DISCUSSION PAPERS IN STATISTICS
AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 3/07

A Generalization of Tyler's M-Estimators to the Case of Incomplete Data

by

Gabriel Frahm
Uwe Jaekel

7^{th} version
July 9, 2009

DISKUSSIONSBEITRÄGE ZUR
STATISTIK UND ÖKONOMETRIE

SEMINAR FÜR WIRTSCHAFTS- UND SOZIALSTATISTIK
UNIVERSITÄT ZU KÖLN

Albertus-Magnus-Platz, D-50923 Köln, Deutschland

This page intentionally left blank.

DISCUSSION PAPERS IN STATISTICS AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 3/07

A Generalization of Tyler's M-Estimators to the Case of Incomplete Data[1]

by

Gabriel Frahm[2]
Uwe Jaekel[3]

7^{th} version
July 9, 2009

Abstract

Many different robust estimation approaches for the covariance or shape matrix of multivariate data have been established until today. Tyler's M-estimator has been recognized as the 'most robust' M-estimator for the shape matrix of elliptically symmetric distributed data. Tyler's M-estimators for location and shape are generalized by taking account of incomplete data. It is shown that the shape matrix estimator remains distribution-free under the class of generalized elliptical distributions. Its asymptotic distribution is also derived and a fast algorithm, which works well even for high-dimensional data, is presented. A simulation study with clean and contaminated data covers the complete-data as well as the incomplete-data case, where the missing data are assumed to be MCAR, MAR, and NMAR.

Keywords: Covariance matrix; distribution-free estimation; missing data; robust estimation; shape matrix; sign-based estimator; Tyler's M-estimator.

AMS Subject Classification: Primary 62H12, Secondary 62H20.

[1]Previous versions of the manuscript have been circulated under the title *Distribution-free shape matrix estimation for incomplete data*.

[2]University of Cologne, Department of Economic and Social Statistics, Albertus-Magnus-Platz, D-50923 Cologne, Germany. Phone: +49 221 470-4267, email: frahm@statistik.uni-koeln.de.

[3]University of Applied Sciences Koblenz, RheinAhrCampus, Department of Mathematics and Technology, Südallee 2, D-53424 Remagen, Germany. Phone: +49 2642 932-334, email: jaekel@rheinahrcampus.de.

This page intentionally left blank.

A Generalization of Tyler's M-Estimators to the Case of Incomplete Data

Gabriel Frahm[1]

University of Cologne, Department of Economic and Social Statistics
Albertus-Magnus-Platz, D-50923 Cologne, Germany

Uwe Jaekel

University of Applied Sciences Koblenz, RheinAhrCampus Remagen
Department of Mathematics and Technology
Südallee 2, D-53424 Remagen, Germany

Abstract

Many different robust estimation approaches for the covariance or shape matrix of multivariate data have been established until today. Tyler's M-estimator has been recognized as the 'most robust' M-estimator for the shape matrix of elliptically symmetric distributed data. Tyler's M-estimators for location and shape are generalized by taking account of incomplete data. It is shown that the shape matrix estimator remains distribution-free under the class of generalized elliptical distributions. Its asymptotic distribution is also derived and a fast algorithm, which works well even for high-dimensional data, is presented. A simulation study with clean and contaminated data covers the complete-data as well as the incomplete-data case, where the missing data are assumed to be MCAR, MAR, and NMAR.

Key words: Covariance matrix, distribution-free estimation, missing data, robust estimation, shape matrix, sign-based estimator, Tyler's M-estimator.

1. Introduction

During the last three decades, robust covariance matrix estimation has become a popular branch of robust statistics. Many different estimation approaches have been established until today. For a broad overview on robust statistics see Hampel et al. (1986), Huber (2003), and Maronna et al. (2006). In the present work we will concentrate on a specific robust covariance matrix estimator, namely Tyler's celebrated M-estimator (Tyler, 1983, 1987a). Many authors have demonstrated its nice statistical properties and advantages compared to other covariance matrix estimators.

[1]Corresponding author. University of Cologne, Department of Economic and Social Statistics, Chair for Statistics & Econometrics, Meister-Ekkehart-Str. 9, D-50937 Cologne, Germany. Phone: +49 221 470-4267, fax: +49 221 470-5084, e-mail: frahm@statistik.uni-koeln.de.

However, a remaining question is how to deal with incomplete data. From our own academic and professional work we have come to know that in almost any practical situation multivariate data are incomplete. Therefore the problem of robust covariance matrix estimation under incomplete data is highly relevant both from a practical and academic perspective. There might exist several reasons for missing data and thus different kinds of missingness mechanisms can be found in the background of the data-generating process. Modern estimation procedures of missing-data analysis (Little and Rubin, 2002; Schafer and Graham, 2002) could be efficiently applied for estimating the covariance matrix if the true data-generating process was known. Traditional ML-theory works only if the proposed model is correct. If the suggested model does not correspond to the true one, the asymptotic distribution of covariance matrix estimators can be calculated on the basis of M-theory.

Nevertheless, there are some remaining difficulties regarding robust covariance matrix estimation. For example, the asymptotic distribution of an M-, R-, or S-estimator in general is determined by unknown quantities which depend on the true data-generating process (Frahm, 2009). Other robust estimation procedures which can be found in the literature are based on geometrical approaches (Visuri, 2001, Ch. 3) and a 'canonical' generalization to the missing-data problem does not seem to exist. To the best of our knowledge, M-estimators for incomplete data have been only discussed by Little (1988). We have not found any other approach to robust covariance matrix estimation taking missing data into consideration.

The paper is organized as follows. In Section 2 we describe the class of generalized elliptical distributions (Frahm, 2004, Ch. 3), which plays a fundamental role when analyzing Tyler's M-estimators for the location and shape of a distribution. We discuss Tyler's M-estimators in Section 3. More precisely, Section 3.1 starts with the complete-data case, where Tyler's estimator for the shape matrix is characterized both as an M-estimator and an ML-estimator. In Section 3.2 we turn to the incomplete-data case. After presenting our necessary instruments of missing data analysis, we motivate our generalization of Tyler's M-estimator for the shape matrix by means of likelihood theory. Here we also derive its asymptotic distribution. Further, in Section 3.3 we discuss the problem of estimating the location vector and present a generalized version of Tyler's corresponding M-estimator.

Section 4 contains a formal representation of the generalized M-estimators for location and shape. We derive a fast algorithm for calculating the estimates, which works well even for high-dimensional data with large numbers of missing values and give some practical advice for its numerical implementation. In Section 5 we provide a simulation study covering the complete-data as well as the incomplete-data case, using clean and contaminated data under different missingness mechanisms. Finally, Section 6 summarizes the results of the present work.

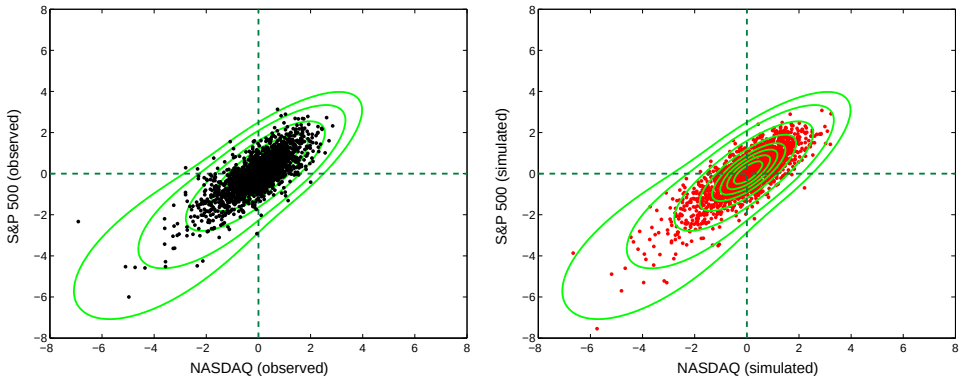

Figure 1: Observed GARCH(1,1)-residuals of NASDAQ and S&P 500 daily log-returns from 1993-01-01 to 2000-06-30 (left hand) and simulated generalized elliptically distributed residuals ($n = 1892$) (right hand). The density contours of the chosen model (Frahm and Jaekel, 2007) are given by the green curves.

2. Generalized Elliptical Distributions

Consider a d-dimensional elliptically symmetric distributed random vector X. This means X can be represented by

$$X = \mu + \Lambda \mathcal{R} U \,, \tag{1}$$

where U is a k-dimensional random vector, uniformly distributed on the unit hypersphere, $\mathcal{R}$ is a nonnegative random variable being stochastically independent of U, $\mu \in \mathbb{R}^d$, and $\Lambda \in \mathbb{R}^{d \times k}$ (Cambanis et al., 1981; Fang et al., 1990, p. 42). The parameter μ denotes the location vector and $\Sigma := \Lambda\Lambda^{\mathsf{T}}$ is referred to as the dispersion matrix of X. The random variable $\mathcal{R}$ is called the generating variate of X.

A d-dimensional random vector X is said to be generalized elliptically distributed (Frahm, 2004, Ch. 3) if it can be represented by Eq. 1, but now $\mathcal{R}$ may depend on U and its realizations can be negative. This class of distributions contains many well-known multivariate distributions, e.g. the entire class of elliptically symmetric distributions, skew-elliptical distributions (Liu and Dey, 2004), and distributions with elliptical directions (Randles, 1989). In the latter case $\mathcal{R}$ might depend on U but it has to be positive (a.s.). For example, consider the class of multi-tail generalized elliptical distributions, which has been recently introduced by Kring et al. (2009) for analyzing financial data. Figure 1 demonstrates that, due to the flexibility of generalized elliptical distributions, it is possible to obtain a fairly nice fit to financial data.

Now we will present a generalized elliptical distribution which will play a major role in the following discussion.

Theorem 1. *Let Λ be a $d \times k$ matrix with $\operatorname{rk}\Lambda = d$ and U a k-dimensional random vector, uniformly distributed on the unit hypersphere. The density of the unit random vector $V = \Lambda U / \|\Lambda U\|_2$ – with respect to the uniform measure*

on the unit hypersphere – corresponds to

$$\psi(v) = \frac{\Gamma(d/2)}{2\pi^{d/2}} \cdot \sqrt{\det \Sigma^{-1}}\, \sqrt{v^{\mathsf{T}} \Sigma^{-1} v}^{\,-d},$$

where $\|v\|_2 = 1$ *and* $\Sigma = \Lambda\Lambda^{\mathsf{T}}$.

Proof. Frahm (2004, pp. 59–60). ∎

Note that the coefficient $\Gamma(d/2)/(2\pi^{d/2})$ corresponds to the uniform density (with respect to the uniform measure) on the unit hypersphere in $\mathbb{R}^d$. Moreover, the random vector V is generalized elliptically distributed with generating variate $\|\Lambda U\|_2^{-1}$. Its distribution is sometimes referred to as the angular central Gaussian distribution on the sphere (Tyler, 1987b; Kent and Tyler, 1988; Mardia and Jupp, 2000, p. 182) or the offset normal distribution (Mardia and Jupp, 2000, p. 178). In case $d = 2$ it is also known as the wrapped Cauchy distribution (Kent and Tyler, 1988). However, we will call it characteristic density function, since it can be shown that the eigenvectors and eigenvalues of Σ are characterized by the stationary points of ψ (Frahm and Jaekel, 2007).

3. Tyler's M-Estimators

Let X be a d-dimensional random vector and $\mathcal{P}^d$ the set of all symmetric positive-definite $d \times d$ matrices. In the following $\Gamma^{\frac{1}{2}}$ denotes the symmetric root of a symmetric positive-definite $d \times d$ matrix Γ. This means $\Gamma^{\frac{1}{2}}$ is the unique symmetric $d \times d$ matrix such that $\Gamma^{\frac{1}{2}}\Gamma^{\frac{1}{2}} = \Gamma$. Accordingly, $\Gamma^{-\frac{1}{2}}$ is the unique symmetric $d \times d$ matrix such that $\Gamma^{-\frac{1}{2}}\Gamma^{-\frac{1}{2}} = \Gamma^{-1}$.

Now let $\mu \in \mathbb{R}^d$ and $\Omega \in \mathcal{P}^d$ be such that

$$\mathbb{E}\left\{ \frac{\Omega^{-\frac{1}{2}}(X-\mu)(X-\mu)^{\mathsf{T}}\Omega^{-\frac{1}{2}}}{\|\Omega^{-\frac{1}{2}}(X-\mu)\|_2^2} \right\} = \frac{I_d}{d} \tag{2}$$

with $\sigma^2(\Omega) = 1$. Here $\sigma^2 \colon \mathcal{P}^d \to \mathbb{R}^+$ is some predefined scale function. This means σ^2 is such that $\sigma^2(\alpha\Sigma) = \alpha\sigma^2(\Sigma) > 0$ for all $\alpha > 0$, $\sigma^2(I_d) = 1$, and σ^2 is differentiable at any point $\Sigma \in \mathcal{P}^d$ (Frahm, 2009; Paindaveine, 2008). The matrix Ω will be called the shape matrix of X (with respect to the scale function σ^2). Moreover, if $\mu \in \mathbb{R}^d$ is such that

$$\mathbb{E}\left\{ \frac{\Omega^{-\frac{1}{2}}(X-\mu)}{\|\Omega^{-\frac{1}{2}}(X-\mu)\|_2} \right\} = 0\,, \tag{3}$$

the vector μ is said to be the multivariate median of X. The following theorem asserts that μ and Ω are well-defined if X is generalized elliptically distributed.

Theorem 2. *Let X be a d-dimensional generalized elliptically distributed random vector with location vector $\mu \in \mathbb{R}^d$, positive-definite dispersion matrix $\Sigma \in \mathbb{R}^{d\times d}$, and generating variate $\mathcal{R}$ with $\mathbb{P}(\mathcal{R} = 0) = 0$. Then $\Omega = \Sigma/\sigma^2(\Sigma)$ is the shape matrix of X with respect to the scale function σ^2. Further, if the sign of $\mathcal{R}$ is stochastically independent of U or $\mathcal{R}$ is positive (a.s.), the location vector μ represents the multivariate median of X.*

Proof. By the definition of generalized elliptical distributions it follows that

$$\frac{\Omega^{-\frac{1}{2}}(X-\mu)(X-\mu)^{\mathsf{T}}\Omega^{-\frac{1}{2}}}{\|\Omega^{-\frac{1}{2}}(X-\mu)\|_2^2} \overset{\text{a.s.}}{=} \frac{\Omega^{-\frac{1}{2}}\Lambda UU^{\mathsf{T}}\Lambda^{\mathsf{T}}\Omega^{-\frac{1}{2}}}{U^{\mathsf{T}}\Lambda^{\mathsf{T}}\Omega^{-1}\Lambda U} = \frac{\Omega^{-\frac{1}{2}}\Lambda UU^{\mathsf{T}}\Lambda^{\mathsf{T}}\Omega^{-\frac{1}{2}}}{\sigma^2(\Sigma)} \tag{4}$$

and due to $\mathbb{E}(UU^{\mathsf{T}}) = I_d/d$ (Fang et al., 1990, p. 34),

$$\mathbb{E}\left\{\frac{\Omega^{-\frac{1}{2}}\Lambda UU^{\mathsf{T}}\Lambda^{\mathsf{T}}\Omega^{-\frac{1}{2}}}{\sigma^2(\Sigma)}\right\} = \frac{\Omega^{-\frac{1}{2}}\Lambda\mathbb{E}(UU^{\mathsf{T}})\Lambda^{\mathsf{T}}\Omega^{-\frac{1}{2}}}{\sigma^2(\Sigma)} = \frac{I_d}{d}\,.$$

Moreover, note that $\sigma^2(\Omega) = \sigma^2\{\Sigma/\sigma^2(\Sigma)\} = 1$ due to the homogeneity of σ^2. Now

$$\frac{\Omega^{-\frac{1}{2}}(X-\mu)}{\|\Omega^{-\frac{1}{2}}(X-\mu)\|_2} \overset{\text{a.s.}}{=} \frac{\operatorname{sgn}(\mathcal{R})\,\Omega^{-\frac{1}{2}}\Lambda U}{\sqrt{U^{\mathsf{T}}\Lambda^{\mathsf{T}}\Omega^{-1}\Lambda U}} = \frac{\operatorname{sgn}(\mathcal{R})\,\Omega^{-\frac{1}{2}}\Lambda U}{\sqrt{\sigma^2(\Sigma)}}\,.$$

If $\mathcal{R} > 0$ (a.s.) it holds that $\operatorname{sgn}(\mathcal{R}) = 1$ (a.s.) and thus $\mathbb{E}\{\operatorname{sgn}(\mathcal{R})\,\Omega^{-\frac{1}{2}}\Lambda U\} = \Omega^{-\frac{1}{2}}\Lambda\mathbb{E}(U) = 0$. If $\mathcal{R}$ and U are stochastically independent, also it turns out that $\mathbb{E}\{\operatorname{sgn}(\mathcal{R})\,\Omega^{-\frac{1}{2}}\Lambda U\} = \mathbb{E}\{\operatorname{sgn}(\mathcal{R})\}\,\Omega^{-\frac{1}{2}}\Lambda\mathbb{E}(U) = 0$. ■

If X is generalized elliptically distributed, outliers are produced by extreme realizations of the generating variate $\mathcal{R}$. Note that by definition such values can be clustered in arbitrary directions in $\mathbb{R}^d$ since $\mathcal{R}$ may depend on U. Hence, the class of generalized elliptical distributions is huge. While this gives the flexibility to adapt to specific characteristics of the data (Kring et al., 2009; Frahm, 2004, Section 3.4), it can be a problem in many practical situations where neither the distribution family of $\mathcal{R}$ nor the dependence structure between $\mathcal{R}$ and U are known. In the following we will focus on estimating the shape matrix Ω *without* specifying the distribution of the generating variate.

For a start it is supposed that the location vector μ is known or that it can be properly estimated. This assumption will be discussed later on in Section 3.3. For the time being we restrict on centered random vectors $X_1, \ldots, X_n$ and their corresponding realizations $x_1, \ldots, x_n$ for the sake of simplicity and without loss of generality.

3.1. The Complete-Data Case

3.1.1. Tyler's M-Estimation Approach

Tyler originally derived his shape matrix estimator as an M-estimator by using a Huber-type weight function (Tyler, 1983, 1987a). It is defined via the fixed point equation

$$T = \frac{d}{n}\sum_{t=1}^{n}\frac{X_t X_t^{\mathsf{T}}}{X_t^{\mathsf{T}}T^{-1}X_t}\,. \tag{5}$$

An alternative way of representing Tyler's M-estimator is given by

$$\frac{1}{n}\sum_{t=1}^{n}\frac{T^{-\frac{1}{2}}X_t X_t^{\mathsf{T}}T^{-\frac{1}{2}}}{\|T^{-\frac{1}{2}}X_t\|_2^2} = \frac{I_d}{d}\,,$$

which can be regarded as the 'sample version' of the shape matrix of X defined by (2). The transformed random vector $T^{-\frac{1}{2}}X_t/\|T^{-\frac{1}{2}}X_t\|_2$ is said to be the multivariate sign of X_t.

Let Σ be positive-definite and $\mathbb{P}(\mathcal{R}=0)=0$. Due to the definition of X it holds that

$$S := \frac{X}{\|X\|_2} = \frac{\mathcal{R}\Lambda U}{\|\mathcal{R}\Lambda U\|_2} \overset{\text{a.s.}}{=} \frac{\operatorname{sgn}(\mathcal{R})\Lambda U}{\|\Lambda U\|_2} = \operatorname{sgn}(\mathcal{R})V\,, \qquad V := \frac{\Lambda U}{\|\Lambda U\|_2}\,. \tag{6}$$

The key observation is that the random vector V does not depend on the absolute value of $\mathcal{R}$. In particular, it is completely robust against extreme outcomes of the generating variate. However, the sign of $\mathcal{R}$ still remains in Eq. 6 and indeed this might further depend on U.

The unit random vector S represents the direction of X on the unit hypersphere. It contains all necessary information for estimating the shape matrix. In the univariate case the concept of shape is void and so it is always required that $d>1$ in the following discussion. Since

$$T = \frac{d}{n}\sum_{t=1}^{n}\frac{S_tS_t^{\mathsf{T}}}{S_t^{\mathsf{T}}T^{-1}S_t} = \frac{d}{n}\sum_{t=1}^{n}\frac{V_tV_t^{\mathsf{T}}}{V_t^{\mathsf{T}}T^{-1}V_t}$$

with $S_t := X_t/\|X_t\|_2$ and $V_t := \Lambda U_t/\|\Lambda U_t\|_2$, Tyler's M-estimator is invariant under any change of $\mathcal{R}$, i.e. it is distribution-free under the class of generalized elliptical distributions (for any given dispersion matrix Σ). This distribution-free property is a typical advantage of sign-based estimators and hypothesis tests which are frequently used in robust statistics (Hallin and Paindaveine, 2006; Hallin et al., 2006; Randles, 1989, 2000). Moreover, T is strongly consistent and asymptotically normally distributed, provided X has a continuous distribution on $\mathbb{R}^d$ (Tyler, 1987a).

Important results concerning the existence of Tyler's M-estimator for *any* kind of distributions were established by Tyler (1987a) as well as Kent and Tyler (1988, 1991). For instance, if the data are contaminated at some point in $\mathbb{R}^d$, the rate of contamination must not exceed $1/d$ (Kent and Tyler, 1988). Further, Kent and Tyler (1988) proved that for any given sample $x_1,\ldots,x_n\neq 0$, the fixed-point solution T exists and the sequence $\{T^{(i)}\}_{i=0,1,\ldots}$ defined by the fixed-point iteration scheme

$$T^{(i+1)} = \frac{d}{n}\sum_{t=1}^{n}\frac{x_tx_t^{\mathsf{T}}}{x_t^{\mathsf{T}}T^{(i)-1}x_t} \tag{7}$$

converges to τ^2T provided the data stem from a continuous distribution on $\mathbb{R}^d$ and $n>d$. The initial value $T^{(0)}$ can be any symmetric positive-definite $d\times d$ matrix and $\tau^2>0$ is a scaling constant depending on the initial value $T^{(0)}$.

Since T is defined only up to scale, it has to be fixed by some additional constraint like $\det T=1$. Of course, any other constraint like $\Sigma_{11}=1$ (Frahm, 2004, p. 64) or $\operatorname{tr}\Sigma=d$ (Tyler, 1987a) would also work. However, the determinant-based normalization has several statistical advantages which are discussed by Frahm (2009) and Paindaveine (2008). The chosen normalization (for example

$T \to T/(\det T)^{1/d}$ can be applied at the end of all iterations. This means it is not necessary to normalize $T^{(i+1)}$ after each step $i+1 = 1, 2, \ldots$ (Kent and Tyler, 1991). If the data are contaminated at some point in $\mathbb{R}^d$, the convergence of this algorithm is guaranteed provided the rate of contamination is smaller than $1/d$ (Kent and Tyler, 1988).

Tyler's M-estimator is a robust estimator and its robustness properties such as its breakdown point, maximum asymptotic bias and variance have been already investigated by Adrover (1998), Dümbgen and Tyler (2005), Maronna and Yohai (1990), as well as Tyler (1983, 1987a). In particular, it has been shown that the Dirac contamination breakdown point of T corresponds to $1/d$ (Maronna and Yohai, 1990) whereas for *any* kind of contamination it is between $1/(d+1)$ and $1/d$ (Adrover, 1998) if the data are elliptically symmetric distributed. Due to the arguments given above, the same conclusion holds for generalized elliptical distributions, too. Moreover, Tyler (1987a) showed that his estimator is 'most robust' over the class of continuous elliptically symmetric distributions. This means there is no other consistent and asymptotically normally distributed shape matrix estimator whose maximum asymptotic variance is lower than the asymptotic variance of T.

The asymptotic distribution of the eigenvalues of Tyler's M-estimator in the context of high-dimensional data (i.e. $n, d \to \infty$) has been studied by Dümbgen (1998) for $n/d \to \infty$ as well as Frahm and Jaekel (2007) for $n/d \to q < \infty$. The authors showed that Tyler's M-estimator has several nice properties in high dimensions, which have been previously found for the sample covariance matrix if the data are multivariate normally distributed. The point is that the asymptotic distribution of the eigenvalues of Tyler's M-estimator is invariant under any change of the generating variate of the generalized elliptical distribution but the results concerning the sample covariance matrix require the normal distribution assumption.

3.1.2. The ML-Estimation Approach

Although Tyler's shape matrix estimator is usually interpreted as an M-estimator, T also turns out to be an ML-estimator after taking the characteristic density function given by Eq. 1 into consideration. This important fact has been already noticed by Tyler (1987b) as well as Kent and Tyler (1988) for the case of elliptical distributions. The following theorem (Frahm, 2004) states that the same conclusion holds for generalized elliptical distributions.

Theorem 3. *Let $X_1, \ldots, X_n$ be a sample of $n > d$ independent copies of a d-dimensional generalized elliptically distributed and centered random vector X with positive-definite dispersion matrix $\Sigma \in \mathbb{R}^{d \times d}$ and generating variate $\mathcal{R}$ such that $\mathbb{P}(\mathcal{R} = 0) = 0$. Consider the unit random vector $V = \operatorname{sgn}(\mathcal{R})\, X/\|X\|_2$ (a.s.) and the corresponding sample $V_1, \ldots, V_n$. Then Tyler's M-estimator T exists almost surely and it is an ML-estimator with respect to the likelihood function $\mathcal{L}(\Sigma\,; V_1, \ldots, V_n) = \prod_{t=1}^{n} \psi(V_t\,; \Sigma)$. Here $V_1, \ldots, V_n$ are defined according to Eq. 6 and ψ is the characteristic density function given by Theorem 1. Fur-*

thermore, T satisfies the ML-equation

$$\frac{\partial \log \mathcal{L}(T\,; V_1, \ldots, V_n)}{\partial T} = \sum_{t=1}^{n} \frac{\partial \log \psi(V_t\,; T)}{\partial T} = 0$$

and it is unique up to a scaling constant.

Proof. The arguments for the existence and thus positive definiteness of Tyler's M-estimator in case $n > d$ can be found in Kent and Tyler (1988) by taking into consideration that ψ represents a continuous distribution with respect to the uniform measure on the unit hypersphere. Since the set of all symmetric positive-definite $d \times d$ matrices is open, the maximizer $\widehat{\Sigma}$ of the (log-)likelihood function is a stationary point. Now consider the log-likelihood function

$$\begin{aligned} \log \mathcal{L}(\Sigma\,; V_1, \ldots, V_n) &= \sum_{t=1}^{n} \log \psi(V_t\,; \Sigma) \\ &= c + \frac{n}{2} \cdot \log \det \Sigma^{-1} - \frac{d}{2} \sum_{t=1}^{n} \log\bigl(S_t^{\mathsf{T}} \Sigma^{-1} S_t\bigr), \end{aligned}$$

where c is a constant and note that $\psi(V_t) = \psi(S_t)$, since ψ is an even function. The partial derivative of $\log \mathcal{L}(\Sigma\,; V_1, \ldots, V_n)$ with respect to the inverse Σ^{-1} is given by

$$\frac{\partial \log \mathcal{L}}{\partial \Sigma^{-1}} = \frac{n}{2} \cdot \bigl(2\Sigma - \operatorname{diag} \Sigma\bigr) - \frac{d}{2} \sum_{t=1}^{n} \left(\frac{2 S_t S_t^{\mathsf{T}} - \operatorname{diag}(S_t S_t^{\mathsf{T}})}{S_t^{\mathsf{T}} \Sigma^{-1} S_t} \right) = M - \operatorname{diag} M/2\,,$$

where

$$M := n\Sigma - d \cdot \sum_{t=1}^{n} \frac{S_t S_t^{\mathsf{T}}}{S_t^{\mathsf{T}} \Sigma^{-1} S_t} = n\Sigma - d \cdot \sum_{t=1}^{n} \frac{X_t X_t^{\mathsf{T}}}{X_t^{\mathsf{T}} \Sigma^{-1} X_t}\,.$$

Thus it follows that

$$n\widehat{\Sigma} - d \cdot \sum_{t=1}^{n} \frac{X_t X_t^{\mathsf{T}}}{X_t^{\mathsf{T}} \widehat{\Sigma}^{-1} X_t} = 0\,,$$

which is equivalent to Eq. 5 after substituting $\widehat{\Sigma}$ by T. Finally, the arguments for the uniqueness of T can be found in Tyler (1987a) and transfer immediately to a sample from a generalized elliptical distribution, since V is continuously distributed on the unit hypersphere in $\mathbb{R}^d$. ∎

Recall that ψ is an even function, i.e. $\psi(-s) = \psi(s)$ for every s with $\|s\|_2 = 1$. This means Tyler's M-estimator indeed maximizes the likelihood function for a sample $V_1, \ldots, V_n$ of n independent copies of the unit random vector V even though the corresponding realizations of V are given only up to the corresponding signs.

3.2. The Incomplete-Data Case

Now Tyler's M-estimator for the shape matrix will be generalized to the case of incomplete data by using the well-developed likelihood theory for missing data. This is not a trivial generalization since Tyler originally argued on the basis of M-estimation theory (Tyler, 1983). The key observation is that Tyler's M-estimator in fact is an *ML-estimator* (Frahm, 2004; Tyler, 1987b) and then methods of missing-data analysis have to be applied carefully. The difficult part is to derive the score function under incomplete data and to formulate an appropriate algorithm for finding its root. First of all we will recapitulate the fundamental background of missing-data analysis, which is necessary for understanding the subsequent derivations. A comprehensive introduction to that topic is given by Little and Rubin (2002) as well as Schafer (1997).

3.2.1. Ignorable Missingness Patterns

Let x be some realized data and m an ensemble of zeros and ones indicating which part of x is observed and which is missing. According to the missingness pattern m let x_o be the observed and x_m the unobserved data. The observed part O of the complete data X is a random index, whereas o denotes some realization of O according to the missingness pattern m, which is a realization of M. Sometimes it is helpful to interpret m as a function $m\colon x \mapsto x_o$. Further, M and X are random quantities possessing the joint distribution $p(m, x\,;\theta)$. Here $\theta \in \Theta \subset \mathbb{R}^k$ is some unknown parameter. The marginal distribution of m and x_o corresponds to

$$p(m, x_o\,;\theta) = \int p(m, x_o, x_m\,;\theta)\, dx_m\,.$$

Suppose that the parameter θ has to be estimated. All available information are given by m and x_o, though $p(m, x\,;\theta)$ is the underlying sampling distribution of the experiment. However, under the standard assumptions of likelihood theory, the likelihood function $\mathcal{L}(\theta\,; m, x_o) = p(m, x_o\,;\theta)$ turns out to be Fisher-consistent for θ. Note that

$$\mathcal{L}(\theta\,; m, x_o) = p(m\,;\theta)\, p(x_o \,|\, m\,;\theta) = p(x_o\,;\theta)\, p(m \,|\, x_o\,;\theta)\,,$$

where $p(x_o\,;\theta)$ denotes the marginal distribution of the observed data X_o and o is the realized index of observations.

Now suppose that the missingness pattern is not determined by the model parameter under the observed data. This means $p(m \,|\, x_o\,;\theta)$ is invariant under a change of θ. In that case the missingness pattern is not relevant and it can be ignored for maximum likelihood estimation, since

$$\mathcal{L}(\theta\,; m, x_o) \propto p(x_o\,;\theta) = \mathcal{L}(\theta\,; x_o)\,.$$

Hence, for estimating θ it is sufficient to concentrate on the marginal distribution of X_o. This is the so-called ignorability assumption of missing-data analysis and $\mathcal{L}(\theta\,; x_o)$ is the observed-data likelihood function (Schafer, 1997,

Section 2.3.1). Estimators which are obtained by maximizing the observed-data likelihood function are referred to as observed-data maximum-likelihood (ODML-)estimators.

To justify the ignorability assumption, the conditional distribution

$$p(m \,|\, x_o \,; \theta) = \int p(m \,|\, x_o, x_m \,; \theta)\, p(x_m \,|\, x_o \,; \theta)\, dx_m$$

has to be examined carefully. In many circumstances it can be assumed that the distribution of M depends on the complete data X but not on the specific parameter θ. For example, non-responses in questionnaires might be determined by the individual outcomes x_o and x_m but it is unlikely that the missingness pattern depends on the model parameter θ per se. The so-called distinctness assumption of missing-data analysis conveys that $p(m \,|\, x_o, x_m \,; \theta)$ is not determined by θ. If the distinctness assumption can be accepted, it follows that

$$p(m \,|\, x_o \,; \theta) = \int p(m \,|\, x_o, x_m)\, p(x_m \,|\, x_o \,; \theta)\, dx_m \,.$$

Now there are two non-excluding cases where the ignorability assumption is satisfied, viz

a. $p(x_m \,|\, x_o \,; \theta)$ is not determined by θ or

b. $p(m \,|\, x_o, x_m)$ is not determined by x_m .

ad a. The distribution of the complete data X is determined by θ. However, if $p(x_o, x_m \,; \theta) = p(x_o \,; \theta)\, p(x_m)$, then $p(x_m \,|\, x_o \,; \theta)$ is not driven by θ and the ignorability assumption is satisfied. This means if the unobserved data are independent of the observed data and do not contain any information about the unknown parameter, the missing data can be ignored.

ad b. If $p(m \,|\, x_o, x_m) = p(m \,|\, x_o)$, i.e. M is stochastically independent of X_M, then

$$p(m \,|\, x_o \,; \theta) = \int p(m \,|\, x_o)\, p(x_m \,|\, x_o \,; \theta)\, dx_m = p(m \,|\, x_o) \,.$$

In that case x_m is said to be missing at random (MAR) (Little and Rubin, 2002, p. 12). Moreover, if M is not only independent of the unobserved data X_M but also of the observed data X_O, the missing data are said to be missing completely at random (MCAR) (Little and Rubin, 2002, p. 12).

In the next section the characteristic density approach will be adapted to incomplete data, but before that we have to discuss an important drawback of missing-data analysis. Let $y = g(x_o)$ be some measurable function of the observed data and $q(y \,; \theta)$ the corresponding density. A naive application of missing-data analysis would suggest to estimate θ by using the observed-data likelihood function related to $q(y \,; \theta)$, i.e. $\mathcal{L}(\theta \,; y) = q(y \,; \theta)$, instead of $\mathcal{L}(\theta \,; x_o)$ if x_m is MAR. For example, this approach is suitable if the transformation of x_o leads to a robust or even distribution-free estimator for θ (see Section 3.2.2). In that case it has to be guaranteed that

$$\mathcal{L}(\theta \,; m, y) = q(y \,; \theta)\, p(m \,|\, y \,; \theta) \propto q(y \,; \theta) = \mathcal{L}(\theta \,; y)$$

with $p(m \,|\, y\,; \theta) = p(m \,|\, g(x_o)\,; \theta)$. Hence, the remaining question is whether the ignorability assumption can be justified for the transformed data y if it is satisfied for the data x_o which has been originally observed.

Suppose that the transformation g is not injective and that the distribution of m given x is determined by the observed part x_o of the data. Then $p(m \,|\, y\,; \theta)$ might be essentially determined by the parameter θ, since the distribution of $X_o \,|\, y$ in general will depend on θ and so this parameter has also an impact on the distribution of $M \,|\, y$. This means even if the distinctness assumption as well as the MAR (but *not* the MCAR) assumption are satisfied for the original data x, the former is usually violated under the transformed data y when working with many-to-one transformations. Due to the fact that the distinctness assumption is a necessary condition for the ignorability of the missingness pattern, standard likelihood-based inference from missing-data analysis would fail. Only if the missing data are MCAR, the distinctness assumption remains plausible (for it is simply assumed that $p(m \,|\, y\,; \theta) = p(m\,; \theta) = p(m)$) and so the corresponding ODML-estimator for θ is still consistent.

3.2.2. The ODML-Estimation Approach

Lemma 1. *Let X be a d-dimensional generalized elliptically distributed and centered random vector with dispersion matrix $\Sigma \in \mathbb{R}^{d \times d}$. Consider the partitions*

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad \textit{and} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where X_1 is an r-dimensional sub-vector of X. It holds that X_1 is a generalized elliptically distributed and centered random vector with dispersion matrix $\Sigma_{11} \in \mathbb{R}^{r \times r}$.

Proof. Write $X_1 = \mathcal{I} X$ with $\mathcal{I} := [\, I_r \; 0 \,]$ $(r \times d)$. This means $X_1 = \mathcal{I} \Lambda \mathcal{R} U$ and note that $\mathcal{I} \Lambda \Lambda^{\mathsf{T}} \mathcal{I}^{\mathsf{T}} = \mathcal{I} \Sigma \mathcal{I}^{\mathsf{T}} = \Sigma_{11}$. ∎

More generally, let $X = (X_o, X_m)$ be a generalized elliptically distributed random vector which is divided into an observed and an unobserved part according to some fixed missingness pattern m. Correspondingly, the vector $x = (x_o, x_m)$ denotes a realization of the complete data, where x_o is an r-dimensional sub-vector of x. Further, let $s_o = x_o / \|x_o\|_2$ be the observed data projected onto the unit hypersphere in $\mathbb{R}^r$ and $S_o = X_o / \|X_o\|_2$ the corresponding random vector. From the preceding lemma it is known that the distribution of $V_o = \operatorname{sgn}(\mathcal{R})\, S_o$ (a.s.) is given by

$$\psi(v_o) = \frac{\Gamma(r/2)}{2\pi^{r/2}} \cdot \sqrt{\det \Sigma_o^{-1}} \, \sqrt{s_o^{\mathsf{T}} \Sigma_o^{-1} s_o}^{\,-r},$$

where Σ_o denotes that part of Σ which is related to the observation x_o. Once again the generating variate of X_o does not play any role for estimating Σ since it is canceled out by the projection onto the unit hypersphere (see Eq. 6).

Now consider a sample of possibly dependent and not identically generalized elliptically distributed random vectors $X_1 \ldots, X_n$, where only the realizations $x_{o1}, \ldots, x_{on}$ of the sub-vectors $X_{o1}, \ldots, X_{on}$ can be observed. More precisely,

it is assumed that X_t $(t = 1, \ldots, n)$ can be represented according to Eq. 1, where $\mu = 0$ without loss of generality, $\Sigma = \Lambda\Lambda^\mathsf{T}$ is positive-definite, and the distribution of X_t has no atom at zero. Hence, the observed data can be written as

$$X_{ot} = \mathcal{I}_{ot} X_t = \mathcal{I}_{ot} \Lambda \mathcal{R}_t U_t = \Lambda_{ot} \mathcal{R}_t U_t \,, \qquad t = 1, \ldots, n \,,$$

where $\mathcal{I}_{ot}$ is a matrix which converts X_t into X_{ot} and $\Lambda_{ot} := \mathcal{I}_{ot}\Lambda$. Now it is only assumed that the angular parts $U_1, \ldots, U_n$ are serially independent, whereas the joint distribution of the radial parts $\mathcal{R}_1, \ldots, \mathcal{R}_n$ is irrelevant. This means the generating variates might depend on each other and do not need to be identically distributed. That feature allows for several kinds of serial dependence imposed by the variation of $\mathcal{R}$ in time.

Let Σ_{ot} be the sub-matrix of Σ associated to the observation x_{ot} and $s_{ot} = x_{ot}/\|x_{ot}\|_2$ $(t = 1, \ldots, n)$ the corresponding projection onto the unit hypersphere. Moreover, let $d_t > 1$ be the number of components of that observation. Then the observed-data likelihood function is given by

$$\mathcal{L}(\Sigma \,; v_{o1}, \ldots, v_{on}) = \prod_{t=1}^{n} \psi\big(v_{ot} \,; \Sigma_{ot}\big) \propto \prod_{t=1}^{n} \sqrt{\det \Sigma_{ot}^{-1}} \, \sqrt{s_{ot}^\mathsf{T} \Sigma_{ot}^{-1} s_{ot}}^{\,-d_t} \,. \qquad (8)$$

Note that $v_{ot} = \pm s_{ot}$ is not a one-to-one function of x_{ot} and due to the arguments given at the end of Section 3.2.1 we must suppose that the missing data $x_{m1}, \ldots, x_{mn}$ are MCAR.

The observed-data log-likelihood function

$$\log \mathcal{L}(\Sigma \,; v_{o1}, \ldots, v_{on}) = c + \frac{1}{2} \sum_{t=1}^{n} \log \det \Sigma_{ot}^{-1} - \frac{1}{2} \sum_{t=1}^{n} d_t \log\big(s_{ot}^\mathsf{T} \Sigma_{ot}^{-1} s_{ot}\big)$$

can be used alternatively, where c is some constant. Since $\mathcal{L}$ is scale-invariant, i.e. $\mathcal{L}(\alpha\Sigma) = \mathcal{L}(\Sigma)$ for every $\alpha > 0$, the corresponding ODML-estimator has to be fixed by some additional constraint (cf. Section 3.1). In the following we will consider the determinant-based normalization (Frahm, 2009; Paindaveine, 2008) which leads to the estimator $\widehat{\Omega} = \widehat{\Sigma}/(\det \widehat{\Sigma}\,)^{1/d}$ for the shape matrix Ω, where $\widehat{\Sigma}$ denotes an unconstraint ODML-estimator for Σ.

Figure 2 contains an outcome of our generalization of Tyler's M-estimator for a sample of multivariate t-distributed data with 2 degrees of freedom, possessing a monotone missingness pattern (Little and Rubin, 2002, p. 5). This can be compared with the corresponding ODML-estimate based on the normal distribution assumption and the factored likelihood method described by Little and Rubin (2002, Ch. 7.4) as well as Schafer (1997, Ch. 6.5). Obviously, the Gauss-type estimator is not robust against extreme realizations of the multivariate t-distribution. In Figure 3 the same experiment is done with multivariate normally distributed data. Our estimator looks pretty much the same as the Gaussian one in agreement with the simulation study discussed in Section 5, showing that the loss of efficiency is small for normally distributed data.

Little (1988) suggests to maximize the observed-data likelihood functions of heavy-tailed or contaminated data if the normal distribution assumption is evidently violated. Usually this leads to robust estimates of the shape matrix and

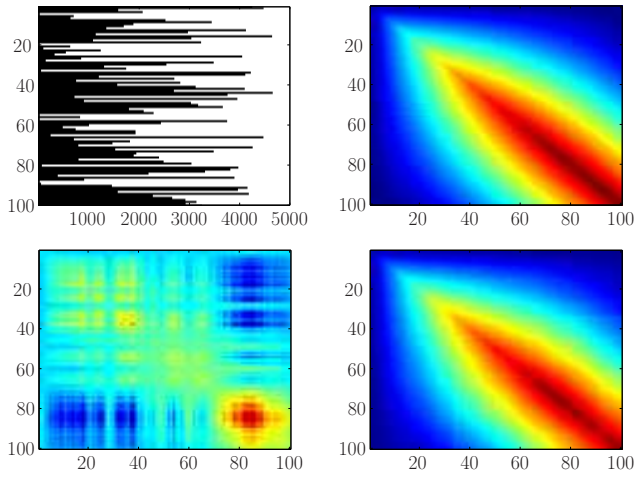

Figure 2: Missingness pattern (upper left) of a simulated sample of 5000 independent copies of a 100-dimensional t-distributed random vector with 2 degrees of freedom, location vector $\mu = 0$, and dispersion matrix Σ proportional to the shape matrix given on the upper right (violet cells indicate small numbers and red cells large numbers). There are 215184 missing values (43% of the sample). Our generalization of Tyler's M-estimator leads to the lower right, whereas the Gauss-type estimate can be found on the lower left. The computational time for the Tyler-type estimate on a standard PC (3 GHz CPU) amounts to 25 seconds.

obviously the method is similar to our characteristic density approach. However, Little's estimators are based on a multivariate t-distribution or a contaminated normal-distribution assumption. This is a parametrical approach and thus not distribution-free under the class of generalized elliptical distributions. With such a parametrical approach one has only limited information about the asymptotic distribution of the estimators if the model is misspecified. This drawback can be avoided for the most part by using our estimator due to its invariance property discussed above. The only conditions which have to be guaranteed are that

(1) the sample consists of data which are generalized elliptically distributed (serial dependence is allowed under the weak conditions described above),

(2) the missing part of the sample is MCAR, and

(3) μ either is known or can be approximated by a consistent estimator.

Before estimating the shape matrix it has to be guaranteed that the MCAR condition is satisfied. In most cases this can be done by examining the data-generating process. For instance, consider a sample of stock return data which have been observed over a relatively long sample period. Typically such kind of time series exhibit missing values possibly caused by bank holidays, system failures, initial public offerings (IPO's) or mergers and acquisitions (M&A's). For the sake of simplicity assume that the stock returns are serially independent although it is well-known that this is not true for real stock market data.

The missingness of the stock returns before a firm's IPO cannot depend on the missing data (since there are no missing data at all). This means the MCAR assumption is clearly satisfied in that case. The same holds of course if the missing data are due to bank holidays or system failures. By contrast, if some stock

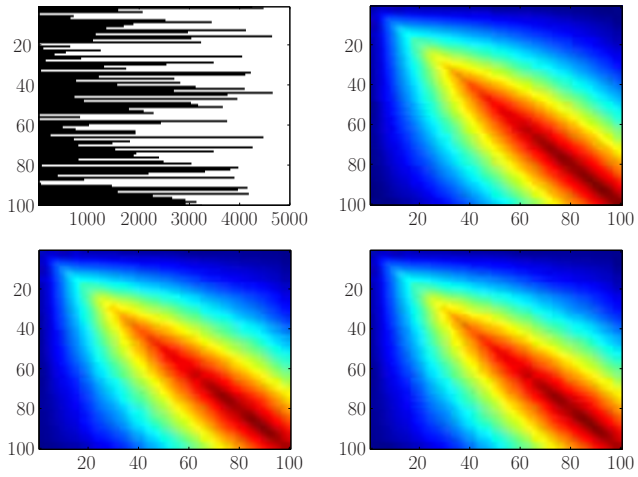

Figure 3: Missingness pattern (upper left) of a simulated sample of 5000 independent copies of a 100-dimensional normally distributed random vector with location vector $\mu = 0$ and dispersion matrix Σ proportional to the shape matrix given on the upper right (see also Figure 2). Our generalization of Tyler's M-estimator leads to the lower right, whereas the Gauss-type estimate can be found on the lower left. The computational time for the Tyler-type estimate on a standard PC (3 GHz CPU) amounts to 26 seconds.

market data become missing after an M&A, i.e. after the corresponding firm has been acquired by another firm, it is reasonable to assume that the missing data are only MAR *without* being MCAR, since the date of the merger typically depends on the observed historical stock prices of the firms which are involved. Finally, the missing data could also occur before an M&A. In that case the missing data have to be considered as NMAR, since the unobservable historical stock prices of the merged firm might have determined the date of the merger. In that case conventional methods of missing data analysis are inappropriate at all. Similar arguments can be found in many fields of applications like in physics, biology, and hydrology.

3.2.3. Asymptotic Distribution

In the complete-data case it can be shown (Frahm, 2009) that

$$\sqrt{n}\,\big(\widehat{\Omega} - \Omega\big) \xrightarrow{\mathrm{d}} \mathcal{N}_{d\times d}\big\{0, V(\Omega)\big\}\,, \qquad n \longrightarrow \infty\,,$$

with

$$V(\Omega) = \frac{d+2}{d} \cdot \Psi(I_{d^2} + K_{d^2})(\Omega \otimes \Omega)\Psi^{\mathsf{T}}\,,$$

where K_{d^2} denotes the $d^2 \times d^2$ commutation matrix (Schott, 1997, p. 277). Further, the $d^2 \times d^2$ matrix Ψ is defined as $\Psi := I_{d^2} - \operatorname{vec}\Omega\,\{\partial\sigma^2(\Omega)/\partial\operatorname{vec}\Omega\}^{\mathsf{T}}$, where $\operatorname{vec}\Omega$ is obtained by stacking the columns of Ω on top of each other. If σ^2 represents the 'canonical' scale function (Paindaveine, 2008), i.e. $\sigma^2(\Sigma) = (\det\Sigma)^{1/d}$, it follows that $\Psi = I_{d^2} - (\operatorname{vec}\Omega)(\operatorname{vec}\Omega^{-1})^{\mathsf{T}}/d$ (Frahm, 2009).

In the incomplete-data case, $\log\mathcal{L}$ is a proper log-likelihood function under the conditions given in Section 3.2.2 and so our estimator turns out to be

asymptotically unbiased, normally distributed, and consistent. For calculating its asymptotic covariance matrix, the Fisher-information has to be calculated either by the score function or the Hessian of $\log \mathcal{L}$. The following proposition can be used for calculating the score function.

Proposition 1. *Let v be a d-dimensional vector with unit length and $\Sigma \in \mathcal{P}^d$. The partial derivative of $\log \psi(v\,;\Sigma)$ with respect to Σ is given by*

$$\frac{\partial \log \psi(v\,;\Sigma)}{\partial \Sigma} = \left(d \cdot \frac{\Sigma^{-1} v v^{\mathsf{T}} \Sigma^{-1}}{v^{\mathsf{T}} \Sigma^{-1} v} - \Sigma^{-1} \right) - \frac{1}{2} \cdot \operatorname{diag}\left(d \cdot \frac{\Sigma^{-1} v v^{\mathsf{T}} \Sigma^{-1}}{v^{\mathsf{T}} \Sigma^{-1} v} - \Sigma^{-1} \right).$$

Proof. Frahm (2004, p. 70). ∎

The Fisher-information of an observed data point $S_{ot} = X_{ot}/\|X_{ot}\|_2$ $(t = 1, \ldots, n)$ is given by the $\binom{d+1}{2} \times \binom{d+1}{2}$ matrix

$$\mathcal{F}_t(\Sigma) = \mathbb{E}\left\{ \operatorname{vech}\left(\frac{\partial \log \psi(V_{ot}\,;\Sigma_{ot})}{\partial \Sigma} \right) \operatorname{vech}\left(\frac{\partial \log \psi(V_{ot}\,;\Sigma_{ot})}{\partial \Sigma} \right)^{\mathsf{T}} \right\}, \tag{9}$$

where the vech-operator converts the lower triangular part of a symmetric matrix to a column vector. Note that S_{ot} refers only to the observed part of the d-dimensional random vector X_t for $t \in \{1, \ldots, n\}$ and thus $\log \psi(V_{ot}\,;\Sigma_{ot})$ is invariant under changing that part of Σ which is not related to the available observation. This means there exists a $d_t \times d_t$ matrix

$$\begin{aligned} \frac{\partial \log \psi(V_{ot}\,;\Sigma_{ot})}{\partial \Sigma_{ot}} \;=\; & \left(d_t \cdot \frac{\Sigma_{ot}^{-1} X_{ot} X_{ot}^{\mathsf{T}} \Sigma_{ot}^{-1}}{X_{ot}^{\mathsf{T}} \Sigma_{ot}^{-1} X_{ot}} - \Sigma_{ot}^{-1} \right) - \\ & \frac{1}{2} \cdot \operatorname{diag}\left(d_t \cdot \frac{\Sigma_{ot}^{-1} X_{ot} X_{ot}^{\mathsf{T}} \Sigma_{ot}^{-1}}{X_{ot}^{\mathsf{T}} \Sigma_{ot}^{-1} X_{ot}} - \Sigma_{ot}^{-1} \right), \end{aligned} \tag{10}$$

but here the $d \times d$ matrix $\partial \log \psi(V_{ot}\,;\Sigma_{ot})/\partial \Sigma$ has to be considered. The latter contains zeros according to each element of Σ which does not belong to the sub-matrix Σ_{ot}. This is denoted by

$$\frac{\partial \log \psi(V_{ot}\,;\Sigma_{ot})}{\partial \Sigma} = \left[\frac{\partial \log \psi(V_{ot}\,;\Sigma_{ot})}{\partial \Sigma_{ot}} \right]_{mt},$$

where the operator $[\cdot]_{mt}$ inflates a given array A_{ot} by zeros according to the missing part of x_t such that $[A_{ot}]_{mt}$ gets the same dimension as $A = (A_{ot}, A_{mt})$.

Now

$$\sqrt{n}\left(\operatorname{vech} \widehat{\Sigma} - \operatorname{vech} \Sigma \right) \xrightarrow{\mathrm{d}} \mathcal{N}_{\binom{d+1}{2}}\left\{ 0, \mathcal{F}(\Sigma)^{-1} \right\}, \qquad n \longrightarrow \infty,$$

where

$$\mathcal{F}(\Sigma) := \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \mathcal{F}_t(\Sigma)$$

denotes the asymptotic average Fisher-information. Following the usual notation of multivariate analysis we will use the vec-operator, which converts Σ into a column vector. This can be obtained after a preceding vech-operation by

$\operatorname{vec}\Sigma = D_d \operatorname{vech}\Sigma$, where D_d represents the $d^2 \times d\,(d+1)/2$ duplication matrix (Schott, 1997, p. 283). Hence, $D_d \mathcal{F}(\Sigma)^{-1} D_d^{\mathsf{T}}$ corresponds to the asymptotic covariance matrix of $\sqrt{n}\,\big(\operatorname{vec}\widehat{\Sigma} - \operatorname{vec}\Sigma\big)$.

However, so far we considered some artificial derivations since we missed to account for the normalization $\Sigma \to \Omega$. Interestingly, such a normalization has a substantial impact on the asymptotic distribution (Frahm, 2009; Paindaveine, 2008). Following the arguments given by Frahm (2009) it can be concluded that

$$\sqrt{n}\,\big(\widehat{\Omega} - \Omega\big) \xrightarrow{\mathrm{d}} \mathcal{N}_{d\times d}\big\{0, \Psi D_d \mathcal{F}(\Omega)^{-1} D_d^{\mathsf{T}} \Psi^{\mathsf{T}}\big\}\,, \qquad n \longrightarrow \infty\,.$$

Note that in contrast to other M-estimators the asymptotic distribution of our estimator is only determined by the dimensions $d_1, \ldots, d_n$ of the observed data $x_{o1}, \ldots, x_{on}$ and the true shape matrix Ω. More precisely, it is possible to assess the asymptotic distribution of our estimator without any parametrical assumption concerning the distribution of $\mathcal{R}$ or, more generally, the stochastic process $\{\mathcal{R}_t\}_{t=1,2,\ldots}$. Hence, there are no nuisance parameters which have to be estimated for the purpose of statistical inference. This is an important advantage compared to other covariance matrix estimators (Frahm, 2009).

From Section 3.2.1 it becomes clear that under the MCAR assumption both the score functions and the Hessians belonging to $\mathcal{L}(\Sigma\,; m, v_{o1}, \ldots, v_{on})$ and $\mathcal{L}(\Sigma\,; v_{o1}, \ldots, v_{on})$ correspond to each other. For the application of large-sample theory in the context of missing data we follow the arguments given by Kenward and Molenberghs (1998) as well as Schafer and Graham (2002). That is we suggest to estimate the Fisher-information in a nonparametric way by the *observed data* rather than calculating the expectations given by Eq. 9 analytically or numerically. Hence, an appropriate estimate for $\mathcal{F}_t(\Omega)$ is

$$\widehat{\mathcal{F}}_t\big(\Omega\big) = \mathcal{F}_t\big(\widehat{\Omega}\big) = \operatorname{vech}\!\left(\frac{\partial \log \psi(v_{ot}\,; \widehat{\Omega}_{ot})}{\partial \widehat{\Omega}}\right) \operatorname{vech}\!\left(\frac{\partial \log \psi(v_{ot}\,; \widehat{\Omega}_{ot})}{\partial \widehat{\Omega}}\right)^{\mathsf{T}},$$

where v_{ot} can be replaced by the observed data point $s_{ot} = x_{ot}/\|x_{ot}\|_2$ and $\widehat{\Omega}_{ot}$ once again is that part of $\widehat{\Omega}$ which is related to the observation at $t \in \{1, \ldots, n\}$. The asymptotic average Fisher-information can be consistently estimated by

$$\widehat{\mathcal{F}}(\Omega) := \frac{1}{n} \sum_{t=1}^{n} \widehat{\mathcal{F}}_t(\Omega)$$

and so a large-sample approximation of the distribution of $\widehat{\Omega}$ is given by

$$\widehat{\Omega} \mathrel{\dot\sim} \mathcal{N}_{d\times d}\big\{\Omega, \widehat{\Psi} D_d\, \widehat{\mathcal{F}}(\Omega)^{-1} D_d^{\mathsf{T}} \widehat{\Psi}^{\mathsf{T}}/n\big\}\,,$$

where $\widehat{\Psi} := I_{d^2} - \operatorname{vec}\widehat{\Omega}\,\big\{\partial\sigma^2(\widehat{\Omega})/\partial \operatorname{vec}\widehat{\Omega}\big\}^{\mathsf{T}}$.

3.3. Estimating the Location Vector

The problem of estimating the location vector μ has been already investigated by Tyler (1987a) under quite general conditions. Suppose that X has a

continuous distribution on $\mathbb{R}^d$. In a first step μ might be estimated by a consistent estimator $\hat{\mu}_n$. Now the corresponding estimate is used for centralizing the data and $\widehat{\Omega}$ can be calculated in a second step by the centralized data. If $\hat{\mu}_n$ converges to μ at an appropriate rate as $n \to \infty$ and X is not too much concentrated around μ, T is still consistent and asymptotically normally distributed (Tyler, 1987a). By contrast, if X is too much concentrated around μ even small perturbations of $\hat{\mu}_n$ would lead to wrong projections of X onto the unit hypersphere, i.e. $(X - \hat{\mu}_n)/\|X - \hat{\mu}_n\|_2$ would be far away from $(X - \mu)/\|X - \mu\|_2$. In case the required conditions hold, Tyler's M-estimator possesses the same asymptotic covariance matrix as if μ was known. In particular, Tyler (1987a) showed that the sample mean and the componentwise sample median are suitable estimators for the location vector under quite general conditions.

One of the referees of an earlier version of the present paper pointed out that choosing the sample mean or the Gauss-type estimator in the context of missing data based on the factored likelihood method described by Little and Rubin (2002, Ch. 7) as well as Schafer (1997, Section 6.5) is not appropriate. In fact these estimators represent the (OD)ML-estimators for the location vector under the normal distribution assumption. By contrast, following the philosophy of Tyler's M-estimator it is more desirable to use a robust estimator for μ such as the componentwise sample median or the multivariate sample median $\hat{\mu}$, i.e. the solution of

$$\frac{1}{n}\sum_{t=1}^{n}\frac{X_t - \hat{\mu}}{\sqrt{(X_t - \hat{\mu})^{\mathsf{T}} T^{-1}(X_t - \hat{\mu})}} = 0\,. \tag{11}$$

The latter has been introduced by Tyler (1987a) as a by-product of his M-estimator for the shape matrix. It has been also investigated by Hettmansperger and Randles (2002), who use an equivalent representation of (11) based on multivariate signs, namely

$$\frac{1}{n}\sum_{t=1}^{n}\frac{T^{-\frac{1}{2}}(X_t - \hat{\mu})}{\left\|T^{-\frac{1}{2}}(X_t - \hat{\mu})\right\|_2} = 0\,.$$

This can be viewed as the 'sample version' of the multivariate median defined by (3). A simple generalization of this estimator to the case of incomplete data is given by

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\left[\widehat{\Sigma}_{ot}^{-\frac{1}{2}}(X_{ot} - \hat{\mu}_{ot})\right]_{mt}}{\left\|\widehat{\Sigma}_{ot}^{-\frac{1}{2}}(X_{ot} - \hat{\mu}_{ot})\right\|_2} = 0\,. \tag{12}$$

However, we admit that it is not easy to find a robust and consistent estimator for μ if the distribution of X is asymmetric. Some of Tyler's conditions for the asymptotic normality of T are violated if X has an asymmetric distribution. Nevertheless, if μ is unknown, the characteristic density approach can be clearly defended if the data are elliptically symmetric distributed and μ is estimated by some of the estimators mentioned above. Moreover, if X is generalized elliptically distributed, there exists a clear answer if the sign of $\mathcal{R}$ is stochastically independent of U or $\mathcal{R}$ is positive (a.s.), i.e. X follows a distribution with elliptical directions (Randles, 1989). In such cases μ indeed corresponds to its

componentwise median (Frahm, 2004, p. 67) as well as to its multivariate median (cf. Theorem 2). This means the generalized multivariate sample median represented by Eq. 12 is a robust and consistent estimator for μ under quite weak regularity conditions on X (provided the missing data are MCAR).

4. Numerical Implementation

In the following it is assumed that there exists a symmetric positive-definite matrix $\widehat{\Sigma}$ which maximizes the observed-data likelihood function given by (8). A necessary condition for the existence under a monotone missingness pattern can be found later on in Theorem 4. Note that the dispersion matrix Σ is symmetric and therefore half of the main diagonal of M has to be subtracted in Theorem 3. Since the set of all symmetric positive-definite $d \times d$ matrices is open, the maximizer $\widehat{\Sigma}$ must be a stationary point of the observed-data likelihood function and thus also the main diagonal part from Eq. 10 can be omitted. Hence, the ODML-equation can be written as

$$\frac{1}{n}\sum_{t=1}^{n}\left[d_t \cdot \frac{\widehat{\Sigma}_{ot}^{-1}(X_{ot}-\hat{\mu}_{ot})(X_{ot}-\hat{\mu}_{ot})^{\mathsf{T}}\widehat{\Sigma}_{ot}^{-1}}{(X_{ot}-\hat{\mu}_{ot})^{\mathsf{T}}\widehat{\Sigma}_{ot}^{-1}(X_{ot}-\hat{\mu}_{ot})}\right]_{mt} = \frac{1}{n}\sum_{t=1}^{n}\left[\widehat{\Sigma}_{ot}^{-1}\right]_{mt}, \qquad (13)$$

where $\hat{\mu}$ is the generalized multivariate sample median. Now (12) and (13) form a system of generalized M-equations. Thus $\hat{\mu}$ and $\widehat{\Omega} = \widehat{\Sigma}/(\det\widehat{\Sigma})^{1/d}$ can be interpreted as generalized M-estimators for location and shape taking account of incomplete data.

Now define a function $g\colon (\mathbb{R}^d, \mathcal{P}^d) \to \mathbb{R}^{d\times d}$ by

$$g(\mu,\Sigma) := \Sigma\left(\frac{1}{n}\sum_{t=1}^{n}\left[d_t \cdot \frac{\Sigma_{ot}^{-1}(x_{ot}-\mu_{ot})(x_{ot}-\mu_{ot})^{\mathsf{T}}\Sigma_{ot}^{-1}}{(x_{ot}-\mu_{ot})^{\mathsf{T}}\Sigma_{ot}^{-1}(x_{ot}-\mu_{ot})}\right]_{mt} - \frac{1}{n}\sum_{t=1}^{n}\left[\Sigma_{ot}^{-1}\right]_{mt}\right)\Sigma \qquad (14)$$

for all $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathcal{P}^d$. Further, consider the function $G(\mu,\Sigma) := \Sigma + g(\mu,\Sigma)$. Hence, the estimate $\widehat{\Sigma}$ solves the fixed-point equation

$$G\big(\hat{\mu},\widehat{\Sigma}\big) = \widehat{\Sigma} + g\big(\hat{\mu},\widehat{\Sigma}\big) = \widehat{\Sigma}\,. \qquad (15)$$

Note that in the complete-data case it follows that

$$G(\mu,\Sigma) = \frac{d}{n}\sum_{t=1}^{n}\frac{(x_t-\mu)(x_t-\mu)^{\mathsf{T}}}{(x_t-\mu)^{\mathsf{T}}\Sigma^{-1}(x_t-\mu)} \qquad (16)$$

and so the fixed-point problem given by (15) is equivalent to (5), i.e. finding Tyler's M-estimator (after centralizing the data).

Now we will concentrate on the numerical evaluation of $g(\mu,\Sigma)$. Beforehand it is worth pointing out that any observation $x_t = \mu \in \mathbb{R}^d$ or $x_t \in \mathbb{R}$ should be

discarded. The former argument is clear from the preceding discussion and the latter argument follows immediately by noting that

$$\left[d_t \cdot \frac{\Sigma_{ot}^{-1}(x_{ot} - \mu_{ot})(x_{ot} - \mu_{ot})^\mathsf{T}\Sigma_{ot}^{-1}}{(x_{ot} - \mu_{ot})^\mathsf{T}\Sigma_{ot}^{-1}(x_{ot} - \mu_{ot})} \right]_{mt} = \left[\Sigma_{ot}^{-1}\right]_{mt}$$

in case $x_t \in \mathbb{R}$. This means univariate data do not contain any valuable information for solving the ODML-equation and n shall quantify the number of *useful* observations.

If there are many observations sharing the same missing components, these observations should be put together. This holds especially if the missingness pattern is monotone. Then the matrix Σ_{ot}^{-1} in Eq. 14 has to be calculated only once for all observations possessing the same missingness. However, from a numerical perspective it is rather inefficient to compute the inverse Σ_{ot}^{-1} if d_t is large, particularly if there are many observations with only a few missing values. If there are e.g. 1000 realizations of a 100-dimensional random vector with 10% of its components missing at random, 1000 inverses of 90×90 sub-matrices of Σ have to be computed. This is computationally expensive, especially if the missingness pattern is irregular. In that case it is unlikely that two realizations share the same missingness and so each inverse could not be re-used. However, the inverses of the sub-matrices can be derived more efficiently from the full inverse of Σ as follows. Consider the partition

$$\Sigma^{-1} = \begin{bmatrix} A & B^\mathsf{T} \\ B & C \end{bmatrix},$$

where the $d_t \times d_t$ matrix A occupies the same range in Σ^{-1} as Σ_{ot} in Σ. Then the inverse of Σ_{ot} can be calculated by the Schur complement $\Sigma_{ot}^{-1} = A - B^\mathsf{T}C^{-1}B$. This means instead of calculating the inverse of the $d_t \times d_t$ matrix Σ_{ot}, only the inverse of the $(d - d_t) \times (d - d_t)$ matrix C has to be calculated. In the case discussed above, this corresponds to the solutions of 10×10 rather than 90×90 linear systems for 1000 observations.

Of course, this is only recommended for each observation where $d - d_t$, i.e. the number of missing values is small, since otherwise it could be more efficient to calculate the inverse of Σ_{ot} by another method, e.g. by using the sweep operator (Beaton, 1964; Goodnight, 1979). At least if the missingness pattern is monotone, we suggest to use the sweep operator for calculating the inverses of the sub-matrices. A sweep operation on a symmetric positive-definite $d \times d$ matrix Σ is a simple manipulation of Σ (Little and Rubin, 2002, p. 221) which produces another symmetric positive-definite matrix. There also exists an inverse function which can be used for reversing a previous sweep operation. By applying the sweep and the reverse sweep operator iteratively, the inverse of a sub-matrix Σ_{ot} can be efficiently calculated from the inverse of another sub-matrix of Σ which is already given by a preceding step.

The next proposition guarantees that not only Σ but also $G(\mu, \Sigma)$ is symmetric and positive-definite if the random vector X is continuously distributed.

Proposition 2. *Let $x_1, \ldots, x_n$ be a realized sample of n independent copies of a d-dimensional random vector X possessing a continuous distribution on $\mathbb{R}^d$.*

Further, let $x_{o1}, \ldots, x_{on}$ be the corresponding sample of observations following an arbitrary missingness pattern. Denote the number of complete observations by $m \leq n$ and consider the map

$$G(\mu, \Sigma) = \Sigma + g(\mu, \Sigma)\,,$$

where $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathcal{P}^d$. If $m > d$, the $d \times d$ matrix $G(\mu, \Sigma)$ is symmetric and positive-definite, too.

Proof. Since the data are continuously distributed and $m > d$, the $d \times d$ matrix

$$\Sigma\left(\frac{1}{n}\sum_{t=1}^{n}\left[d_t \cdot \frac{\Sigma_{ot}^{-1}(x_{ot}-\mu_{ot})(x_{ot}-\mu_{ot})^{\mathsf{T}}\Sigma_{ot}^{-1}}{(x_{ot}-\mu_{ot})^{\mathsf{T}}\Sigma_{ot}^{-1}(x_{ot}-\mu_{ot})}\right]_{mt}\right)\Sigma$$

is symmetric and positive-definite (a.s.). Hence, it suffices to prove that

$$\Sigma - \Sigma\left(\frac{1}{n}\sum_{t=1}^{n}\left[\Sigma_{ot}^{-1}\right]_{mt}\right)\Sigma = \frac{1}{n}\sum_{t=1}^{n}\left(\Sigma - \Sigma\left[\Sigma_{ot}^{-1}\right]_{mt}\Sigma\right) \tag{17}$$

is positive-semidefinite. Note that, without loss of generality,

$$\begin{aligned}\Sigma - \Sigma\begin{bmatrix}\Sigma_{11}^{-1} & 0\\ 0 & 0\end{bmatrix}\Sigma &= \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix} - \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\end{bmatrix}\\ &= \begin{bmatrix}0 & 0\\ 0 & \Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\end{bmatrix}.\end{aligned}$$

Since $\Sigma \in \mathcal{P}^d$, the remaining Schur complement is symmetric and positive-definite, too, i.e. the matrix given by Eq. 17 is positive-semidefinite. ∎

Now consider the function $f\colon (\mathbb{R}^d, \mathcal{P}^d) \to \mathbb{R}^d$ which is defined by

$$f(\mu, \Sigma) := \Sigma^{\frac{1}{2}}\left(\frac{1}{n}\sum_{t=1}^{n}\frac{\left[\Sigma_{ot}^{-\frac{1}{2}}(x_{ot}-\mu_{ot})\right]_{mt}}{\left\|\Sigma_{ot}^{-\frac{1}{2}}(x_{ot}-\mu_{ot})\right\|_2}\right).$$

Accordingly, by setting $F(\mu, \Sigma) := \mu + f(\mu, \Sigma)$ a fixed-point representation of the generalized multivariate sample median $\hat{\mu}$ given by Eq. 12 can be formulated as

$$F(\hat{\mu}, \widehat{\Sigma}) = \hat{\mu} + f(\hat{\mu}, \widehat{\Sigma}) = \hat{\mu}\,. \tag{18}$$

Similar to (16) it can be seen that

$$F(\mu, \Sigma) = \mu + \frac{1}{n}\sum_{t=1}^{n}\frac{x_t - \mu}{\sqrt{(x_t-\mu)^{\mathsf{T}}\Sigma^{-1}(x_t-\mu)}}$$

if there are no missing values. Thus $\hat{\mu}$ represents a solution of Tyler's original M-estimation equation (11) for the multivariate sample median in the complete-data case.

For solving the estimation equations 15 and 18 we propose the following fixed-point iteration scheme. First of all let $\mu^{(0)} = 0$ and $\Sigma^{(0)} = I_d$ be the initial

values for the location vector and dispersion matrix. Now consider the sequence $\{(\mu^{(i)}, \Sigma^{(i)})\}_{i=0,1,\ldots}$ defined by

$$\begin{bmatrix} \mu^{(i+1)} \\ \Sigma^{(i+1)} \end{bmatrix} = \begin{bmatrix} \mu^{(i)} + f(\mu^{(i)}, \Sigma^{(i)}) \\ \Sigma^{(i)} + g(\mu^{(i)}, \Sigma^{(i)}) \end{bmatrix}. \tag{19}$$

After performing a sufficiently large number N of iterations, the desired estimate for the shape matrix can be approximated by $\Omega^{(N)} = \Sigma^{(N)}/(\det \Sigma^{(N)})^{1/d}$, relying on the determinant-based normalization.

Proposition 2 guarantees that the fixed-point iteration scheme represented by (19) can only produce symmetric and positive-definite matrices, analytically, but our own experience shows that the numerical computation of inverses leads to roundoff errors that make the iterations slightly asymmetric. The asymmetric component is tiny in the beginning, but can blow up especially in higher dimensions after 50 or 100 iterations. This can be easily avoided by symmetrizing $g(\mu^{(i)}, \Sigma^{(i)}) \to \{g(\mu^{(i)}, \Sigma^{(i)}) + g(\mu^{(i)}, \Sigma^{(i)})^\mathsf{T}\}/2$ in every step.

Of course, our fixed-point algorithm can only work if there are not 'too many' missing values. In the following we give a necessary condition for the existence of the shape matrix estimate provided the missingness pattern is monotone.

Theorem 4. *Let $x_1, \ldots, x_n \in \mathbb{R}^d$ be a realized sample and $\hat{\mu} \in \mathbb{R}^d$. Further, let $x_{o1}, \ldots, x_{on}$ be the corresponding sample of observations following a monotone missingness pattern. Denote the number of complete observations by $m \leq n$. Then a solution $\widehat{\Sigma}$ of Eq. 13 exists only if $m \geq d$.*

Proof. Suppose that $m < d$ and the solution $\widehat{\Sigma}$ exists. Then Eq. 13 can be written as

$$\frac{1}{n}\sum_{t=1}^{n}\left[\widehat{\Sigma}_{ot}^{-1} y_{ot} y_{ot}^\mathsf{T} \widehat{\Sigma}_{ot}^{-1}\right]_{mt} = \frac{1}{n}\sum_{t=1}^{n}\left[\widehat{\Sigma}_{ot}^{-1}\right]_{mt}$$

with $y_{ot} := \sqrt{d_t}\,(x_{ot} - \hat{\mu}_{ot})/\left\|\widehat{\Sigma}_{ot}^{-\frac{1}{2}}(x_{ot} - \hat{\mu}_{ot})\right\|_2$. This is an ODML-equation under the normal distribution assumption given the observations $y_{o1}, \ldots, y_{on}$. Hence, the Tyler-type estimate can only exist for $x_{o1}, \ldots, x_{on}$ if the Gauss-type estimate exists for $y_{o1}, \ldots, y_{on}$. The latter can be obtained by factorizing the observed-data likelihood function (Schafer, 1997, Ch. 6.5.1) and following the method described by Schafer (1997, Ch. 6.5.2). Due to elementary properties of the sweep operator it becomes clear that the first sweep operation leads to a singular Schur complement in case $m < d$. Hence, after finishing all necessary sweep operations the final reverse sweep operation cannot produce a nonsingular covariance matrix estimate. This means $\widehat{\Sigma}$ cannot exist which contradicts the assertion at the beginning of the proof. ∎

In a separate work we will discuss necessary and sufficient conditions for the existence of the generalized M-estimator for the shape matrix given more general missingness patterns. The mathematical details are rather complicated and would go beyond the scope of the present work. For convenience we would like to mention that for the existence of such an estimate it is sufficient to have a continuous distribution on $\mathbb{R}^d$ possessing an arbitrary missingness pattern with $m > d$ complete observations.

5. Simulation Study

In the following simulation study our generalized M-estimator for the shape matrix is compared with the shape matrix estimator based on the normal distribution assumption. Let $X_1, \ldots, X_n$ be a sample of d-dimensional random vectors and suppose that the data are complete. Then

$$\widehat{\Sigma}_{\mathrm{G}} := \frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X})(X_t - \bar{X})^{\mathsf{T}}$$

represents the sample covariance matrix and $\bar{X}$ is the sample mean vector. If the data are incomplete, the location vector and covariance matrix are estimated by the ODML-estimators based on the normal distribution assumption and the factored likelihood method already mentioned in Section 3.2.2. The associated Gauss-type shape matrix estimator will be denoted by $\widehat{\Omega}_{\mathrm{G}} := \widehat{\Sigma}_{\mathrm{G}}/(\det \widehat{\Sigma}_{\mathrm{G}})^{1/d}$. When calculating Tyler's M-estimator for the shape matrix, the data are centralized by the multivariate sample median. Provided the data are incomplete, the location vector and shape matrix are estimated by the Tyler-type estimators as described in Section 4. Tyler's M-estimator and our generalized M-estimator for Ω will be denoted by $\widehat{\Omega}_{\mathrm{T}}$. In the simulation study we distinguish between situations where the location vector μ is known and where it is unknown.

For simulating heavy-tailed data we use the multivariate t-distribution with $\nu > 0$ degrees of freedom, i.e. the density of X is supposed to be

$$p(x) \propto \left(1 + \frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{\nu}\right)^{-(d+\nu)/2}.$$

It is well-known that the multivariate t-distribution is a heavy-tail distribution with tail index ν and for $\nu \to \infty$ it converges to the multivariate normal distribution. By contrast, a suitable model which allows for simulating light-tailed data is the multivariate power-exponential distribution represented by

$$p(x) \propto \exp\left\{-\left(\frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{\eta}\right)^{\gamma}\right\}$$

with $\gamma > 0$ and $\eta = d\,\Gamma\{d/(2\gamma)\}/\Gamma\{(d+2)/(2\gamma)\}$. In case $\gamma = 1$ it coincides with the multivariate normal distribution. If $\gamma > 1$ its tails are lighter and for $0 < \gamma < 1$ they are heavier than the tails of the multivariate normal distribution. In the following we consider scenarios where the data are multivariate t-distributed with $\nu = 1, 3, 5, \infty$ degrees of freedom and multivariate power-exponentially distributed with $\gamma = 5$. The data are assumed to be independent and identically distributed.

The considered number of dimensions is $d = 5$ and for the parameters we choose $\mu = 0$ and $\Sigma = I_5$. For studying the robustness of the shape matrix estimators we also simulate three additional scenarios where the normal distributions are contaminated at $(10, \ldots, 10) \in \mathbb{R}^5$. More precisely, we add an amount of $k = \lfloor nc/(1-c) \rfloor$ $(0 < c < 1)$ contaminating data to the sample, so that a proportion of $c \approx k/(n+k)$ of all data becomes contaminated. In

our numerical experiments we consider $c = 0.01, 0.05, 0.10$. We also distinguish between three different sample sizes, i.e. a small sample ($n = 100$), a moderate sample ($n = 1000$), and a large sample ($n = 10000$). The number of Monte-Carlo replications is 10000. This is a large number of replications which guarantees that the standard errors of the numerical outcomes are negligible.

The different shape matrix estimators are evaluated by the following quantities. The absolute bias of a shape matrix estimator $\widehat{\Omega}$ is defined as

$$\mathrm{AB}\big(\widehat{\Omega}\big) := \frac{1'|\mathbb{E}(\widehat{\Omega} - \Omega)|\,1}{d^2}\,,$$

where $|A|$ is the matrix of absolute values of a matrix A and 1 is a vector of ones. It is worth pointing out that in general any shape matrix estimator is biased in finite samples although it might be *asymptotically* unbiased. Hence, the absolute bias of a shape matrix estimator can be relatively large for small or moderate sample sizes even if it essentially vanishes in large samples.

The mean squared error (MSE) of the shape matrix estimator is given by the average mean squared error of all components of $\widehat{\Omega}$, i.e.

$$\mathrm{MSE}\big(\widehat{\Omega}\big) := \frac{\mathbb{E}\Big[\mathrm{tr}\big\{(\widehat{\Omega} - \Omega)(\widehat{\Omega} - \Omega)^{\mathsf{T}}\big\}\Big]}{d^2}\,.$$

Finally, the shape matrix estimators $\widehat{\Omega}_{\mathrm{G}}$ and $\widehat{\Omega}_{\mathrm{T}}$ can be compared by the relative efficiency

$$\mathrm{RE}_{\,\mathrm{T/G}} := \frac{\mathrm{MSE}(\widehat{\Omega}_{\mathrm{G}})}{\mathrm{MSE}(\widehat{\Omega}_{\mathrm{T}})}\,.$$

The results of the complete-data case are given in Table 2. Tyler's M-estimator turns out to be always favorable with respect to bias and relative efficiency unless the data are normally distributed or light-tailed. As it might be expected, T is more efficient the heavier the tails of the distribution and especially if the data are contaminated. Note that even in small samples the MSE of Tyler's M-estimator remains constant over the entire class of uncontaminated distributions. This confirms its distribution-freeness which does not depend on the sample size. Interestingly, also the absolute bias of Tyler's M-estimator remains constant if the data are uncontaminated, whereas the Gauss-type estimator becomes more biased the heavier the tails. The impact of estimating the location vector is quite negligible.

If the data are multivariate t-distributed, the asymptotic relative efficiency $\mathrm{ARE}_{\,\mathrm{T/G}} = \lim_{n\to\infty} \mathrm{RE}_{\,\mathrm{T/G}}$ can be calculated analytically (Frahm, 2009) and corresponds to

$$\mathrm{ARE}_{\,\mathrm{T/G}} = \frac{d}{d+2} \cdot \frac{\nu - 2}{\nu - 4}\,, \qquad \nu > 4\,. \tag{20}$$

Moreover, if the data follow a multivariate power-exponential distribution it can be shown that

$$\mathrm{ARE}_{\,\mathrm{T/G}} = \frac{d}{d+2} \cdot \frac{\Gamma\big(\frac{d+2}{2\gamma} + \frac{1}{\gamma}\big)\Gamma\big(\frac{d+2\gamma}{2\gamma}\big)}{\Gamma\big(\frac{d+2\gamma}{2\gamma} + \frac{1}{\gamma}\big)\Gamma\big(\frac{d+2}{2\gamma}\big)}\,, \qquad \gamma > 0\,.$$

This is fairly reflected by the relative efficiencies which can be found in the large-sample panel of Table 2 regarding the parameters $\nu = 5$ ($\mathrm{ARE}_{\mathrm{T/G}} = 2.1429$), $\nu = \infty$ ($\mathrm{ARE}_{\mathrm{T/G}} = 0.7143$), and $\gamma = 5$ ($\mathrm{ARE}_{\mathrm{T/G}} = 0.5735$).

Now we consider the three different missingness mechanisms MCAR, MAR, and NMAR. Let x $(5 \times n)$ be a realized sample of uncontaminated data. Some of the data in the first row of x are missing. This is denoted by $m_t = 1$ if x_{1t} is missing and $m_t = 0$ if it is observed $(t = 1, \ldots, n)$. The missing data are MCAR if the missingness pattern $M = (m_1, \ldots, m_n)$ is stochastically independent of the sample. By contrast, if the distribution of M depends only on the observed part of x, the missing data are MAR and if the missingness is determined by the unobserved part of the sample, the missing data are NMAR. For the MCAR case we simulate a $1 \times n$ vector Y which is stochastically independent of the sample. Each component of Y is t-distributed with ν degrees of freedom ($Y_t \sim t_\nu$) or power-exponentially distributed with parameter $\gamma = 5$ ($Y_t \sim p_5$) and $Y_1, \ldots, Y_n$ are mutually independent. Now the element x_{1t} is considered as missing whenever $y_t < t_\nu^{-1}(0.75)$ or $y_t < p_5^{-1}(0.75)$, respectively. Further, for the MAR case x_{1t} is missing if $x_{2t} < t_\nu^{-1}(0.75)$ or $x_{2t} < p_5^{-1}(0.75)$, whereas for the NMAR case it is unobserved whenever $x_{1t} < t_\nu^{-1}(0.75)$ or $x_{1t} < p_5^{-1}(0.75)$. Hence, approximately 75% of the data in the first row of x are missing for each missingness mechanism. Table 3 contains the results of the MCAR case and, accordingly, Table 4 and Table 5 report the outcomes of the MAR and NMAR case.

In the MCAR case the overall picture is essentially the same as in the complete-data case, i.e. the generalized M-estimator is always favorable except for normally distributed and light-tailed data. Note that the absolute biases of the Gauss-type and Tyler-type estimator increase in small samples compared to the complete-data case. However, Table 3 also reveals that the estimators indeed become unbiased as the sample size tends to infinity. The MSE of the generalized M-estimator is constant over the entire class of uncontaminated distributions, which follows from the arguments given in Section 3.2.3.

If the missing data are only MAR, Table 4 reveals that the Tyler-type shape matrix estimator is asymptotically biased. This has been discussed in Section 3.2.2. By contrast, as a direct consequence of the results given in Section 3.2.1, the Gauss-type shape matrix estimator remains asymptotically unbiased. Nevertheless, as long as the sample size is small or moderate, the Tyler-type estimator remains favorable in terms of the relative efficiency unless the data are normally distributed or light-tailed. In large samples the Gauss-type estimator becomes more efficient only for t-distributed data with $\nu = 5$ degrees of freedom.

Finally, if the missing data are NMAR (see Table 5), both the Gauss-type and the Tyler-type estimator are asymptotically biased. In most cases the Tyler-type estimator turns out to be slightly more biased than the Gauss-type estimator if the data are uncontaminated. By contrast, if the data are contaminated, the bias of the Gauss-type estimator becomes tremendously large compared to the bias of the Tyler-type estimator. The same conclusion can be drawn for any other missingness mechanism. Similarly, all things considered, the impact of estimating the location vector remains negligible whether the data are missing or not.

Table 1: This table indicates whether the Gauss-type (G) or Tyler-type estimator (T) is favorable if the data are light-, normal-, heavy-tailed or contaminated, the sample sizes are small, moderate or large and either there are no missing data (NM) or missing data under the missingness mechanisms MCAR, MAR, and NMAR. The estimators are evaluated with respect to absolute bias (AB) and mean squared error (MSE). In case no estimator is dominated by its competitor, the table contains a line in the corresponding cell.

Small Samples					
		NM	MCAR	MAR	NMAR
Light tails	AB	G	G	G	G
	MSE	G	G	—	G
Normal tails	AB	G	G	G	G
	MSE	G	G	—	G
Heavy tails	AB	T	T	—	—
	MSE	T	T	T	—
Contaminated	AB	T	T	T	T
	MSE	T	T	T	T
Moderate Samples					
		NM	MCAR	MAR	NMAR
Light tails	AB	G	G	G	—
	MSE	G	G	G	—
Normal tails	AB	G	G	G	G
	MSE	G	G	G	G
Heavy tails	AB	T	T	—	—
	MSE	T	T	T	—
Contaminated	AB	T	T	T	T
	MSE	T	T	T	T
Large Samples					
		NM	MCAR	MAR	NMAR
Light tails	AB	G	G	G	G
	MSE	G	G	G	G
Normal tails	AB	G	G	G	G
	MSE	G	G	G	G
Heavy tails	AB	T	T	—	—
	MSE	T	T	—	—
Contaminated	AB	T	T	T	T
	MSE	T	T	T	T

The results of the simulation studies are summarized in Table 1. Our conclusion is that the Tyler-type estimator for the shape matrix is favorable whenever the data are heavy-tailed or contaminated, whereas the Gauss-type estimator serves its purpose if they are clean and multivariate normally distributed or light-tailed.

6. Conclusion

In the present work we derived generalized M-estimators for the multivariate median and shape matrix of a random vector X. The presented M-estimators can be seen as 'natural' generalizations of the multivariate sample median proposed by Hettmansperger and Randles (2002) as well as Tyler's M-estimator for the shape matrix to the case of incomplete data. If X is generalized elliptically distributed our shape matrix estimator retains the most important property of

Tyler's counterpart, namely it is invariant under arbitrary changes of the generating variate. This means the underlying mechanism which is responsible for outliers or clusters can be eliminated and thus the generalized M-estimator for the shape matrix becomes 'most robust'. We also derived its asymptotic distribution under the MCAR assumption of missing-data analysis. An important argument in favor of our estimator is that if the data stem from a generalized elliptical distribution, no nuisance parameters need to be estimated for assessing its asymptotic distribution since its asymptotic covariance matrix solely follows from the estimate itself. Moreover, we developed a fast algorithm for calculating the generalized M-estimates for location and shape and gave some practical advice for its numerical implementation. A simulation study for the complete-data and the incomplete-data case reveals that for heavy-tailed or contaminated data our shape matrix estimator should be always preferred. Only if the data are uncontaminated and normal- or light-tailed, the Gauss-type estimator presented by Little and Rubin (2002, Ch. 7) as well as Schafer (1997, Section 6.5) remains preferable.

Acknowledgements

The authors would like to thank David Tyler and Don Rubin for insightful discussions. G.F. thanks the NEC Laboratories Europe (NEC Europe Ltd.) for their kind support and hospitality. Many thanks belong also to Karl Mosler for his important suggestions and Rainer Dyckerhoff for an excellent introduction to the sweep operator and many fruitful discussions about missing-data analysis.

References

Adrover, J., 1998. Minimax bias-robust estimation of the dispersion matrix of a multivariate distribution. Annals of Statistics 26, 2301–2320.

Beaton, A., 1964. The use of special matrix operations in statistical calculus. Research bulletin RB-64-51, Educational Testing Service, Princeton, NJ.

Cambanis, S., Huang, S., Simons, G., 1981. On the theory of elliptically contoured distributions. Journal of Multivariate Analysis 11, 368–385.

Dümbgen, L., 1998. On Tyler's M-functional of scatter in high dimension. Annals of the Institute of Statistical Mathematics 50, 471–491.

Dümbgen, L., Tyler, D., 2005. On the breakdown properties of some multivariate M-functionals. Scandinavian Journal of Statistics 32, 247–264.

Fang, K., Kotz, S., Ng, K., 1990. Symmetric Multivariate and Related Distributions. Chapman & Hall.

Frahm, G., 2004. Generalized Elliptical Distributions: Theory and Applications. Ph.D. thesis, University of Cologne, Department of Economic and Social Statistics, Germany.

Frahm, G., 2009. Asymptotic distributions of robust shape matrices and scales. Journal of Multivariate Analysis, DOI: 10.1016/j.jmva.2008.11.007.

Frahm, G., Jaekel, U., 2007. Tyler's M-estimator, random matrix theory, and generalized elliptical distributions with applications to finance. Discussion paper, University of Cologne, Department of Economic and Social Statistics, Germany.

Goodnight, J., 1979. A tutorial on the sweep operator. American Statistician 33, 149–158.

Hallin, M., Oja, H., Paindaveine, D., 2006. Semiparametrically efficient rank-based inference for shape. II. Optimal R-estimation of shape. Annals of Statistics 34, 2757–2789.

Hallin, M., Paindaveine, D., 2006. Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity. Annals of Statistics 34, 2707–2756.

Hampel, F., Ronchetti, E., Rousseeuw, P., Stahel, W., 1986. Robust Statistics. John Wiley.

Hettmansperger, T., Randles, R., 2002. A practical affine equivariant multivariate median. Biometrika 89, 851–860.

Huber, P., 2003. Robust Statistics. John Wiley.

Kent, J., Tyler, D., 1988. Maximum likelihood estimation for the wrapped cauchy distribution. Journal of Applied Statistics 15, 247–254.

Kent, J., Tyler, D., 1991. Redescending m-estimates of multivariate location and scatter. Annals of Statistics 19, 2102–2119.

Kenward, M., Molenberghs, G., 1998. Likelihood based frequentist inference when data are missing at random. Statistical Science 13, 236–247.

Kring, S., Rachev, S., Höchstötter, M., Fabozzi, F., Bianchi, M., 2009. Multi-tail generalized elliptical distributions for asset returns. Forthcoming in the Econometrics Journal.

Little, R., 1988. Robust estimation of the mean and covariance matrix from data with missing values. Applied Statistics 37, 23–38.

Little, R., Rubin, D., 2002. Statistical Analysis with Missing Data, 2nd Edition. John Wiley.

Liu, J., Dey, D., 2004. Skew-elliptical distributions. In: Genton, M. (Ed.), Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality. Chapman & Hall, Ch. 3.

Mardia, K., Jupp, P., 2000. Directional Statistics. John Wiley.

Maronna, R., Martin, D., Yohai, V., 2006. Robust Statistics. John Wiley.

Maronna, R., Yohai, V., 1990. The maximum bias of robust covariances. Communications in Statistics: Theory and Methods 19, 3925–3933.

Paindaveine, D., 2008. A canonical definition of shape. Statistics and Probability Letters 78, 2240–2247.

Randles, R., 1989. A distribution-free multivariate sign test based on interdirections. Journal of the American Statistical Association 84, 1045–1050.

Randles, R., 2000. A simpler, affine-invariant, multivariate, distribution-free sign test. Journal of the American Statistical Association 95, 1263–1268.

Schafer, J., 1997. Analysis of Incomplete Multivariate Data. Chapman & Hall.

Schafer, J., Graham, J., 2002. Missing data: our view of the state of the art. Psychological Methods 7, 147–177.

Schott, J., 1997. Matrix Analysis for Statistics. John Wiley.

Tyler, D., 1983. Robustness and efficiency properties of scatter matrices. Biometrika 70, 411–420.

Tyler, D., 1987a. A distribution-free M-estimator of multivariate scatter. Annals of Statistics 15, 234–251.

Tyler, D., 1987b. Statistical analysis for the angular central Gaussian distribution on the sphere. Biometrika 74, 579–589.

Visuri, S., 2001. Array and multichannel signal processing using nonparametric statistics. Ph.D. thesis, Helsinki University of Technology, Signal Processing Laboratory, Finland.

Table 2: Results of the simulation study for the complete-data case, where t_ν indicates a 5-dimensional t-distribution with ν degrees of freedom, t_∞ stands for a clean normal distribution, p_5 for a clean power-exponential distribution with parameter $\gamma = 5$, and t^c_∞ denotes a normal distribution with $c = 1\%$, 5%, and 10% of the data being contaminated.

$n = 100$								
μ known	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	3353.2	.0379	.0132	.0065	.0052	.6580	2.5712	4.9167
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0087	.0090	.0093	.0090	.0091	.0192	.0764	.2226
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$2 \cdot 10^{11}$	.4181	.0321	.0121	.0095	.4542	6.6745	24.308
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0177	.0178	.0176	.0177	.0175	.0175	.0218	.0683
$\mathrm{RE}_{\mathrm{T/G}}$	10^{13}	23.524	1.8221	.6843	.5420	26.013	305.60	356.13
μ unknown	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	1580.7	.0382	.0130	.0063	.0052	.6586	2.4885	4.5522
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0091	.0091	.0091	.0091	.0090	.0202	.0818	.2676
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$4 \cdot 10^{10}$	.3114	.0322	.0123	.0097	.4552	6.2555	20.854
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0179	.0179	.0180	.0181	.0178	.0177	.0231	.0921
$\mathrm{RE}_{\mathrm{T/G}}$	$2 \cdot 10^{12}$	17.376	1.790	.6798	.5422	25.649	271.18	226.42
$n = 1000$								
μ known	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	151.34	.0108	.0018	.0007	.0006	.6406	2.6001	4.8551
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0010	.0010	.0010	.0010	.0009	.0125	.0739	.2202
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$2 \cdot 10^{8}$	.0586	.0031	.0011	.0009	.4256	6.8007	23.630
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0016	.0016	.0016	.0016	.0016	.0017	.0077	.0544
$\mathrm{RE}_{\mathrm{T/G}}$	10^{11}	37.024	1.9781	.7171	.5712	246.64	880.01	434.70
μ unknown	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	134.87	.0115	.0018	.0008	.0006	.6353	2.4939	4.4559
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0009	.0011	.0011	.0010	.0009	.0124	.0779	.2667
MSE($\widehat{\Omega}_{\mathrm{G}}$)	10^{8}	.0831	.0033	.0011	.0009	.4187	6.2589	19.910
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0016	.0016	.0016	.0016	.0016	.0017	.0084	.0783
$\mathrm{RE}_{\mathrm{T/G}}$	$9 \cdot 10^{10}$	52.380	2.0908	.7136	.5726	242.51	743.13	254.30
$n = 10000$								
μ known	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	141.13	.0035	.0003	.0002	.0001	.6443	2.6189	4.8477
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0001	.0001	.0002	.0002	.0001	.0117	.0743	.2202
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$2 \cdot 10^{8}$	.0155	.0003	.0001	.0001	.4298	6.8961	23.551
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0002	.0002	.0002	.0002	.0002	.0003	.0066	.0532
$\mathrm{RE}_{\mathrm{T/G}}$	10^{12}	100.27	2.0754	.7231	.5811	1342.6	1048.1	442.61
μ unknown	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	95.219	.0037	.0002	.0001	.0001	.6387	2.5098	4.4472
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0001	.0002	.0001	.0002	.0002	.0119	.0786	.2669
MSE($\widehat{\Omega}_{\mathrm{G}}$)	10^{8}	.0211	.0003	.0001	.0001	.4225	6.3360	19.826
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0002	.0002	.0002	.0002	.0002	.0003	.0073	.0772
$\mathrm{RE}_{\mathrm{T/G}}$	$8 \cdot 10^{11}$	135.75	2.1578	.7236	.5788	1293.3	864.38	256.88

Table 3: Results of the simulation study for the incomplete-data case, where the missing data are assumed to be MCAR. The symbols are the same as in Table 2.

$n = 100$								
μ known	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	45.522	.0567	.0272	.0156	.0123	.7360	2.7665	5.3924
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0225	.0221	.0223	.0226	.0222	.0470	.1542	.4025
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$5 \cdot 10^6$	.3095	.0677	.0320	.0247	.5770	7.7537	29.349
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0509	.0500	.0506	.0499	.0503	.0477	.0668	.2143
$\mathrm{RE}_{\mathrm{T/G}}$	10^8	6.1840	1.3380	.6415	.4901	12.087	116.13	136.95
μ unknown	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	618.74	.0580	.0287	.0167	.0135	.7424	2.7049	5.0409
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0238	.0242	.0233	.0233	.0234	.0475	.1632	.4661
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$6 \cdot 10^9$	.3503	.0713	.0342	.0266	.5880	7.4220	25.671
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0526	.0524	.0520	.0515	.0510	.0487	.0715	.2738
$\mathrm{RE}_{\mathrm{T/G}}$	10^{11}	6.6897	1.3715	.6647	.5209	12.062	103.87	93.748
$n = 1000$								
μ known	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	186.89	.0135	.0030	.0014	.0011	.6911	2.7169	5.1664
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0021	.0019	.0020	.0020	.0020	.0270	.1369	.3685
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$4 \cdot 10^8$	.0605	.0058	.0023	.0019	.4984	7.4269	26.760
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0034	.0034	.0034	.0034	.0034	.0043	.0262	.1506
$\mathrm{RE}_{\mathrm{T/G}}$	10^{11}	17.861	1.6957	.6836	.5434	115.94	283.79	177.68
μ unknown	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	398.54	.0125	.0034	.0013	.0012	.6859	2.6100	4.7493
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0021	.0023	.0021	.0019	.0020	.0269	.1422	.4256
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$3 \cdot 10^9$	.0491	.0059	.0023	.0018	.4911	6.8562	22.621
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0034	.0034	.0034	.0034	.0034	.0043	.0278	.1970
$\mathrm{RE}_{\mathrm{T/G}}$	$9 \cdot 10^{11}$	14.349	1.7227	.6812	.5453	114.35	246.40	114.84
$n = 10000$								
μ known	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	4215.0	.0038	.0005	.0002	.0002	.6928	2.7313	5.1460
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0003	.0003	.0003	.0003	.0003	.0254	.1366	.3657
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$6 \cdot 10^{11}$	.0117	.0006	.0002	.0002	.4997	7.5009	26.5343
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0003	.0003	.0003	.0003	.0003	.0013	.0237	.1460
$\mathrm{RE}_{\mathrm{T/G}}$	$2 \cdot 10^{15}$	35.590	1.9540	.6885	.5461	371.78	316.29	181.72
μ unknown	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	59.086	.0057	.0005	.0002	.0001	.6866	2.6182	4.7225
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0003	.0003	.0003	.0003	.0002	.0255	.1420	.4221
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$2 \cdot 10^7$	.1535	.0006	.0002	.0002	.4909	6.8947	22.353
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0003	.0003	.0003	.0003	.0003	.0014	.0254	.1913
$\mathrm{RE}_{\mathrm{T/G}}$	$6 \cdot 10^{10}$	464.40	1.9487	.6939	.5483	362.62	271.83	116.82

Table 4: Results of the simulation study for the incomplete-data case, where the missing data are assumed to be MAR. The symbols are the same as in Table 2.

$n = 100$								
μ known	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	142.39	.0490	.0242	.0144	.0151	.7287	2.7524	5.3598
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0238	.0275	.0284	.0297	.0309	.0561	.1683	.4281
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$2 \cdot 10^8$	.3497	.0598	.0291	.0245	.5655	7.6781	29.002
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0484	.0505	.0496	.0500	.0505	.0474	.0734	.2399
$\mathrm{RE}_{\mathrm{T/G}}$	$4 \cdot 10^9$	6.9242	1.2039	.5830	.4859	11.920	104.65	120.90
μ unknown	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	1486.0	.0647	.0351	.0255	.0248	.7966	2.8398	5.2869
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0232	.0299	.0317	.0336	.0350	.0575	.1614	.4560
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$4 \cdot 10^{10}$	.7019	.1071	.0637	.0631	.6990	8.2327	28.374
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0481	.0483	.0482	.0492	.0495	.0469	.0719	.2755
$\mathrm{RE}_{\mathrm{T/G}}$	$8 \cdot 10^{11}$	14.522	2.2242	1.2958	1.2752	14.898	114.57	102.99
$n = 1000$								
μ known	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	66.518	.0162	.0114	.0014	.0088	.6858	2.7009	5.1390
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0148	.0190	.0200	.0214	.0223	.0470	.1594	.3946
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$4 \cdot 10^7$	.0614	.0068	.0022	.0023	.4911	7.3401	26.477
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0046	.0055	.0058	.0061	.0064	.0070	.0328	.1739
$\mathrm{RE}_{\mathrm{T/G}}$	$9 \cdot 10^9$	11.118	1.1711	.3575	.3553	70.256	223.79	152.28
μ unknown	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	713.22	.0174	.0121	.0020	.0085	.7322	2.7425	4.9773
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0155	.0203	.0214	.0232	.0244	.0458	.1494	.4199
MSE($\widehat{\Omega}_{\mathrm{G}}$)	10^{10}	.0536	.0093	.0034	.0036	.5701	7.5970	24.907
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0047	.0057	.0060	.0065	.0069	.0071	.0326	.1999
$\mathrm{RE}_{\mathrm{T/G}}$	$2 \cdot 10^{12}$	9.4042	1.5572	.5259	.5271	80.220	232.93	124.57
$n = 10000$								
μ known	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	159.99	.0235	.0133	.0002	.0081	.6872	2.7156	5.1181
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0139	.0179	.0188	.0203	.0213	.0461	.1602	.3914
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$2 \cdot 10^8$	.0224	.0021	.0002	.0007	.4921	7.4151	26.248
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0017	.0026	.0028	.0032	.0035	.0042	.0306	.1689
$\mathrm{RE}_{\mathrm{T/G}}$	10^{11}	8.6596	.7374	.0669	.1873	118.27	242.65	155.44
μ unknown	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	89.136	.0235	.0134	.0002	.0081	.7323	2.7502	4.9501
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0146	.0192	.0203	.0219	.0233	.0448	.1497	.4165
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$5 \cdot 10^7$	.0232	.0024	.0003	.0008	.5685	7.6330	24.618
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0018	.0029	.0032	.0036	.0040	.0044	.0303	.1942
$\mathrm{RE}_{\mathrm{T/G}}$	$3 \cdot 10^{10}$	7.9997	.7603	.0906	.1935	129.85	252.13	126.75

Table 5: Results of the simulation study for the incomplete-data case, where the missing data are assumed to be NMAR. The symbols are the same as in Table 2.

$n = 100$								
μ known	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	413.29	.0513	.0391	.0432	.0437	.6614	2.4758	4.8249
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0495	.0635	.0662	.0710	.0739	.0913	.1651	.3604
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$2 \cdot 10^9$	.3093	.0881	.0642	.0601	.4874	6.2142	23.461
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.1103	.1396	.1467	.1596	.1630	.1414	.1319	.2254
$\mathrm{RE}_{\mathrm{T/G}}$	$2 \cdot 10^{10}$	2.2154	.6005	.4020	.3690	3.4477	47.117	104.08
μ unknown	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	1588.7	.0897	.0797	.0982	.1157	.8878	3.2100	5.9524
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0435	.0836	.0919	.1082	.1213	.1218	.2280	.5613
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$4 \cdot 10^{10}$	1.6051	.0898	.0756	.0894	.8319	10.498	35.997
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0544	.0748	.0824	.0982	.1133	.0877	.1006	.3747
$\mathrm{RE}_{\mathrm{T/G}}$	$8 \cdot 10^{11}$	21.471	1.0904	.7697	.7886	9.4827	104.35	96.056
$n = 1000$								
μ known	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	110.11	.0423	.0452	.0443	.0432	.6228	2.4410	4.6417
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0501	.0631	.0657	.0696	.0731	.0899	.1638	.3218
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$7 \cdot 10^7$	.2377	.0260	.0217	.0203	.4199	6.0047	21.608
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0299	.0469	.0513	.0578	.0632	.0535	.0538	.1306
$\mathrm{RE}_{\mathrm{T/G}}$	$2 \cdot 10^9$	5.0708	.5067	.3754	.3212	7.8424	111.55	165.43
μ unknown	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	98.316	.0339	.0508	.0815	.1012	.8215	3.0881	5.6010
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0299	.0657	.0742	.0878	.0996	.1035	.2132	.5240
MSE($\widehat{\Omega}_{\mathrm{G}}$)	10^8	.0613	.0211	.0388	.0560	.7016	9.6367	31.622
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0088	.0277	.0338	.0448	.0557	.0402	.0573	.2834
$\mathrm{RE}_{\mathrm{T/G}}$	10^{10}	2.2153	.6263	.8652	1.0056	17.473	168.141	111.57
$n = 10000$								
μ known	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	784.77	.0459	.0462	.0443	.0429	.6243	2.4541	4.6270
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0501	.0630	.0657	.0696	.0726	.0898	.1643	.3186
MSE($\widehat{\Omega}_{\mathrm{G}}$)	10^{10}	.0460	.0211	.0189	.0176	.4206	6.0660	21.462
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0249	.0410	.0450	.0511	.0561	.0473	.0483	.1240
$\mathrm{RE}_{\mathrm{T/G}}$	$4 \cdot 10^{11}$	1.1204	.4683	.3696	.3134	8.8969	125.63	173.01
μ unknown	t_1	t_3	t_5	t_∞	p_5	$t_\infty^{0.01}$	$t_\infty^{0.05}$	$t_\infty^{0.10}$
AB($\widehat{\Omega}_{\mathrm{G}}$)	1458.6	.0201	.0466	.0800	.0429	.8218	3.0968	5.5722
AB($\widehat{\Omega}_{\mathrm{T}}$)	.0287	.0640	.0726	.0860	.0726	.1021	.2134	.5217
MSE($\widehat{\Omega}_{\mathrm{G}}$)	$4 \cdot 10^{12}$	.0148	.0143	.0360	.0176	.7010	9.6842	31.277
MSE($\widehat{\Omega}_{\mathrm{T}}$)	.0059	.0244	.0304	.0410	.0561	.0367	.0546	.2776
$\mathrm{RE}_{\mathrm{T/G}}$	$7 \cdot 10^{14}$	.6060	.4689	.8777	.3134	19.125	177.38	112.67