tilde{w}_{i,\gamma,n} = 1).$$

The result then follows from the fact that conditioning to $\tilde{w}_{i,\gamma,n} = 1$ restricts the integrals involved in the cross-moments of the components of y_i to be computed over a bounded support. □

It is now possible to show consistency of the FS location and scatter estimators when the parameters are known. This result is obtained by expressing the FS estimators as the product of two factors converging almost surely, along the lines of the proof of Theorem 1. The main difference is that the sum $\tilde{W}_{\gamma,n}$ is now fixed by design.

Theorem 3. *Under M_0 with known μ and Σ, when $m, n \to \infty$ so that (6) holds the FS estimators of location and scatter, $\tilde{\mu}_{\gamma,n}$ and $\tilde{\Sigma}_{\gamma,n}$, are strongly consistent.*

Proof. By Lemma 2

$$\mathrm{E}(\tilde{w}_{i,\gamma,n}y_i) \to \gamma\mu,$$

and

$$\mathrm{E}\{\tilde{w}_{i,\gamma,n}(y_i - \mu)(y_i - \mu)'\} \to \gamma c_\gamma^{-1}\Sigma.$$

For what concerns the location estimator, we have that

$$\tilde{\mu}_{\gamma,n} = \left(\frac{n}{\tilde{W}_{\gamma,n}}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{w}_{i,\gamma,n}y_i\right). \tag{24}$$

It follows from (9) that

$$\tilde{W}_{\gamma,n} = \sum_{i=1}^{n}\tilde{w}_{i,\gamma,n} = m. \tag{25}$$

Therefore, when $m, n \to \infty$ and (6) holds

$$\frac{n}{\tilde{W}_{\gamma,n}} \to \frac{1}{\gamma}.$$

For what concerns the second factor in (24), the random variables $\tilde{w}_{i,\gamma,n}y_i$ are now dependent because of constraint (25). Instead of the classical Kolmogorov condition, we thus apply Theorem 2 of [9], which requires two conditions to establish the SLLN for a sequence of dependent random variables.

To state the first, denote with J_s, $s = 1, 2$, two arbitrary subsets of $\{1, \ldots, n\}$ such that, without loss of generality, for any $i_1 \in J_1$ and $i_2 \in J_2$, we have $i_1 < i_2$, and for which

$$\min\{i : i \in J_2\} - \max\{i : i \in J_1\} \ge k. \tag{26}$$

Let a_s, $s = 1, 2$ denote two arbitrary vectors, each of dimension $|J_s|$, whose sum of squares is finite, and

$$L_s(a_s) = \sum_{i \in J_s} a_{si}\tilde{w}_{i,\gamma,n}y_i. \tag{27}$$

According to Bryc and Smolenski [9], we need to show that there exists a finite $k > 0$ such that when $n \to \infty$

$$\sup_{a_1,a_2} \operatorname{cor}(L_1(a_1), L_2(a_2)) < 1. \tag{28}$$

To see (28), note that (26) implies that $k-1$ elements of the vector $\tilde{w}_{i,\gamma,n}y_i$ are removed before computing (28). Condition (28) does not hold if it is possible to predict any observation indexed in J_2 with a linear combination of observations indexed in J_1. Given that y_i comes from M_0 and that for $0 < \gamma < 1$

$$P(\tilde{w}_{i,\gamma,n} = 1) > 0 \quad i = 1, \ldots, n,$$

even if weights $\tilde{w}_{i,\gamma,n}$ are dependent due to constraint (25) it is not possible to exactly predict any $\tilde{w}_{i,\gamma,n}y_i$. Therefore, (28) holds with $k = 1$.

The second condition in Theorem 2 of [9] is the extended Kolmogorov condition

$$\sum_{i=1}^{\infty} \frac{\operatorname{Var}(\tilde{w}_{i,\gamma,n}y_{ij})}{i^{3/2}} < \infty, \quad \forall j = 1, \ldots, v. \tag{29}$$

To see (29), let C' be such that $\max_j \operatorname{Var}(\tilde{w}_{i,\gamma,n}y_{ij}) \le C'$. Since $C' < \infty$ asymptotically, as a consequence of Lemma 2,

$$\lim_n \sum_{i=1}^{n} \frac{\operatorname{Var}(\tilde{w}_{i,\gamma,n}y_{ij})}{i^{3/2}} \le C' \sum_{i=1}^{\infty} \frac{1}{i^{3/2}} < \infty.$$

Therefore, we obtain that

$$\frac{1}{n}\sum_{i=1}^{n} \tilde{w}_{i,\gamma,n}y_i \xrightarrow{a.s.} \gamma\mu,$$

from which strong consistency of $\tilde{\mu}_{\gamma,n}$ follows.

A similar proof can be used to show consistency of $\tilde{\Sigma}_{\gamma,n}$. One difference is that higher order moments are involved in the above conditions, and those are finite as well due to distributional assumptions on y_i and Lemma 2. Finally, we use the continuous mapping theorem to show that $y_i - \tilde{\mu}_{\gamma,n}$ gets closer and closer to $y_i - \mu$:

$$\|y_i - \tilde{\mu}_{\gamma,n} - (y_i - \mu)\| = \|\tilde{\mu}_{\gamma,n} - \mu\| \xrightarrow{a.s.} 0, \tag{30}$$

because $\tilde{\mu}_{\gamma,n}$ is strongly consistent. Therefore, the a.s. consistency result for

$$c_\gamma \sum_{i=1}^{n} \frac{\tilde{w}_{i,\gamma,n}}{\tilde{W}_{\gamma,n}} (y_i - \mu)(y_i - \mu)'$$

holds for $\tilde{\Sigma}_{\gamma,n}$ as well. □

A consequence of Theorem 3 is that

$$\operatorname{Var}(\tilde{\mu}_{\gamma,n}) \to 0 \qquad \operatorname{Var}(\tilde{\Sigma}_{\gamma,n}) \to 0, \tag{31}$$

and the same applies for the idealized estimators as a consequence of Theorem 1. We can also show that the FS and the idealized estimators get closer and closer as the sample size grows.

Corollary 1. *For fixed γ and $m, n \to \infty$ in such a way that (6) holds*

$$\|\tilde{\mu}_{\gamma,n} - \tilde{\mu}^*_{\gamma,n}\| \xrightarrow{a.s} 0$$
$$\|\tilde{\Sigma}_{\gamma,n} - \tilde{\Sigma}^*_{\gamma,n}\| \xrightarrow{a.s.} 0.$$

Proof. This is a consequence of Theorems 1 and 3, given that both estimators are strongly consistent and converge to the same limit. □

We now derive some results concerning the generalized Mahalanobis squared distances

$$\tilde{d}^2_{i,\gamma,n} = (y_i - \tilde{\mu}_{\gamma,n})'\tilde{\Sigma}^{-1}_{\gamma,n}(y_i - \tilde{\mu}_{\gamma,n}) \quad i = 1, \ldots, n, \tag{32}$$

which specialize the general formula (5) to the case of estimators $\tilde{\mu}_{\gamma,n}$ and $\tilde{\Sigma}_{\gamma,n}$. These results are important because they show that the distances $\tilde{d}^2_{i,\gamma,n}$ obtained from the FS estimators are asymptotically equivalent to the population Mahalanobis distances (7), even if they are not independent for $i = 1, \ldots, n$. Furthermore, they provide key ingredients for the case of unknown parameters, to be addressed in the next section.

Theorem 4. *Under M_0 with known μ and Σ, for any $i = 1, \ldots, n$, the sequence of squared distances for observation y_i converges a.s. to the population Mahalanobis distance defined in Eq. (7):*

$$|\tilde{d}^2_{i,\gamma,n} - d^2_{i,n}| \xrightarrow{a.s.} 0$$

when $m, n \to \infty$ so that (6) holds.

Proof. See the Appendix. $\square$

Theorem 5. *Assume that model M_0 holds with known μ and Σ. Let*

$$\tilde{F}_{\gamma,n}(x) = \frac{1}{n}\sum_{i=1}^{n} I(\tilde{d}^2_{i,\gamma,n} \leq x) \tag{33}$$

be the empirical distribution function of the generalized Mahalanobis squared distances (32). Then, for each $x > 0$

$$\tilde{F}_{\gamma,n}(x) \xrightarrow{a.s.} F_{\chi^2_v}(x)$$

when $m, n \to \infty$ so that (6) holds.

Proof. See the Appendix. $\square$

We conclude this section by stating two immediate and useful consequences of Theorems 4 and 5.

Corollary 2. *Under M_0 with known μ and Σ, for any $i = 1, \ldots, n$,*

$$\tilde{d}^2_{i,\gamma,n} \xrightarrow{d} \chi^2_v.$$

Proof. The result is a straightforward implication of Theorem 4. $\square$

Corollary 3. *Let $\tilde{d}^2_{(1),\gamma,n}, \tilde{d}^2_{(2),\gamma,n}, \ldots, \tilde{d}^2_{(n),\gamma,n}$ be the order statistics of the squared distances (32). Take $0 < \gamma^* < 1$ and*

$$\tilde{\delta}^2_{\gamma^*,\gamma,n} = \tilde{d}^2_{(m^*),\gamma,n}, \tag{34}$$

with $m^* = \lfloor n\gamma^* \rfloor$ as in (2). Under M_0 with known μ and Σ,

$$\tilde{\delta}^2_{\gamma^*,\gamma,n} \xrightarrow{a.s.} \chi^2_{v,\gamma^*}, \tag{35}$$

where χ^2_{v,γ^*} is the γ^*-th quantile of the χ^2_v distribution.

Proof. The result follows from Theorem 5 and the fact that χ^2_{v,γ^*} is the unique solution of $F_{\chi^2_v}(x-) \leq \gamma^* \leq F_{\chi^2_v}(x)$ for any $0 < \gamma^* < 1$. $\square$

3. Case 2: fixed step with unknown parameters

We are now in the position to establish the properties of the FS estimators in the case where the model parameters are unknown. For this purpose, we repeat most of the steps described in Section 2, with the classical convergence result (20) replaced by (35) or by similar results. In the proofs that follow we thus only highlight the differences with respect to Section 2.

Given the adaptive nature of the FS, when the parameters in (1) are unknown their estimators at step γ must be based on the results of a previous step. As a consequence, we fix γ_0 and γ such that $0 < \gamma_0 < \gamma < 1$. We also take

$$m_0 = \lfloor n\gamma_0 \rfloor \qquad m = \lfloor n\gamma \rfloor, \tag{36}$$

as in (2). Let $\hat{\mu}_{\gamma_0,n}$ and $\hat{\Sigma}_{\gamma_0,n}$ be the estimators of the unknown parameters μ and Σ at step γ_0. We make the following assumption.

Assumption 1. Under M_0 with unknown μ and Σ,

$$\hat{\mu}_{\gamma_0,n} \xrightarrow{a.s.} \mu \qquad \hat{\Sigma}_{\gamma_0,n} \xrightarrow{a.s.} \Sigma, \tag{37}$$

when $m_0, n \to \infty$ so that $m_0/n \to \gamma_0$.

In the setting of the FS, γ_0 may represent either the initial step, so that $m^* = m_0$, or any step prior to γ. Therefore, given a sequence of steps $0 < \gamma_0 < \gamma_1 < \cdots < \gamma_t < 1$, we can iterate the arguments of this section to show that consistency at step γ_{t_1} implies consistency at step γ_{t_2}, if $t_1 < t_2$. The only substantial implication of Assumption 1 is the requirement of strong consistency for the estimators adopted in the initial step.

Note that García-Escudero and Gordaliza [19] assume weak convergence of $\hat{\mu}_{\gamma_0,n}$ and $\hat{\Sigma}_{\gamma_0,n}$ with rate $n^{-1/2}$, when $m_0 = m^*$, in order to derive the asymptotic distribution of the radius process. This process is defined through the estimators $\hat{\mu}_{\gamma_0,n}$ and $\hat{\Sigma}_{\gamma_0,n}$ for any trimming level $1-\gamma$, even if $\gamma \neq \gamma_0$. For our purposes, Assumption 1 suffices with no requirements on the rate. It thus also includes estimators with lower convergence rates, such as the Minimum Volume Ellipsoid. Our contribution is to show that the strong consistency property (37) is maintained for any sequence of subsequent steps $\gamma_0 < \gamma_1 < \gamma_2 \ldots$ of the FS, when new step-specific estimators are obtained.

The assumption of consistent initial estimators, with rate $n^{-1/2}$, is also made by Johansen and Nielsen [23] for the FS in the regression context. This assumption is refined by Johansen and Nielsen [24], who show that a somewhat weaker condition is sufficient for convergence of the univariate process defined by the FS model residuals. We may expect that similar conditions on the rate of convergence of the initial estimators are needed in order to show weak convergence also in the multivariate framework of this paper. However, an advantage of Assumption 1, which does not put any condition on the rate of convergence, is to broaden the class of potential initial estimators. This point is important in practice, where the use of computationally intensive estimators, like the MCD, as a starting point for the FS might be too much time-consuming. Even more important is the fact that the diagnostic power of the FS is enhanced when $m^* \ll n$. In this case simple and consistent estimators, but with unknown converge rate, might provide the only feasible starting point. Indeed, if we assume no contamination so that model (1) holds for all the data, many robust estimators may be shown to be still consistent if $m^*/n \to \infty$, but their convergence rates may be more difficult to derive. A relevant application arises when the FS for multivariate data is applied for the purpose of detecting unknown group structures in the data. In this case the FS is run many times from different small starting subsets chosen at random [7], so that computational issues may become crucial. Another useful consequence of Assumption 1 is that under this umbrella we can potentially use different estimators at subsequent steps of the FS. For instance, we could legitimately alternate steps of the FS where the estimators are computed following the standard formulae (3) and (4), with weights defined as in (39) below, with steps where μ and Σ are estimated in a different way, e.g. by robust S-estimators or by the Minimum Volume Ellipsoid, or even where the estimates are kept constant, like in the radius process of [19].

If μ and Σ are unknown, progression in the FS from γ_0 to γ is based on the generalized Mahalanobis squared distances

$$\hat{d}^2_{i,\gamma_0,n} = (y_i - \hat{\mu}_{\gamma_0,n})' \hat{\Sigma}^{-1}_{\gamma_0,n} (y_i - \hat{\mu}_{\gamma_0,n}) \quad i = 1, \ldots, n. \tag{38}$$

Therefore, the FS estimators at step γ are given by expressions (3) and (4), with the weights $w_{i,\gamma,n}$ now defined as

$$\hat{w}_{i,\gamma,\gamma_0,n} = I(\hat{d}^2_{i,\gamma_0,n} \le \hat{\delta}^2_{\gamma,\gamma_0,n}) \quad i = 1, \ldots, n, \tag{39}$$

where $\hat{\delta}^2_{\gamma,\gamma_0,n}$ is the γ-th quantile among the n squared distances (38). Following (9), we take

$$\hat{\delta}^2_{\gamma,\gamma_0,n} = \hat{d}^2_{(m),\gamma_0,n},$$

where $\hat{d}^2_{(1),\gamma_0,n}, \hat{d}^2_{(2),\gamma_0,n}, \ldots, \hat{d}^2_{(n),\gamma_0,n}$ are the order statistics of the squared distances (38), and m is defined as in (2) and in (36). The FS estimators with weights (39) are denoted by $\hat{\mu}_{\gamma,\gamma_0,n}$ and $\hat{\Sigma}_{\gamma,\gamma_0,n}$. The scaling factor c_γ in $\hat{\Sigma}_{\gamma,\gamma_0,n}$ is the same as in $\tilde{\Sigma}_{\gamma,n}$, and is given by (10).

We start by summarizing some useful properties of the generalized Mahalanobis squared distances $\hat{d}^2_{i,\gamma_0,n}$. These properties arise from Assumption 1. They specialize Theorems 4 and 5, and the subsequent Corollaries, to a form which is directly applicable to the FS estimators $\hat{\mu}_{\gamma,\gamma_0,n}$ and $\hat{\Sigma}_{\gamma,\gamma_0,n}$.

Lemma 3. *Under M_0 with unknown μ and Σ, when $m_0, n \to \infty$ so that $m_0/n \to \gamma_0$, the following holds if Assumption 1 is true:*

(a) *for any $i = 1, \ldots, n$, the sequence of squared distances for observation y_i converges a.s. to the population Mahalanobis distance defined in Eq. (7):*

$$|\hat{d}^2_{i,\gamma_0,n} - d^2_{i,n}| \xrightarrow{a.s.} 0;$$

(b) *if $\hat{F}_{\gamma_0,n}(x) = \frac{1}{n}\sum_{i=1}^n I(\hat{d}^2_{i,\gamma_0,n} \le x)$, then*

$$\hat{F}_{\gamma_0,n}(x) \xrightarrow{a.s.} F_{\chi^2_v}(x), \quad \textit{for each } x > 0;$$

(c) *for any $i = 1, \ldots, n$,*

$$\hat{d}^2_{i,\gamma_0,n} \xrightarrow{d} \chi^2_v;$$

d) *if* $0 < \gamma^* < 1$, $\hat{\delta}^2_{\gamma^*,\gamma_0,n} = \hat{d}^2_{(m^*),\gamma_0,n}$ *and* $m^* = \lfloor n\gamma^* \rfloor$ *as in* (36), *then*

$$\hat{\delta}^2_{\gamma^*,\gamma_0,n} \xrightarrow{a.s.} \chi^2_{v,\gamma^*}.$$

Proof. Replace γ with γ_0, $\tilde{\mu}_{\gamma,n}$ and $\tilde{\Sigma}_{\gamma,n}$ with $\hat{\mu}_{\gamma_0,n}$ and $\hat{\Sigma}_{\gamma_0,n}$, $\tilde{d}^2_{i,\gamma,n}$ with $\hat{d}^2_{i,\gamma_0,n}$, in the proofs of Theorems 4 and 5, and in those of Corollaries 2 and 3. □

The next two results consider the asymptotic properties of the empirical weights $\hat{w}_{i,\gamma,\gamma_0,n}$, $i = 1, \ldots, n$, and their use in elliptical truncation.

Theorem 6. *Under Assumption 1 and model M_0 with unknown μ and Σ, for any $i = 1, \ldots, n$, the sequence of empirical weights for observation y_i converges a.s. to the weight (14):*

$$|\hat{w}_{i,\gamma,\gamma_0,n} - \tilde{w}^*_{i,\gamma,n}| \xrightarrow{a.s.} 0$$

when $m, m_0, n \to \infty$ so that $m/n \to \gamma$ and $m_0/n \to \gamma_0$.

Proof. Repeat the proof of Theorem 2 with $\hat{d}^2_{i,\gamma_0,n}$ in place of $d^2_{i,n}$, $i = 1, \ldots, n$, and convergence (20) replaced by Lemma 3(d) with $\gamma^* = \gamma$. □

Lemma 4. *Under Assumption 1 and model M_0 with unknown μ and Σ, if $m, m_0, n \to \infty$ so that $m/n \to \gamma$ and $m_0/n \to \gamma_0$,*

$$\mathrm{E}(y_i|\hat{w}_{i,\gamma,\gamma_0,n} = 1) \to \mu$$
$$\mathrm{Var}(y_i|\hat{w}_{i,\gamma,\gamma_0,n} = 1) \to c_\gamma^{-1}\Sigma.$$

Furthermore, all moments of $y_i\hat{w}_{i,\gamma,\gamma_0,n}$ are finite.

Proof. The first two statements follow from Lemma 3(d) and the continuous mapping theorem. For the last statement, repeat the proof of Lemma 2 with $\hat{w}_{i,\gamma,\gamma_0,n}$ in place of $\tilde{w}_{i,\gamma,n}$. □

It is now possible to show consistency of the FS location and scatter estimators $\hat{\mu}_{\gamma,\gamma_0,n}$ and $\hat{\Sigma}_{\gamma,\gamma_0,n}$, along the lines of Theorem 3.

Theorem 7. *Under Assumption 1 and model M_0 with unknown μ and Σ, if $m, m_0, n \to \infty$ so that $m/n \to \gamma$ and $m_0/n \to \gamma_0$, the FS estimators of location and scatter $\hat{\mu}_{\gamma,\gamma_0,n}$ and $\hat{\Sigma}_{\gamma,\gamma_0,n}$ are strongly consistent.*

Proof. Repeat the proof of Theorem 3 with $\hat{w}_{i,\gamma,\gamma_0,n}$ in place of $\tilde{w}_{i,\gamma,n}$. Note that the additional source of dependence among the weights $\hat{w}_{i,\gamma,\gamma_0,n}$ implied by parameter estimation does not affect the applicability of Theorem 2 of [9]. In fact, Assumption 1 ensures that, for any pair of linear combinations

$$\hat{L}_s(a_s) = \sum_{i\in J_s} a_{si}\hat{w}_{i,\gamma,\gamma_0,n}y_i \quad s = 1, 2$$

such that $\sum_{i\in J_s} a^2_{si} < \infty$,

$$\mathrm{cor}(\hat{L}_1(a_1), \hat{L}_2(a_2)) \to \mathrm{cor}(L_1(a_1), L_2(a_2)),$$

where $L_1(a_1)$ and $L_2(a_2)$ are the corresponding linear combinations when the parameters are known, as given in (27). Therefore, we still have asymptotically

$$\sup_{a_1,a_2} \mathrm{cor}(L_1(a_1), L_2(a_2)) < 1$$

for a finite $k > 0$. □

We note that, thanks to Theorem 7, also the squared distances

$$\hat{d}^2_{i,\gamma,\gamma_0,n} = (y_i - \hat{\mu}_{\gamma,\gamma_0,n})'\hat{\Sigma}^{-1}_{\gamma,\gamma_0,n}(y_i - \hat{\mu}_{\gamma,\gamma_0,n}) \quad i = 1, \ldots, n \tag{40}$$

share the good asymptotic properties given in Lemma 3, when $m, n \to \infty$ so that $m/n \to \gamma$. This extension of Lemma 3 is straightforward to derive and is not detailed here. Nevertheless, it is crucial both for progressing in the search, when moving from γ to a subsequent step, and for providing formal justification of the use of distances (40) for the detection of outliers from model M_0 [30].

We conclude this section by discussing two issues that outline possible extensions of our work. Our first remark is that the consistency properties derived in this paper can be seen as a first step towards studying the weak convergence of the empirical process defined by $\hat{\mu}_{\gamma,\gamma_0,n}$, as indexed by $\gamma \in (0, 1]$, and similarly for $\hat{\Sigma}_{\gamma,\gamma_0,n}$. For this purpose, we would need

- distributional convergence for each joint distribution over $\gamma_1, \ldots, \gamma_T$, where T is finite;
- asymptotic equicontinuity.

The latter is shown in [24] for the FS estimator in a univariate regression setting. Extension to the multivariate case is straightforward, since condition (6) ensures that

$$\hat{\mu}_{\gamma_{t_1},\gamma_0,n} = \hat{\mu}_{\gamma_{t_2},\gamma_0,n} \qquad \hat{\Sigma}_{\gamma_{t_1},\gamma_0,n} = \hat{\Sigma}_{\gamma_{t_2},\gamma_0,n},$$

if $|\gamma_{t_1} - \gamma_{t_2}| < 1/n$. Our estimators are thus stepwise constant, from which equicontinuity follows. On the other hand, when $\gamma_{t_2} > \gamma_{t_1}$, there is a steady dependence between estimators obtained with γ_{t_2} on those obtained with γ_{t_1}, making it hard to study the behavior of finite dimensional distributions. Uniform convergence over the intervals of continuity would be implied by weak convergence. Another possibility is to establish uniform convergence directly. In this case we would anyway need results about rates of convergence, which unfortunately do not seem straightforward to obtain in this context.

Our second remark is that we have stated the results relying on the multivariate normal model (1), given that we have explicit expressions of all the relevant quantities in this setting. We speculate that our results could be extended to elliptically contoured distributions with minor adjustments. Indeed, the analytic formula for the scaling factor of the scatter estimate, which is given by expression (10) at the normal model, is also available for the general case of elliptically contoured distributions [14, p. 165]. What is still unclear is the link between this general expression and results on elliptical truncation under non-normal models. For instance, properties like (15), which play a crucial role in our framework, to the best of our knowledge have been analytically derived only for the multivariate normal distribution. An important research goal to be pursued in the future would thus be the systematic study of the effect of truncation under non-normal distributional models, which would provide the missing link.

4. Global robustness of the estimators

We now give a global robustness property of the FS estimators $\hat{\mu}_{\gamma,\gamma_0,n}$ and $\hat{\Sigma}_{\gamma,\gamma_0,n}$. In particular, we show that the maximum proportion of nasty outliers that the estimators at step γ can tolerate is $1-\gamma$. This proportion is given by the finite sample breakdown point [26], which for the location estimator $\hat{\mu}_{\gamma,\gamma_0,n}$ and the sample $y = \{y_1, \ldots, y_n\}$ is defined as

$$\epsilon(\hat{\mu}_{\gamma,\gamma_0,n}, y) = \min_{m^* \le n^* \le n} \left\{ \frac{n^*}{n} : \sup \|\hat{\mu}_{\gamma,\gamma_0,n} - \hat{\mu}^{(n^*)}_{\gamma,\gamma_0,n}\| = \infty \right\}. \tag{41}$$

In (41), $\hat{\mu}^{(n^*)}_{\gamma,\gamma_0,n}$ is the FS estimator of μ computed on a modified sample $y^{(n^*)}$ in which a subset of n^* among the original n observations is replaced by arbitrary values, the supremum is taken over all possible modified samples $y^{(n^*)}$ and $\|\|$ is the Euclidean norm, as before. When considering the robustness of the estimator of scatter a different norm is required:

$$D(\hat{\Sigma}_{\gamma,\gamma_0,n}; \hat{\Sigma}^{(n^*)}_{\gamma,\gamma_0,n}) = \max \left\{ |\lambda_1(\hat{\Sigma}_{\gamma,\gamma_0,n}) - \lambda_1(\hat{\Sigma}^{(n^*)}_{\gamma,\gamma_0,n})|, |\lambda_v(\hat{\Sigma}_{\gamma,\gamma_0,n})^{-1} - \lambda_v(\hat{\Sigma}^{(n^*)}_{\gamma,\gamma_0,n})^{-1}| \right\},$$

where $\hat{\Sigma}^{(n^*)}_{\gamma,\gamma_0,n}$ is the FS estimator of Σ computed on the modified sample $y^{(n^*)}$ and $\lambda_j(\cdot), j = 1, \ldots, v$, is the jth largest eigenvalue of the corresponding matrix. The finite sample breakdown point of $\hat{\Sigma}_{\gamma,\gamma_0,n}$ at y is

$$\epsilon(\hat{\Sigma}_{\gamma,\gamma_0,n}, y) = \min_{m^* \le n^* \le n} \left\{ \frac{n^*}{n} : \sup D(\hat{\Sigma}_{\gamma,\gamma_0,n}; \hat{\Sigma}^{(n^*)}_{\gamma,\gamma_0,n}) = \infty \right\}. \tag{42}$$

We suppose that the sample y is in general position, which means that no subset of $v+1$ observations from y lies in a hyperplane of dimension smaller than v. This condition is verified with probability 1 under Model M_0. Furthermore, we need an additional assumption about the estimators computed at step γ_0.

Assumption 2. Take $m_0 \ge \lfloor (n+v+1)/2 \rfloor$ and let $\gamma_0 = m_0/n$. We assume that

$$\min \left\{ \epsilon(\hat{\mu}_{\gamma_0,n}, y), \epsilon(\hat{\Sigma}_{\gamma_0,n}, y) \right\} \ge 1 - \gamma_0.$$

This assumption intuitively means that at step m_0 we estimate both location and scatter with a degree of robustness which is not smaller than $1-\gamma_0$. Since $1-\gamma_0$ represents the proportion of points trimmed by the FS at step m_0, Assumption 2 implies that gross outliers do not enter the fitting subset of the FS unless their number is very close to m_0. In fact, if $m_0 = \lfloor (n+v+1)/2 \rfloor$, the value $1-\gamma_0$ is the upper bound for the breakdown point of any affine equivariant estimator of Σ [17]. Therefore, $\hat{\mu}_{\gamma,\gamma_0,n}, y$ and $\hat{\Sigma}_{\gamma,\gamma_0,n}, y$ will not be affected by such gross outliers if Assumption 2 holds. This condition might thus be seen as the "robust" complement to Assumption 1 under contamination in the modified sample $y^{(n^*)}$.

Theorem 8. *Let $m > m_0$ and $\gamma = m/n$. Under Assumption 2, if y is in general position*

$$\epsilon(\hat{\mu}_{\gamma,\gamma_0,n}, y) = \epsilon(\hat{\Sigma}_{\gamma,\gamma_0,n}, y) = 1 - \gamma.$$

Proof. Without loss of generality, suppose that contamination takes place in the first n^* units. We thus write the modified sample as

$$y^{(n^*)} = \{y_1^*, \ldots, y_{n^*}^*, y_{n^*+1}, \ldots, y_n\},$$

where $y_{i^*}^*$, $i^* = 1, \ldots, n^*$, denotes an arbitrary value in the contaminated subset. For what concerns the location estimator, we have that

$$\sup \|\hat{\mu}_{\gamma,\gamma_0,n} - \hat{\mu}^{(n^*)}_{\gamma,\gamma_0,n}\| = \infty \quad \text{iff} \quad \|y_{i^*}^*\| = \infty \tag{43}$$

for at least one contaminated unit i^* for which $\hat{w}_{i^*,\gamma,\gamma_0,n} = 1$, since $\sum_i \hat{w}_{i,\gamma,\gamma_0,n} = m > 0$ by design. Condition (43) implies that

$$\hat{d}^2_{i^*,\gamma_0,n} = \infty \tag{44}$$

for these contaminated units, thanks to Assumption 2. Write $n^*_\infty \leq n^*$ for the number of units for which (44) holds. By definition (39), we obtain that the corresponding weights

$$\hat{w}_{i^*,\gamma,\gamma_0,n} = I(\hat{d}^2_{i^*,\gamma_0,n} \leq \hat{\delta}^2_{\gamma,\gamma_0,n}) = I(\hat{d}^2_{i^*,\gamma_0,n} \leq \hat{d}^2_{(m),\gamma_0,n}) = 0 \quad i^* = 1, \ldots, n^*_\infty,$$

unless $n^*_\infty \geq n-m$. Letting (44) hold for all contaminated units, i.e. considering the worst case $n^* = n^*_\infty$, yields the statement on $\epsilon(\hat{\mu}_{\gamma,\gamma_0,n}, y)$.

A similar argument can also be applied to obtain $\epsilon(\hat{\Sigma}_{\gamma,\gamma_0,n}, y)$. In fact, for any v-dimensional vector $b \neq 0$,

$$\lambda_1(\hat{\Sigma}_{\gamma,\gamma_0,n}) = \sup_b \frac{b'\hat{\Sigma}^{-1}_{\gamma,\gamma_0,n}b}{b'b}.$$

Therefore, under Assumption 2,

$$\sup\left\{|\lambda_1(\hat{\Sigma}_{\gamma,\gamma_0,n}) - \lambda_1(\hat{\Sigma}^{(n^*)}_{\gamma,\gamma_0,n})|\right\} = \infty$$

only if $\|y_{i^*}^*\| = \infty$ for at least one unit for which $\hat{w}_{i^*,\gamma,\gamma_0,n} = 1$. Again, this cannot happen unless $n^* = n^*_\infty \geq n - m$. Similarly, under Assumption 2,

$$\sup\left\{|\lambda_v(\hat{\Sigma}_{\gamma,\gamma_0,n})^{-1} - \lambda_v(\hat{\Sigma}^{(n^*)}_{\gamma,\gamma_0,n})^{-1}|\right\} = \infty$$

only if, for at least $n^* = m - v$ units for which $\hat{w}_{i^*,\gamma,\gamma_0,n} = 1$, either we have

$$y_{i^*}^* \propto u,$$

where $u = (1, \ldots, 1)'$, or it holds that

$$\exists i \in \{n^* + 1, \ldots, n\} : y_{i^*}^* \propto y_i.$$

But this cannot happen if $n^* < n - m$, because y is in general position. □

An immediate consequence of Theorem 8 is that the squared generalized Mahalanobis distances $\hat{d}^2_{i,\gamma,\gamma_0,n}$, defined in (40), share the same global robustness property of the estimators, since

$$\epsilon(\hat{d}^2_{i,\gamma,\gamma_0,n}, y) = \min\left\{\epsilon(\hat{\mu}_{\gamma,\gamma_0,n}, y), \epsilon(\hat{\mu}_{\gamma,\gamma_0,n}, y)\right\} = 1 - \gamma \quad \text{for } i = 1, \ldots, n.$$

Therefore, these distances are a natural ingredient for building robust diagnostic methods that are able to highlight multivariate outliers, without suffering from masking [21]. Empirical evidence of their performance for this purpose is given by Riani et al. [30].

5. Empirical performance of the FS estimators

In the previous section we have shown theoretically the convergence of the FS estimators when $m, n \to \infty$. We now investigate empirically the speed of convergence as a function of the sample size n, and of the fraction $\gamma = m/n$ of units belonging to the fitting subset. Due to affine invariance of the squared distances $\hat{d}^2_{i,\gamma,\gamma_0,n}$, and thus of the weights $\hat{w}_{i,\gamma,\gamma_0,n}$, we restrict our attention to data from the standard multivariate normal distribution $N(0, I_v)$. In what follows we take $v = 5$. The results for other values of v are very similar and are not reported here.

In order to evaluate the performance of $\hat{\mu}_{\gamma,\gamma_0,n}$ and $\hat{\Sigma}_{\gamma,\gamma_0,n}$ as estimators of μ and Σ, we consider the following measures:

- the squared bias of the location estimator, defined as $e_\mu(\text{bias}) = \|\hat{\mu}_{\gamma,\gamma_0,n}\|^2$;
- the variance of the location estimator, defined as $e_\mu(\text{var}) = \|\hat{\mu}_{\gamma,\gamma_0,n} - \bar{\hat{\mu}}_{\gamma,\gamma_0,n}\|^2/(v-1)$, where $\bar{\hat{\mu}}_{\gamma,\gamma_0,n}$ is the overall mean across the simulations;

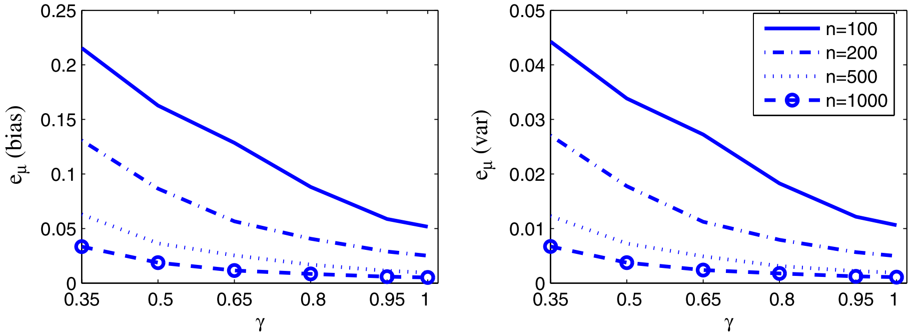

Fig. 1. Squared bias e_μ (bias) (left panel) and variance e_μ (var) (right panel) of the FS location estimator, as a function of γ, for $n = 100$ (solid line), $n = 200$ (dash dotted line), $n = 500$ (dotted lines), $n = 1000$ (dashed lines with round marker). Averages over 500 replications for each value of n.

- an error measure for the scatter estimator, defined as the logarithm of its condition number:

$$e_\Sigma = \log_{10}(\mathrm{cond}(\hat{\Sigma}_{\gamma,\gamma_0,n})),$$

where $\mathrm{cond}(\hat{\Sigma}_{\gamma,\gamma_0,n}) = \lambda_1(\hat{\Sigma}_{\gamma,\gamma_0,n})/\lambda_v(\hat{\Sigma}_{\gamma,\gamma_0,n})$ is the condition number of $\hat{\Sigma}_{\gamma,\gamma_0,n}$ (see, e.g., [22]).

Clearly we would like these quantities to be as small as possible, when the data come from $N(0, I_v)$. In order to evaluate the error measures for a particular value of γ, we stop the search at γ and use the weights (39) to estimate both the centroid $\hat{\mu}_{\gamma,\gamma_0,n}$ and the covariance matrix $\hat{\Sigma}_{\gamma,\gamma_0,n}$. In practice, the FS estimators are computed for all integers $m = v + 1, \ldots, n$, with $\gamma = m/n$. At each step we take the estimators obtained at the previous step as $\hat{\mu}_{\gamma_0,n}$ and $\hat{\Sigma}_{\gamma_0,n}$. As customary, the algorithm is initialized by means of robust bivariate projections. The error measures are computed, and then recorded, on each simulated sample. In our study we run 500 replications of the FS for each value of n.

Fig. 1 shows the simulation average of the squared bias (left panel) and of the variance (right panel) of the location estimator, when $\gamma = 0.35, 0.5, 0.65, 0.80, 0.95, 1$, and $n = 100, 200, 500, 1000$. This figure clearly shows that the error measures for the estimate of μ decrease both as a function of γ and as a function of the sample size. Furthermore, it is interesting to note that for $n = 100$ the curves for two error measures flatten only when $\gamma = 0.95$. On the other hand, when $n = 1000$, the curves seem to be virtually horizontal from $\gamma = 0.65$ onwards, but this effect is mainly due to the use of the same scale in the vertical axis of the plots for different values of n.

Fig. 2 shows the monitoring of e_Σ for the same values of γ and the four sample sizes considered in Fig. 1. We now show the boxplots instead of the simulation average, in order to have an idea also of the distribution of the values throughout the simulations. Clearly, the condition number decreases as n increases. The different scale in the four panels enables us to understand that the shape of the curves of the condition number is the same independently of the sample size. We thus see that the slope of the curve of the condition number levels off only when $\gamma = 0.95$.

The message which comes from Figs. 1 and 2 is that, while for large values of n accurate estimates of the parameters can be found just considering a fixed percentage of central data (typically a bit larger than 50%), for small and moderate sample sizes it is extremely important to use as many good observations as possible in order to reduce the variability of the error estimates. In any case, independently from the sample size, the shape of the error measures remains the same. In order to better understand the implications of this aspect, in Fig. 3 we show the boxplots of the distribution of values of e_μ (bias) for $n = 100$ (left panel) and $n = 200$ (right panel). In this figure the mean values of squared bias computed on all simulations are shown in the boxplots as circles. In addition, in the plots we display as horizontal lines the values of squared bias for the Minimum Covariance Determinant (MCD) estimator, based on a subset of $\lfloor (n + v + 1)/2 \rfloor$ observations (upper horizontal line), and for the reweighted MCD estimator, which uses the quantile $\chi^2_{0.975,v}$ as a threshold to determine the units forming the new centroid. In the computation of the reweighted MCD, both the asymptotic consistency factor and the small sample correction factor of [29] have been applied to the estimated covariance matrix at the first step. As expected, the precision of the reweighted MCD estimator, given that it uses more observations, is greater than that of the raw MCD. Fig. 3 also shows that the squared bias of the raw MCD estimator of location is comparable with that of the FS estimator with $\gamma = 0.5$. Finally, it is interesting to note that the horizontal lines associated to the error level of the reweighted MCD estimator of μ are below those of the FS when $\gamma \leq 0.8$, but are above those of the FS based on $\gamma \geq 0.95$. This effect is more pronounced in the left panel, where the sample size is smaller, rather than in the right panel.

In spite of its limited size, the conclusion that can be drawn from this simulation study is twofold. First, it confirms the consistency properties of the FS estimators of μ and Σ, and it provides us with some guidance about the estimation error as a function of both the sample size n and the ratio m/n. Second, it suggests the importance (especially for small and moderate sample sizes) of adopting a procedure which uses an adaptive level of trimming, instead of considering just a fixed percentage of data to estimate location and scatter.

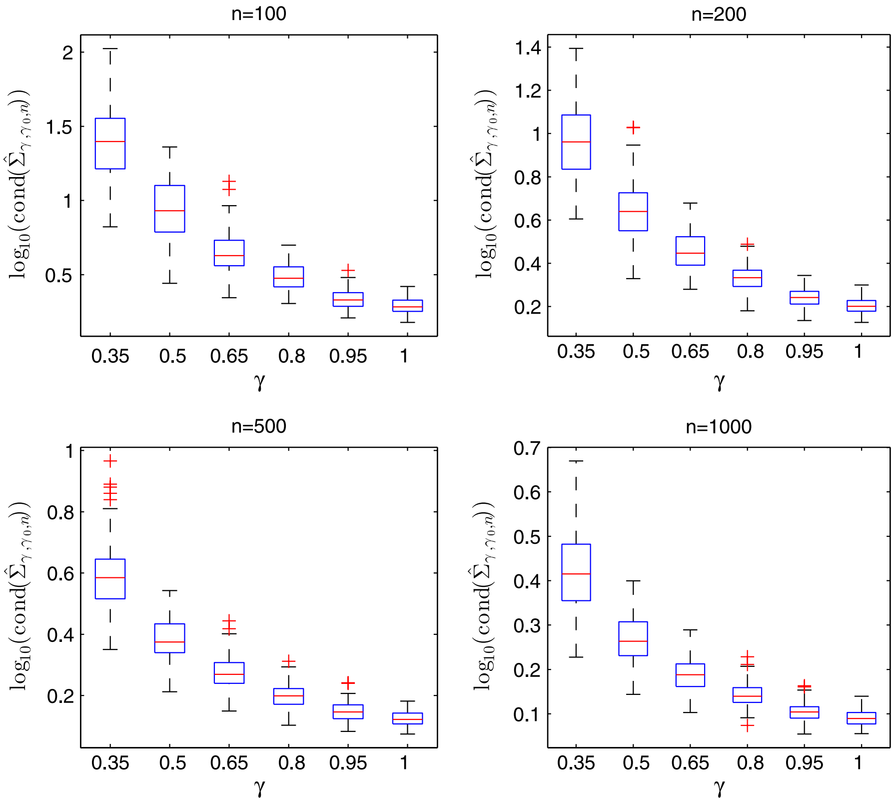

Fig. 2. Boxplots of the distribution of the logarithm of the condition number $\log_{10}(\mathrm{cond}(\hat{\Sigma}_{\gamma,\gamma_0,n}))$, as a function of γ, for $n = 100$ (top left panel), $n = 200$ (top right panel), $n = 500$ (bottom left panel) and $n = 1000$ (bottom right panel). The boxplots are obtained from 500 replications for each value of n.

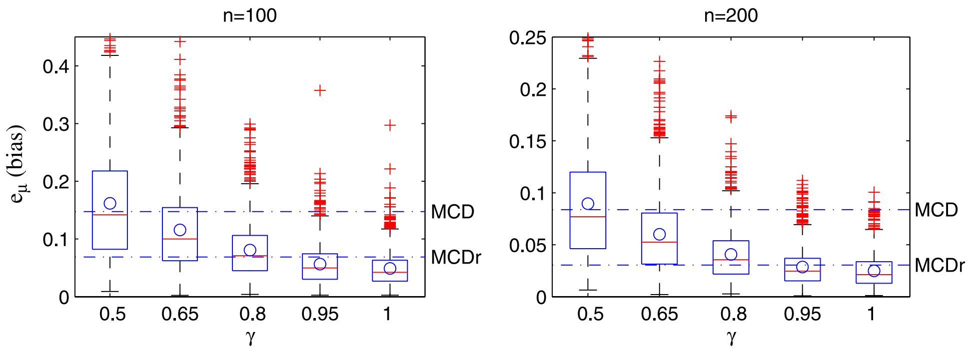

Fig. 3. Boxplots of the values of the squared bias for the FS estimator of location, as a function of γ, for $n = 100$ (left panel) and $n = 200$ (right panel). The circles over the boxplots denote the average values. The horizontal dotted–dashed lines are associated with the squared bias for the MCD location estimator (upper line) and the reweighted MCD (MCDr) location estimator (lower line). The boxplots and averages are obtained from 500 replications for each value of n.

6. Discussion and open issues

In this work we have studied some asymptotic properties of the Forward Search, a powerful general method for detecting anomalies in structured data, whose diagnostic power has been shown in many statistical contexts, but for which theoretical results are lacking in the multivariate case. Specifically, we have shown that the estimators of location and scatter obtained from the Forward Search are strongly consistent when the multivariate normal model holds. We have started from the ideal case where the model parameters are known, and then we have considered the effect of their estimation. We have also obtained the finite sample breakdown point of these estimators.

Our results serve the purpose of motivating the use of the Forward Search in robust applied statistics and of providing theoretical justification for its very good diagnostic properties. Furthermore, they allow us to compare the Forward Search estimators with other well known multivariate high-breakdown techniques, for which similar properties have been established in the past. A preliminary study involving the Minimum Covariance Determinant estimator, and its reweighted version, is described in this paper. More extensive comparisons are left for further research.

Our work does not complete the asymptotic analysis of the Forward Search for multivariate data, since there remain at least two important questions to be addressed. The first one concerns the asymptotic distribution of the estimators of location and scatter at a fixed step of the Forward Search. A way to achieve this goal could be to extend the results of [14,25,37], among others, to the Forward Search context. Another major open problem is a detailed asymptotic analysis of the behavior of Forward Search estimators under the general contamination model

$$M_1 : y_i \sim \xi G_0 + (1-\xi)G_1 \quad i = 1, \dots, n, \tag{45}$$

where G_0 stands for the $N_v(\mu, \Sigma)$ distribution which defines model (1), G_1 is a contaminant distribution, and $0.5 \leq \xi \leq 1$.

In fact, if the multivariate normal population defined by G_0 and the contaminant population defined by G_1 are "well separated", we can see that the asymptotic breakdown point of the Forward Search estimators of μ and Σ becomes

$$\begin{aligned} &1-\gamma \quad \text{if } \xi \geq \gamma, \\ &0 \quad \text{if } \xi < \gamma, \end{aligned}$$

when $m, n \to \infty$ so that $m/n \to \gamma$. Cerioli et al. [12] provide some simple examples where such a "separation" can occur. Therefore, it is crucial to obtain a sound estimate of ξ under model M_1, which would yield the best trimming level $1-\gamma = 1-\xi$ for the Forward Search estimators $\hat{\mu}_{\gamma,\gamma_0,n}$ and $\hat{\Sigma}_{\gamma,\gamma_0,n}$.

Despite its limited size, the simulation study reported in this paper has clearly shown the effect of different choices of γ on the properties of the resulting estimators of multivariate location and scatter. A proposal for selecting ξ, and thus γ, is made in [30]. However, further theoretical research on this topic is needed. We have studied the asymptotic properties of the Forward Search estimators at each of a finite number of fixed steps, but we have not addressed the problem of coupling these estimators with a stopping rule for the choice of the optimal γ, which would make the number of steps a random variable. Furthermore, we could allow the trimming level $1-\gamma$ in $\hat{\mu}_{\gamma,\gamma_0,n}$ and $\hat{\Sigma}_{\gamma,\gamma_0,n}$ to vary continuously in a subset of the unit interval, as in model (45). A careful study of the resulting asymptotic properties, as well as a comparison with the corresponding radius process approach of [19], would require different tools and will be tackled in future work.

Acknowledgments

The authors warmly thank Anthony C. Atkinson, Lucio Barabesi, Bent Nielsen and the participants in the *Scientific Meeting on Robust Cluster Analysis and Trimming Methods* and in the workshop on *Advances in Robust Data Analysis and Clustering*, both held at the Joint Research Centre of the European Commission, Ispra (Italy), for stimulating discussions. Kind and constructive comments of two anonymous referees and an Associate Editor are also gratefully acknowledged.

This work was partly supported by the project MIUR PRIN "*MISURA – Multivariate models for risk assessment*".

Appendix. Proofs

Proof of Theorem 4. Theorem 3 ensures that the condition

$$\tilde{\Sigma}^{-1}_{\gamma,n} \xrightarrow{a.s.} \Sigma^{-1} \tag{A.1}$$

is satisfied, because the mapping f defined by $f(\tilde{\Sigma}_{\gamma,n}) = \tilde{\Sigma}^{-1}_{\gamma,n}$ is continuous with probability 1 at $\tilde{\Sigma}_{\gamma,n} = \Sigma$ when Σ is positive definite. For the same reason

$$\tilde{\Sigma}^{-1/2}_{\gamma,n} \xrightarrow{a.s.} \Sigma^{-1/2}, \tag{A.2}$$

where $\{\tilde{\Sigma}^{-1/2}_{\gamma,n}\}'\tilde{\Sigma}^{-1/2}_{\gamma,n} = \tilde{\Sigma}^{-1}_{\gamma,n}$ and $\{\Sigma^{-1/2}\}'\Sigma^{-1/2} = \Sigma^{-1}$. Furthermore, (A.2) implies component-wise convergence of the matrix elements:

$$\tilde{s}_{j,l,\gamma,n} \xrightarrow{a.s.} s_{j,l} \quad \forall j, l = 1, \dots, v, \tag{A.3}$$

where $\tilde{s}_{j,l,\gamma,n}$ is the generic element of $\{\tilde{\Sigma}^{-1/2}_{\gamma,n}\}$ and $s_{j,l}$ is the generic element of Σ^{-1}.

We continue by writing the quadratic form (32) as

$$\tilde{d}^2_{i,\gamma,n} = (y_i - \tilde{\mu}_{\gamma,n})'\{\tilde{\Sigma}^{-1/2}_{\gamma,n}\}'\{\tilde{\Sigma}^{-1/2}_{\gamma,n}\}(y_i - \tilde{\mu}_{\gamma,n})$$

and by considering the elements of the $1 \times v$ vector

$$z'_{i,n} = (y_i - \tilde{\mu}_{\gamma,n})'\{\tilde{\Sigma}^{-1/2}_{\gamma,n}\}'. \tag{A.4}$$

For this purpose, let $\tilde{\mu}_{j,\gamma,n}$ and $z_{i,j,n}$, $j = 1, \dots, v$, be the jth element of vectors $\tilde{\mu}_{\gamma,n}$ and $z_{i,n}$, respectively. Direct multiplication shows that

$$z_{i,j,n} = \sum_{l=1}^{v} (y_{i,l} - \tilde{\mu}_{j,\gamma,n}) \tilde{s}_{j,l,\gamma,n}. \tag{A.5}$$

From Theorem 3 we have that $y_{i,j} - \tilde{\mu}_{j,\gamma,n}$ gets closer and closer to $y_{i,j} - \mu_j$ as

$$|y_{i,j} - \tilde{\mu}_{j,\gamma,n} - (y_{i,j} - \mu_j)| = |\tilde{\mu}_{j,\gamma,n} - \mu_j| \xrightarrow{a.s.} 0, \tag{A.6}$$

by the continuous mapping theorem. We can thus apply the continuous mapping theorem again to show that

$$\left| (y_{i,j} - \tilde{\mu}_{j,\gamma,n}) \tilde{s}_{j,l,\gamma,n} - (y_{i,j} - \mu_j) s_{j,l} \right| \xrightarrow{a.s.} 0, \quad \forall j, l = 1, \dots, v, \tag{A.7}$$

and

$$\left| \sum_{l=1}^{v} (y_{i,j} - \tilde{\mu}_{j,\gamma,n}) \tilde{s}_{j,l,\gamma,n} - \sum_{l=1}^{v} (y_{i,j} - \mu_j) s_{j,l} \right| \xrightarrow{a.s.} 0 \quad \text{for } j = 1, \dots, v, \tag{A.8}$$

by virtue of (A.3). Therefore,

$$\left| (y_i - \tilde{\mu}_{j,\gamma,n})' \{\tilde{\Sigma}_{\gamma,n}^{-1/2}\}' - (y_i - \mu)' \{\Sigma^{-1/2}\}' \right| \xrightarrow{a.s.} 0.$$

We can similarly prove that

$$\left| \tilde{\Sigma}_{\gamma,n}^{-1/2} (y_i - \tilde{\mu}_{j,\gamma,n}) - \Sigma^{-1/2} (y_i - \mu) \right| \xrightarrow{a.s.} 0.$$

Hence,

$$\left| (y_i - \tilde{\mu}_{j,\gamma,n})' \{\tilde{\Sigma}_{\gamma,n}^{-1/2}\}' \tilde{\Sigma}_{\gamma,n}^{-1/2} (y_i - \tilde{\mu}_{j,\gamma,n}) - (y_i - \mu)' \{\Sigma^{-1/2}\}' \Sigma^{-1/2} (y_i - \mu) \right| \xrightarrow{a.s.} 0, \tag{A.9}$$

again for the continuous mapping theorem, which completes the proof.

Proof of Theorem 5. The distances $\tilde{d}^2_{i,\gamma,n}$ are identically distributed under M_0. Then, by Theorem 4

$$\mathrm{E}\left\{ \tilde{F}_{\gamma,n}(x) \right\} = P(\tilde{d}^2_{i,\gamma,n} \le x) \to F_{\chi^2_v}(x).$$

Now define $\gamma_x = \tilde{F}_{\gamma,n}(x)$ and let $\tilde{\delta}^2_{\gamma_x,\gamma,n}$ be the γ_x-th quantile of the n squared distances $\tilde{d}^2_{i,\gamma,n}$. For $i = 1, \dots, n$, also define $\tilde{w}_{i,\gamma_x,\gamma,n} = 1$ if $\tilde{d}^2_{i,\gamma,n} \le \tilde{\delta}^2_{\gamma_x,\gamma,n}$, and $\tilde{w}_{i,\gamma_x,\gamma,n} = 0$ otherwise. We can thus write (33) as

$$\tilde{F}_{\gamma,n}(x) = \frac{1}{n} \sum_{i=1}^{n} \tilde{w}_{i,\gamma_x,\gamma,n}.$$

For any $x > 0$, there is a unique value $\gamma^* \in (0, 1)$ such that

$$\tilde{\delta}^2_{\gamma_x,\gamma,n} = \tilde{d}^2_{(m^*),\gamma,n},$$

where, following definitions (2) and (9), $m^* = \lfloor \gamma^* n \rfloor$ and $\tilde{d}^2_{(m^*),\gamma,n}$ is the m^*-th order statistic among the squared distances $\tilde{d}^2_{i,\gamma,n}$. Therefore, we have

$$\tilde{w}_{i,\gamma_x,\gamma,n} = \tilde{w}_{i,\gamma^*,\gamma,n} = I(\tilde{d}^2_{i,\gamma,n} \le \tilde{\delta}^2_{\gamma^*,\gamma,n}). \tag{A.10}$$

The weights $\tilde{w}_{i,\gamma_x,\gamma,n}$ are dependent because they must satisfy the constraint

$$\sum_{i=1}^{n} \tilde{w}_{i,\gamma_x,\gamma,n} = m^*,$$

as a consequence of (A.10). Furthermore, if $\gamma_x = \gamma^*$ they must satisfy the additional constraint

$$\sum_{i=1}^{n} \tilde{w}_{i,\gamma_x,\gamma,n} \tilde{d}^2_{i,\gamma,n} = \frac{m^* v}{c_{\gamma^*}}, \tag{A.11}$$

which arises from the fact that the observations for which $\tilde{w}_{i,\gamma^*,\gamma,n} = 1$ are those used in computing $\tilde{\mu}_{\gamma^*,n}$ and $\tilde{\Sigma}_{\gamma^*,n}$. For these observations constraint (A.11) is a consequence of parameter estimation [6, p. 86]. We can thus apply a reasoning similar to that of Theorem 3 to show that the correlation condition (28) is verified with either $k = 2$ or $k = 3$. Finally, we note that the support of $\tilde{w}_{i,\gamma_x,\gamma,n}$ is bounded, and thus $\mathrm{Var}(\tilde{w}_{i,\gamma_x,\gamma,n})$ is finite for any $x > 0$. Furthermore, for given x, it is constant for $i = 1, \ldots, n$. Hence,

$$\mathrm{Var}(\tilde{w}_{i,\gamma_x,\gamma,n}) \sum_{i=1}^{\infty} \frac{1}{i^{3/2}} < \infty,$$

and we can apply Theorem 2 of [9] to obtain the result.

References

[1] J. Arismendi, Multivariate truncated moments, J. Multivariate Anal. 117 (2013) 41–75.
[2] A.C. Atkinson, M. Riani, Robust Diagnostic Regression Analysis, Springer–Verlag, New York, 2000.
[3] A.C. Atkinson, M. Riani, Forward search added variable t tests and the effect of masked outliers on model selection, Biometrika 89 (2002) 939–946.
[4] A.C. Atkinson, M. Riani, Distribution theory and simulations for tests of outliers in regression, J. Comput. Graph. Statist. 15 (2006) 460–476.
[5] A.C. Atkinson, M. Riani, Exploratory tools for clustering multivariate data, Comput. Statist. Data Anal. 52 (2007) 272–285.
[6] A.C. Atkinson, M. Riani, A. Cerioli, Exploring Multivariate Data with the Forward Search, Springer–Verlag, New York, 2004.
[7] A.C. Atkinson, M. Riani, A. Cerioli, Random start forward searches with envelopes for detecting clusters in multivariate data, in: S. Zani, A. Cerioli, M. Riani, M. Vichi (Eds.), Data Analysis, Classification and the Forward Search, Springer-Verlag, Berlin, 2006, pp. 163–171.
[8] A.C. Atkinson, M. Riani, A. Cerioli, The forward search: theory and data analysis. With discussion, J. Korean Stat. Soc. 39 (2010) 117–134.
[9] W. Bryc, W. Smolenski, Moment conditions for almost sure convergence of random variables, Proc. Amer. Math. Soc. 119 (1993) 629–635.
[10] E.A. Cator, H.P. Lopuhaä, Asymptotic expansion of the minimum covariance determinant estimator, J. Multivariate Anal. 101 (2010) 2372–2388.
[11] A. Cerioli, Multivariate outlier detection with high-breakdown estimators, J. Amer. Statist. Assoc. 105 (2010) 147–156.
[12] A. Cerioli, A. Farcomeni, M. Riani, Robust distances for outlier-free goodness-of-fit testing, Comput. Statist. Data Anal. 65 (2013) 29–45.
[13] A. Cerioli, M. Riani, The ordering of spatial data and the detection of multiple outliers, J. Comput. Graph. Statist. 8 (1999) 239–258.
[14] H. Croux, G. Haesbroeck, Influence function and efficiency of the minimum covariance determinant scatter matrix estimator, J. Multivariate Anal. 71 (1999) 161–190.
[15] H. Croux, G. Haesbroeck, Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies, Biometrika 87 (2000) 603–618.
[16] J.A. Cuesta-Albertos, C. Matrán, A. Mayo-Iscar, Trimming and likelihood: robust location and dispersion estimation in the elliptical model, Ann. Statist. 36 (2008) 2284–2318.
[17] L. Davies, Asymptotic behaviour of S-estimates of multivariate location parameters and disperion matrices, Ann. Statist. 15 (1987) 1269–1292.
[18] F. De Battisti, S. Salini, Robust analysis of bibliometric data, Stat. Methods Appl. 22 (2013) 269–283.
[19] L.A. García-Escudero, A. Gordaliza, Generalized radius processes for elliptically contoured distributions, J. Amer. Statist. Assoc. 100 (2005) 1036–1045.
[20] O. Hössjer, C. Croux, P. Rousseeuw, Asymptotics of generalized S-estimators, J. Multivariate Anal. 61 (1994) 148–177.
[21] M. Hubert, P.J. Rousseeuw, S. Van Aelst, High-breakdown robust multivariate methods, Statist. Sci. 23 (2008) 92–119.
[22] M. Hubert, P.J. Rousseeuw, T. Verdonck, A deterministic algorithm for robust location and scatter, J. Comput. Graph. Statist. 21 (2012) 618–637.
[23] S. Johansen, B. Nielsen, Discussion: the forward search: theory and data analysis, J. Korean Stat. Soc. 39 (2010) 137–145.
[24] S. Johansen, B. Nielsen, Asymptotic analysis for the Forward Search. Technical Report Nuffield Discussion Paper 2013-W02, Aarhus University, 2013.
[25] H.P. Lopuhaä, Asymptotics of reweighted estimators of multivariate location and scatter, Ann. Statist. 27 (1999) 1638–1665.
[26] H.P. Lopuhaä, P.J. Rousseeuw, Breakdown points of affine equivariant estimators of multivariate location and covariance matrices, Ann. Statist. 19 (1991) 229–248.
[27] D. Mavridis, I. Moustaki, The forward search algorithm for detecting aberrant response patterns in factor analysis for binary data, J. Comput. Graph. Statist. 18 (2009) 1016–1034.
[28] D. Peña, F.J. Prieto, Multivariate outlier detection and robust covariance matrix estimation (with discussion), Technometrics 43 (2001) 286–310.
[29] G. Pison, S. Van Aelst, G. Willems, Small sample corrections for LTS and MCD, Metrika 55 (2002) 111–123.
[30] M. Riani, A.C. Atkinson, A. Cerioli, Finding an unknown number of multivariate outliers, J. Roy. Statist. Soc. Ser. B 71 (2009) 447–466.
[31] M. Riani, A. Cerioli, F. Torti, On consistency factors and efficiency of robust S-estimators, Test 23 (2014) http://dx.doi.org/10.1007/s11749-014-0357-7.
[32] R.J. Serfling, Approximation Theorems of Mathematical Statistics, Wiley, New York, 1980.
[33] G.M. Tallis, Elliptical and radial truncation in normal samples, Ann. Math. Stat. 34 (1963) 940–944.
[34] S. Van Aelst, E. Vandervieren, G. Willems, A Stahel Donoho estimator based on huberized outlyingness, Comput. Statist. Data Anal. 56 (2012) 531–542.
[35] S. Van Aelst, G. Willems, Robust and efficient one-way MANOVA tests, J. Amer. Statist. Assoc. 106 (2011) 706–718.
[36] S. Zani, M. Riani, A. Corbellini, Robust bivariate boxplots and multiple outlier detection, Comput. Statist. Data Anal. 28 (1998) 257–270.
[37] Y. Zuo, H. Cui, Depth weighted scatter estimators, Ann. Statist. 33 (2005) 381–413.