AP 2016–01

Chair of Applied Stochastics and
Risk Management

HELMUT SCHMIDT
UNIVERSITÄT
Universität der Bundeswehr Hamburg
Faculty of Economic and Social Sciences
Department of Mathematics and Statistics

Working Paper

M-Estimation with Incomplete and Dependent Multivariate Data

Gabriel Frahm, Klaus Nordhausen and Hannu Oja

January 20, 2018

Working Paper AP 2016–01
January 20, 2018

M-Estimation with Incomplete and Dependent Multivariate Data

Gabriel Frahm

Helmut Schmidt University
Faculty of Economic and Social Sciences
Department of Mathematics and Statistics
Chair of Applied Stochastics and
Risk Management
D-22043 Hamburg, Germany

URL: www.hsu-hh.de/stochastik
Phone: +49 (0)40 6541-2791
E-mail: frahm@hsu-hh.de

Klaus Nordhausen

TU Wien
Institute of Statistics and Mathematical
Methods in Economics
A-1040 Wien, Austria

E-mail: Klaus.Nordhausen@tuwien.ac.at

Hannu Oja

University of Turku
Department of Mathematics and Statistics
FI-20014 Turku, Finland

E-mail: hannu.oja@utu.fi

Supervised by: Prof. Dr. Gabriel Frahm
Chair of Applied Stochastics and
Risk Management

URL: www.hsu-hh.de/stochastik

M-Estimation with Incomplete and Dependent Multivariate Data

Gabriel Frahm*
Helmut Schmidt University
Department of Mathematics and Statistics

Klaus Nordhausen†
TU Wien
Institute of Statistics and
Mathematical Methods in Economics

Hannu Oja‡
University of Turku
Department of Mathematics and Statistics

January 20, 2018

Abstract

We extend the theory of M-estimation to incomplete and dependent multivariate data. ML-estimation can still be considered a special case of M-estimation in this context. In order to guarantee the consistency of an M-estimator, the unobserved data must be missing completely at random but not only missing at random, which is a typical assumption of ML-estimation. We show that the weight functions for scatter must satisfy a critical scaling condition, which is implicitly fulfilled by the Gaussian and Tyler's weight function. We generalize this result by introducing the class of power weight functions. The aforementioned weight functions represent two extreme examples of a power weight function. A simulation study confirms our theoretical findings. If the data are heavy tailed or contaminated, the M-estimators turn out to be favorable compared to the ML-estimators based on the normal-distribution assumption.

Keywords: Dependent data; Incomplete data; Location; M-estimation; Missing data; Scatter; Spatial data; Time-series data.

*Phone: +49 40 6541-2791, e-mail: Gabriel.Frahm@hsu-hh.de.
†Phone: +43 1 58801-10563, e-mail: Klaus.Nordhausen@tuwien.ac.at.
‡Phone: +358 2 333-5441, e-mail: Hannu.Oja@utu.fi.

1. Motivation

In multivariate data analysis, practitioners often deal with high-dimensional, incomplete, serially or spatially dependent, and heavy-tailed data that can even be contaminated by measurement errors. For example, this is a typical situation in modern portfolio optimization (Markowitz, 1959, 1987), where it is a stylized fact that asset returns are not serially independent. However, the aforementioned problems frequently occur in many other scientific disciplines like meteorology, environmental sciences, life sciences, geophysics, signal and image processing, etc. The reason for this might be attributed to modern computer networks, which lead to an enormous flood of information. Nowadays, under the buzzword “big data,” people try to find significant and useful patterns in large data sets. Hence, the development of appropriate statistical procedures is highly relevant both from a practical and a theoretical point of view.

The robust-statistics literature regarding complete and independent data is overwhelming. See, for example, the well-known textbooks of Hampel et al. (1986) and Maronna et al. (2006). Robust estimation procedures for complete but dependent data are less widespread.[1] We have found only Gastwirth and Rubin (1975), which refers to univariate data, and Yuanxi (1994) in the context of multivariate data. Robust procedures that are typically applied in time-series analysis, such as HAC estimation and the stationary bootstrap (see, e.g., Hansen, 1992, Politis, 2003), aim at computing the asymptotic covariance matrix or the standard error of some given estimator. This is not the focus of our work. Indeed, nowadays missing-data analysis is a well-established branch of statistics. Some very nice and exciting textbooks on that field are written by Little and Rubin (2002) as well as Schafer (1997). However, most estimation procedures that are presented in those textbooks are not robust, and the robust alternatives to the standard methods of missing-data analysis that can be found in the literature typically presume that the data are independent. In this work, we focus on multivariate data analysis. More precisely, we discuss the estimation of location and scatter. Of course, this includes the univariate case.

Estimating location and scatter is an essential task of multivariate data analysis, but there exist only a few contributions on robust analysis of incomplete multivariate data. For example, Frahm and Jaekel (2010) generalize Tyler’s M-estimator for scatter (Tyler, 1987a) to the case of incomplete data.[2] Wang (1999) discusses M-estimation for censored data, whereas Danilov et al. (2012) refer to S-estimation with incomplete data. Flossmann (2010) and Wooldridge (2007) propose inverse probability weighting. This requires us to specify selection probabilities, which might lead to inconsistent parameter estimates if the propensity model is misspecified. Other authors focus on regression analysis (Boente et al., 2009, Han, 2014, Sued and Yohai, 2013) or principal component analysis (Serneels and Verdonck, 2008). However, the general scope of regression or principal component analysis seems to be quite different from ours.

The traditional estimation approach for location and scatter of incomplete multivariate data is based on multiple imputation (Little and Rubin, 2002, Schafer, 1997). This estimation approach

[1] Of course, this assertion holds no longer true if we drop our restriction to robust statistics.
[2] Their estimator requires only the angular but not the radial part of elliptically distributed data to be independent.

typically presumes that the data are multivariate normally distributed and so the resulting estimators are not robust. For this reason, Branden and Verboven (2009), Hron et al. (2010) and Templ et al. (2011) develop robust imputation algorithms. By contrast, Little (1988a) assumes that the data are contaminated or multivariate t-distributed. He estimates location and scatter by maximum likelihood, whereas Little and Smith (1987) propose an estimation method based on imputation. Cheng and Victoria-Feser (2002) improve the algorithm used by Little and Smith (1987) for high-dimensional data by applying high-breakdown estimators (e.g., Rousseeuw's minimum-volume ellipsoid estimator) and hybrid algorithms.[3] Another promising alternative is presented by Han (2016), who combines multiple imputation with inverse probability weighting.

All aforementioned contributions presume that the data are independent. By contrast, Palma and del Pino (1999) consider incomplete (long-range) dependent data. However, they refer to univariate time series and their methods are not robust. Although the given list of contributions is not exhaustive, our observation is that most authors do not take serial or spatial dependence into account and, at least in some cases, the scope of their contributions seems to be somewhat limited. In our opinion, the main challenge for incomplete and dependent multivariate data is not only to guarantee the consistency of the estimators but also to obtain their asymptotic covariance matrices. Of course, this is an essential task if one is interested in confidence regions or wants to conduct hypothesis tests. To the best of our knowledge, a general theory of M-estimation with incomplete and dependent multivariate data is still missing. This work tries to fill this gap. We demonstrate our method by deriving M-estimators for location and scatter. We also conduct a simulation study in order to confirm our theoretical findings.

2. Notation and Definitions

Let X be a sample, i.e., an $m \times n$ real-valued random matrix. For example, X may consist of m attributes of n individuals at a specific point in time (cross-sectional data), of m attributes of a specific individual at n points in time (time-series data), or of a specific attribute of m individuals at n points in time (panel data). In the following, we assume without loss of generality that X is a sample of cross-sectional data. Let R be a response indicator, i.e., an $m \times n$ matrix of Bernoulli variables. It indicates which part of X is missing (0) and which one is observed (1). As is usual in the statistics literature, random quantities are denoted by capital letters, whereas their realizations are symbolized by small letters. For example, x and r are realizations of the random matrices X and R, respectively. This means that x is an $m \times n$ matrix of real numbers and r is an $m \times n$ matrix of zeros and ones.

The components of X and R may depend on each other. Let X_i and R_i be the ith column of X and R, respectively.[4] The joint probability distribution of X_i and R_i is symbolized by f

[3] See also Copt and Victoria-Feser (2004), who propose a modified algorithm and use the orthogonalized Gnanadesikan-Kettenring estimator as a starting point for an adapted S-estimator.

[4] The m-tuple $X_i = (X_{1i}, X_{2i}, \ldots, X_{mi})$ denotes an m-dimensional *column* vector and thus $X = [X_1\ X_2\ \cdots\ X_n]$ is an $m \times n$ random matrix. For this reason, we have that $X_i' = [X_{1i}\ X_{2i}\ \cdots\ X_{mi}] \neq (X_{1i}, X_{2i}, \ldots, X_{mi})$ for $m > 1$.

with $f(x_i, r_i; \theta) = f(x_i; \theta) f(r_i \mid X_i = x_i; \theta)$ $(i = 1, 2, \ldots, n)$, where $\theta \in \Theta \subseteq \mathbb{R}^p$ is some unknown parameter and Θ is an open subset of $\mathbb{R}^p$.[5] The distribution of X_i can either be discrete or continuous, whereas the distribution of R_i is always discrete by definition. It is assumed that the joint distribution is identical for each individual, i.e., $f(x_i, r_i; \theta) = f(x_j, r_j; \theta)$ for $i, j = 1, 2, \ldots, n$. Let x_r be the observed and $x_{\bar{r}}$ the missing data of all individuals. This means that $\bar{r} = 1 - r$, where 1 denotes an $m \times n$ matrix of ones. Further, let r_i be the response, x_{r_i} the observed, and $x_{\bar{r}_i}$ the missing data of Individual i. Actually, x_{r_i} is a shorthand notation for $x_{r_i i}$, i.e., it denotes the observed components of the vector x_i.

The joint distribution of X_{R_i} and R_i is

$$\begin{aligned} f(x_{r_i}, r_i; \theta) &= \int f(x_{r_i}, x_{\bar{r}_i}, r_i; \theta)\, dx_{\bar{r}_i} \\ &= \int \underbrace{f(x_{r_i}, x_{\bar{r}_i}; \theta)}_{= f(x_i; \theta)} \underbrace{f(r_i \mid X_{r_i} = x_{r_i}, X_{\bar{r}_i} = x_{\bar{r}_i}; \theta)}_{= f(r_i \mid X_i = x_i; \theta)}\, dx_{\bar{r}_i}. \end{aligned} \tag{1}$$

Further, we often encounter the conditional distribution $f(r_i \mid X_{r_i} = x_{r_i}; \theta)$. This can be interpreted as the probability that Individual i provides the response r_i given that his *observed* data are x_{r_i}. Here, the response of Individual i, i.e., r_i, is considered *fixed*.

In the following, we use the typical abbreviations "ML" for maximum likelihood and "M" for maximum-likelihood type. Suppose that we want to estimate the unknown parameter θ by ML. The problem is that we can observe only x_r and r. Thus, our (composite) likelihood is

$$L(\theta; X_R, R) = \prod_{i=1}^{n} f(X_{R_i}, R_i; \theta). \tag{2}$$

In many practical applications the random vectors $(X_{R_1}, R_1), (X_{R_2}, R_2), \ldots, (X_{R_n}, R_n)$ are serially or spatially dependent.[6] This depends on whether we work with cross-sectional, time-series, or panel data. Treating the data as independent is a standard approach in econometrics (Hansen, 1982). In general, the resulting ML-estimator is asymptotically inefficient but not necessarily inconsistent. Of course, if the statistician *knows* the sort of serial or spatial dependence he or she should take the dependence structure into consideration when estimating θ. In this work, we do neither presume that the dependence structure is known nor that it is unknown. Our primary focus is on consistency and robustness rather than asymptotic efficiency.

Typically, it is not possible to use the likelihood function $L(\cdot; X_R, R)$. This is not because we are unable to specify the distribution of X_i, i.e., $f(x_i; \theta)$. The problem is that we do not know the distribution of R_i given $X_i = x_i$, and so we cannot calculate $f(x_{r_i}, r_i; \theta)$ (see (1)). Indeed, the dependence structure between X_i and R_i can be quite complicated. Of course, the same holds true for the serial or spatial dependence structure. Here, we focus on the dependence between X_i and R_i. Thus, we ignore the dependence between (X_i, R_i) and (X_j, R_j) for $i \neq j$. The

[5] For notational convenience we will omit the enumeration "$i = 1, 2, \ldots, n$" in the subsequent analysis if it is clear from the context that the given statement refers to each individual.

[6] More precisely, the joint distribution $f(x_{r_1}, x_{r_2}, \ldots, x_{r_n}, r_1, r_2, \ldots, r_n; \theta)$ does not correspond to $\prod_{i=1}^{n} f(x_{r_i}, r_i; \theta)$.

distribution of R_i given $X_i = x_i$ is said to be the missingness mechanism of our experiment.

Estimating θ would be much easier if we could ignore the missingness mechanism and use the observed-data likelihood (Schafer, 1997, p. 12)

$$L(\theta; X_R) = \prod_{i=1}^{n} f(X_{R_i}; \theta), \tag{3}$$

where $f(x_{r_i}; \theta)$ is the distribution of the subvector of X_i that is observed for Individual i. This is possible if the so-called ignorability condition

$$f(r_i \,|\, X_{r_i} = x_{r_i}; \theta) = f(r_i \,|\, X_{r_i} = x_{r_i}) \tag{4}$$

is satisfied (Schafer, 1997, Section 2.3.1). Under these circumstances we have that $f(x_{r_i}, r_i; \theta) = f(x_{r_i}; \theta)\, f(r_i \,|\, X_{r_i} = x_{r_i}) \propto f(x_{r_i}; \theta)$ and thus $L(\theta; X_R, R) \propto L(\theta; X_R)$.

The resulting ML-estimator $\hat{\theta}$ is the solution of the estimating equation

$$\Phi_R\big(\hat{\theta}; X_R\big) = \frac{1}{n} \sum_{i=1}^{n} \phi_{R_i}\big(\hat{\theta}; X_{R_i}\big) = 0 \tag{5}$$

with $\phi_{R_i}(\theta; X_{R_i}) = \partial \log f\big(X_{R_i}; \theta\big) / \partial \theta$.[7] Note that $f\big(x_{r_i}; \theta\big)$ denotes a *marginal* distribution. More precisely, it is the distribution of the observed part of the random vector X_i, i.e., X_{r_i}. Whenever we refer to the ML-estimator, we suppose that the ignorability condition is satisfied. In Section 3.1 we discuss typical assumptions that guarantee that this important requirement is met.

Let F be the joint cumulative distribution function of X_i and $\vartheta = T(F) \in \mathbb{R}^q$ some parameter such that

$$\mathrm{E}\big(\psi_{r_i}(\vartheta; X_{r_i})\big) = 0 \tag{6}$$

for every *fixed* response r_i that is possible for Individual i. Here $\psi_{r_i}(\cdot\,; X_{r_i})$ is a function from $\mathbb{R}^q$ to $\mathbb{R}^q$. An M-estimator $\hat{\vartheta}$ is the solution of the estimating equation

$$\Psi_R\big(\hat{\vartheta}; X_R\big) = \frac{1}{n} \sum_{i=1}^{n} \psi_{R_i}\big(\hat{\vartheta}; X_{R_i}\big) = 0\,, \tag{7}$$

where $\Psi_R\big(\hat{\vartheta}; X_R\big)$ represents a (composite) score. Hence, the ML-estimating equation (5) is just a special case of the M-estimating equation (7). In the ML-case we have that $\psi_{R_i}(\vartheta; X_{R_i}) = \partial \log f(X_{R_i}; \vartheta) / \partial \vartheta$ with $\vartheta \equiv \theta$. If integral and differential are interchangeable, it turns out that

$$\begin{aligned}
\mathrm{E}\big(\psi_{r_i}(\vartheta; X_{r_i})\big) &= \mathrm{E}\left(\frac{\partial \log f(X_{r_i}; \vartheta)}{\partial \vartheta}\right) = \int \frac{\partial \log f(x_{r_i}; \vartheta)}{\partial \vartheta} f(x_{r_i}; \vartheta)\, dx_{r_i} \\
&= \int \frac{\partial f(x_{r_i}; \vartheta) / \partial \vartheta}{f(x_{r_i}; \vartheta)} f(x_{r_i}; \vartheta)\, dx_{r_i} = \frac{\partial}{\partial \vartheta} \int f(x_{r_i}; \vartheta)\, dx_{r_i} = \frac{\partial}{\partial \vartheta} 1 = 0\,.
\end{aligned}$$

Hence, in the context of ML-estimation, the orthogonality condition (6) is always satisfied.

[7] We write "$\hat{\theta}$" instead of "$\hat{\theta}_n$" just to avoid an abundant use of subscripts.

The asymptotic results presented later are based on the following regularity conditions:

A1: We have that

$$n^{\frac{1}{2}}\Big(\Psi_R\big(\vartheta;X_R\big)-\mathrm{E}\big(\psi_{R_i}(\vartheta;X_{R_i})\big)\Big)\longrightarrow N_q(0,F_\vartheta)\,,\qquad n\longrightarrow\infty\,,$$

with

$$F_\vartheta=\lim_{n\to\infty}\frac{1}{n}\sum_{i,j=1}^{n}\mathrm{Cov}\big(\psi_{R_i}(\vartheta;X_{R_i}),\psi_{R_j}(\vartheta;X_{R_j})\big).$$

A2: The $q\times q$ matrix $\partial\Psi_R(\vartheta;X_R)/\partial\vartheta^\top$ is regular and it holds that

$$H_\vartheta=\operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\psi_{R_i}(\vartheta;X_{R_i})}{\partial\vartheta^\top}=\mathrm{E}\left(\frac{\partial\psi_{R_i}(\vartheta;X_{R_i})}{\partial\vartheta^\top}\right)$$

with $\det H_\vartheta\neq 0$.

A3: We can apply the Taylor expansion

$$\Psi_R\big(\hat\vartheta;X_R\big)=\Psi_R\big(\vartheta;X_R\big)+\frac{\partial\Psi_R\big(\vartheta;X_R\big)}{\partial\vartheta^\top}\big(\hat\vartheta-\vartheta\big)+O_{\mathrm{p}}\big(n^{-1}\big)\,.$$

In the ML-case, we denote F_ϑ and H_ϑ by F_θ and H_θ, respectively.

A1 and A2 can be motivated by ergodic theory. More precisely, in many practical applications we observe an ergodic stationary process $\{Z_t\}$ with $Z_t\sim Z$. This means that for every measurable function h with $\mathrm{E}\big(|h(Z)|\big)<\infty$ we have that $n^{-1}\sum_{t=1}^{n}h(Z_t)\to\mathrm{E}\big(h(Z)\big)$ as $n\to\infty$. The mode of convergence may depend on the chosen framework, but at least the convergence is in probability. We can assume also that the process $\{Z_t\}$ is strong mixing, which means that the one-sided processes $\{Z_t\}_{t\le 0}$ and $\{Z_t\}_{t\ge l}$ ($l\in\mathbb{N}$) are asymptotically independent, i.e., they are independent as the lag l between $\{Z_t\}_{t\le 0}$ and $\{Z_t\}_{t\ge l}$ grows to infinity. If the convergence rate is high enough, $n^{-\frac{1}{2}}\sum_{t=1}^{n}\big(Z_t-\mathrm{E}(Z_t)\big)$ is asymptotically normally distributed as $n\to\infty$. For more details on that topic see, e.g., the review article by Bradley (2005).

The aforementioned properties are typically used in the context of time-series analysis, but they have a meaningful interpretation also if we deal with cross-sectional data. In the latter case, strong mixing says that the attributes of any individual become independent from the attributes of another individual as the distance between the two individuals grows to infinity. Here, the term "distance" can be understood, e.g., in a social or regional sense. Hence, the regularity conditions A1 and A2 are satisfied if the process $\{\psi_{R_i}(\vartheta;X_{R_i})\}$ is strong mixing (with a sufficiently high convergence rate) and $\{\partial\psi_{R_i}(\vartheta;X_{R_i})/\partial\vartheta^\top\}$ is ergodic stationary. We hope that these minimal conditions are satisfied in most practical applications.

A3 is a typical regularity condition of asymptotic theory (van der Vaart, 1998, Ch. 5). There are many possibilities to guarantee that the remainder of the Taylor expansion expressed by A3 is bounded in probability at the rate n^{-1}. Sufficient conditions can be found, e.g., in Huber (2003). The same arguments can be applied to the incomplete-data case. In order to keep things as general as possible we avoid any specific requirement on the score function $\Psi_R(\cdot\,;X_R)$.

3. Theory of M-Estimation for Incomplete and Dependent Data

3.1. Maximum-Likelihood Estimation

Now, we discuss typical assumptions for the ignorability condition (4). Indeed, we have that

$$f(r_i \,|\, X_{r_i} = x_{r_i};\theta) = \int \underbrace{f(r_i \,|\, X_{r_i} = x_{r_i}, X_{\bar{r}_i} = x_{\bar{r}_i};\theta)}_{=\, f(r_i \,|\, X_i = x_i;\theta)} f(x_{\bar{r}_i} \,|\, X_{r_i} = x_{r_i};\theta)\, dx_{\bar{r}_i}.$$

It is assumed that the missingness mechanism is not determined by the parameter θ.

DIS: $f(r_i \,|\, X_{r_i} = x_{r_i}, X_{\bar{r}_i} = x_{\bar{r}_i};\theta) = f(r_i \,|\, X_{r_i} = x_{r_i}, X_{\bar{r}_i} = x_{\bar{r}_i})$.

This is the so-called distinctness assumption. The missingness mechanism may be parametric, too, but we are interested only in θ. Thus, we ignore the parameter in $f(r_i \,|\, X_i = x_i)$.

Additionally, it is typically assumed that, conditional on his *observed* data, the response of Individual i does not depend on his missing data.

MAR: $f(r_i \,|\, X_{r_i} = x_{r_i}, X_{\bar{r}_i} = x_{\bar{r}_i};\theta) = f(r_i \,|\, X_{r_i} = x_{r_i};\theta)$.

In this case, we say that $x_{\bar{r}}$ is missing at random. MAR only requires that the response of each individual is conditionally independent of his *own* missing data. If MAR is violated, we say that the unobserved data are not missing at random (NMAR).

From DIS and MAR it follows that $f(r_i \,|\, X_i = x_i;\theta) = f(r_i \,|\, X_i = x_i) = f(r_i \,|\, X_{r_i} = x_{r_i})$ and thus

$$\begin{aligned} f(r_i \,|\, X_{r_i} = x_{r_i};\theta) &= \int f(r_i \,|\, X_{r_i} = x_{r_i})\, f(x_{\bar{r}_i} \,|\, X_{r_i} = x_{r_i};\theta)\, dx_{\bar{r}_i} \\ &= f(r_i \,|\, X_{r_i} = x_{r_i}) \int f(x_{\bar{r}_i} \,|\, X_{r_i} = x_{r_i};\theta)\, dx_{\bar{r}_i} = f(r_i \,|\, X_{r_i} = x_{r_i})\,. \end{aligned}$$

This means that the ignorability condition is satisfied and so we can substitute (2) with (3).

The following interchangeability assumption is familiar in ML-theory.

INT: The integrals and differentials are twice interchangeable, i.e.,

$$\int\!\!\int \frac{\partial f(r_i, x_{r_i};\theta)}{\partial\theta}\, dx_{r_i}\, dr_i = \frac{\partial}{\partial\theta} \int\!\!\int f(r_i, x_{r_i};\theta)\, dx_{r_i}\, dr_i$$

and

$$\int\!\!\int \frac{\partial^2 f(r_i, x_{r_i};\theta)}{\partial\theta\,\partial\theta^\top}\, dx_{r_i}\, dr_i = \frac{\partial}{\partial\theta^\top} \int\!\!\int \frac{\partial f(r_i, x_{r_i};\theta)}{\partial\theta}\, dx_{r_i}\, dr_i\,.$$

The following proposition guarantees that the observed-data likelihood function given by (3) leads to a consistent ML-estimator, provided DIS, MAR, and INT are satisfied.[8]

Proposition 1. *Under the assumptions* DIS, MAR, *and* INT *the score function* $\Phi_R(\cdot\,; X_R)$ *is Fisher consistent for* θ, *i.e.,* $\mathrm{E}\big(\phi_{R_i}(\theta; X_{R_i})\big) = 0$.

We usually wish to guarantee also that $\hat{\theta}$ is asymptotically normally distributed after its usual standardization. Asymptotic normality and efficiency are established by the following theorem.

[8] All proofs can be found in the appendix.

Theorem 1. *Under the assumptions* A1–A3, DIS, MAR, *and* INT *we have that*

$$n^{\frac{1}{2}}\big(\hat{\theta}-\theta\big) \longrightarrow N_p\Big(0, H_\theta^{-1}F_\theta H_\theta^{-1}\Big), \qquad n \longrightarrow \infty.$$

Moreover, if $f(X_{R_1};\theta), f(X_{R_2};\theta), \ldots, f(X_{R_n};\theta)$ *are mutually independent, it holds that*

$$n^{\frac{1}{2}}\big(\hat{\theta}-\theta\big) \longrightarrow N_p\Big(0, F_\theta^{-1}\Big), \qquad n \longrightarrow \infty,$$

and $\hat{\theta}$ *is an asymptotically efficient estimator for* θ.

3.2. Maximum-Likelihood-Type Estimation

Now, we aim at estimating the parameter $\vartheta = T(F)$ by solving (7). Although $\mathrm{E}\big(\psi_{r_i}(\vartheta; X_{r_i})\big) = 0$ is satisfied by definition, it can happen that $\mathrm{E}\big(\psi_{R_i}(\vartheta; X_{R_i})\big) \neq 0$. In this case, $\Psi_R(\cdot\,; X_R)$ is not Fisher consistent for ϑ. Hence, we need an appropriate regularity condition in the context of M-estimation.

MCAR: $f(r_i \,|\, X_i = x_i;\theta) = f(r_i;\theta)$.

This means that the response of Individual i must not depend both on his observed and on his missing data.[9] In this case, we say that $x_{\bar{r}}$ is missing completely at random. It is only required that the response of each individual is independent of his *own* data.

Proposition 2. *Under the assumption* MCAR *the score function* $\Psi_R(\cdot\,; X_R)$ *is Fisher consistent for* ϑ*, i.e.,* $\mathrm{E}\big(\psi_{R_i}(\vartheta; X_{R_i})\big) = 0$.

MCAR guarantees that the M-estimator $\hat{\vartheta}$ is consistent. Since the ML-estimator $\hat{\theta}$ is an M-estimator, DIS is no longer required for the consistency of $\hat{\theta}$ if MCAR is satisfied.

MCAR implies MAR. However, MCAR alone does not guarantee the asymptotic efficiency of $\hat{\theta}$: Besides A1–A3, Theorem 1 requires the additional assumptions DIS and INT.

DIS together with MAR implies that we can ignore the missingness mechanism when we calculate the likelihood (2). However, this is not sufficient for the consistency of an M-estimator, which can be seen as follows: Under the ignorability condition it turns out that

$$\mathrm{E}\big(\psi_{R_i}(\vartheta; X_{R_i})\big) = \int\!\!\int f(r_i \,|\, X_{r_i} = x_{r_i})\,\psi_{r_i}(\vartheta; x_{r_i})\, f(x_{r_i};\theta)\, dx_{r_i}\, dr_i\,,$$

but the term $f(r_i \,|\, X_{r_i} = x_{r_i})$ is still determined by x_{r_i}, unless the unobserved data are MCAR. That is, it cannot be extracted from the inner integral in order to make use of the fact that

$$\int \psi_{r_i}(\vartheta; x_{r_i})\, f(x_{r_i};\theta)\, dx_{r_i} = \mathrm{E}\big(\psi_{r_i}(\vartheta; X_{r_i})\big) = 0.$$

This means that we do not make the assumption MCAR just because it is a sufficient condition for the consistency of an M-estimator. An M-estimator, in general, fails to be consistent if the

[9] Note that MCAR does not require DIS.

missing data are MAR (but not MCAR). Hence, MCAR is a requirement that must not be ignored in practical applications of missing-data analysis. This is because in many real-life situations the parametric family of X_i is unknown to us and then applying an ML-estimator can be misleading.

The following theorem completes our general results on M-estimation with incomplete and dependent data.

Theorem 2. *Under the assumptions* A1–A3 *and* MCAR *we have that*

$$n^{\frac{1}{2}}(\hat{\vartheta}-\vartheta) \longrightarrow N_q\Big(0, H_\vartheta^{-1} F_\vartheta H_\vartheta^{-1}\Big), \qquad n \longrightarrow \infty.$$

Whether or not MCAR might be violated in a real-life situation often follows from practical considerations. For example, if some respondents in a questionnaire refuse to answer a question because the value of their answer would exceed a critical threshold, we can expect that MCAR is violated. Before using an M-estimator for incomplete data, one can apply some test for the null hypothesis that the missing data are MCAR. A well-known test for MCAR is presented by Little (1988b). However, this requires the data to be multivariate normally distributed and thus it is not robust. Jamshidian and Jalal (2010) propose two hypothesis tests for MCAR. One is based on the normal-distribution assumption and the other is distribution-free.[10] Listing and Schlittgen (2003) present a nonparametric test for MCAR that combines several Wilcoxon rank sum tests. Since MCAR is a relatively simple independence assumption, we can imagine several other parametric and nonparametric testing procedures (see, e.g., Allison, 2001, p. 3).

Now, one might ask why not to use an ML-estimator right from the start. Indeed, M-estimation requires MCAR, whereas ML-estimation needs only MAR. The problem is that Theorem 1 is valid only if we *know* the parametric family of X_i. Unfortunately, if this is unknown to us, in general we cannot guarantee that the ML-estimator $\hat{\theta}$ is consistent if the missing data are MAR. Hence, it does not help much to conclude that MAR is weaker than MCAR if we use an ML-estimator but our distributional assumption is wrong. Nonetheless, if the prerequisites of Theorem 2 are satisfied, the ML-estimator $\hat{\theta}$ turns into an M-estimator $\hat{\vartheta}$. The technical details are elaborated in the subsequent analysis, where we concentrate on the estimation of location and scatter.

4. Estimation of Location and Scatter

Let U be a d-dimensional random vector that is uniformly distributed on the unit hypersphere $S^{d-1} = \{u \in \mathbb{R}^d : u^\top u = 1\}$. A random vector X is said to be elliptically distributed if and only if there exist a vector $\mu \in \mathbb{R}^m$, a matrix $\Lambda \in \mathbb{R}^{m \times d}$, a non-negative random variable V, and a random vector $U \sim S^{d-1}$ being independent of V such that

$$X = \mu + \Lambda V U.$$

[10] Jamshidian et al. (2014) provide an R package based on Jamshidian and Jalal (2010).

The parameter μ is called the location vector, whereas $\Sigma = \Lambda\Lambda^\top$ is referred to as the scatter matrix of X. Further, V is the generating variate of X. This general approach allows us to consider factor or seemingly unrelated regression models, etc. In our context, we have to guarantee only that $X - \mu$ is not concentrated on a linear subspace of $\mathbb{R}^m$. Hence, we assume that rank $\Lambda = m \leq d$ and thus $\Sigma > 0$. The distribution of an elliptically distributed random vector depends on Λ only through $\Sigma = \Lambda\Lambda^\top$ and so we may choose $d = m$. Further, we assume that V has no atom at 0, i.e., $\mathrm{P}(V = 0) = 0$. Now, let $X_1, X_2, \ldots, X_n$ be identically elliptically distributed.

4.1. Maximum-Likelihood Estimation

In the context of ML-estimation, we assume that X has an absolutely continuous distribution. The density of X is $f(x;\mu,\Sigma) = \det\Sigma^{-\frac{1}{2}}\, g\big((x-\mu)^\top\Sigma^{-1}(x-\mu)\big)$, where

$$g\colon\ \xi \longmapsto \frac{\Gamma(\frac{m}{2})}{\pi^{\frac{m}{2}}}\, f_{V^2}(\xi)\, \xi^{-\left(\frac{m}{2}-1\right)}, \qquad \xi > 0\,,$$

represents its density generator and f_{V^2} is the density function of V^2 (Tyler, 1982). For example, $V = (mF_{m,\nu})^{\frac{1}{2}}$ is the generating variate of the multivariate t-distribution with $\nu > 0$ degrees of freedom, where $F_{m,\nu}$ is an F-distributed random variable with m numerator and ν denominator degrees of freedom. Correspondingly, its density generator reads

$$g\colon\ \xi \longmapsto \frac{\Gamma(\frac{m+\nu}{2})}{\Gamma(\frac{\nu}{2})}\, \frac{1}{(\nu\pi)^{\frac{m}{2}}} \left(1 + \frac{\xi}{\nu}\right)^{-\frac{m+\nu}{2}}.$$

Another example is the multivariate power-exponential distribution, whose generating variate is $V = G_{\alpha,\beta}^{\frac{1}{2\gamma}}$ with parameter $\gamma > 0$, where $G_{\alpha,\beta}$ is a Gamma-distributed random variable with shape $\alpha = \frac{m}{2\gamma}$ and rate $\beta = m\Gamma(\frac{m}{2\gamma})/\Gamma((m+2)/(2\gamma))$. Its density generator is

$$g\colon\ \xi \longmapsto \frac{\Gamma(\frac{m}{2})}{\Gamma(\frac{m}{2\gamma})}\, \frac{\gamma}{\beta^{\frac{m}{2\gamma}}\pi^{\frac{m}{2}}} \exp\left(-\frac{\xi^\gamma}{\beta}\right).$$

4.1.1. Complete-Data Case

First of all, we consider the complete-data case. As is shown in Section 4.1 of a 2004 Cologne University PhD thesis by G. Frahm (http://kups.ub.uni-koeln.de/1319/), we have that

$$\phi_\mu(\mu,\Sigma;X_i) = \frac{\partial \log f(X_i;\mu,\Sigma)}{\partial\mu} = w\big((X_i-\mu)^\top\Sigma^{-1}(X_i-\mu)\big)\,\Sigma^{-1}(x-\mu)$$

and

$$\phi_\Sigma(\mu,\Sigma;X_i) = \frac{\partial \log f(X_i;\mu,\Sigma)}{\partial\Sigma} = A_i - \frac{1}{2}\operatorname{diag} A_i$$

with

$$A_i = w\big((X_i-\mu)^\top\Sigma^{-1}(X_i-\mu)\big)\,\Sigma^{-1}(X_i-\mu)(X_i-\mu)^\top\Sigma^{-1} - \Sigma^{-1},$$

where $w(\xi) = -2\,\partial \log g(\xi)/\partial\xi$. In order to obtain the corresponding ML-estimators for μ and Σ we have to solve the equation

$$\Phi\big(\hat\mu,\widehat\Sigma;X\big) = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix}\phi_\mu\big(\hat\mu,\widehat\Sigma;X_i\big)\\ \phi_\Sigma\big(\hat\mu,\widehat\Sigma;X_i\big)\end{bmatrix} = 0.$$

This leads to the usual ML-estimating equations

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^{n} w(\xi_i)\,(X_i-\hat\mu)\\ \widehat\Sigma &= \frac{1}{n}\sum_{i=1}^{n} w(\xi_i)\,(X_i-\hat\mu)(X_i-\hat\mu)^\top \end{aligned} \tag{8}$$

with $\xi_i = (X_i-\hat\mu)^\top\widehat\Sigma^{-1}(X_i-\hat\mu)$.

4.1.2. Incomplete-Data Case

In the incomplete-data case, we can observe only x_{r_i} for Individual i. We denote the number of attributes that are observable for Individual i by m_i. Actually, m_i represents a realization of a random variable $M_i \in \{1,2,\ldots,m\}$. Since the number of observations, m_i, may change with each individual, it is not appropriate to choose the same weight function for $i = 1,2,\ldots,n$. Otherwise, the resulting ML-estimators for μ and Σ might be inconsistent. Thus, we have

$$\begin{aligned} \phi_{\mu,R_i}(\mu,\Sigma;X_{R_i}) &= \frac{\partial \log f(X_{R_i};\mu,\Sigma)}{\partial\mu}\\ &= w_i\big((X_{R_i}-\mu_{R_i})^\top\Sigma_{R_i}^{-1}(X_{R_i}-\mu_{R_i})\big)\Big\langle\Sigma_{R_i}^{-1}(X_{R_i}-\mu_{R_i})\Big\rangle \end{aligned}$$

and

$$\phi_{\Sigma,R_i}(\mu,\Sigma;X_{R_i}) = \frac{\partial \log f(X_{R_i};\mu,\Sigma)}{\partial\Sigma} = \Big\langle A_{R_i} - \frac{1}{2}\,\mathrm{diag}\,A_{R_i}\Big\rangle$$

with

$$A_{R_i} = w_i\big((X_{R_i}-\mu_{R_i})^\top\Sigma_{R_i}^{-1}(X_{R_i}-\mu_{R_i})\big)\,\Sigma_{R_i}^{-1}(X_{R_i}-\mu_{R_i})(X_{R_i}-\mu_{R_i})^\top\Sigma_{R_i}^{-1} - \Sigma_{R_i}^{-1}.$$

Here, $f(x_{r_i};\mu,\Sigma)$ is the density of X_{r_i}, i.e., of the *observed* part of X_i, which we express in terms of the parameters μ and Σ. The symbols μ_{r_i} and Σ_{r_i} denote those parts of μ and Σ that are relevant for calculating $f(x_{r_i};\mu,\Sigma)$, i.e., that are associated with the response of Individual i.

For example, if we observe only the first component of X_i, i.e., X_{1i}, we have that $\mu_{r_i} = \mu_1$ and $\Sigma_{r_i} = \Sigma_{11}$. Since X_{r_i} is a subvector of X_i, the entries in μ and Σ that are *not* associated with any response of Individual i are redundant for $f(x_{r_i};\mu,\Sigma)$. Because we express the density of X_{r_i} in terms of μ and Σ, we have to use the inflation operator $\langle\cdot\rangle$. It inflates an array ("$\cdot$") by inserting zeros at those parts of the array that are associated with the non-response of the corresponding individual. In the previous example, $\partial \log f(X_{r_i};\mu,\Sigma)/\partial\mu$ is an $m\times 1$ vector that contains $\partial \log f(X_{1i};\mu,\Sigma)/\partial\mu_1$ at first place and zeros elsewhere (because $\partial \log f(X_{1i};\mu,\Sigma)/\partial\mu_i = 0$ for

$i = 2,3,\ldots,m$), which can be written as $\langle \partial \log f(X_{1i};\mu,\Sigma)/\partial\mu_1\rangle$. Similarly, $\partial \log f(X_{r_i};\mu,\Sigma)/\partial\Sigma$ is an $m\times m$ matrix that contains $\partial\log f(X_{1i};\Sigma)/\partial\Sigma_{11}$ on the upper left and zeros elsewhere (because $\partial\log f(X_{1i};\Sigma)/\partial\Sigma_{ij} = 0$ for $i = 2,3,\ldots,m$ or $j = 2,3,\ldots,m$), i.e., $\langle\partial\log f(X_{1i};\Sigma)/\partial\Sigma_{11}\rangle$. The inflation operator $\langle\cdot\rangle$ guarantees that the M-estimating equations are properly specified and thus it is an essential instrument of our estimation approach.

Suppose that $X_i = (X_{r_i}, X_{\bar r_i})$ for the sake of simplicity but without loss of generality. Further, we can assume that Λ is a lower triangular matrix, so that

$$\Lambda = \begin{bmatrix}\Lambda_{11} & 0\\ \Lambda_{21} & \Lambda_{22}\end{bmatrix}$$

with $\Lambda_{11}\in\mathbb{R}^{m_i\times m_i}$ and $\Sigma_{r_i} = \Lambda_{11}\Lambda_{11}^\top$. According to Cambanis et al. (1981), we have the stochastic representation $X_{r_i} = \mu_{r_i} + \Lambda_{11}V\beta^{\frac12}U$, where $\beta\sim\mathrm{Beta}\big(m_i/2,(m-m_i)/2\big)$ and $U\sim S^{m_i-1}$. Moreover, V, β, and U are mutually independent. This means that X_{r_i} has not the same generating variate as X_i. Hence, the corresponding weight function is $w_i:\xi\mapsto -2\,\partial\log g_i(\xi)/\partial\xi$ with

$$g_i(\xi) = \frac{\Gamma(\frac{m_i}{2})}{\pi^{\frac{m_i}{2}}}\, f_{V^2\beta}(\xi)\,\xi^{-\left(\frac{m_i}{2}-1\right)},\qquad \xi>0\,.$$

Now, the ML-estimators $\hat\mu$ and $\widehat\Sigma$ represent the solution of

$$\Phi\big(\hat\mu,\widehat\Sigma;X_R\big) = \frac1n\sum_{i=1}^n\begin{bmatrix}\phi_{\mu,R_i}\big(\hat\mu,\widehat\Sigma;X_{R_i}\big)\\ \phi_{\Sigma,R_i}\big(\hat\mu,\widehat\Sigma;X_{R_i}\big)\end{bmatrix} = 0\,,$$

which leads to the ML-estimating equations

$$\begin{aligned} 0 &= \frac1n\sum_{i=1}^n w_i(\xi_i)\left\langle\widehat\Sigma_{R_i}^{-1}(X_{R_i}-\hat\mu_{R_i})\right\rangle\\ \frac1n\sum_{i=1}^n\left\langle\widehat\Sigma_{R_i}^{-1}\right\rangle &= \frac1n\sum_{i=1}^n w_i(\xi_i)\left\langle\widehat\Sigma_{R_i}^{-1}(X_{R_i}-\hat\mu_{R_i})(X_{R_i}-\hat\mu_{R_i})^\top\widehat\Sigma_{R_i}^{-1}\right\rangle\end{aligned}\tag{9}$$

with $\xi_i = (X_{R_i}-\hat\mu_{R_i})^\top\widehat\Sigma_{R_i}^{-1}(X_{R_i}-\hat\mu_{R_i})$, where $\hat\mu_{R_i}$ and $\widehat\Sigma_{R_i}$ are those parts of $\hat\mu$ and $\widehat\Sigma$ that are associated with the observations of Individual i. Moreover, we have that $w_i(\xi) = -2\,\partial\log g_i(\xi)/\partial\xi$. If the data are complete, (9) simplifies to (8).[11]

This completes the ML-estimation of location and scatter with incomplete and dependent data. The next section proceeds further with M-estimation. We maintain our assumption that X_i is elliptically distributed and still focus on μ and Σ.

[11] Actually, the weight function for Individual i depends on his number of observations, i.e., m_i. However, we write "w_i" instead of "w_{m_i}" for notational convenience.

4.2. Maximum-Likelihood-Type Estimation

4.2.1. Complete-Data Case

Now, we can drop the assumption that X has an absolutely continuous distribution. If the data are complete, we have that

$$\psi_\mu\big(\mu,\Sigma;X_i\big) = v\Big(\big\{(X_i-\mu)^\top\Sigma^{-1}(X_i-\mu)\big\}^{\frac{1}{2}}\Big)\,\Sigma^{-1}(x-\mu)$$

and $\psi_\Sigma\big(\mu,\Sigma;X_i\big) = A_i - \frac{1}{2}\operatorname{diag} A_i$ with

$$A_i = w\big((X_i-\mu)^\top\Sigma^{-1}(X_i-\mu)\big)\,\Sigma^{-1}(X_i-\mu)(X_i-\mu)^\top\Sigma^{-1} - \Sigma^{-1}.$$

The corresponding score function Ψ is

$$\Psi\big(\hat\mu,\widehat\Sigma;X\big) = \frac{1}{n}\sum_{i=1}^n \begin{bmatrix}\psi_\mu\big(\hat\mu,\widehat\Sigma;X_i\big)\\ \psi_\Sigma\big(\hat\mu,\widehat\Sigma;X_i\big)\end{bmatrix}$$

and thus we obtain the M-estimating equations

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^n v\big(\xi_i^{\frac{1}{2}}\big)(X_i-\hat\mu) \\ \widehat\Sigma &= \frac{1}{n}\sum_{i=1}^n w\big(\xi_i\big)\,(X_i-\hat\mu)(X_i-\hat\mu)^\top, \end{aligned} \tag{10}$$

where v and w must satisfy a set of regularity conditions (see, e.g., Maronna, 1976).

The following weight functions for Σ can frequently be found in the literature (see, e.g., Tyler, 1987a, Kent and Tyler, 1991): The Gaussian weight function $w\colon \xi\mapsto 1$, Tyler's weight function $w\colon\xi\mapsto m/\xi$, Student's weight function $w\colon\xi\mapsto(m+\nu)/(\xi+\nu)$ with $\nu>0$, and Huber's weight function

$$w\colon\ \xi \longmapsto \begin{cases}\kappa, & \xi<\lambda\\ \kappa\lambda/\xi, & \xi\geq\lambda,\end{cases}$$

where the parameters $\kappa>0$ and $\lambda>0$ are such that $\mathrm{E}\big(w(\chi^2_m)\,\chi^2_m\big)=m$. See Dümbgen et al. (2015) for a comprehensive survey on M-estimation of scatter.

In the context of M-estimation, the distribution of the generating variate V is unknown. Since $\Lambda VU = (\sigma\Lambda)\big(V/\sigma\big)U$ for all $\sigma>0$, we have a well-known identification problem regarding the scatter matrix Σ. This problem is typically solved by the choice of the weight function w. Another alternative is to require $\sigma^2(\Sigma)=1$, where σ^2 represents a certain scale function. This point will be detailed below. In fact, the population version of the second part of (10) reads

$$\Sigma = \mathrm{E}\Big(w\big((X-\mu)^\top\Sigma^{-1}(X-\mu)\big)\,(X-\mu)(X-\mu)^\top\Big).$$

From $(X-\mu)^\top\Sigma^{-1}(X-\mu)=V^2$, $(X-\mu)(X-\mu)^\top = V^2\Lambda UU^\top\Lambda^\top$, and $\mathrm{E}(UU^\top)=I_m/m$ it follows that

$$\mathrm{E}\big(\varphi(V^2)\big) = m, \qquad \varphi(\xi) = w(\xi)\,\xi. \tag{11}$$

Hence, by applying the second M-estimating equation in (10), we implicitly assume that the generating variate V satisfies the scaling condition (11). Condition C of Maronna (1976) usually guarantees that there is no positive number $\sigma \neq 1$ such that $\mathrm{E}\big(\varphi(V^2/\sigma^2)\big) = m$. Otherwise, we could substitute the generating variate V with V/σ and Λ with $\sigma\Lambda$ without changing the distribution of X. Maronna's Condition C fails to solve the identification problem if there exists a threshold $\zeta \geq 0$ such that $\varphi(\xi)$ is constant for all $\xi > \zeta$. For example, Tyler's weight function implies that $\varphi(\xi) = m$ for all $\xi > 0$. In this case, we have to fix the scale of $\widehat{\Sigma}$, i.e., the M-estimator $\widehat{\Sigma}$ has to be normalized. This can be done, e.g., by requiring that $\operatorname{tr}\widehat{\Sigma} = m$ or $\det\widehat{\Sigma} = 1$ (Frahm, 2009, Paindaveine, 2008).

4.2.2. Incomplete-Data Case

Things become more complicated in case of incomplete data, where the M-estimating equations are similar to (9), i.e.,

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^{n} v_i\big(\xi_i^{\frac{1}{2}}\big)\left\langle \widehat{\Sigma}_{R_i}^{-1}(X_{R_i} - \hat{\mu}_{R_i})\right\rangle \\ \frac{1}{n}\sum_{i=1}^{n}\left\langle \widehat{\Sigma}_{R_i}^{-1}\right\rangle &= \frac{1}{n}\sum_{i=1}^{n} w_i(\xi_i)\left\langle \widehat{\Sigma}_{R_i}^{-1}(X_{R_i} - \hat{\mu}_{R_i})(X_{R_i} - \hat{\mu}_{R_i})^{\top}\widehat{\Sigma}_{R_i}^{-1}\right\rangle. \end{aligned} \tag{12}$$

A keynote of this work is that the choice of the weight functions $w_1, w_2, \ldots, w_n$ is not arbitrary. More precisely, we have to guarantee that the basic condition expressed by (6) is satisfied. Thus, we have to guarantee that

$$\mathrm{E}\left(v_i\left(\big\{(X_{r_i} - \mu_{r_i})^{\top}\Sigma_{r_i}^{-1}(X_{r_i} - \mu_{r_i})\big\}^{\frac{1}{2}}\right)\Sigma_{r_i}^{-1}(X_{r_i} - \mu_{r_i})\right) = 0 \tag{13}$$

and

$$\mathrm{E}\Big(w_i\big((X_{r_i} - \mu_{r_i})^{\top}\Sigma_{r_i}^{-1}(X_{r_i} - \mu_{r_i})\big)\Sigma_{r_i}^{-1}(X_{r_i} - \mu_{r_i})(X_{r_i} - \mu_{r_i})^{\top}\Sigma_{r_i}^{-1} - \Sigma_{r_i}^{-1}\Big) = 0 \tag{14}$$

for every fixed response r_i that is possible for Individual i. To the best of our knowledge, this issue has not yet been considered in the literature.

Now, we have that $X_{r_i} - \mu_{r_i} = \Lambda_{11}VU$, $(X_{r_i} - \mu_{r_i})^{\top}\Sigma_{r_i}^{-1}(X_{r_i} - \mu_{r_i}) = V^2\beta$, and

$$(X_{r_i} - \mu_{r_i})(X_{r_i} - \mu_{r_i})^{\top} = V^2\beta\Lambda_{11}UU^{\top}\Lambda_{11}^{\top}$$

with $\mathrm{E}(U) = 0$ and $\mathrm{E}(UU^{\top}) = I_{m_i}/m_i$, where the random quantities V, $\beta \sim \mathrm{Beta}\big(m_i/2, (m - m_i)/2\big)$, and $U \sim S^{m_i - 1}$ are mutually independent. Hence, the condition expressed by (13) is always satisfied, but (14) leads to the critical scaling condition

$$\mathrm{E}\big(\varphi_i(V^2\beta)\big) = m_i, \qquad \varphi_i(\xi) = w_i(\xi)\,\xi. \tag{15}$$

There exist two well-known weight functions that satisfy this scaling condition implicitly: The Gaussian weight function and Tyler's weight function. For the Gaussian weight function we have that $\varphi_i(\xi) = \xi$. From (11) we already know that $\mathrm{E}(V^2) = \mathrm{E}\big(\varphi(V^2)\big) = m$ and thus we obtain

$$\mathrm{E}\big(\varphi_i(V^2\beta)\big) = \mathrm{E}\big(V^2\beta\big) = \mathrm{E}(V^2)\,\mathrm{E}(\beta) = m \cdot \frac{m_i}{m} = m_i\,.$$

Moreover, Tyler's weight function leads to $\varphi_i(\xi) = m_i$ and thus $\mathrm{E}(\varphi_i(V^2\beta)) = m_i$. Hence, in our context, these weight functions can be considered canonical.

As pointed out by Frahm and Jaekel (2010), the Gauss- and Tyler-type M-estimators for scatter can always be considered ML-estimators, irrespective of whether the data are complete or incomplete: The Gauss-type M-estimator is an (observed-data) ML-estimator under the assumption that the data are multivariate normally distributed, whereas the Tyler-type M-estimator maximizes the likelihood function after projecting the observed data of each individual onto the unit hypersphere, in which case we obtain an angular central Gaussian distribution (Tyler, 1987b). When applying the Tyler-type M-estimator it has to be assumed only that the data are *generalized* elliptically distributed, which is shown in a 2004 Cologne University PhD thesis by G. Frahm (http://kups.ub.uni-koeln.de/1319/). Moreover, the finite-sample distribution of the Tyler-type M-estimator does not depend on the generating variate of each observation.

In this work, we generalize the insights given by Frahm and Jaekel (2010) by changing from ML-estimation to M-estimation of location and scatter.

5. The Power Weight Functions

5.1. Theoretical Properties

In order to obtain M-estimators for location and scatter that are consistent in the case of incomplete elliptically distributed data, we construct a class of weight functions that satisfy the critical scaling condition (15). The Gaussian and Tyler's weight function represent two extreme elements of this class. Hence, this work closes a gap left open by Frahm and Jaekel (2010).

In the following, $\mathrm{B}(a,b)$ denotes Euler's beta function with parameters $a,b>0$. We define $\mathrm{B}(a,0)/\mathrm{B}(b,0) = 1$ for all $a,b>0$. It can easily be seen that $\mathrm{B}(a,x)/\mathrm{B}(b,x) = 1$ as $x \searrow 0$.

Theorem 3. *Consider any real number* $0 \le \alpha \le 1$ *and suppose that*

$$\mathrm{E}\left(\left(\frac{V^2}{m}\right)^{-\alpha} V^2\right) = m. \tag{16}$$

Then, for every $d \in \{1,2,\ldots,m\}$*, we have that*

$$\mathrm{E}\left(\frac{\mathrm{B}\left(\frac{d}{2}+1, \frac{m-d}{2}\right)}{\mathrm{B}\left(\frac{d}{2}+1-\alpha, \frac{m-d}{2}\right)}\left(\frac{V^2\beta}{m}\right)^{-\alpha} V^2\beta\right) = d,$$

where V and $\beta \sim \mathrm{Beta}\big(d/2, (m-d)/2\big)$ are assumed to be stochastically independent.

Hence, a natural weight function for scatter is

$$w_i\colon \xi \longmapsto \frac{\mathrm{B}\left(\frac{m_i}{2}+1, \frac{m-m_i}{2}\right)}{\mathrm{B}\left(\frac{m_i}{2}+1-\alpha, \frac{m-m_i}{2}\right)} \left(\frac{\xi}{m}\right)^{-\alpha}, \qquad 0 \le \alpha \le 1.$$

In the complete-data case, we simply have that $w_i(\xi) = (\xi/m)^{-\alpha}$. Hence, the scaling condition expressed by (16) is an immediate consequence of (11).

The parameter α can be considered a tail index. For $\alpha = 0$ we obtain the Gaussian weight $w_i(\xi) = 1$. Moreover, since

$$\mathrm{B}\left(\frac{m_i}{2}+1, \frac{m-m_i}{2}\right) = \frac{m_i}{m}\,\mathrm{B}\left(\frac{m_i}{2}, \frac{m-m_i}{2}\right),$$

Tyler's weight $w_i(\xi) = m_i/\xi$ can readily be obtained by setting α equal to 1. Since (13) is always satisfied, the choice of the weight function for μ is quite arbitrary, but it is tempting to choose $v_i : \xi^{\frac{1}{2}} \mapsto \xi^{-\frac{\alpha}{2}}$ $(0 \le \alpha \le 1)$. If the data are complete, $\alpha = 0$ leads to the empirical mean vector, whereas for $\alpha = 1$ we obtain the M-estimator for location proposed by Hettmansperger and Randles (2002). Similarly, if we choose $\alpha = 0$, the resulting M-estimator for Σ corresponds to the empirical covariance matrix, whereas for $\alpha = 1$ we obtain Tyler's M-estimator for scatter (Tyler, 1987a). In the following, v_i and w_i will be referred to as power weight functions and each M-estimator that is based on a power weight function will be called power M-estimator.

5.2. Asymptotic Distributions

The following theorems establish the joint asymptotic distribution of $n^{\frac{1}{2}}(\hat{\mu}-\mu)$ and $n^{\frac{1}{2}}(\widehat{\Sigma}-\Sigma)$ given that the M-estimators $\hat{\mu}$ and $\widehat{\Sigma}$ are based on the power weight functions with common tail index α. It is straightforward to obtain similar results if the tail indices of v_i and w_i differ from each other. If the data are complete and independent, we can apply the standard results given by Huber (2003) and Maronna (1976). Alternatively, the asymptotic covariance matrices can be derived by Theorem 2, which can be used even if the data are incomplete or dependent.

We concentrate on the case of complete and independent data in order to obtain closed-form expressions. The $m^2 \times m^2$ identity matrix is symbolized by I_{m^2}. Let e_{ij} be the $m \times m$ matrix with 1 in the ijth position and zeros elsewhere. The $m^2 \times m^2$ commutation matrix is defined as $K_{m^2} = \sum_{i,j=1}^{m} e_{ij} \otimes e_{ji}$, where $\otimes$ denotes the Kronecker product. For any $m \times m$ random matrix M, the m^2-dimensional vector $\operatorname{vec} M$ is obtained by stacking the columns of M on top of each other. Further, $n^{\frac{1}{2}} M \to N_{m\times m}(0, C)$ $(n \to \infty)$ means that $n^{\frac{1}{2}} \operatorname{vec} M$ is asymptotically normally distributed with asymptotic covariance matrix $C \in \mathbb{R}^{m^2 \times m^2}$.

Theorem 4. *Let $\hat{\mu}$ and $\widehat{\Sigma}$ be the power M-estimators for location and scatter with common tail index $0 \le \alpha < 1$. Suppose that $X_1, X_2, \ldots, X_n$ are complete, independent, and identically elliptically* *distributed. Further, let the assumptions* A1–A3 *be satisfied. Then we have that*

$$n^{\frac{1}{2}}\big(\hat{\mu}-\mu\big) \longrightarrow N_m\left(0, \frac{m}{(m-\alpha)^2}\frac{\mathrm{E}\big(V^{2(1-\alpha)}\big)}{\mathrm{E}^2\big(V^{-\alpha}\big)}\,\Sigma\right), \qquad n\longrightarrow\infty,$$

and $n^{\frac{1}{2}}\big(\hat{\Sigma}-\Sigma\big) \to N_{m\times m}(0,C)$ *as* $n\to\infty$ *with*

$$C = \gamma_1\big(I_{m^2}+K_{m^2}\big)\big(\Sigma\otimes\Sigma\big) + \gamma_2(\operatorname{vec}\Sigma)(\operatorname{vec}\Sigma)^\top.$$

The numbers γ_1 *and* γ_2 *are*

$$\gamma_1 = \frac{(m+2)^2\tau_1}{(m+2\tau_2)^2} \quad \textit{and} \quad \gamma_2 = \frac{(\tau_1-1)-2\tau_1(\tau_2-1)\big\{m+(m+4)\,\tau_2\big\}/(m+2\tau_2)^2}{\tau_2^2}$$

with

$$\tau_1 = \frac{m^{2\alpha}}{m(m+2)}\,\mathrm{E}\big(V^{4(1-\alpha)}\big) \qquad \textit{and} \qquad \tau_2 = \frac{1-\alpha}{m^{1-\alpha}}\,\mathrm{E}\big(V^{2(1-\alpha)}\big).$$

Moreover, $\hat{\mu}$ *and* $\hat{\Sigma}$ *are asymptotically independent.*

Theorem 4 does not cover the limiting case $\alpha = 1$ because Tyler's M-estimator for scatter has to be normalized. Let σ^2 be an appropriate scale function. This means that σ^2 is a differentiable homogeneous function, i.e., $\sigma^2(\tau\Sigma) = \tau\sigma^2(\Sigma) > 0$ for all $\tau > 0$ and every positive definite $m\times m$ matrix Σ, that is such that $\sigma^2(I_m) = 1$. Each positive definite $m\times m$ matrix Ω that is such that $\sigma^2(\Omega) = 1$ is said to be a shape matrix. Now, Tyler's M-estimator can be normalized by $\hat{\Omega} = \hat{\Sigma}/\sigma^2(\hat{\Sigma})$, which represents an estimator for the shape matrix $\Omega = \Sigma/\sigma^2(\Sigma)$ (Frahm, 2009, Paindaveine, 2008). Apparently, estimating the shape matrix makes sense only if $m > 1$, i.e., for multivariate data. In the following, $\Upsilon = I_{m^2} - (\operatorname{vec}\Omega)J_{\sigma^2}$ symbolizes an $m^2\times m^2$ matrix, where $J_{\sigma^2} = \partial\sigma^2(\Omega)/\partial(\operatorname{vec}\Omega)^\top$ denotes the Jacobian of the chosen scale function σ^2 (Frahm, 2009).

Theorem 5. *Let* $\hat{\mu}$ *and* $\hat{\Sigma}$ *be the power M-estimators for location and scatter with common tail index* $0\le\alpha\le 1$. *Further, let* $\hat{\Omega} = \hat{\Sigma}/\sigma^2(\hat{\Sigma})$ *be the corresponding shape matrix estimator. Suppose that* $X_1, X_2, \ldots, X_n$ *are complete, independent, and identically elliptically distributed with* $m > 1$. *Further, let the assumptions* A1–A3 *be satisfied. Then we have that*

$$n^{\frac{1}{2}}\big(\hat{\mu}-\mu\big) \longrightarrow N_m\left(0, \frac{m}{(m-\alpha)^2}\frac{\mathrm{E}\big(V^{2(1-\alpha)}\big)}{\mathrm{E}^2\big(V^{-\alpha}\big)}\,\Sigma\right), \qquad n\longrightarrow\infty,$$

and $n^{\frac{1}{2}}\big(\hat{\Omega}-\Omega\big) \to N_{m\times m}(0,C)$ *as* $n\to\infty$ *with*

$$C = \frac{m+2}{m}\,\frac{\mathrm{E}\big(V^{4(1-\alpha)}\big)}{\big(m^{1-\alpha}+2(1-\alpha)\mathrm{E}\big(V^{2(1-\alpha)}\big)/m\big)^2}\,\Upsilon\big(I_{m^2}+K_{m^2}\big)\big(\Omega\otimes\Omega\big)\Upsilon^\top.$$

Moreover, $\hat{\mu}$ *and* $\hat{\Omega}$ *are asymptotically independent.*

According to Theorem 4 and Theorem 5, the asymptotic normality of $n^{\frac{1}{2}}\big(\hat{\Sigma}-\Sigma\big)$ and $n^{\frac{1}{2}}\big(\hat{\Omega}-\Omega\big)$

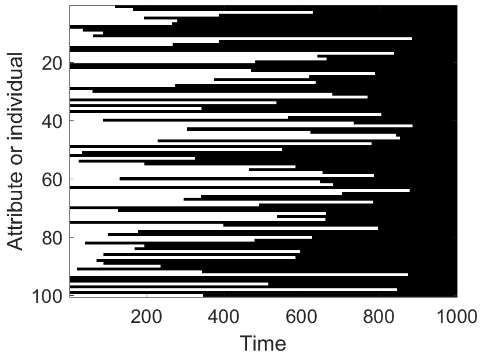

Figure 1: Responses (black lines) and non-responses (white space) in a sample with $m = 100$ dimensions and size $n = 1000$. There are 36727 missing values, i.e., about 37% of the entire sample cannot be observed.

requires the generating variate, V, to have a finite moment of order $4(1-\alpha)$. For this reason we recommend to choose a sufficiently high tail index $\alpha \in [0,1]$ if the data are heavy tailed. Conversely, if the data are light tailed, the tail index should be sufficiently low. In fact, if α is close to 1, the data must not be too heavily concentrated around μ. More precisely, for the asymptotic normality of the power M-estimators for location and scatter, we must guarantee that $\mathrm{E}(V^{-\alpha}) < \infty$. In the case of $\alpha = 1$, this phenomenon is already observed by Tyler (1987a).

The power M-estimates can be computed by the fixed-point algorithm developed by Frahm and Jaekel (2010). This algorithm turns out to be very fast and reliable even if the number of dimensions is high. Since every ML-estimator for location and scatter is an M-estimator, of course, the same estimation procedure can be applied in order to compute the ML-estimator. However, in quite simple cases (for example, if the data are multivariate normally distributed), it could be more efficient to apply some standard algorithm (e.g., the EM-algorithm). For example, in the case of $\alpha = 0$, i.e., when computing the Gauss-type M-estimates, we recommend an algorithm based on the sweep operator (Schafer, 1997, Ch. 6.5). This leads to exact solutions and is even faster than our fixed-point algorithm. Dümbgen et al. (2016) propose an alternative procedure based on a Taylor expansion, which could be useful also in the case of missing data.

5.3. Graphical Illustration

Now, we want to illustrate the impact of the tail index α. For this purpose, we simulate two samples of $n = 1000$ multivariate t-distributed data with $m = 100$ dimensions. In the first sample the number of degrees of freedom of the t-distribution amounts to $\nu = 2$, which means that the data are heavy tailed. By contrast, in the second sample the data are multivariate normally distributed ($\nu = \infty$). The location vector corresponds to $\mu = 0$ and the shape matrix is given by

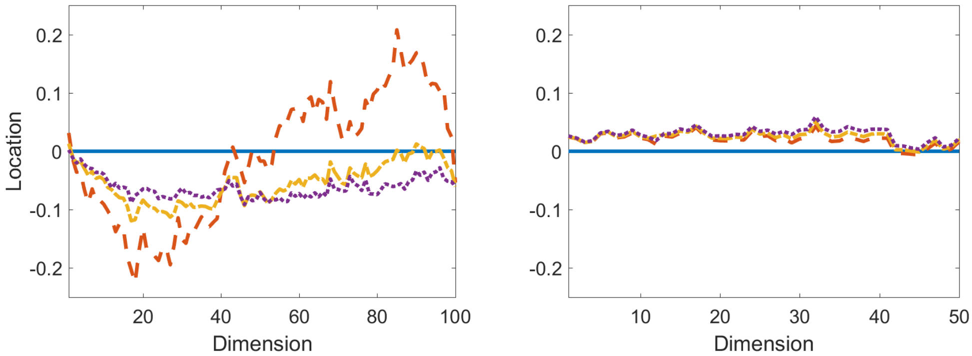

Figure 2: Power M-estimates of μ for multivariate t-distributed data (left) and multivariate normally distributed data (right): The Gauss-type M-estimates (dashes), the power M-estimates with $\alpha = 0.5$ (dots and dashes), and the Tyler-type M-estimates (dots). The horizontal lines indicate the true location vector.

the Toeplitz matrix

$$\Omega = \begin{bmatrix} 1 & 0.99 & \cdots & 0.02 & 0.01 \\ 0.99 & 1 & & & 0.02 \\ \vdots & & \ddots & & \vdots \\ 0.02 & & & 1 & 0.99 \\ 0.01 & 0.02 & \cdots & 0.99 & 1 \end{bmatrix}.$$

Figure 1 indicates which part of both samples is observed (black) and which part is missing (white). We can see that this is a monotone missingness pattern, which typically occurs when analyzing time-series or panel data. The unobserved data are MCAR.

The different power M-estimates for the location vector μ are depicted in Fig. 2. If the data are heavy tailed, a higher tail index is preferable. By contrast, if the data are not heavy tailed, a lower tail index leads to an outcome that is slightly better. For making the power M-estimates for scatter with different choices of α comparable, we need to normalize $\widehat{\Sigma}$. Here, we choose the scale function $\sigma^2 \colon \Sigma \mapsto \operatorname{tr}\Sigma/m$ and consider the shape matrix estimator $\widehat{\Omega} = \widehat{\Sigma}/\sigma^2(\widehat{\Sigma})$. Due to this normalization, we have that $\operatorname{tr}\widehat{\Omega} = m$. Note that the true shape matrix Ω already satisfies the condition $\operatorname{tr}\Omega = m$ by construction.

The results for the multivariate t-distributed data are depicted in Fig. 3, whereas the results for the multivariate normally distributed data are given by Fig. 4. The Gauss-type M-estimate on the upper right of Fig. 3 is heavily distorted. This is due to the fact that $\nu = 2$. That is, the number of degrees of freedom of the multivariate t-distribution is very low. Hence, we can expect to get a better estimate by choosing a higher tail index α for the power weight functions. If we choose the common tail index $\alpha = 0.5$ the result looks a little bit better than the Gauss-type M-estimate (see the lower left of Fig. 3). However, the Tyler-type M-estimator, i.e., the power M-estimator with $\alpha = 1$, clearly provides the best result (see the lower right of Fig. 3). By contrast, if the data are not heavy tailed but multivariate normally distributed (see Fig. 4), Theorem 1 implies that

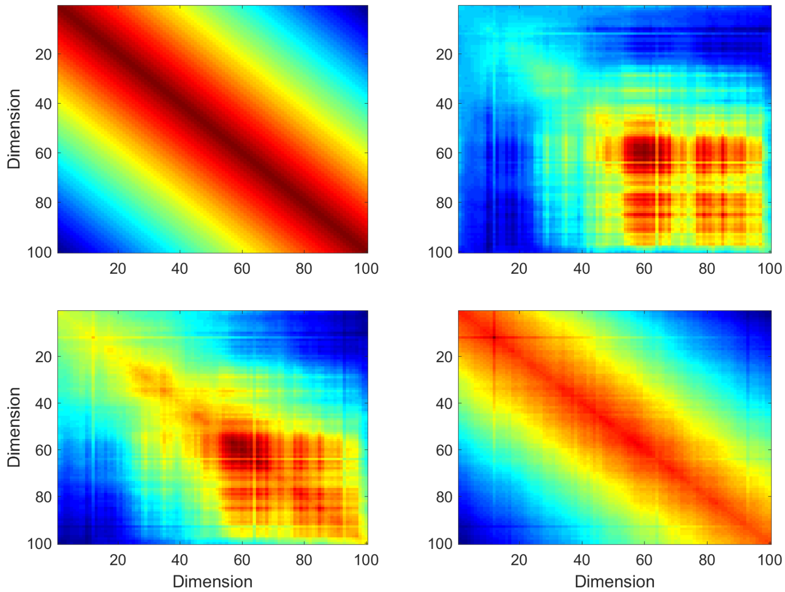

Figure 3: Power M-estimates of Ω for multivariate t-distributed data: The Gauss-type M-estimate (upper right), the power M-estimate with $\alpha = 0.5$ (lower left), and the Tyler-type M-estimate (lower right). The upper left is the true shape matrix. Violet cells indicate small numbers and red cells represent large numbers.

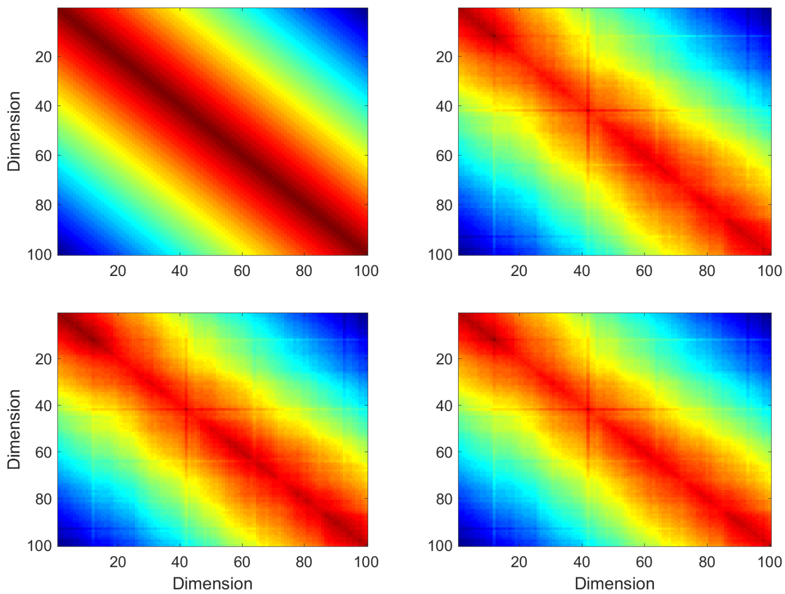

Figure 4: Power M-estimates of Ω for multivariate normally distributed data: The Gauss-type M-estimate (upper right), the power M-estimate with $\alpha = 0.5$ (lower left), and the Tyler-type M-estimate (lower right). The upper left is the true shape matrix. Violet cells indicate small numbers and red cells represent large numbers.

the Gaussian weight functions are superior because, in this case, they lead to an ML-estimator. Does this mean that the power M-estimates with $\alpha > 0$ are much worse? The answer is "No!" As we can see on the right-hand side of Fig. 2, as well as throughout Fig. 4, the power M-estimates with tail index $\alpha > 0$ are almost indistinguishable from the Gauss-type M-estimates.

6. Simulation Study

Of course, the graphical illustration of the power M-estimates does not say much about which M-estimator should be favored in different real-life situations. In order to answer this question we have to conduct an extensive simulation study.

6.1. Design of the Study

To be able to compare the power M-estimators for scatter, we apply the canonical scale function $\sigma^2 : \Sigma \mapsto (\det \Sigma)^{1/m}$ (Paindaveine, 2008). Hence, we focus on the M-estimator $\widehat{\Omega} = \widehat{\Sigma}/(\det \widehat{\Sigma})^{1/m}$ for the shape matrix $\Omega = \Sigma/(\det \Sigma)^{1/m}$. Apart from its theoretical advantages—which have been thoroughly investigated by Paindaveine (2008)—the canonical scale function turns out to be convenient from a *numerical* perspective: It guarantees that $\det \widehat{\Omega} = 1$ and thus $\widehat{\Omega}$ can never be singular, even if the number of dimensions, m, is very high. In our simulation study, we always consider μ unknown when estimating Ω. This is in contrast to the simulation study conducted by Frahm and Jaekel (2010), where the location vector is considered known.

The power M-estimators for μ and Ω are symbolized by $\hat{\mu}_\alpha$ and $\widehat{\Omega}_\alpha$, respectively, where α represents the common tail index of the power weight functions v_i and w_i. In the simulation study, we take the tail indices $\alpha = 0, 0.25, 0.50, 0.75, 1$ into account. In the limiting cases $\alpha = 0$ and $\alpha = 1$ we obtain the Gauss-type and the Tyler-type M-estimators for location and shape. Those power M-estimators are symbolized by $\hat{\mu}_{\mathrm{G}}, \widehat{\Omega}_{\mathrm{G}}, \hat{\mu}_{\mathrm{T}}, \widehat{\Omega}_{\mathrm{T}}$. If the data are complete, $\hat{\mu}_{\mathrm{G}}$ is the empirical mean vector. Moreover, in this case, we have that $\widehat{\Omega}_{\mathrm{G}} = \widehat{\Sigma}_{\mathrm{G}}/(\det \widehat{\Sigma}_{\mathrm{G}})^{1/m}$, where $\widehat{\Sigma}_{\mathrm{G}}$ represents the empirical covariance matrix. Analogously, $\hat{\mu}_{\mathrm{T}}$ is the M-estimator for location proposed by Hettmansperger and Randles (2002), whereas $\widehat{\Sigma}_{\mathrm{T}}$ is Tyler's (normalized) M-estimator for scatter (Tyler, 1987a). Actually, Tyler (1987a) uses the scale function $\sigma^2 : \Sigma \mapsto \mathrm{tr}\,\Sigma/m$ for normalization, but this does not alter the conclusions of our simulation study. A similar study can be found in Frahm and Jaekel (2010), but they investigate only the Gauss-type and Tyler-type M-estimators for shape. Hence, the study presented here can be considered a substantial generalization of the results documented by Frahm and Jaekel (2010).

For simulating heavy-tailed data, we use the multivariate t-distribution with $\nu > 0$ degrees of freedom. The multivariate t-distribution converges to the multivariate normal distribution as $\nu \to \infty$, which is indicated by $\nu = \infty$. By contrast, for simulating light-tailed data, we apply the multivariate power-exponential distribution with parameter $\gamma > 0$. In the case of $\gamma = 1$ it coincides with the multivariate normal distribution. By contrast, for $\gamma > 1$ its tails are lighter and for $0 < \gamma < 1$ they are heavier than the tails of the multivariate normal distribution. We consider scenarios in which the data are multivariate t-distributed with $\nu = 1,2,3,4,\infty$ degrees of freedom ($t_1, t_2, t_3, t_4, t_\infty$) and multivariate power-exponentially distributed with $\gamma = 5$ (p_5). The random vectors $X_1, X_2, \ldots, X_n$ are independent and identically distributed. Due to space limitations we concentrate on independent data. We do not think that the comparative result would be surprisingly different for dependent data. Our working hypothesis is that the components of $\{(X_{R_i}, R_i)\}$ are independent and thus, according to Theorem 2, serial or spatial dependence first and foremost blows up the asymptotic covariance matrices of the M-estimators.

The number of dimensions amounts to $m = 5$ and the true parameters of location and scatter are $\mu = 0$ and $\Sigma = I_5$, respectively. Besides the six aforementioned scenarios, $t_1, t_2, t_3, t_4, t_\infty, p_5$, we consider three additional scenarios in which the data are contaminated. More precisely, we substitute a number of cn $(0 < c < 1)$ multivariate normally distributed observations with the outlier $(10,10,\ldots,10) \in \mathbb{R}^5$. We consider the contamination rates $c = 0.01, 0.05, 0.10$. We also distinguish between a small sample ($n = 100$), a moderate sample ($n = 1000$), and a large sample ($n = 10000$). The number of Monte Carlo replications is always 10000.

As is done in Frahm and Jaekel (2010), the estimators are evaluated by their absolute bias (AB), i.e.,

$$\mathrm{AB}(\hat{\mu}) = \frac{1^\top |\mathrm{E}(\hat{\mu} - \mu)|}{m} \quad \text{and} \quad \mathrm{AB}(\widehat{\Omega}) = \frac{1^\top |\mathrm{E}(\widehat{\Omega} - \Omega)|\, 1}{m^2},$$

where $|A|$ is the matrix of absolute values of A and 1 is an appropriate vector of ones. The absolute bias can be relatively large for small or moderate sample sizes, although it is supposed to vanish in large samples. Our second quantity of interest is the mean squared error (MSE). This is the average mean squared error of all components, i.e.,

$$\mathrm{MSE}(\hat{\mu}) = \frac{\mathrm{E}\big((\hat{\mu} - \mu)^\top (\hat{\mu} - \mu)\big)}{m} \quad \text{and} \quad \mathrm{MSE}(\widehat{\Omega}) = \frac{\mathrm{E}\Big(\mathrm{tr}\big((\widehat{\Omega} - \Omega)(\widehat{\Omega} - \Omega)^\top\big)\Big)}{m^2}.$$

Finally, we investigate the relative efficiency (RE) of the Tyler-type M-estimators with respect to the Gauss-type M-estimators by

$$\mathrm{RE}(\hat{\mu}_\mathrm{T}) = \frac{\mathrm{MSE}(\hat{\mu}_\mathrm{G})}{\mathrm{MSE}(\hat{\mu}_\mathrm{T})} \qquad \text{and} \qquad \mathrm{RE}(\widehat{\Omega}_\mathrm{T}) = \frac{\mathrm{MSE}(\widehat{\Omega}_\mathrm{G})}{\mathrm{MSE}(\widehat{\Omega}_\mathrm{T})}.$$

The reader can easily derive the relative efficiency of some power M-estimator with respect to any other power M-estimator by the mean squared errors reported in the appendix.

The complete-data case is denoted by COM. For investigating the performance of the power M-estimators in the case of incomplete data, we simulate three different missingness mechanisms that satisfy MCAR, MAR (but not MCAR), and NMAR. Let x_i be a realization of X_i. We allow only the first component of x_i, i.e., x_{1i}, to be missing. More precisely, we have that $r_{1i} = 0$ if x_{1i} is missing and $r_{1i} = 1$ if x_{1i} is observed. This means that r_{1i} is the realization of the first component of the response vector R_i. The unobserved data are MCAR if R_i is stochastically independent of X_i. It is worth emphasizing that in principle we need *not* assume that R_i is stochastically independent of X_j with $j \neq i$, but this assumption is implicitly satisfied in our simulation study.

If the distribution of R_i depends only on the observed part of X_i, the missing data are MAR, and if the response is determined by the unobserved part of X_i, the missing data are NMAR. For the MCAR case, we simulate n mutually independent Bernoulli variables $R_{11}, R_{12}, \ldots, R_{1n}$ with probability of success $\pi = 0.5$, where R_{1i} is stochastically independent of X_i. In the MAR case, we have that $r_{1i} = 0$, i.e., x_{1i} is considered missing, if $x_{2i} < 0$. Finally, in the NMAR case we set $r_{1i} = 0$ if $x_{1i} < 0$. This procedure guarantees that approximately 50% of the data in the first row of the sample are missing for each missingness mechanism.

6.2. Numerical Results

The results of our simulation study can be found in the appendix (see Tables 2–9). Tables 2,4,6,8 provide the results regarding the location vector, whereas Tables 3,5,7,9 refer to the shape matrix. It is well-known that the generating variate of a multivariate t-distribution with ν degrees of freedom has finite moments only of orders lower than ν. In Section 5.2 we demonstrated that the joint asymptotic normality of the standardized power M-estimators with common tail index α requires the generating variate to have a finite moment of order $4(1-\alpha)$. In fact, our fixed-point algorithm sometimes diverges if $\alpha \leq 1 - \nu/4$. This could be due to the fact that a solution of the power M-estimating equations may not exist if the data are heavy tailed but the tail index of the power weight functions is too low. Hence, for $\nu = 1$ we do not apply the power M-estimators with tail indices $\alpha = 0.25, 0.50, 0.75$. Similarly, for $\nu = 2$ we ignore the power M-estimators with $\alpha = 0.25, 0.50$, etc. Nonetheless, we always compute the Gauss-type M-estimators by using the sweep operator (Schafer, 1997, Ch. 6.5), which cannot diverge by construction.

First of all we refer to Table 2 and Table 3, which contain the results of the complete-data case. If the data are clean, the absolute bias of the power M-estimators with $\alpha > 0$ decreases with the sample size, n, and eventually vanishes in the large samples. The Gauss-type M-estimators remain biased if ν is too low, i.e., $\nu < 2$ (Table 2) and $\nu < 4$ (Table 3). If the data are contaminated, the absolute bias essentially decreases with α, but it does not vanish with n. Indeed, the power M-estimators cannot be expected to be consistent if the data are contaminated, but in this case the Tyler-type M-estimators turn out to be always preferable in terms of bias.

Table 4 and Table 5 provide the results under the assumption that the unobserved data are MCAR. The overall findings do not differ essentially from Table 2 and Table 3. Our numerical results confirm that the power M-estimators are consistent if the unobserved data are MCAR. This picture changes substantially in Table 6 and 7, which cover the case in which the unobserved data are MAR. In the previous sections, we have argued that in this case the M-estimators can be inconsistent—even if the data are uncontaminated. This is confirmed by the results of our simulation study. The only exception are the Gauss-type M-estimators in the case of multivariate normally distributed data. Indeed, in this special case the Gauss-type M-estimators represent ML-estimators and their consistency under MAR is guaranteed by Theorem 1. Otherwise, we must accept that the power M-estimators are biased even in large samples. Finally, Table 8 and Table 9 are based on the assumption that the unobserved data are NMAR. In this case, also the

Table 1: Most favorable tail index of the power M-estimators under different scenarios.

		Location vector				Shape matrix			
Small sample		COM	MCAR	MAR	NMAR	COM	MCAR	MAR	NMAR
Light tails	AB	0	0	0	0	0	0	0	0
	MSE	0	0	1	0	0	0	0	0
Normal tails	AB	0	0.50	0	1	0	0	0	0
	MSE	0	0	1	1	0	0	0	0
Heavy tails	AB	1	1	0	1	0.75	0.75	1	0.50
	MSE	1	1	1	1	0.75	0.75	1	0.75
Contaminated									
1%	AB	1	1	1	1	1	1	1	1
	MSE	0.75	0.75	1	1	0.75	0.75	1	0.75
5%	AB	1	1	1	1	1	1	1	1
	MSE	1	1	1	1	1	1	1	1
10%	AB	1	1	1	1	1	1	1	1
	MSE	1	1	1	1	1	1	1	1
Moderate sample		COM	MCAR	MAR	NMAR	COM	MCAR	MAR	NMAR
Light tails	AB	0	0	0	0	0	0	0	0
	MSE	0	0	0	0	0	0	0	0
Normal tails	AB	0.75	0.25	0	1	0.25	0	0	0
	MSE	0	0	0	1	0	0	0	0
Heavy tails	AB	1	0.75	0	1	0.75	0.75	0.75	0
	MSE	1	1	1	1	0.75	0.75	1	0.50
Contaminated									
1%	AB	1	1	1	1	1	1	1	1
	MSE	1	1	1	1	1	1	1	0.75
5%	AB	1	1	1	1	1	1	1	1
	MSE	1	1	1	1	1	1	1	1
10%	AB	1	1	1	1	1	1	1	1
	MSE	1	1	1	1	1	1	1	1
Large sample		COM	MCAR	MAR	NMAR	COM	MCAR	MAR	NMAR
Light tails	AB	0	0	0	0	0	0	0	0
	MSE	0	0	0	0	0	0	0	0
Normal tails	AB	0	0	0	1	0	0	0.25	0
	MSE	0	0	0	1	0	0	0	0
Heavy tails	AB	1	1	0	1	0.75	0.75	0.25	0
	MSE	1	1	0.25	1	0.75	0.75	0.75	0
Contaminated									
1%	AB	1	1	1	1	1	1	1	1
	MSE	1	1	1	1	1	1	1	1
5%	AB	1	1	1	1	1	1	1	1
	MSE	1	1	1	1	1	1	1	1
10%	AB	1	1	1	1	1	1	1	1
	MSE	1	1	1	1	1	1	1	1

Gauss-type M-estimators are inconsistent, even if the data stem from a clean multivariate normal distribution. Similar arguments hold for the mean squared error of the power M-estimators.

The results of our simulation study are summarized in Table 1. This table indicates which power M-estimator is favorable in different situations. The M-estimators have been evaluated by means of their absolute bias and mean squared error. Each number in Table 1 represents the optimal choice of the tail index α with respect to the efficiency of the power M-estimator relative to the Gauss-type M-estimator. We can see that the Tyler-type M-estimators are nearly always preferable for contaminated data, whereas the Gauss-type M-estimators should be preferred only if the data have normal or light tails. If the data are heavy tailed one should choose a large tail index α. However, regarding the shape matrix it is usually better *not* to choose $\alpha = 1$, i.e., to avoid the Tyler-type M-estimator. This holds for complete and incomplete data, irrespective of whether the missing data are MCAR, MAR or NMAR.

Acknowledgment

Klaus Nordhausen and Hannu Oja would like to thank very much the Academy of Finland for their financial support. Gabriel Frahm would like to thank very much his co-authors for their kind invitations to Finland.

A. Proofs

A.1. Proof of Proposition 1

We have that

$$\begin{aligned}
\mathrm{E}\left(\frac{\partial \log f(X_{R_i};\theta)}{\partial\theta}\right) &= \int\int \frac{\partial \log f(x_{r_i};\theta)}{\partial\theta}\, f(r_i, x_{r_i};\theta)\, dx_{r_i}\, dr_i \\
&= \int\int f(r_i \,|\, X_{r_i} = x_{r_i})\, \frac{\partial \log f(x_{r_i};\theta)}{\partial\theta}\, f(x_{r_i};\theta)\, dx_{r_i}\, dr_i \\
&= \int\int \frac{\partial f(r_i \,|\, X_{r_i} = x_{r_i})\, f(x_{r_i};\theta)}{\partial\theta}\, dx_{r_i}\, dr_i \\
&= \int\int \frac{\partial f(r_i, x_{r_i};\theta)}{\partial\theta}\, dx_{r_i}\, dr_i \\
&= \frac{\partial}{\partial\theta}\int\int f(r_i, x_{r_i};\theta)\, dx_{r_i}\, dr_i = \frac{\partial}{\partial\theta}1 = 0\,.
\end{aligned}$$

A.2. Proof of Theorem 1

The ML-estimator of θ is the solution of

$$\Phi_R(\hat\theta; X_R) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log f\big(X_{R_i};\hat\theta\big)}{\partial\theta} = 0\,.$$

A3 guarantees that

$$\Phi_R\big(\hat\theta;X_R\big)=\Phi_R\big(\theta;X_R\big)+\frac{\partial\Phi_R\big(\theta;X_R\big)}{\partial\theta^\top}\big(\hat\theta-\theta\big)+O_{\mathrm{p}}\big(n^{-1}\big).$$

Hence, we have that

$$n^{\frac12}\big(\hat\theta-\theta\big)=-\left(\frac{\partial\Phi_R\big(\theta;X_R\big)}{\partial\theta^\top}\right)^{-1}n^{\frac12}\Phi_R\big(\theta;X_R\big)+O_{\mathrm{p}}\big(n^{-\frac12}\big).$$

From Proposition 1 we know that $\mathrm{E}\big(\phi_{R_i}(\theta;X_{R_i})\big)=0$. Due to A1 we conclude that $n^{\frac12}\Phi_R\big(\theta;X_R\big)\to N_p(0,F_\theta)$ as $n\to\infty$. A2 in connection with Slutsky's theorem implies that

$$n^{\frac12}\big(\hat\theta-\theta\big)\longrightarrow N_p\big(0,H_\theta^{-1}F_\theta H_\theta^{-1}\big),\qquad n\longrightarrow\infty.$$

Moreover, due to DIS, MAR, and INT, we obtain

$$\begin{aligned}
0&=\frac{\partial^2}{\partial\theta\,\partial\theta^\top}\underbrace{\int\!\!\int f(r_i,x_{r_i};\theta)\,dx_{r_i}\,dr_i}_{=1}\\
&=\frac{\partial}{\partial\theta^\top}\int\!\!\int f(r_i\,|\,X_{r_i}=x_{r_i})\,\frac{\partial f(x_{r_i};\theta)}{\partial\theta}\,dx_{r_i}\,dr_i\\
&=\int\!\!\int f(r_i\,|\,X_{r_i}=x_{r_i})\,\frac{\partial}{\partial\theta^\top}\left(\frac{\partial\log f(x_{r_i};\theta)}{\partial\theta}\,f(x_{r_i};\theta)\right)dx_{r_i}\,dr_i\,,
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial}{\partial\theta^\top}\left(\frac{\partial\log f(x_{r_i};\theta)}{\partial\theta}\,f(x_{r_i};\theta)\right)=\;&\frac{\partial^2\log f(x_{r_i};\theta)}{\partial\theta\,\partial\theta^\top}\,f(x_{r_i};\theta)+\\
&\frac{\partial\log f(x_{r_i};\theta)}{\partial\theta}\,\frac{\partial\log f(x_{r_i};\theta)}{\partial\theta^\top}\,f(x_{r_i};\theta)
\end{aligned}$$

and thus

$$\begin{aligned}
&\int\!\!\int\frac{\partial^2\log f(x_{r_i};\theta)}{\partial\theta\,\partial\theta^\top}\,f(r_i,x_{r_i};\theta)\,dx_{r_i}\,dr_i+\\
&\qquad\int\!\!\int\frac{\partial\log f(x_{r_i};\theta)}{\partial\theta}\,\frac{\partial\log f(x_{r_i};\theta)}{\partial\theta^\top}\,f(r_i,x_{r_i};\theta)\,dx_{r_i}\,dr_i=0\,.
\end{aligned}$$

Hence, if $f(X_{R_1};\theta),f(X_{R_2};\theta),\ldots,f(X_{R_n};\theta)$ are mutually independent then Fisher's information equality $F_\theta=-H_\theta$ is satisfied and we obtain $n^{\frac12}\big(\hat\theta-\theta\big)\to N_p\big(0,F_\theta^{-1}\big)$ as $n\to\infty$. Now, the Cramér-Rao Theorem implies that $\hat\theta$ is asymptotically efficient.

A.3. Proof of Proposition 2

MCAR implies that

$$f(r_i,x_{r_i};\theta)=\int f(r_i,x_{r_i},x_{\bar r_i};\theta)\,dx_{\bar r_i}=f(r_i;\theta)\int f(x_{r_i},x_{\bar r_i};\theta)\,dx_{\bar r_i}=f(r_i;\theta)\,f(x_{r_i};\theta)$$

and thus

$$\begin{aligned}
\mathrm{E}\big(\psi_{R_i}(\vartheta; X_{R_i})\big) &= \int\int \psi_{r_i}(\vartheta; x_{r_i})\, f(r_i, x_{r_i}; \theta)\, dx_{r_i}\, dr_i \\
&= \int f(r_i;\theta) \int \psi_{r_i}(\vartheta; x_{r_i})\, f(x_{r_i};\theta)\, dx_{r_i}\, dr_i \\
&= \int f(r_i;\theta)\, \mathrm{E}\big(\psi_{r_i}(\vartheta; X_{r_i})\big)\, dr_i = \int f(r_i;\theta)\, 0\, dr_i = 0\,.
\end{aligned}$$

A.4. Proof of Theorem 2

This proof is skipped, since it follows from the same arguments as those in the first part of the proof of Theorem 1. The second part of that proof, which refers to the asymptotic efficiency of the ML-estimator, is void in the case of M-estimation.

A.5. Proof of Theorem 3

We obtain

$$\mathrm{E}\left(\frac{\mathrm{B}\left(\frac{d}{2}+1, \frac{m-d}{2}\right)}{\mathrm{B}\left(\frac{d}{2}+1-\alpha, \frac{m-d}{2}\right)} \left(\frac{V^2\beta}{m}\right)^{-\alpha} V^2\beta\right) = \frac{\mathrm{B}\left(\frac{d}{2}+1, \frac{m-d}{2}\right)}{\mathrm{B}\left(\frac{d}{2}+1-\alpha, \frac{m-d}{2}\right)}\, m^{\alpha}\mathrm{E}\big((V^2\beta)^{1-\alpha}\big),$$

where $\mathrm{E}\big((V^2\beta)^{1-\alpha}\big) = \mathrm{E}\big(V^{2(1-\alpha)}\big)\mathrm{E}(\beta^{1-\alpha})$. From (16) it follows that $\mathrm{E}\big(V^{2(1-\alpha)}\big) = m^{1-\alpha}$. Thus, we have that

$$\frac{\mathrm{B}\left(\frac{d}{2}+1, \frac{m-d}{2}\right)}{\mathrm{B}\left(\frac{d}{2}+1-\alpha, \frac{m-d}{2}\right)}\, m^{\alpha}\mathrm{E}\big((V^2\beta)^{1-\alpha}\big) = \frac{\mathrm{B}\left(\frac{d}{2}+1, \frac{m-d}{2}\right)}{\mathrm{B}\left(\frac{d}{2}+1-\alpha, \frac{m-d}{2}\right)}\, m\mathrm{E}(\beta^{1-\alpha}).$$

Moreover, since $\mathrm{B}\left(\frac{d}{2}, \frac{m-d}{2}\right) = \frac{m}{d}\,\mathrm{B}\left(\frac{d}{2}+1, \frac{m-d}{2}\right)$ and thus

$$\mathrm{E}(\beta^{1-\alpha}) = \frac{\mathrm{B}\left(\frac{d}{2}+1-\alpha, \frac{m-d}{2}\right)}{\mathrm{B}\left(\frac{d}{2}, \frac{m-d}{2}\right)} = \frac{d}{m}\,\frac{\mathrm{B}\left(\frac{d}{2}+1-\alpha, \frac{m-d}{2}\right)}{\mathrm{B}\left(\frac{d}{2}+1, \frac{m-d}{2}\right)},$$

we conclude that

$$\frac{\mathrm{B}\left(\frac{d}{2}+1, \frac{m-d}{2}\right)}{\mathrm{B}\left(\frac{d}{2}+1-\alpha, \frac{m-d}{2}\right)}\, m\mathrm{E}(\beta^{1-\alpha}) = d.$$

A.6. Proof of Theorem 4

The asymptotic covariance matrix of $n^{\frac{1}{2}}(\hat{\mu}-\mu)$ can be obtained by Theorem 6 of Maronna (1976). We have that $v\big(\xi^{\frac{1}{2}}\big) = \xi^{-\frac{\alpha}{2}}$ and thus $\varphi_\mu\big(\xi^{\frac{1}{2}}\big) = v\big(\xi^{\frac{1}{2}}\big)\,\xi^{\frac{1}{2}} = \xi^{\frac{1-\alpha}{2}}$. It follows that $\varphi_\mu^2\big(\xi^{\frac{1}{2}}\big) = \xi^{1-\alpha}$ and $\varphi'_\mu\big(\xi^{\frac{1}{2}}\big) = (1-\alpha)\,\xi^{-\frac{\alpha}{2}}$, where φ'_μ denotes the first derivative of φ_μ with respect to $\xi^{\frac{1}{2}}$. Further, we already know that $\xi = V^2$. This means that $a = \mathrm{E}\big(\varphi_\mu^2(V)\big)/m = \mathrm{E}\big(V^{2(1-\alpha)}\big)/m$ and

$$b = \mathrm{E}\left(\upsilon(V)\left(1-\frac{1}{m}\right)+\frac{\varphi_\mu'(V)}{m}\right) = \frac{\mathrm{E}\big((m-1)V^{-\alpha}+(1-\alpha)V^{-\alpha}\big)}{m} = \Big(\frac{m-\alpha}{m}\Big)\mathrm{E}\big(V^{-\alpha}\big).$$

Hence, we obtain

$$\frac{a}{b^2} = \frac{m}{(m-\alpha)^2}\,\frac{\mathrm{E}\big(V^{2(1-\alpha)}\big)}{\mathrm{E}^2\big(V^{-\alpha}\big)},$$

which leads to the given asymptotic covariance matrix of $n^{\frac{1}{2}}(\hat{\mu}-\mu)$. Now, we turn to the asymptotic covariance matrix of $n^{\frac{1}{2}}(\widehat{\Sigma}-\Sigma)$. The numbers γ_1 and γ_2 are given by Tyler (1982, p. 432). We have that $w(\xi) = (\xi/m)^{-\alpha}$ and thus $\varphi_\Sigma(\xi) = w(\xi)\,\xi = m^\alpha \xi^{1-\alpha}$. This implies $\varphi_\Sigma^2(\xi) = m^{2\alpha}\xi^{2(1-\alpha)}$ and $\varphi_\Sigma'(\xi) = (1-\alpha)\,m^\alpha \xi^{-\alpha}$, where φ_Σ' is the first derivative of φ_Σ with respect to ξ. For calculating the numbers γ_1 and γ_2 we need

$$\tau_1 = \frac{\mathrm{E}\big(\varphi_\Sigma^2(V^2)\big)}{m(m+2)} = \frac{m^{2\alpha}}{m(m+2)}\mathrm{E}\big(V^{4(1-\alpha)}\big) \quad \text{and} \quad \tau_2 = \frac{\mathrm{E}\big(\varphi_\Sigma'(V^2)\,V^2\big)}{m} = \frac{1-\alpha}{m^{1-\alpha}}\mathrm{E}\big(V^{2(1-\alpha)}\big).$$

Finally, according to Theorem 6 of Maronna (1976), $\hat{\mu}$ and $\widehat{\Sigma}$ are asymptotically independent.

A.7. Proof of Theorem 5

For $0 \le \alpha < 1$, the asymptotic covariance matrix of $n^{\frac{1}{2}}(\hat{\mu}-\mu)$ has already been established by Theorem 4. The case $\alpha = 1$ is investigated by Hettmansperger and Randles (2002). It can be verified that the resulting asymptotic covariance matrix is covered by Theorem 4. Furthermore, according to Frahm (2009), the asymptotic covariance matrix of $n^{\frac{1}{2}}(\widehat{\Omega}-\Omega)$ corresponds to $C = \gamma_1 \Upsilon(I_{m^2}+K_{m^2})(\Omega\otimes\Omega)\Upsilon^\top$, where γ_1 is already given by Theorem 4. The numbers τ_1 and τ_2 have been derived in the proof of Theorem 4 and lead to

$$\gamma_1 = \frac{m+2}{m}\,\frac{\mathrm{E}\big(V^{4(1-\alpha)}\big)}{\big(m^{1-\alpha}+2(1-\alpha)\mathrm{E}\big(V^{2(1-\alpha)}\big)/m\big)^2}.$$

Finally, according to Theorem 4, $\hat{\mu}$ and $\widehat{\Sigma}$ are asymptotically independent in the case of $0 \le \alpha < 1$. Since $\widehat{\Omega}$ is a function of $\widehat{\Sigma}$, we conclude that $\hat{\mu}$ and $\widehat{\Omega}$ are asymptotically independent, too. This holds irrespective of the chosen scale function. Moreover, for $\alpha = 1$ the asymptotic independence of $\hat{\mu}$ and $\widehat{\Omega}$ is proved by Hettmansperger and Randles (2002). We can switch from one scale function to another by re-scaling the shape matrix, and so their result does not depend on the chosen scale function either.

B. Detailed Results of the Simulation Study

Table 2: Results for the location vector with complete data.

Small sample ($n = 100$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\hat\mu_{\mathrm{G}})$	0.5261	0.0023	0.0005	0.0010	0.0011	0.0005	0.0993	0.4994	1.0007
$\mathrm{AB}(\hat\mu_{0.25})$	—	—	—	0.0010	0.0011	0.0005	0.0629	0.3897	0.8863
$\mathrm{AB}(\hat\mu_{0.50})$	—	—	0.0005	0.0010	0.0012	0.0006	0.0351	0.2561	0.7192
$\mathrm{AB}(\hat\mu_{0.75})$	—	0.0007	0.0004	0.0009	0.0013	0.0006	0.0188	0.1284	0.4413
$\mathrm{AB}(\hat\mu_{\mathrm{T}})$	0.0010	0.0007	0.0004	0.0009	0.0013	0.0006	0.0101	0.0604	0.1622
$\mathrm{MSE}(\hat\mu_{\mathrm{G}})$	3×10^3	0.1382	0.0302	0.0202	0.0100	0.0022	0.0198	0.2590	1.0103
$\mathrm{MSE}(\hat\mu_{0.25})$	—	—	—	0.0166	0.0100	0.0023	0.0140	0.1617	0.7948
$\mathrm{MSE}(\hat\mu_{0.50})$	—	—	0.0164	0.0144	0.0102	0.0025	0.0115	0.0760	0.5271
$\mathrm{MSE}(\hat\mu_{0.75})$	—	0.0167	0.0141	0.0132	0.0105	0.0027	0.0110	0.0275	0.2055
$\mathrm{MSE}(\hat\mu_{\mathrm{T}})$	0.0177	0.0142	0.0131	0.0127	0.0111	0.0030	0.0113	0.0154	0.0385
$\mathrm{RE}(\hat\mu_{\mathrm{T}})$	2×10^5	9.7160	2.2957	1.5921	0.8995	0.7311	1.7480	16.8372	26.2520

Moderate sample ($n = 1000$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\hat\mu_{\mathrm{G}})$	0.8176	0.0011	0.0006	0.0004	0.0001	0.0001	0.0999	0.5002	1.0001
$\mathrm{AB}(\hat\mu_{0.25})$	—	—	—	0.0004	0.0001	0.0001	0.0636	0.3906	0.8856
$\mathrm{AB}(\hat\mu_{0.50})$	—	—	0.0004	0.0004	0.0001	0.0001	0.0360	0.2579	0.7195
$\mathrm{AB}(\hat\mu_{0.75})$	—	0.0003	0.0003	0.0004	0.0001	0.0001	0.0196	0.1312	0.4441
$\mathrm{AB}(\hat\mu_{\mathrm{T}})$	0.0003	0.0003	0.0002	0.0004	0.0001	0.0001	0.0108	0.0629	0.1659
$\mathrm{MSE}(\hat\mu_{\mathrm{G}})$	5×10^3	0.0157	0.0003	0.0020	0.0010	0.0002	0.0110	0.2512	1.0011
$\mathrm{MSE}(\hat\mu_{0.25})$	—	—	—	0.0016	0.0010	0.0002	0.0050	0.1535	0.7852
$\mathrm{MSE}(\hat\mu_{0.50})$	—	—	0.0016	0.0014	0.0010	0.0002	0.0023	0.0676	0.5187
$\mathrm{MSE}(\hat\mu_{0.75})$	—	0.0017	0.0014	0.0013	0.0011	0.0003	0.0014	0.0183	0.1984
$\mathrm{MSE}(\hat\mu_{\mathrm{T}})$	0.0017	0.0014	0.0013	0.0012	0.0011	0.0003	0.0012	0.0051	0.0287
$\mathrm{RE}(\hat\mu_{\mathrm{T}})$	3×10^6	11.0379	2.3107	1.5966	0.9023	0.7334	8.9184	49.0186	34.8332

Large sample ($n = 10000$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\hat\mu_{\mathrm{G}})$	0.6699	0.0003	0.0001	0.0002	0.0001	0	0.1000	0.5000	1.0000
$\mathrm{AB}(\hat\mu_{0.25})$	—	—	—	0.0002	0.0001	0	0.0637	0.3903	0.8854
$\mathrm{AB}(\hat\mu_{0.50})$	—	—	0	0.0001	0.0001	0	0.0361	0.2577	0.7194
$\mathrm{AB}(\hat\mu_{0.75})$	—	0.0001	0	0.0001	0.0001	0	0.0197	0.1311	0.4443
$\mathrm{AB}(\hat\mu_{\mathrm{T}})$	0.0001	0.0001	0	0.0001	0.0001	0	0.0108	0.0627	0.1662
$\mathrm{MSE}(\hat\mu_{\mathrm{G}})$	9×10^3	0.0017	0.0003	0.0002	0.0001	0	0.0101	0.2501	1.0001
$\mathrm{MSE}(\hat\mu_{0.25})$	—	—	—	0.0002	0.0001	0	0.0042	0.1524	0.7840
$\mathrm{MSE}(\hat\mu_{0.50})$	—	—	0.0002	0.0001	0.0001	0	0.0014	0.0665	0.5177
$\mathrm{MSE}(\hat\mu_{0.75})$	—	0.0002	0.0001	0.0001	0.0001	0	0.0005	0.0173	0.1975
$\mathrm{MSE}(\hat\mu_{\mathrm{T}})$	0.0002	0.0001	0.0001	0.0001	0.0001	0	0.0002	0.0041	0.0277
$\mathrm{RE}(\hat\mu_{\mathrm{T}})$	5×10^7	11.9105	2.2830	1.5822	0.9047	0.7327	44.2474	61.7401	36.0633

Table 3: Results for the shape matrix with complete data.

Small sample ($n = 100$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{G}})$	67.3426	0.1971	0.0393	0.0195	0.0065	0.0051	0.6654	2.5962	4.6046
$\mathrm{AB}(\widehat{\Omega}_{0.25})$	—	—	—	0.0129	0.0066	0.0057	0.3246	1.8825	4.1346
$\mathrm{AB}(\widehat{\Omega}_{0.50})$	—	—	0.0108	0.0092	0.0070	0.0064	0.1225	1.0068	3.2077
$\mathrm{AB}(\widehat{\Omega}_{0.75})$	—	0.0095	0.0082	0.0083	0.0077	0.0074	0.0445	0.3185	1.5368
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{T}})$	0.0092	0.0097	0.0089	0.0093	0.0090	0.0090	0.0206	0.0865	0.2739
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{G}})$	2×10^7	61.9094	0.6033	0.0672	0.0122	0.0097	0.4645	6.8070	21.3399
$\mathrm{MSE}(\widehat{\Omega}_{0.25})$	—	—	—	0.0306	0.0124	0.0108	0.1199	3.5933	17.2155
$\mathrm{MSE}(\widehat{\Omega}_{0.50})$	—	—	0.0273	0.0177	0.0132	0.0123	0.0270	1.0445	10.3807
$\mathrm{MSE}(\widehat{\Omega}_{0.75})$	—	0.0184	0.0161	0.0157	0.0148	0.0145	0.0158	0.1192	2.4132
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{T}})$	0.0182	0.0180	0.0179	0.0180	0.0179	0.0180	0.0180	0.0247	0.0977
$\mathrm{RE}(\widehat{\Omega}_{\mathrm{T}})$	9×10^8	3×10^3	33.6585	3.7324	0.6828	0.5409	25.8446	275.8591	218.3199

Moderate sample ($n = 1000$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{G}})$	102.0371	0.1098	0.0102	0.0029	0.0007	0.0007	0.6413	2.5189	4.4619
$\mathrm{AB}(\widehat{\Omega}_{0.25})$	—	—	—	0.0014	0.0007	0.0007	0.3119	1.8297	4.0141
$\mathrm{AB}(\widehat{\Omega}_{0.50})$	—	—	0.0012	0.0009	0.0007	0.0008	0.1154	0.9804	3.1178
$\mathrm{AB}(\widehat{\Omega}_{0.75})$	—	0.0010	0.0010	0.0009	0.0008	0.0009	0.0379	0.3104	1.5005
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{T}})$	0.0010	0.0010	0.0010	0.0010	0.0009	0.0010	0.0127	0.0795	0.2676
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{G}})$	5×10^7	8.9332	0.0800	0.0060	0.0011	0.0009	0.4264	6.3844	19.9645
$\mathrm{MSE}(\widehat{\Omega}_{0.25})$	—	—	—	0.0024	0.0011	0.0010	0.1050	3.3803	16.1651
$\mathrm{MSE}(\widehat{\Omega}_{0.50})$	—	—	0.0019	0.0015	0.0012	0.0011	0.0161	0.9826	9.7658
$\mathrm{MSE}(\widehat{\Omega}_{0.75})$	—	0.0016	0.0015	0.0014	0.0013	0.0013	0.0029	0.1043	2.2812
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{T}})$	0.0016	0.0016	0.0016	0.0016	0.0016	0.0016	0.0017	0.0088	0.0790
$\mathrm{RE}(\widehat{\Omega}_{\mathrm{T}})$	3×10^{10}	6×10^3	50.1581	3.8112	0.7120	0.5693	245.0330	727.1088	252.7565

Large sample ($n = 10000$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{G}})$	128.8058	0.0755	0.0032	0.0005	0.0001	0.0001	0.6390	2.5112	4.4480
$\mathrm{AB}(\widehat{\Omega}_{0.25})$	—	—	—	0.0002	0.0001	0.0001	0.3108	1.8243	4.0024
$\mathrm{AB}(\widehat{\Omega}_{0.50})$	—	—	0.0002	0.0001	0.0001	0.0001	0.1147	0.9777	3.1092
$\mathrm{AB}(\widehat{\Omega}_{0.75})$	—	0.0002	0.0001	0.0001	0.0001	0.0001	0.0373	0.3096	1.4971
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{T}})$	0.0002	0.0002	0.0001	0.0001	0.0002	0.0001	0.0120	0.0788	0.2671
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{G}})$	2×10^8	4.6930	0.0106	0.0009	0.0001	0.0001	0.4228	6.3431	19.8327
$\mathrm{MSE}(\widehat{\Omega}_{0.25})$	—	—	—	0.0002	0.0001	0.0001	0.1036	3.3595	16.0654
$\mathrm{MSE}(\widehat{\Omega}_{0.50})$	—	—	0.0002	0.0002	0.0001	0.0001	0.0151	0.9766	9.7083
$\mathrm{MSE}(\widehat{\Omega}_{0.75})$	—	0.0002	0.0001	0.0001	0.0001	0.0001	0.0018	0.1029	2.2690
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{T}})$	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0003	0.0074	0.0773
$\mathrm{RE}(\widehat{\Omega}_{\mathrm{T}})$	1×10^{12}	3×10^4	67.0743	6.0001	0.7146	0.5721	1×10^3	861.3457	256.4386

Table 4: Results for the location vector with incomplete data (MCAR).

Small sample ($n = 100$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\hat{\mu}_{\mathrm{G}})$	0.6890	0.0020	0.0019	0.0018	0.0006	0.0003	0.0911	0.4968	1.0000
$\mathrm{AB}(\hat{\mu}_{0.25})$	—	—	—	0.0017	0.0006	0.0003	0.0583	0.3794	0.8653
$\mathrm{AB}(\hat{\mu}_{0.50})$	—	—	0.0013	0.0017	0.0006	0.0003	0.0340	0.2439	0.6753
$\mathrm{AB}(\hat{\mu}_{0.75})$	—	0.0008	0.0012	0.0016	0.0006	0.0004	0.0192	0.1242	0.3921
$\mathrm{AB}(\hat{\mu}_{\mathrm{T}})$	0.0009	0.0008	0.0012	0.0015	0.0006	0.0004	0.0107	0.0611	0.1552
$\mathrm{MSE}(\hat{\mu}_{\mathrm{G}})$	5×10^3	0.1256	0.0341	0.0240	0.0122	0.0026	0.0215	0.2606	1.0116
$\mathrm{MSE}(\hat{\mu}_{0.25})$	—	—	—	0.0201	0.0123	0.0028	0.0161	0.1574	0.7609
$\mathrm{MSE}(\hat{\mu}_{0.50})$	—	—	0.0202	0.0176	0.0125	0.0030	0.0138	0.0730	0.4690
$\mathrm{MSE}(\hat{\mu}_{0.75})$	—	0.0207	0.0176	0.0162	0.0130	0.0033	0.0134	0.0292	0.1682
$\mathrm{MSE}(\hat{\mu}_{\mathrm{T}})$	0.0219	0.0176	0.0163	0.0155	0.0135	0.0036	0.0137	0.0179	0.0396
$\mathrm{RE}(\hat{\mu}_{\mathrm{T}})$	2×10^5	7.1467	2.0978	1.5453	0.9024	0.7391	1.5728	14.5373	25.5596

Moderate sample ($n = 1000$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\hat{\mu}_{\mathrm{G}})$	2.8973	0.0019	0.0006	0.0002	0.0003	0.0002	0.0994	0.4998	1.0000
$\mathrm{AB}(\hat{\mu}_{0.25})$	—	—	—	0.0002	0.0003	0.0002	0.0626	0.3818	0.8651
$\mathrm{AB}(\hat{\mu}_{0.50})$	—	—	0.0005	0.0001	0.0003	0.0002	0.0355	0.2462	0.6764
$\mathrm{AB}(\hat{\mu}_{0.75})$	—	0.0006	0.0005	0.0001	0.0003	0.0002	0.0195	0.1260	0.3959
$\mathrm{AB}(\hat{\mu}_{\mathrm{T}})$	0.0003	0.0005	0.0005	0.0002	0.0003	0.0002	0.0107	0.0618	0.1584
$\mathrm{MSE}(\hat{\mu}_{\mathrm{G}})$	1×10^5	0.0160	0.0035	0.0024	0.0012	0.0003	0.0111	0.2510	1.0011
$\mathrm{MSE}(\hat{\mu}_{0.25})$	—	—	—	0.0020	0.0012	0.0003	0.0052	0.1470	0.7497
$\mathrm{MSE}(\hat{\mu}_{0.50})$	—	—	0.0019	0.0017	0.0013	0.0003	0.0025	0.0619	0.4588
$\mathrm{MSE}(\hat{\mu}_{0.75})$	—	0.0020	0.0017	0.0016	0.0013	0.0003	0.0017	0.0172	0.1581
$\mathrm{MSE}(\hat{\mu}_{\mathrm{T}})$	0.0021	0.0017	0.0016	0.0015	0.0014	0.0003	0.0015	0.0052	0.0266
$\mathrm{RE}(\hat{\mu}_{\mathrm{T}})$	7×10^7	9.2500	2.2195	1.5890	0.8925	0.7270	7.5619	47.9879	37.6154

Large sample ($n = 10000$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\hat{\mu}_{\mathrm{G}})$	1.1483	0.0003	0.0002	0.0001	0.0001	0	0.1000	0.4999	1.0000
$\mathrm{AB}(\hat{\mu}_{0.25})$	—	—	—	0.0001	0.0001	0	0.0631	0.3819	0.8651
$\mathrm{AB}(\hat{\mu}_{0.50})$	—	—	0.0001	0.0001	0.0001	0	0.0358	0.2465	0.6764
$\mathrm{AB}(\hat{\mu}_{0.75})$	—	0.0001	0.0001	0	0.0001	0	0.0197	0.1264	0.3960
$\mathrm{AB}(\hat{\mu}_{\mathrm{T}})$	0	0.0001	0.0001	0	0.0001	0	0.0108	0.0621	0.1584
$\mathrm{MSE}(\hat{\mu}_{\mathrm{G}})$	2×10^4	0.0021	0.0003	0.0002	0.0001	0	0.0101	0.2501	1.0001
$\mathrm{MSE}(\hat{\mu}_{0.25})$	—	—	—	0.0002	0.0001	0	0.0041	0.1460	0.7485
$\mathrm{MSE}(\hat{\mu}_{0.50})$	—	—	0.0002	0.0002	0.0001	0	0.0014	0.0609	0.4576
$\mathrm{MSE}(\hat{\mu}_{0.75})$	—	0.0002	0.0002	0.0002	0.0001	0	0.0005	0.0161	0.1570
$\mathrm{MSE}(\hat{\mu}_{\mathrm{T}})$	0.0002	0.0002	0.0002	0.0002	0.0001	0	0.0003	0.0040	0.0253
$\mathrm{RE}(\hat{\mu}_{\mathrm{T}})$	7×10^7	12.0377	2.2789	1.5840	0.8926	0.7259	40.0740	62.5207	39.6063

Table 5: Results for the shape matrix with incomplete data (MCAR).

Small sample ($n = 100$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{G}})$	287.2054	0.1696	0.0413	0.0236	0.0092	0.0074	0.5704	2.6018	4.6651
$\mathrm{AB}(\widehat{\Omega}_{0.25})$	—	—	—	0.0165	0.0093	0.0081	0.2856	1.8437	4.0901
$\mathrm{AB}(\widehat{\Omega}_{0.50})$	—	—	0.0148	0.0125	0.0098	0.0092	0.1171	0.9574	3.0229
$\mathrm{AB}(\widehat{\Omega}_{0.75})$	—	0.0129	0.0121	0.0116	0.0109	0.0108	0.0471	0.3110	1.3323
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{T}})$	0.0132	0.0132	0.0136	0.0132	0.0130	0.0133	0.0243	0.0924	0.2682
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{G}})$	6×10^{8}	11.7595	0.2093	0.0769	0.0178	0.0140	0.4394	6.9380	21.9404
$\mathrm{MSE}(\widehat{\Omega}_{0.25})$	—	—	—	0.0398	0.0181	0.0155	0.1239	3.5027	16.8824
$\mathrm{MSE}(\widehat{\Omega}_{0.50})$	—	—	0.0318	0.0257	0.0192	0.0178	0.0352	0.9679	9.2563
$\mathrm{MSE}(\widehat{\Omega}_{0.75})$	—	0.0270	0.0238	0.0232	0.0218	0.0212	0.0230	0.1228	1.8437
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{T}})$	0.0270	0.0270	0.0269	0.0270	0.0269	0.0269	0.0269	0.0341	0.1033
$\mathrm{RE}(\widehat{\Omega}_{\mathrm{T}})$	2×10^{10}	435.1656	7.7806	2.8476	0.6612	0.5210	16.3347	203.2369	212.4660

Moderate sample ($n = 1000$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{G}})$	2×10^{3}	0.1030	0.0121	0.0041	0.0010	0.0008	0.6373	2.5212	4.4656
$\mathrm{AB}(\widehat{\Omega}_{0.25})$	—	—	—	0.0021	0.0010	0.0009	0.3069	1.7903	3.9256
$\mathrm{AB}(\widehat{\Omega}_{0.50})$	—	—	0.0016	0.0014	0.0010	0.0010	0.1147	0.9305	2.9055
$\mathrm{AB}(\widehat{\Omega}_{0.75})$	—	0.0014	0.0013	0.0013	0.0011	0.0012	0.0385	0.2981	1.2843
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{T}})$	0.0014	0.0014	0.0014	0.0014	0.0014	0.0014	0.0133	0.0798	0.2529
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{G}})$	3×10^{10}	9.3080	0.0689	0.0086	0.0015	0.0012	0.4230	6.3970	20.0003
$\mathrm{MSE}(\widehat{\Omega}_{0.25})$	—	—	—	0.0033	0.0016	0.0013	0.1030	3.2381	15.4641
$\mathrm{MSE}(\widehat{\Omega}_{0.50})$	—	—	0.0027	0.0021	0.0017	0.0015	0.0166	0.8874	8.4878
$\mathrm{MSE}(\widehat{\Omega}_{0.75})$	—	0.0022	0.0020	0.0019	0.0018	0.0018	0.0035	0.0971	1.6778
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{T}})$	0.0022	0.0022	0.0022	0.0022	0.0022	0.0022	0.0024	0.0094	0.0715
$\mathrm{RE}(\widehat{\Omega}_{\mathrm{T}})$	1×10^{13}	4×10^{3}	31.1264	3.8627	0.6919	0.5533	177.9689	678.5405	279.5977

Large sample ($n = 10000$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{G}})$	123.2153	0.0829	0.0032	0.0007	0.0001	0.0001	0.6386	2.5113	4.4484
$\mathrm{AB}(\widehat{\Omega}_{0.25})$	—	—	—	0.0003	0.0001	0.0001	0.3080	1.7838	3.9110
$\mathrm{AB}(\widehat{\Omega}_{0.50})$	—	—	0.0002	0.0002	0.0001	0.0001	0.1145	0.9276	2.8946
$\mathrm{AB}(\widehat{\Omega}_{0.75})$	—	0.0002	0.0002	0.0002	0.0002	0.0002	0.0377	0.2971	1.2790
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{T}})$	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0122	0.0789	0.2509
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{G}})$	1×10^{8}	9.4276	0.0086	0.0011	0.0002	0.0001	0.4225	6.3439	19.8366
$\mathrm{MSE}(\widehat{\Omega}_{0.25})$	—	—	—	0.0003	0.0002	0.0001	0.1019	3.2128	15.3420
$\mathrm{MSE}(\widehat{\Omega}_{0.50})$	—	—	0.0003	0.0002	0.0002	0.0002	0.0151	0.8804	8.4186
$\mathrm{MSE}(\widehat{\Omega}_{0.75})$	—	0.0002	0.0002	0.0002	0.0002	0.0002	0.0019	0.0951	1.6611
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{T}})$	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0004	0.0074	0.0686
$\mathrm{RE}(\widehat{\Omega}_{\mathrm{T}})$	5×10^{11}	4×10^{4}	39.5784	4.9955	0.6949	0.5560	1×10^{3}	853.3625	289.3145

Table 6: Results for the location vector with incomplete data (MAR).

Small sample ($n = 100$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\hat\mu_{\mathrm{G}})$	0.8378	0.0031	0.0015	0.0027	0.0008	0.0008	0.1281	0.4305	0.9339
$\mathrm{AB}(\hat\mu_{0.25})$	—	—	—	0.0052	0.0032	0.0019	0.0964	0.3082	0.7961
$\mathrm{AB}(\hat\mu_{0.50})$	—	—	0.0074	0.0077	0.0059	0.0032	0.0598	0.2112	0.6100
$\mathrm{AB}(\hat\mu_{0.75})$	—	0.0112	0.0100	0.0103	0.0089	0.0047	0.0296	0.1202	0.3356
$\mathrm{AB}(\hat\mu_{\mathrm{T}})$	0.0153	0.0131	0.0126	0.0124	0.0120	0.0065	0.0186	0.0536	0.1576
$\mathrm{MSE}(\hat\mu_{\mathrm{G}})$	6×10^3	0.1599	0.0505	0.0368	0.0210	0.0050	0.0378	0.2222	0.9056
$\mathrm{MSE}(\hat\mu_{0.25})$	—	—	—	0.0312	0.0214	0.0053	0.0334	0.1331	0.6695
$\mathrm{MSE}(\hat\mu_{0.50})$	—	—	0.0311	0.0282	0.0225	0.0057	0.0271	0.0678	0.4106
$\mathrm{MSE}(\hat\mu_{0.75})$	—	0.0346	0.0299	0.0286	0.0246	0.0063	0.0241	0.0367	0.1530
$\mathrm{MSE}(\hat\mu_{\mathrm{T}})$	0.0237	0.0188	0.0171	0.0166	0.0146	0.0038	0.0147	0.0184	0.0414
$\mathrm{RE}(\hat\mu_{\mathrm{T}})$	2×10^5	8.4895	2.9493	2.2188	1.4384	1.2917	2.5810	12.1041	21.8539

Moderate sample ($n = 1000$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\hat\mu_{\mathrm{G}})$	1.9746	0.0011	0.0003	0.0003	0.0003	0.0001	0.1271	0.4309	0.9342
$\mathrm{AB}(\hat\mu_{0.25})$	—	—	—	0.0030	0.0026	0.0012	0.0952	0.3084	0.7962
$\mathrm{AB}(\hat\mu_{0.50})$	—	—	0.0063	0.0059	0.0052	0.0025	0.0583	0.2122	0.6111
$\mathrm{AB}(\hat\mu_{0.75})$	—	0.0098	0.0092	0.0089	0.0081	0.0041	0.0292	0.1218	0.3389
$\mathrm{AB}(\hat\mu_{\mathrm{T}})$	0.0146	0.0129	0.0124	0.0123	0.0113	0.0060	0.0189	0.0552	0.1610
$\mathrm{MSE}(\hat\mu_{\mathrm{G}})$	7×10^4	0.0382	0.0086	0.0047	0.0019	0.0005	0.0208	0.2065	0.8916
$\mathrm{MSE}(\hat\mu_{0.25})$	—	—	—	0.0032	0.0020	0.0005	0.0160	0.1166	0.6542
$\mathrm{MSE}(\hat\mu_{0.50})$	—	—	0.0032	0.0027	0.0021	0.0006	0.0088	0.0503	0.3951
$\mathrm{MSE}(\hat\mu_{0.75})$	—	0.0035	0.0031	0.0029	0.0025	0.0007	0.0033	0.0173	0.1361
$\mathrm{MSE}(\hat\mu_{\mathrm{T}})$	0.0032	0.0026	0.0024	0.0023	0.0020	0.0005	0.0019	0.0053	0.0287
$\mathrm{RE}(\hat\mu_{\mathrm{T}})$	2×10^7	14.8351	3.5916	2.0513	0.9573	0.8537	10.7778	38.7109	31.0715

Large sample ($n = 10000$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\hat\mu_{\mathrm{G}})$	0.7493	0.0005	0.0007	0.0002	0.0001	0	0.1269	0.4311	0.9341
$\mathrm{AB}(\hat\mu_{0.25})$	—	—	—	0.0029	0.0024	0.0012	0.0951	0.3087	0.7961
$\mathrm{AB}(\hat\mu_{0.50})$	—	—	0.0061	0.0058	0.0050	0.0025	0.0582	0.2119	0.6110
$\mathrm{AB}(\hat\mu_{0.75})$	—	0.0098	0.0090	0.0088	0.0079	0.0040	0.0291	0.1215	0.3390
$\mathrm{AB}(\hat\mu_{\mathrm{T}})$	0.0142	0.0128	0.0123	0.0120	0.0113	0.0058	0.0188	0.0555	0.1612
$\mathrm{MSE}(\hat\mu_{\mathrm{G}})$	1×10^4	0.0192	0.0025	0.0007	0.0002	0	0.0192	0.2050	0.8901
$\mathrm{MSE}(\hat\mu_{0.25})$	—	—	—	0.0004	0.0002	0.0001	0.0143	0.1150	0.6525
$\mathrm{MSE}(\hat\mu_{0.50})$	—	—	0.0005	0.0004	0.0003	0.0001	0.0071	0.0485	0.3935
$\mathrm{MSE}(\hat\mu_{0.75})$	—	0.0008	0.0007	0.0006	0.0005	0.0001	0.0014	0.0153	0.1344
$\mathrm{MSE}(\hat\mu_{\mathrm{T}})$	0.0012	0.0010	0.0009	0.0009	0.0008	0.0002	0.0007	0.0041	0.0274
$\mathrm{RE}(\hat\mu_{\mathrm{T}})$	1×10^7	19.5087	2.7638	0.8221	0.2507	0.2256	28.1668	50.4076	32.5334

Table 7: Results for the shape matrix with incomplete data (MAR).

Small sample ($n = 100$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{G}})$	142.4627	0.1917	0.0473	0.0264	0.0110	0.0093	0.7344	2.7633	4.9146
$\mathrm{AB}(\widehat{\Omega}_{0.25})$	—	—	—	0.0189	0.0112	0.0101	0.3871	2.0270	4.4403
$\mathrm{AB}(\widehat{\Omega}_{0.50})$	—	—	0.0165	0.0147	0.0119	0.0113	0.1694	1.1269	3.4790
$\mathrm{AB}(\widehat{\Omega}_{0.75})$	—	0.0155	0.0138	0.0139	0.0135	0.0132	0.0681	0.4085	1.7448
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{T}})$	0.0126	0.0124	0.0126	0.0124	0.0126	0.0125	0.0264	0.1124	0.3501
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{G}})$	2×10^{8}	26.2678	0.6868	0.0781	0.0219	0.0183	0.5733	7.7329	24.3710
$\mathrm{MSE}(\widehat{\Omega}_{0.25})$	—	—	—	0.0440	0.0223	0.0202	0.1765	4.1801	19.9067
$\mathrm{MSE}(\widehat{\Omega}_{0.50})$	—	—	0.0460	0.0302	0.0240	0.0231	0.0516	1.3168	12.2466
$\mathrm{MSE}(\widehat{\Omega}_{0.75})$	—	0.0326	0.0296	0.0287	0.0279	0.0276	0.0295	0.1988	3.1250
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{T}})$	0.0255	0.0253	0.0256	0.0255	0.0253	0.0253	0.0251	0.0368	0.1554
$\mathrm{RE}(\widehat{\Omega}_{\mathrm{T}})$	6×10^{9}	1×10^{3}	26.7828	3.0648	0.8625	0.7231	22.7975	210.1198	156.8042

Moderate sample ($n = 1000$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{G}})$	703.3148	0.1459	0.0110	0.0043	0.0013	0.0010	0.6990	2.6460	4.7007
$\mathrm{AB}(\widehat{\Omega}_{0.25})$	—	—	—	0.0022	0.0013	0.0011	0.3667	1.9444	4.2560
$\mathrm{AB}(\widehat{\Omega}_{0.50})$	—	—	0.0017	0.0015	0.0014	0.0012	0.1564	1.0811	3.3354
$\mathrm{AB}(\widehat{\Omega}_{0.75})$	—	0.0016	0.0013	0.0016	0.0017	0.0015	0.0565	0.3884	1.6715
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{T}})$	0.0032	0.0031	0.0028	0.0031	0.0032	0.0031	0.0174	0.1003	0.3313
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{G}})$	5×10^{9}	40.4568	0.0393	0.0096	0.0018	0.0015	0.5113	7.0528	22.1746
$\mathrm{MSE}(\widehat{\Omega}_{0.25})$	—	—	—	0.0038	0.0019	0.0017	0.1491	3.8234	18.1861
$\mathrm{MSE}(\widehat{\Omega}_{0.50})$	—	—	0.0030	0.0025	0.0020	0.0019	0.0317	1.1982	11.1862
$\mathrm{MSE}(\widehat{\Omega}_{0.75})$	—	0.0026	0.0024	0.0023	0.0023	0.0022	0.0063	0.1652	2.8333
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{T}})$	0.0022	0.0022	0.0022	0.0022	0.0022	0.0022	0.0024	0.0140	0.1205
$\mathrm{RE}(\widehat{\Omega}_{\mathrm{T}})$	2×10^{12}	2×10^{4}	18.2494	4.4252	0.8462	0.7076	210.9266	504.7486	183.9650

Large sample ($n = 10000$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{G}})$	326.2852	0.1450	0.0069	0.0008	0.0001	0.0001	0.6957	2.6343	4.6813
$\mathrm{AB}(\widehat{\Omega}_{0.25})$	—	—	—	0.0003	0.0001	0.0002	0.3648	1.9363	4.2393
$\mathrm{AB}(\widehat{\Omega}_{0.50})$	—	—	0.0004	0.0005	0.0004	0.0005	0.1552	1.0766	3.3224
$\mathrm{AB}(\widehat{\Omega}_{0.75})$	—	0.0010	0.0010	0.0011	0.0010	0.0011	0.0556	0.3866	1.6650
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{T}})$	0.0023	0.0023	0.0023	0.0023	0.0022	0.0023	0.0169	0.0993	0.3295
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{G}})$	7×10^{8}	115.5289	0.2165	0.0012	0.0002	0.0002	0.5058	6.9877	21.9816
$\mathrm{MSE}(\widehat{\Omega}_{0.25})$	—	—	—	0.0004	0.0002	0.0002	0.1467	3.7893	18.0352
$\mathrm{MSE}(\widehat{\Omega}_{0.50})$	—	—	0.0003	0.0002	0.0002	0.0002	0.0300	1.1872	11.0936
$\mathrm{MSE}(\widehat{\Omega}_{0.75})$	—	0.0003	0.0002	0.0002	0.0002	0.0002	0.0045	0.1624	2.8084
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{T}})$	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0005	0.0121	0.1175
$\mathrm{RE}(\widehat{\Omega}_{\mathrm{T}})$	3×10^{12}	5×10^{5}	893.0252	4.9298	0.7389	0.6126	927.1899	576.8705	186.9984

Table 8: Results for the location vector with incomplete data (NMAR).

Small sample ($n = 100$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\hat{\mu}_{\mathrm{G}})$	1.5783	0.2621	0.2137	0.1967	0.1600	0.0768	0.2600	0.6543	1.1454
$\mathrm{AB}(\hat{\mu}_{0.25})$	—	—	—	0.1883	0.1581	0.0769	0.2218	0.5374	1.0147
$\mathrm{AB}(\hat{\mu}_{0.50})$	—	—	0.1895	0.1799	0.1560	0.0769	0.1925	0.4037	0.8365
$\mathrm{AB}(\hat{\mu}_{0.75})$	—	0.1940	0.1782	0.1717	0.1539	0.0769	0.1733	0.2818	0.5672
$\mathrm{AB}(\hat{\mu}_{\mathrm{T}})$	0.2037	0.1771	0.1674	0.1635	0.1513	0.0768	0.1611	0.2144	0.3216
$\mathrm{MSE}(\hat{\mu}_{\mathrm{G}})$	9×10^3	0.4533	0.2560	0.2123	0.1368	0.0314	0.1798	0.5321	1.4060
$\mathrm{MSE}(\hat{\mu}_{0.25})$	—	—	—	0.1927	0.1338	0.0316	0.1610	0.3942	1.1248
$\mathrm{MSE}(\hat{\mu}_{0.50})$	—	—	0.1947	0.1752	0.1307	0.0318	0.1478	0.2707	0.7975
$\mathrm{MSE}(\hat{\mu}_{0.75})$	—	0.2047	0.1721	0.1596	0.1277	0.0320	0.1380	0.1906	0.4254
$\mathrm{MSE}(\hat{\mu}_{\mathrm{T}})$	0.2261	0.1707	0.1527	0.1455	0.1243	0.0322	0.1299	0.1593	0.2258
$\mathrm{RE}(\hat{\mu}_{\mathrm{T}})$	4×10^4	2.6550	1.6761	1.4589	1.1011	0.9753	1.3842	3.3412	6.2255

Moderate sample ($n = 1000$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\hat{\mu}_{\mathrm{G}})$	1.5171	0.2759	0.2193	0.1996	0.1599	0.0764	0.2601	0.6538	1.1457
$\mathrm{AB}(\hat{\mu}_{0.25})$	—	—	—	0.1890	0.1576	0.0765	0.2217	0.5367	1.0143
$\mathrm{AB}(\hat{\mu}_{0.50})$	—	—	0.1893	0.1793	0.1553	0.0765	0.1923	0.4035	0.8369
$\mathrm{AB}(\hat{\mu}_{0.75})$	—	0.1921	0.1770	0.1704	0.1529	0.0766	0.1729	0.2820	0.5697
$\mathrm{AB}(\hat{\mu}_{\mathrm{T}})$	0.2006	0.1748	0.1661	0.1621	0.1504	0.0765	0.1609	0.2146	0.3247
$\mathrm{MSE}(\hat{\mu}_{\mathrm{G}})$	2×10^3	0.3952	0.2429	0.2008	0.1283	0.0294	0.1710	0.5229	1.3983
$\mathrm{MSE}(\hat{\mu}_{0.25})$	—	—	—	0.1799	0.1247	0.0294	0.1518	0.3842	1.1152
$\mathrm{MSE}(\hat{\mu}_{0.50})$	—	—	0.1805	0.1618	0.1211	0.0295	0.1380	0.2606	0.7886
$\mathrm{MSE}(\hat{\mu}_{0.75})$	—	0.1856	0.1578	0.1461	0.1174	0.0295	0.1275	0.1798	0.4174
$\mathrm{MSE}(\hat{\mu}_{\mathrm{T}})$	0.2026	0.1537	0.1391	0.1323	0.1137	0.0295	0.1193	0.1479	0.2157
$\mathrm{RE}(\hat{\mu}_{\mathrm{T}})$	1×10^4	2.5712	1.7463	1.5173	1.1286	0.9952	1.4340	3.5361	6.4809

Large sample ($n = 10000$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\hat{\mu}_{\mathrm{G}})$	2.3292	0.2810	0.2204	0.2000	0.1596	0.0764	0.2600	0.6537	1.1456
$\mathrm{AB}(\hat{\mu}_{0.25})$	—	—	—	0.1891	0.1573	0.0764	0.2216	0.5366	1.0142
$\mathrm{AB}(\hat{\mu}_{0.50})$	—	—	0.1892	0.1792	0.1550	0.0765	0.1921	0.4033	0.8368
$\mathrm{AB}(\hat{\mu}_{0.75})$	—	0.1917	0.1768	0.1702	0.1526	0.0765	0.1727	0.2819	0.5699
$\mathrm{AB}(\hat{\mu}_{\mathrm{T}})$	0.2001	0.1743	0.1659	0.1619	0.1501	0.0765	0.1608	0.2146	0.3249
$\mathrm{MSE}(\hat{\mu}_{\mathrm{G}})$	2×10^4	0.3956	0.2429	0.2001	0.1274	0.0292	0.1701	0.5219	1.3974
$\mathrm{MSE}(\hat{\mu}_{0.25})$	—	—	—	0.1788	0.1238	0.0292	0.1508	0.3831	1.1142
$\mathrm{MSE}(\hat{\mu}_{0.50})$	—	—	0.1791	0.1605	0.1201	0.0292	0.1370	0.2595	0.7876
$\mathrm{MSE}(\hat{\mu}_{0.75})$	—	0.1837	0.1564	0.1448	0.1164	0.0293	0.1264	0.1786	0.4165
$\mathrm{MSE}(\hat{\mu}_{\mathrm{T}})$	0.2003	0.1519	0.1377	0.1310	0.1126	0.0293	0.1181	0.1467	0.2147
$\mathrm{RE}(\hat{\mu}_{\mathrm{T}})$	8×10^4	2.6043	1.7638	1.5267	1.1319	0.9973	1.4393	3.5570	6.5092

Table 9: Results for the shape matrix with incomplete data (NMAR).

Small sample ($n = 100$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{G}})$	283.6810	0.1847	0.0774	0.0657	0.0674	0.0723	0.8097	3.0516	5.4274
$\mathrm{AB}(\widehat{\Omega}_{0.25})$	—	—	—	0.0603	0.0679	0.0735	0.4329	2.2332	4.8967
$\mathrm{AB}(\widehat{\Omega}_{0.50})$	—	—	0.0561	0.0583	0.0688	0.0751	0.2087	1.2313	3.8240
$\mathrm{AB}(\widehat{\Omega}_{0.75})$	—	0.0479	0.0559	0.0590	0.0702	0.0771	0.1140	0.4461	1.8948
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{T}})$	0.0341	0.0512	0.0587	0.0616	0.0727	0.0801	0.0850	0.1684	0.4211
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{G}})$	1×10^{9}	13.8695	0.3339	0.0952	0.0416	0.0420	0.6881	9.4293	29.7487
$\mathrm{MSE}(\widehat{\Omega}_{0.25})$	—	—	—	0.0572	0.0422	0.0443	0.2033	5.0664	24.2249
$\mathrm{MSE}(\widehat{\Omega}_{0.50})$	—	—	0.0494	0.0423	0.0438	0.0474	0.0605	1.5629	14.7967
$\mathrm{MSE}(\widehat{\Omega}_{0.75})$	—	0.0377	0.0394	0.0402	0.0470	0.0521	0.0434	0.2208	3.6744
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{T}})$	0.0323	0.0389	0.0437	0.0452	0.0535	0.0597	0.0513	0.0591	0.2103
$\mathrm{RE}(\widehat{\Omega}_{\mathrm{T}})$	4×10^{10}	356.3258	7.6470	2.1062	0.7778	0.7030	13.4144	159.5002	141.4512

Moderate sample ($n = 1000$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{G}})$	61.6154	0.1250	0.0452	0.0455	0.0591	0.0653	0.7708	2.9225	5.1907
$\mathrm{AB}(\widehat{\Omega}_{0.25})$	—	—	—	0.0467	0.0599	0.0661	0.4136	2.1441	4.6965
$\mathrm{AB}(\widehat{\Omega}_{0.50})$	—	—	0.0441	0.0486	0.0606	0.0669	0.1978	1.1836	3.6719
$\mathrm{AB}(\widehat{\Omega}_{0.75})$	—	0.0387	0.0465	0.0504	0.0613	0.0678	0.1052	0.4296	1.8222
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{T}})$	0.0247	0.0418	0.0483	0.0516	0.0619	0.0688	0.0749	0.1578	0.4047
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{G}})$	2×10^{7}	24.4671	0.1149	0.0215	0.0222	0.0261	0.6168	8.6056	27.0691
$\mathrm{MSE}(\widehat{\Omega}_{0.25})$	—	—	—	0.0167	0.0227	0.0267	0.1759	4.6451	22.1659
$\mathrm{MSE}(\widehat{\Omega}_{0.50})$	—	—	0.0147	0.0166	0.0232	0.0275	0.0429	1.4316	13.5643
$\mathrm{MSE}(\widehat{\Omega}_{0.75})$	—	0.0117	0.0153	0.0173	0.0238	0.0283	0.0228	0.1899	3.3639
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{T}})$	0.0062	0.0131	0.0164	0.0183	0.0245	0.0294	0.0233	0.0325	0.1689
$\mathrm{RE}(\widehat{\Omega}_{\mathrm{T}})$	4×10^{9}	2×10^{3}	7.0127	1.1762	0.9076	0.8873	26.4378	265.1185	160.2652

Large sample ($n = 10000$)

	t_1	t_2	t_3	t_4	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{G}})$	322.1516	0.0729	0.0385	0.0416	0.0582	0.0646	0.7674	2.9098	5.1690
$\mathrm{AB}(\widehat{\Omega}_{0.25})$	—	—	—	0.0450	0.0590	0.0654	0.4119	2.1353	4.6783
$\mathrm{AB}(\widehat{\Omega}_{0.50})$	—	—	0.0428	0.0476	0.0596	0.0662	0.1968	1.1788	3.6583
$\mathrm{AB}(\widehat{\Omega}_{0.75})$	—	0.0378	0.0456	0.0494	0.0602	0.0670	0.1044	0.4277	1.8158
$\mathrm{AB}(\widehat{\Omega}_{\mathrm{T}})$	0.0238	0.0408	0.0472	0.0506	0.0606	0.0678	0.0741	0.1565	0.4032
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{G}})$	1×10^{9}	4.3763	0.4222	0.0122	0.0205	0.0247	0.6107	8.5269	26.8312
$\mathrm{MSE}(\widehat{\Omega}_{0.25})$	—	—	—	0.0131	0.0210	0.0252	0.1735	4.6047	21.9843
$\mathrm{MSE}(\widehat{\Omega}_{0.50})$	—	—	0.0120	0.0144	0.0214	0.0258	0.0413	1.4187	13.4568
$\mathrm{MSE}(\widehat{\Omega}_{0.75})$	—	0.0095	0.0133	0.0154	0.0218	0.0263	0.0211	0.1869	3.3371
$\mathrm{MSE}(\widehat{\Omega}_{\mathrm{T}})$	0.0041	0.0110	0.0142	0.0160	0.0221	0.0269	0.0211	0.0301	0.1654
$\mathrm{RE}(\widehat{\Omega}_{\mathrm{T}})$	2×10^{11}	399.5492	29.7291	0.7632	0.9286	0.9174	29.0087	282.9399	162.2019

References

Allison, P. (2001): *Missing Data*, Sage Publications.

Boente, G., González-Manteiga, W., Pérez-González, A. (2009): "Robust nonparametric estimation with missing data," *Journal of Statistical Planning and Inference* **139**, pp. 571–592.

Bradley, R. (2005): "Basic properties of strong mixing conditions. a survey and some open questions," *Probability Surveys* **2**, pp. 107–144.

Branden, K., Verboven, S. (2009): "Robust data imputation," *Computational Biology and Chemistry* **33**, pp. 7–13.

Cambanis, S., Huang, S., Simons, G. (1981): "On the theory of elliptically contoured distributions," *Journal of Multivariate Analysis* **11**, pp. 368–385.

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

Copt, S., Victoria-Feser, M.P. (2004): "Fast algorithms for computing high breakdown covariance matrices with missing data," in M. Hubert, G. Pison, A. Struyf, S. van Aelst (editors), "Theory and Applications of Recent Robust Methods," Birkhäuser, pp. 71–82.

Danilov, M., Yohai, V., Zamar, R. (2012): "Robust estimation of multivariate location and scatter in the presence of missing data," *Journal of the American Statistical Association* **107**, pp. 1178–1186.

Dümbgen, L., Nordhausen, K., Schuhmacher, H. (2016): "New algorithms for M-estimation of multivariate scatter and location," *Journal of Multivariate Analysis* **144**, pp. 200–217.

Dümbgen, L., Pauly, M., Schweizer, T. (2015): "M-functionals of multivariate scatter," *Statistics Surveys* **9**, pp. 32–105.

Flossmann, A. (2010): "Accounting for missing data in M-estimation: a general matching approach," *Empirical Economics* **38**, pp. 85–117.

Frahm, G. (2009): "Asymptotic distributions of robust shape matrices and scales," *Journal of Multivariate Analysis* **100**, pp. 1329–1337.

Frahm, G., Jaekel, U. (2010): "A generalization of Tyler's M-estimators to the case of incomplete data," *Computational Statistics and Data Analysis* **54**, pp. 374–393.

Gastwirth, J., Rubin, H. (1975): "The behavior of robust estimators on dependent data," *Annals of Statistics* **3**, pp. 1070–1100.

Hampel, F., Ronchetti, E., Rousseeuw, P., Stahel, W. (1986): *Robust Statistics*, John Wiley.

Han, P. (2014): “Multiply robust estimation in regression analysis with missing data,” *Journal of the American Statistical Association* **109**, pp. 1159–1173.

Han, P. (2016): “Combining inverse probability weighting and multiple imputation to improve robustness of estimation,” *Scandinavian Journal of Statistics* **43**, pp. 246–260.

Hansen, B. (1992): “Consistent covariance matrix estimation for dependent heterogeneous processes,” *Econometrica* **60**, pp. 967–972.

Hansen, L. (1982): “Large sample properties of generalized method of moments estimators,” *Econometrica* **50**, pp. 1029–1054.

Hettmansperger, T., Randles, R. (2002): “A practical affine equivariant multivariate median,” *Biometrika* **89**, pp. 851–860.

Hron, K., Templ, M., Filzmoser, P. (2010): “Imputation of missing values for compositional data using classical and robust methods,” *Computational Statistics and Data Analysis* **54**, pp. 3095–3107.

Huber, P. (2003): *Robust Statistics*, John Wiley.

Jamshidian, M., Jalal, S. (2010): “Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data,” *Psychometrika* **75**, pp. 649–674.

Jamshidian, M., Jalal, S., Jansen, C. (2014): “MissMech: An R package for testing homoscedasticity, multivariate normality, and missing completely at random (MCAR),” *Journal of Statistical Software* **56**, pp. 1–31.

Kent, J., Tyler, D. (1991): “Redescending M-estimates of multivariate location and scatter,” *Annals of Statistics* **19**, pp. 2102–2119.

Listing, J., Schlittgen, R. (2003): “A nonparametric test for random dropouts,” *Biometrical Journal* **45**, pp. 113–127.

Little, R. (1988a): “Robust estimation of the mean and covariance matrix from data with missing values,” *Journal of the Royal Statistical Society, Series C* **37**, pp. 23–38.

Little, R. (1988b): “A test of missing completely at random for multivariate data with missing values,” *Journal of the American Statistical Association* **83**, pp. 1198–1202.

Little, R., Rubin, D. (2002): *Statistical Analysis with Missing Data*, John Wiley, 2nd edition.

Little, R., Smith, P. (1987): “Editing and imputation for quantitative survey data,” *Journal of the American Statistical Association* **82**, pp. 58–68.

Markowitz, H. (1959): *Portfolio Selection: Efficient Diversification of Investments*, Wiley.

Markowitz, H. (1987): *Mean-Variance Analysis in Portfolio Choice and Capital Markets*, Basil Blackwell.

Maronna, R. (1976): “Robust M-estimators of multivariate location and scatter,” *Annals of Statistics* **4**, pp. 51–67.

Maronna, R., Martin, D., Yohai, V. (2006): *Robust Statistics*, John Wiley.

Paindaveine, D. (2008): “A canonical definition of shape,” *Statistics and Probability Letters* **78**, pp. 2240–2247.

Palma, W., del Pino, G. (1999): “Statistical analysis of incomplete long-range dependent data,” *Biometrika* **86**, pp. 965–972.

Politis, D. (2003): “The impact of bootstrap methods on time series analysis,” *Statistical Science* **18**, pp. 219–230.

Schafer, J. (1997): *Analysis of Incomplete Multivariate Data*, Chapman & Hall.

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

Sued, M., Yohai, V. (2013): “Robust location estimation with missing data,” *Canadian Journal of Statistics* **41**, pp. 111–132.

Templ, M., Kowarik, A., Filzmoser, P. (2011): “Iterative stepwise regression imputation using standard and robust methods,” *Computational Statistics and Data Analysis* **55**, pp. 2793–2806.

Tyler, D. (1982): “Radial estimates and the test for sphericity,” *Biometrika* **69**, pp. 429–436.

Tyler, D. (1987a): “A distribution-free M-estimator of multivariate scatter,” *Annals of Statistics* **15**, pp. 234–251.

Tyler, D. (1987b): “Statistical analysis for the angular central Gaussian distribution on the sphere,” *Biometrika* **74**, pp. 579–589.

van der Vaart, A. (1998): *Asymptotic Statistics*, Cambridge University Press.

Wang, J.L. (1999): “Asymptotic properties of M-estimators based on estimating equations and censored data,” *Scandinavian Journal of Statistics* **26**, pp. 297–318.

Wooldridge, J. (2007): “Inverse probability weighted estimation for general missing data problems,” *Journal of Econometrics* **141**, pp. 1281–1301.

Yuanxi, Y. (1994): “Robust estimation for dependent observations,” *Manuscripta Geodaetica* **19**, pp. 10–17.