DOTTORATO IN ECONOMIA E METODI QUANTITATIVI

CURRICULUM IN METODI STATISTICI APPLICATI ALL'ECONOMIA E ALL'IMPRESA

XXIX CICLIO

Tesi di dottorato

LIKELIHOOD–FREE METHODS FOR LARGE DIMENSIONAL SPARSE NON-GAUSSIAN MODELS

Dottorando: Paola Stolfi

Relatore: Prof. Caterina Conigliani

Coordinatore: Prof. Luca Salvatici

Acknowledgement

I am deeply thankful to my mentor and guide Dr. Mauro Bernardi from University of Padova. With his guidance, research attitude and enthusiasm he was a cornerstone for the development and completion of this project. His inspirations and meticulous supervision helped me in transforming a simple idea into a valuable research project. I am also very grateful to him for his timely advices and his co–operation that where fundamental for me to learn how to carry out valued research.

Thesis Committee: Prof. C. Conigliani, Prof. L. Petrella

To Blunt

Contents

Chapter 1

Introduction

Model–based statistical inference primarily deals with parameters estimation. Under the usual assumption of data being generated from a fully specified model belonging to a given family of distributions F_ϑ indexed by a parameter $\vartheta \subset \Theta \in \mathbb{R}^p$, inference on the true unknown parameter ϑ_0 can be easily performed by maximum likelihood. However, in some pathological situations the maximum likelihood estimator (MLE) is difficult to compute either because of the model complexity or because the probability density function is not analytically available. For example, the computation of the log–likelihood may involve numerical approximations or integrations that highly deteriorate the quality of the resulting estimates. Moreover, as the dimension of the parameter space increases the computation of the likelihood or its maximisation in a reasonable amount of time becomes even more prohibitive.

In all those circumstances, the researcher should resort to alternative solutions. The method of moments or its generalised versions (GMM), Hansen (1982) or (EMM), Gallant and Tauchen (1996), may constitute feasible solutions when expressions for some moment conditions that uniquely identify the parameters of interest are analytically available. When this is not the case, simulation–based methods, such as, the method of simulated moments (MSM), McFadden (1989), the method of simulated maximum likelihood (SML), Gouriéroux and Monfort (1996) and its nonparametric version Kristensen and Shin (2012) or the indirect inference (II) method Gouriéroux et al. (1993), are the only viable solutions to the inferential problem. Jiang and Turnbull (2004) give a comprehensive review of indirect inference from a statistical point of view. Despite their appealing characteristics of only requiring to be able to simulate from the specified DGP, some of those methods suffer from serious drawbacks. The MSM, for example, requires that the existence of the moments of the postulated DGP is guaranteed, while, the II method relies on an alternative, necessarily misspecified, auxiliary model as well as on a strong from of identification between the parameters of interests and those of the auxiliary model. The quantile–matching estimation method (QM), Koenker (2005), exploits the same idea behind the method of moments without imposing any conditions on the moment finiteness. The QM approach estimates model parameters by matching the empirical percentiles with their theoretical counterparts thereby requiring only the existence of a closed form expression for the quantile function.

All those approaches do not effectively deal with the curse of dimensionality problem, i.e., the situation where the number of parameters grows quadratically or exponentially with the dimension of the problem. Indeed, the right identification of the sparsity patterns becomes crucial because it reduces the number of parameters to be estimated. Those reasonings motivate the use of sparse estimators that automatically shrink to zero some parameters, such as, for example, the off diagonal elements of the variance–covariance matrix. Several works related to sparse estimation of the covariance matrix are available in literature; most of them are related to the graphical models, where the precision matrix, e.g., the inverse of the covariance matrix, represents the conditional dependence structure of the graph. Friedman et al. (2008) propose a fast algorithm based on coordinate–wise updating scheme in order to estimate a sparse graph using the least absolute shrinkage and selection operator (LASSO) ℓ_1–penalty of Tibshirani (1996). Meinshausen and Bühlmann (2006) propose a method for neighbourhood selection using the LASSO ℓ_1–penalty as an alternative to covariance selection for Gaussian graphical models where the number of observations is less than the number of variables. Gao and Massam (2015) estimate the variance–covariance matrix of symmetry–constrained Gaussian models using three different ℓ_1–type penalty functions, i.e., the LASSO, the smoothly clipped absolute deviation (SCAD) of Fan and Li (2001) and the minimax concave penalty (MCP) of Zhang (2010). Bien and Tibshirani (2011) proposed a penalised version of the log–likelihood function, using the LASSO penalty, in order to estimate a sparse covariance matrix of a multivariate Gaussian distribution. Sparse estimation has been proposed mainly either within the regression framework or in the context of Gaussian graphical models. In boh those cases, sparsity patterns are imposed by penalising a Gaussian log–likelihood.

In this work we handle the lack of the model–likelihood or the existence of valid moment conditions together with the curse of dimensionality problem within a high–dimensional non–Gaussian framework. Specifically, our approach penalises the objective function of simulation–based inferential procedures such as the II method of Gouriéroux et al. (1993) and the Method of Simulated Quantiles (MSQ) of Dominicy and Veredas (2013). The II method replaces the maximum likelihood estimator of the model parameters with a quasi–maximum likelihood estimator which relies on an alternative auxiliary model and then exploits simulations from the original model to correct for inconsistency. The MSQ instead estimates parameters by minimising a quadratic distance between a vector of quantile–based summary statistics calculated on the available sample of observations and those calculated on synthetic observations generated from the model. Of course, there are many similarities between the two methods. MSQ only implies to be able to simulate from the model, relaxing any assumption about its analytical tractability or the existence of moments or any specification of the quantile function. However, unlike II, MSQ only relies on appropriately chosen functions of quantiles that drive information from data to the parameters of interest, while II uses the likelihood of the auxiliary model as a replacement to the intractable generative model likelihood. Another interesting property of MSQ estimates is that they inherits the robust properties of quantiles while retaining levels of efficiency comparable with the II estimators. Moreover, it is not trivial to identify appropriate quantile–based statistics in the context of conditional models where the effect of exogenous regressors or latent variables is also included. The II method instead handles unconditional models as much as conditional ones because it relies on the statistics of the auxiliary model. Those reasons motivate our solution to extend the MSQ to a multivariate non–Gaussian framework where data are assumed to be independently and identically distributed as the postulated parametric model. As concerns the II method, instead, we also consider conditional univariate and multivariate models.

The multivariate Method of Simulated Quantiles (MMSQ) and its sparse counterpart (S–MMSQ) are introduced in Chapter 3. We establish consistency and asymptotic normality of the proposed MMSQ estimators under mild conditions on the underlying true data generating process. Few illustrative examples detail how we calculate all the quantities involved in the asymptotic variance–covariance matrix are provided. The asymptotic variance–covariance matrix of the MMSQ estimators are helpful to derive their efficient versions.

The sparse II method (S–II) instead is addressed in Chapter 4. The organisation of the chapter follows along the same lines as Chapter 3. Specifically, we first introduce the penalised Indirect Inference method and then we derive the asymptotic properties of the proposed estimator under mild conditions on the data generating process.

The proposed methods can be effectively used to make inference on the parameters of large–dimensional distributions such as, for example, Stable, Elliptical Stable, Skew–Elliptical Stable, Copula, Multivariate Gamma and Tempered Stable. Among those, the Stable distribution allows for infinite variance, skewness and heavy–tails that exhibit power decay allowing extreme events to have higher probability mass than in Gaussian model. For a summary of the properties of the stable distributions see Zolotarev (1964) and Samorodnitsky and Taqqu (1994), which provide a good theoretical background on heavy–tailed distributions. Univariate Stable laws have been studied in many branches of the science and their theoretical properties have been deeply investigated from multiple perspectives, therefore many tools are now available for estimation and inference on parameters, to evaluate the cumulative density or the quantile function, or to perform fast simulation. Stable distribution plays an interesting role in modelling multivariate data. Its peculiarity of having heavy tailed properties and its closeness under summation make it appealing in the financial contest. Nevertheless, multivariate Stable laws pose several challenges that go further beyond the lack of closed form expression for the density. Although general expressions for the multivariate density have been provided by Abdul-Hamid and Nolan (1998), Byczkowski et al. (1993) and Matsui and Takemura (2009), their computations is still not feasible in dimension larger than two. A recent overview of multivariate Stable distributions can be found in Nolan (2008). Chapter 2 is devoted to the Elliptical Stable and Skew Elliptical distribution previously introduced by Branco and Dey (2001) as an interesting application of the MMSQ.

As regards applications to real data considered in Chapter 5, we first consider the well–known portfolio optimisation problem, where the performances of the ESD and SESD distributions are compared to those obtained under alternative assumption for the underlying DGP. Portfolio optimisation has a long tradition in finance since the seminal paper of Markowitz (1952) that introduced the mean–variance (MVO) approach. The MVO approach relies on quite restrictive conditions about the underlying DGP that are relaxed by the assuming multivariate Stable distributions. Furthermore, the assumption of elliptically contoured joint returns has important implications about the risk–measures than can be considered in the portfolio optimisation problem. The usual consequence of the elliptical assumption is that Value–at–Risk and variance are coherent risk measures, see Artzner et al. (1999). However, since Stable distributions do not have finite second moment, we consider a portfolio allocation problem where the expected return is traded–off against higher Value–at–Risk profiles that make investment less attractive. Portfolio optimisation under Stable distributed asset returns has been first considered by Fama (1965), and, more recently, by Gamba (1999), Meerschaert and Scheffler (2003), Ortobelli et al. (2004).

Chapter 2

The multivariate Stable distributions

2.1 Introduction

Stable distributions have been introduced in many fields such as hydrology, telecommunications, physics and finance as a generalisation of the Gaussian distribution to model data that exhibits a high degree of heterogeneity. Stable distributions allow also for infinite variance, skewness and heavy–tails that are characterised by power decay allowing extreme events to have higher probability mass than in Gaussian model. For a summary of the properties of the Stable distributions see Zolotarev (1964) and Samorodnitsky and Taqqu (1994), which provide a good theoretical background on heavy–tailed distributions. The practical use of heavy–tailed distributions in many different fields is well documented in the book of Adler et al. (1998), which also reviews the estimation techniques. In finance, the first studies on the hypothesis of Stable distributed stock prices dates back to the pioneering works of Mandelbrot (2012), Fama (2012), Fama and Roll (1968) and Fama (1965). They propose Stable distributions and give some statistical instruments for the inference on the characteristic exponent. The empirical evidence from financial markets motivates in Brenner (1974) the use of Stable distributions as innovations terms in dynamic models for stationary time series of stock returns. In the related context of financial risk modelling Stable distributions have been used by Bradley and Taqqu (2003) and Mikosch (2003). The works of Mittnik et al. (1998) and Rachev and Mittnik (2000) provide a complete analysis of the theoretical and empirical aspects of the Stable distributions in finance. In survival models, Qiou et al. (1999) model the heterogeneity within survival times of a population through common latent factors which follow Stable distributions. Stable distributions are also used to model heterogeneity over time and non–linear dependencies exhibited by the data. For an introduction to time series models with Stable noises, see Ravishanker and Qiou (1998) and Mikosch (2003).

Theoretical properties of univariate Stable laws have been deeply investigated from multiple perspectives, therefore many tools are now available for estimation and inference on parameters, to evaluate the cumulative density or the quantile function, or to perform fast simulation. Different estimation methods for Stable distributions have been proposed in the literature. Buckle (1995) earlier proposed a Bayesian approach, while the maximum likelihood estimator has been proposed along with the asymptotic theory by DuMouchel (1973). More recently, Godsill (1999) considered a Monte Carlo Expectation Maximisation (MCEM) approach with application to time series with symmetric Stable innovations. A full Bayesian approach accounting also for the selection of number of components in mixture models has been proposed by Salas-Gonzalez et al. (2009).

However, Stable distributions play an interesting role even in modelling multivariate data. Nevertheless, multivariate Stable laws pose several challenges that go further beyond the lack of closed form expression for the density. Although general expressions for the multivariate density have been provided by Abdul-Hamid and Nolan (1998), Byczkowski et al. (1993) and Matsui and Takemura (2009), their computations is still not feasible in dimension larger than two. A recent overview of multivariate Stable distributions can be found in Nolan (2008).

Despite its characteristics, estimation of parameters has been always challenging and this has limited its use in applied works requiring a simulation–based methods. Quite a few papers deal with parameters estimation in large dimensions: Nolan (2013) makes use of projection methods to compute the likelihood, Dominicy et al. (2013) introduce a measure of co–dispersion to estimate the covariation function of Elliptical Stable distributions, while Lombardi and Veredas (2009) consider indirect inference method, while Tsionas (2013), Tsionas (2016) approach the problem from a Bayesian perspective.

The remainder of this Chapter is organised as follows. In Section 2.2 we recall the definition and the main properties of the multivariate Elliptical Stable distribution (ESD) already considered by Lombardi and Veredas (2009), while in Section 2.3 we consider the Skew Elliptical Stable distribution firstly mentioned by Branco and Dey (2001). In particular, as another interesting contribution we work with a parameterisation similar to that employed by Azzalini (2013) and provide many theoretical results, such as the closure with respect to linear combination and marginalisation and the cumulative distribution function.

2.2 Multivariate Elliptical Stable distribution

A random vector $\mathbf{Y} \in \mathbb{R}^m$ is elliptically distributed if

$$\mathbf{Y} =^d \boldsymbol{\xi} + \mathcal{R}\boldsymbol{\Gamma}\mathbf{U}, \tag{2.1}$$

where $\boldsymbol{\xi} \in \mathbb{R}^m$ is a vector of location parameters, $\boldsymbol{\Gamma}$ is a matrix such that $\boldsymbol{\Omega} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'$ is a $m \times m$ full rank matrix of scale parameters, $\mathbf{U} \in \mathbb{R}^m$ is a random vector uniformly distributed in the unit sphere

$$\mathbb{S}^{m-1} = \{\mathbf{u} \in \mathbb{R}^m \ : \ \mathbf{u}'\mathbf{u} = 1\} \tag{2.2}$$

and $\mathcal{R}$ is a non–negative random variable stochastically independent of $\mathbf{U}$, called generating variate of $\mathbf{Y}$.

If $\mathcal{R} = \sqrt{Z_1}\sqrt{Z_2}$ where $Z_1 \sim \chi^2_m$ and $Z_2 \sim \mathcal{S}_{\frac{\alpha}{2}}(\xi, \omega, \delta)$ is a positive

Stable distributed random variable with kurtosis parameter equal to $\frac{\alpha}{2}$ for $\alpha \in (0,2]$, location parameter $\xi = 0$, scale parameter $\omega = 1$ and asymmetry parameter $\delta = 1$, stochastically independent of χ^2_m, then the random vector $\mathbf{Y}$ has Elliptical Stable distribution, denoted as follows

$$\mathbf{Y} \sim \mathcal{ESD}_m\left(\alpha, \boldsymbol{\xi}, \boldsymbol{\Omega}\right), \tag{2.3}$$

with characteristic function

$$\begin{aligned}\psi_{\mathbf{Y}}\left(\mathbf{t}\right) &= \mathbb{E}\left(\exp\left\{i\mathbf{t}'\mathbf{Y}\right\}\right)\\ &= \exp\left\{i\mathbf{t}'\boldsymbol{\xi} - \left(\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}\right)^{\frac{\alpha}{2}}\right\}.\end{aligned} \tag{2.4}$$

See Samorodnitsky and Taqqu (1994) for more details on the positive Stable distribution and Nolan (2013) for the recent developments on multivariate elliptically contoured stable distributions.

Among the properties that the class of elliptical distribution possesses, the most relevant are the closure with respect to affine transformations, conditioning and marginalisation, see Fang et al. (1990) and Embrechts et al. (2005) end McNeil et al. (2015) for further details. Simulating from an ESD is straightforward, indeed let $\bar{\omega}_\alpha = \left(\cos\frac{\pi\alpha}{4}\right)^{\frac{2}{\alpha}}$, then $\mathbf{Y} \sim \mathcal{ESD}_m\left(\alpha, \boldsymbol{\xi}, \boldsymbol{\Omega}\right)$ if and only if $\mathbf{Y}$ has the following stochastic representation as a scale mixture of Gaussian distributions

$$\mathbf{Y} = \boldsymbol{\xi} + \zeta^{\frac{1}{2}}\mathbf{X}, \tag{2.5}$$

where $\zeta \sim \mathcal{S}_{\frac{\alpha}{2}}\left(0, \bar{\omega}_\alpha, 1\right)$ and $\mathbf{X} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Omega}\right)$ independent of ζ. Following the Proposition 2.5.2 of Samorodnitsky and Taqqu (1994), the characteristic function of $\mathbf{Y}$ is

$$\begin{aligned}\psi_{\mathbf{Y}}(\mathbf{t}) &= \mathbb{E}\left(\exp\left\{i\mathbf{t}'\mathbf{Y}\right\}\right)\\ &= \mathbb{E}\left(\exp\left\{i\mathbf{t}'\boldsymbol{\xi} + i\zeta^{\frac{1}{2}}\mathbf{t}'\mathbf{X}\right\} \mid \zeta\right)\\ &= \mathbb{E}\left(\exp\left\{i\mathbf{t}'\boldsymbol{\xi} - \frac{\zeta\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}}{2}\right\} \mid \zeta\right)\\ &= \exp\left\{i\mathbf{t}'\boldsymbol{\xi} - \left(\frac{1}{2}\right)^{\frac{\alpha}{2}}\left(\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}\right)^{\frac{\alpha}{2}}\right\}, \qquad \alpha \neq 1, \end{aligned} \tag{2.6}$$

which is the characteristic function of an Elliptical Stable distribution with scale matrix $\boldsymbol{\Omega}/2$. The last equation follows the fact that the Laplace transform of $\zeta \sim \mathcal{S}_{\frac{\alpha}{2}}(0, \bar{\omega}_\alpha, 1)$ with $0 < \alpha \leq 2$ is

$$\begin{aligned}\psi^*_\zeta(A) &= \mathbb{E}\left(\exp\left\{-A\zeta\right\}\right)\\ &= \begin{cases}\exp\left\{-\frac{(\bar{\omega}_\alpha)^{\frac{\alpha}{2}}}{\cos\frac{\pi\alpha}{4}}A^{\frac{\alpha}{2}}\right\}, & \alpha \neq 1\\ \exp\left\{\frac{2\bar{\omega}_\alpha}{\pi}A\log(A)\right\}, & \alpha = 1.\end{cases}\end{aligned} \tag{2.7}$$

The Elliptical Stable distribution is a particular case of multivariate Stable distribution so it admits finite moments if $\mathbb{E}[\zeta^p] < \infty$ for $p < \alpha$. For $\alpha \in (1,2)$, $\mathbb{E}\left(\zeta^{\frac{1}{2}}\right) < \infty$, so that by the law of iterated expectations $\mathbb{E}(\mathbf{Y}) = \boldsymbol{\xi}$, while the second moment never exists. Except for few cases, $\alpha = 2$ (Gaussian), $\alpha = 1$ (Cauchy) and $\alpha = \frac{1}{2}$ (Lévy), the density function cannot be represented in closed form. Those characteristics of the Stable distribution motivate the use of simulations methods in order to make inference on the parameters of interest. In particular we concentrate our interest on the use of the multivariate method of simulated quantile to make inference on the Elliptical Stable distributions since alternative likelihood–based or moments–based methods are not analytically available.

2.3 Multivariate Skew Elliptical Stable distribution

The skew extension of the multivariate Elliptical Stable distribution has been briefly introduced by Branco and Dey (2001) as a special case of a more general class of multivariate Skew Elliptical distributions. The class of Skew Elliptical distributions has been further examined by Lachos et al. (2010) with emphasis on model fitting, while Kim and Genton (2011) obtain the characteristic function for this class and other related distributions. For a general and up to date introduction to Skew Elliptical distributions, we refer to the book of Genton (2004) and to chaper 6 of Azzalini (2013).

Here, we follow the approach of Azzalini (2013) and we consider a slightly different parameterisation from Branco and Dey (2001). Our parameterisation of the SESD has the interesting property that the diagonal elements of the scale matrix do not affect the overall skewness of the distribution. This property is highly appealing for the purposes of the present paper.

As for the Elliptical Stable considered in the previous Section 2.2 that can be obtained as scale mixture of Normal distributions, the Skew–Elliptical Stable distribution can be obtained as a scale mixture of Skew Normals. In what follow we first introduce the definition of multivariate Skew Elliptical Stable distribution, then we provide a stochastic parameterisation which is useful to deal with simulation of random variates and we consider the expression of the cumulative density function which is useful to evaluate the quantiles. Furthermore, we provide an expression for the moment generating function and we exploit it to get the characterisation of linear transformations of SESD random variable which are particularly useful in our applications.

Proposition 1. *Let* $(\mathbf{X}', Y)'$ *be a Normal random vector of dimension* $(d+1)$ *conditional on* $\zeta \sim \mathcal{S}_{\frac{\alpha}{2}}(\bar{\omega}_\alpha, 1, 0)$, *with* $\bar{\omega}_\alpha = \left(\cos\frac{\pi\alpha}{2}\right)^{\frac{2}{\alpha}}$, $\alpha \in (0, 2)$, *i.e.*

$$\begin{bmatrix} \mathbf{X} \\ Y \end{bmatrix} \mid \zeta \sim \mathcal{N}_{d+1}\left(\mathbf{0}, \zeta \boldsymbol{\Omega}_\delta\right), \tag{2.8}$$

where $\boldsymbol{\Omega}_\delta = \begin{bmatrix} \bar{\boldsymbol{\Omega}} & \boldsymbol{\delta} \\ \boldsymbol{\delta}' & 1 \end{bmatrix}$ *and* $\bar{\boldsymbol{\Omega}}$ *is a proper correlation matrix, symmetric and positive definite with* $|\sigma_{ij}| < 1$*, for* $i, j = 1, 2, \ldots, d$ *and* $i \neq j$*, then the distribution of the variable* $\mathbf{Z} = (\mathbf{X} \mid Y > 0)$ *is Skew Elliptical Stable distributed, i.e.,* $\mathbf{Z} \sim \mathcal{SESD}_d\left(\alpha, \bar{\boldsymbol{\Omega}}, \boldsymbol{\lambda}\right)$ *with density*

$$f_{\mathbf{X}}\left(\mathbf{x}, \alpha, \mathbf{0}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\lambda}\right) = 2 \int_0^{+\infty} \phi_d\left(\mathbf{x}, \mathbf{0}, \zeta \bar{\boldsymbol{\Omega}}\right) \Phi_1\left(\frac{\boldsymbol{\lambda}'\mathbf{x}}{\sqrt{\zeta}}\right) h(\zeta)\, d\zeta, \tag{2.9}$$

where $\phi_d(\cdot)$ *and* $\Phi_1(\cdot)$ *denote the density of the multivariate Normal distribution and the cumulative density function of the univariate Normal distribution, respectively, and* $h(\zeta)$ *is the density of the mixing variable. The shape parameter* $\boldsymbol{\lambda}$ *is defined as*

$$\boldsymbol{\lambda} = \left(1 - \boldsymbol{\delta}'\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\delta}\right)^{-\frac{1}{2}} \bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\delta} \in \mathbb{R}^d \tag{2.10}$$

$$\boldsymbol{\delta} = \left(1 + \boldsymbol{\lambda}'\bar{\boldsymbol{\Omega}}\boldsymbol{\lambda}\right)^{-\frac{1}{2}} \bar{\boldsymbol{\Omega}}\boldsymbol{\lambda} \in [-1, 1]^d. \tag{2.11}$$

Proof. See Appendix A. □

Remark 2. *Let* $\mathbf{X} \sim \mathcal{SESD}_d\left(\alpha, 0, \bar{\boldsymbol{\Omega}}, \boldsymbol{\lambda}\right)$*, then the random variable* $\mathbf{X}$ *has a stochastic representation in terms of a scale mixture of Skew Normals, i.e.,* $\mathbf{X} = \zeta^{\frac{1}{2}}\mathbf{Z}$ *where* $\mathbf{Z} \sim \mathcal{SN}_d\left(\mathbf{0}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\lambda}\right)$ *and* $\zeta \sim \mathcal{S}_{\frac{\alpha}{2}}(\bar{\omega}_\alpha, 1, 0)$*, with* $\bar{\omega}_\alpha = \left(\cos\frac{\pi\alpha}{2}\right)^{\frac{2}{\alpha}}$*,* $\alpha \in (0, 2)$*.*

Remark 3. *Let* $\mathbf{X} \sim \mathcal{SESD}_d\left(\alpha, 0, \bar{\boldsymbol{\Omega}}, \boldsymbol{\lambda}\right)$*, then the transformation* $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\mathbf{X}$ *where diag*$\{\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_d\}$ *is* $\mathbf{Y} \sim \mathcal{SESD}_d(\alpha, \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda})$*, with density*

$$f_{\mathbf{Y}}(\mathbf{y}, \alpha, \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda}) = 2 \int_0^{+\infty} \phi_d(\mathbf{y}, \boldsymbol{\xi}, \zeta\boldsymbol{\Omega}) \Phi_1\left(\frac{\boldsymbol{\lambda}'\boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi})}{\sqrt{\zeta}}\right) h(\zeta)\, d\zeta, \tag{2.12}$$

where $\boldsymbol{\xi} \in \mathbb{R}^d$ *is a* d*–dimensional vector of location parameters,* $\boldsymbol{\Omega} = \boldsymbol{\omega}\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}$

is a positive definite square matrix of dimension d, and $\boldsymbol{\lambda} \in \mathbb{R}^d$ *is the shape parameter.*

The following proposition provides a stochastic representation of the Skew–Elliptical Stable distribution which becomes extremely relevant to efficiently simulate random variates.

Proposition 4. *Let* $\mathbf{X} \sim \mathcal{SESD}_d\left(\alpha, \bar{\boldsymbol{\Omega}}, \boldsymbol{\lambda}\right)$*, then* $\mathbf{X}$ *has the following stochastic representation*

$$\mathbf{X} = \begin{cases} \mathbf{U}, & \text{if} \quad U_0 > 0 \\ -\mathbf{U}, & \text{if} \quad U_0 \leq 0, \end{cases} \tag{2.13}$$

with

$$\begin{bmatrix} \mathbf{U} \\ U_0 \end{bmatrix} \sim \mathcal{ESD}_{d+1}\left(\mathbf{0}, \boldsymbol{\Omega}_\delta\right), \tag{2.14}$$

and $\boldsymbol{\Omega}_\delta$ *is defined as in Proposition 1.*

The next proposition provides an easy and intuitive way to calculate the SESD cumulative density function in terms of the cumulative density function of the ESD and it extends previous results on the cumulative density function of the Skew Normal and Skew Student–t distribution, see Azzalini (2013).

Proposition 5. *Let* $\mathbf{X} \sim \mathcal{SN}_d\left(\mathbf{0}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\lambda}\right)$ *then the distribution of the random vector* $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\sqrt{\zeta}\mathbf{X}$*, with* $\boldsymbol{\omega} = diag\{\omega_1, \omega_2, \dots, \omega_d\}$ *and* $\zeta \sim \mathcal{S}_{\frac{\alpha}{2}}\left(\bar{\omega}_\alpha, 1, 0\right)$*, is Skew–Elliptical Stable, i.e.,* $\mathbf{Y} \sim \mathcal{SESD}_d\left(\alpha, \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda}\right)$*. The* *multivariate Skew Elliptical Stable cdf can be calculated as follows*

$$
\begin{aligned}
F_{\mathbf{Y}}(\mathbf{y},\alpha,\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\delta}) &= \mathbb{P}(\mathbf{Y}\le\mathbf{y}) \\
&= \mathbb{P}\left(\boldsymbol{\xi}+\boldsymbol{\omega}\sqrt{\zeta}\mathbf{X}\le\mathbf{y}\right) \\
&= \mathbb{P}\left(\sqrt{\zeta}\mathbf{X}\le\boldsymbol{\omega}^{-1}(\mathbf{y}-\boldsymbol{\xi})\right) \\
&= \mathbb{P}\left(\sqrt{\zeta}\mathbf{U}\le\boldsymbol{\omega}^{-1}(\mathbf{x}-\boldsymbol{\xi})\mid U_0>0\right) \\
&= \frac{\mathbb{P}\left(\sqrt{\zeta}\mathbf{U}\le\boldsymbol{\omega}^{-1}(\mathbf{x}-\boldsymbol{\xi}),U_0>0\right)}{\mathbb{P}(U_0>0)} \\
&= 2\left(\mathbb{P}\begin{bmatrix}\mathbf{U}\\-U_0\end{bmatrix}\le\begin{pmatrix}\boldsymbol{\omega}^{-1}(\mathbf{y}-\boldsymbol{\xi})\\0\end{pmatrix}\right) \\
&= 2\left(\Phi^e_{t,d+1}(\mathbf{u}_1,\alpha,0,\boldsymbol{\Omega}_{-\delta})\right), \qquad (2.15)
\end{aligned}
$$

where $\Phi^e_{d+1}(\cdot)$ *denotes the multivariate ESD cdf of dimension* $d+1$*, with*

$$
\mathbf{u}_1=\begin{bmatrix}\boldsymbol{\omega}^{-1}(\mathbf{y}-\boldsymbol{\xi})\\0\end{bmatrix}, \qquad \boldsymbol{\Omega}_{-\delta}=\begin{bmatrix}\bar{\boldsymbol{\Omega}} & -\boldsymbol{\delta}'\\-\boldsymbol{\delta} & 1\end{bmatrix}. \qquad (2.16)
$$

In the univariate case the Skew ESD cdf in equation (2.15), reduces to $F_Y(y,\alpha,\xi,\omega,\delta) = 2\Phi^e_2\left(\mathbf{u}^*_1,\alpha,0,\boldsymbol{\Omega}^*_{-\delta}\right)$*, where* $\mathbf{u}^*_1=\left(\frac{y-\xi}{\omega},0\right)'$ *and* $\boldsymbol{\Omega}^*_{-\delta}=\begin{bmatrix}1 & -\delta\\-\delta & 1\end{bmatrix}$.

Now we introduce the moment generating function of the SESD which is exploited to characterise the distribution of linear combinations of SESD. Before we introduce the expression for the moment generating functions of univariate and multivariate Skew Normal distributions.

Lemma 6. *Let* $X\sim\mathcal{SN}(\xi,\omega,\lambda)$ *be a univariate Skew Normal distribution, and* $\mathbf{X}\sim\mathcal{MSN}(\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\lambda})$ *be a d–dimensional Skew Normal distribution,* *then the MGF of X and $\mathbf{X}$ are respectively*

$$M_X(t) = 2\exp\left\{\xi t + \frac{t^2\omega^2}{2}\right\}\Phi(\delta\omega t) \tag{2.17}$$

$$M_{\mathbf{X}}(\mathbf{t}) = 2\exp\left\{\mathbf{t}'\boldsymbol{\xi} + \frac{\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}}{2}\right\}\Phi(\boldsymbol{\delta}'\boldsymbol{\omega}\mathbf{t}), \tag{2.18}$$

where $\boldsymbol{\delta}$ has been defined in equation (2.11) and $\boldsymbol{\Omega} = \boldsymbol{\omega}\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}$.

Proof. See Bernardi (2013) and Azzalini (2013). □

Lemma 7. *Let $\mathbf{Y} \sim \mathcal{SESD}_d(\alpha, \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda})$ be a d–dimensional Skew Elliptical Stable distribution, then the MGF of $\mathbf{Y}$ is*

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= \int_0^\infty 2\exp\left\{\mathbf{t}'\boldsymbol{\xi} + \frac{\zeta\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}}{2}\right\}\Phi\left(\sqrt{\zeta}\boldsymbol{\delta}'\boldsymbol{\omega}\mathbf{t}\right)h(\zeta)\,d\zeta \\ &= \int_0^\infty M_{\mathbf{X}}\left(\sqrt{\zeta}\mathbf{t}\right)h(\zeta)\,d\zeta, \end{aligned} \tag{2.19}$$

where $\boldsymbol{\delta}$ has been defined in equation (2.11), $\boldsymbol{\Omega} = \boldsymbol{\omega}\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}$ and $M_{\mathbf{X}}\left(\sqrt{\zeta}\mathbf{t}\right)$ is the mfg of the multivariate Skew Normal distribution defined in equation (2.18).

Proof. See Appendix A. □

For the purposes of the present paper it is important to establish the closure property of the Skew Elliptical Stable distribution with respect to marginalisation. As stated in the following proposition, the closure of the SESD with respect to marginalisation follows immediately from the moment generating function defined in equation (A.7).

Theorem 8 (Closure of the SESD with respect to marginalisation). *Let $\mathbf{X}$ be a d–dimensional Skew Elliptical Stable distribution, i.e., $\mathbf{Y} \sim \mathcal{SESD}_d(\alpha, \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda})$, and assume that the random vector $\mathbf{Y}$ is partitioned into $\mathbf{Y} = [\mathbf{Y}_1', \mathbf{Y}_2']'$, where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are of dimension $\dim(\mathbf{Y}_1) = d_1$ and $\dim(\mathbf{Y}_2) = d_2 = d - d_1$, respectively. The location vector $\boldsymbol{\xi}$, the scale* *matrix* $\mathbf{\Omega}$ *and the shape parameters* $\boldsymbol{\lambda}$ *and* $\boldsymbol{\delta}$ *are partitioned accordingly to*

$$\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{bmatrix}, \quad \mathbf{\Omega} = \begin{bmatrix} \mathbf{\Omega}_{11} & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}'_{12} & \mathbf{\Omega}_{22} \end{bmatrix}, \quad \boldsymbol{\lambda} = \begin{bmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \end{bmatrix}, \quad \boldsymbol{\delta} = \begin{bmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \end{bmatrix}, \tag{2.20}$$

where $\mathbf{\Omega}_{jj}$ *for* $j = 1, 2$ *are positive definite matrices. Then*

$$\mathbf{Y}_1 \sim \mathcal{SESD}_{d_1}\left(\alpha, \boldsymbol{\xi}_1, \mathbf{\Omega}_1, \boldsymbol{\lambda}_1^*\right),$$

with

$$\boldsymbol{\lambda}_1^* = \frac{\boldsymbol{\lambda}_1 + \bar{\mathbf{\Omega}}_{11}^{-1}\bar{\mathbf{\Omega}}_{12}\boldsymbol{\lambda}_2}{\sqrt{1 + \boldsymbol{\lambda}_2'\bar{\mathbf{\Omega}}_{22}^*\boldsymbol{\lambda}_2}} \tag{2.21}$$

$$\boldsymbol{\delta}_1^* = \boldsymbol{\delta}_1, \tag{2.22}$$

where $\mathbf{\Omega}_{22}^* = \bar{\mathbf{\Omega}}_{22} - \bar{\mathbf{\Omega}}_{12}'\bar{\mathbf{\Omega}}_{11}^{-1}\bar{\mathbf{\Omega}}_{12}$*, on partitioning* $\bar{\mathbf{\Omega}}$ *similarly to* $\mathbf{\Omega}$*.*

Proof. See Appendix A. □

In some applications, such as portfolio allocation, we are interested in finding the distribution of linear combinations of Skew Elliptical distributions. The following proposition characterises the closure of the SESD with respect to linear combinations.

Theorem 9 (**Linear combinations of multivariate Skew Elliptical Stable distributions**)**.** *Let* $\mathbf{X}$ *be a* d*–dimensional skew normal, i.e.,* $\mathbf{X} \sim \mathcal{SESD}_d\left(\alpha, \boldsymbol{\xi}, \mathbf{\Omega}, \boldsymbol{\lambda}\right)$*, and assume* $\mathbf{d} \in \mathbb{R}^d$ *be a vector of real coefficients and* $\mathbf{C}$ *be a full rank matrix of dimension* $k \times d$ *with* $k \leq d$*, then the linear combination* $\mathbf{Z} = \mathbf{d} + \mathbf{C}\mathbf{X}$ *has density function*

$$\mathbf{Z} \sim \mathcal{SESD}_k\left(\alpha, \boldsymbol{\xi}_Z, \mathbf{\Omega}_Z, \boldsymbol{\lambda}_Z\right),$$

where

$$\boldsymbol{\xi}_Z = \mathbf{d} + \mathbf{C}\boldsymbol{\xi} \tag{2.23}$$

$$\boldsymbol{\Omega}_Z = \mathbf{C}\boldsymbol{\Omega}\mathbf{C}' \tag{2.24}$$

$$\boldsymbol{\lambda}_Z = \frac{\boldsymbol{\omega}_Z \boldsymbol{\Omega}_Z^{-1}\mathbf{C}\boldsymbol{\omega}\boldsymbol{\delta}}{\sqrt{1 - \boldsymbol{\delta}'\boldsymbol{\omega}\mathbf{C}'\boldsymbol{\Omega}_Z^{-1}\mathbf{C}\boldsymbol{\omega}\boldsymbol{\delta}}}, \tag{2.25}$$

with $\boldsymbol{\delta}_Z = \boldsymbol{\omega}_Z^{-1}\mathbf{C}\boldsymbol{\omega}\boldsymbol{\delta}$, *where* $\boldsymbol{\omega}_Z = (\boldsymbol{\Omega}_Z \odot I_h)^{\frac{1}{2}}$ *and* $\odot$ *denotes the Hadamart component–wise multiplication.*

Proof. See Appendix A. □

As for the Elliptical Stable distribution the multivariate Skew Elliptical Stable distribution admits finite moments if $\mathbb{E}[\zeta^p] < \infty$ for $p < \alpha$. For $\alpha \in (1, 2)$, $\mathbb{E}\left(\zeta^{\frac{1}{2}}\right) < \infty$, so that if $\mathbf{Y} \sim \mathcal{SESD}_d(\alpha, \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\lambda})$ then by the law of iterated expectations $\mathbb{E}(\mathbf{Y}) = \boldsymbol{\xi} + \boldsymbol{\omega}\mathbb{E}\left(\zeta^{\frac{1}{2}}\right)\mathbb{E}(\mathbf{X})$, with $\mathbb{E}(\mathbf{X}) = \sqrt{\frac{2}{\pi}}\boldsymbol{\delta}$, while the second moment never exists.

Chapter 3

The Method of Simulated Quantiles

3.1 Introduction

This Chapter focuses on the method of simulated quantiles (MSQ) recently proposed by Dominicy and Veredas (2013) as a simulation–based extension of the quantile – matching method (QM), see Koenker (2005) for an up to date reference. As any other simulation–based methods, the MSQ estimates parameters by minimising a quadratic distance between a vector of quantile–based summary statistics calculated on the available sample of observations and that calculated on synthetic observations generated from the model. Specifically, we extend the method of simulated quantiles to deal with multivariate data, originating the multivariate method of simulated quantiles (MMSQ). The extension of the MSQ to multivariate data is not trivial because it requires the concept of multivariate quantile that is not unique given the lack of a natural ordering in $\mathbb{R}^n$ for $n > 1$. Indeed, only very recently the literature on multivariate quantiles has proliferated, see, e.g., Serfling (2002) for a review of some extensions of univariate quantiles to the multivariate case. The MMSQ relies on the definition of projectional quantile of Hallin et al. (2010) and Kong and Mizera (2012), that is a particular case of directional quantile. This latter definition is particularly appealing since it allows to reduce the dimension of the problem from $\mathbb{R}^n$ to $\mathbb{R}$ by projecting data towards given directions in the plane. Moreover, the projectional quantiles incorporate information on the covariance between the projected variables which is crucial in order to relax the assumption of independence between variables. An important methodological contribution of this thesis concerns the choice of the relevant directions to project data in order to summarise the information for any given parameter of interest. Although the inclusion of more directions can convey more information about the parameters, it comes at a cost of a larger number of expensive quantile evaluations. Of course the number of quantile functions is unavoidably related to the dimension of the observables and strictly depends upon the specific distribution considered. We provide a general solution for Elliptical distributions and for those Skew–Elliptical distributions that are closed under linear combinations.

We also establish consistency and asymptotic normality of the proposed MMSQ estimator under weak conditions on the underlying true DGP. The conditions for consistency and asymptotic Normality of the MMSQ are similar to those imposed by Dominicy and Veredas (2013) with minor changes due to the employed projectional quantiles. Moreover, for the distributions considered in our illustrative examples, full details on how to calculate all the quantities involved in the asymptotic variance–covariance matrix are provided. The asymptotic variance–covariance matrix of the MMSQ estimator is helpful to derive its efficient version, the E–MMSQ.

The remaining of this Chapter is organised as follows. In Section 3.2.1 the basic concepts on directional and projectional quantiles are recalled. Section 3.2.2 introduces the method of simulated quantiles, while in Section 3.2.3 we establish consistency and asymptotic normality of the proposed estimator. In Section 3.3.1 we introduce an important methodological contribution that deals with the curse of dimensionality problem. Specifically, the objective function of the MMSQ is penalised by adding a SCAD ℓ_1–penalisation term that shrinks to zero the off–diagonal elements of the scale matrix and of the Cholesky factor of the postulated distribution proposing two different algorithms. The one related to the scale matrix is a similar to those proposed by Bien and Tibshirani (2011). We extend the asymptotic theory in order to accommodate sparse estimators, and we prove that the resulting sparse–MMSQ estimator enjoys the oracle properties of Fan and Li (2001) under mild regularity conditions. Moreover, since the chosen penalty is a concave function, it is necessary to construct an algorithm to solve the optimisation problem. This is done in Section 3.3.2 while Section 3.3.3 deals with tuning parameter selection and Section 3.3.4 deals with its implementation. The MMSQ is illustrated and its effectiveness is tested through several examples where synthetic datasets are simulated from well known data generating processes for which alternative methods are known to perform poorly in Section 3.4.

3.2 Multivariate method of simulated quantiles

3.2.1 Directional quantiles

The MMSQ requires the prior definition of the concept of multivariate quantile, a notion still vague until quite recently because of the lack of a natural ordering in dimension greater than one. Here, we relies on the definition of directional quantiles and projectional quantiles introduced by Hallin et al. (2010), Paindaveine and Šiman (2011) and Kong and Mizera (2012). In this section we first recall the definition of directional quantile given in Hallin et al. (2010) and then we introduce the main assumptions that we will use to develop MMSQ.

Definition 10. *Let* $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_m)$ *be a* m*–dimensional random vector in* $\mathbb{R}^m$, $\mathbf{u} \in \mathbb{S}^{m-1}$ *be a vector in the unit sphere*

$$\mathbb{S}^{m-1} = \{\mathbf{u} \in \mathbb{R}^m \ : \ \mathbf{u}'\mathbf{u} = 1\} \tag{3.1}$$

and $\tau \in (0,1)$. *The* $\tau\mathbf{u}$*–quantile of* $\mathbf{Y}$ *is defined as any element of the* *collection* $\Pi^{\tau \mathbf{u}}$ *of hyperplanes*

$$\pi^{\tau \mathbf{u}} = \left\{ \mathbf{Y} : \mathbf{b}^{\tau \mathbf{u}\prime} \mathbf{Y} - q^{\tau \mathbf{u}} = 0 \right\},$$

such that

$$\left(q^{\tau \mathbf{u}}, \mathbf{b}^{\tau \mathbf{u}\prime} \right)' \in \left\{ \arg\min_{(q, \mathbf{b})} \Psi^{\tau \mathbf{u}} \left(q, \mathbf{b} \right) \quad s.t. \quad \mathbf{b}' \mathbf{u} = 1 \right\}, \tag{3.2}$$

where

$$\Psi^{\tau \mathbf{u}} \left(q, \mathbf{b} \right) = \mathbb{E} \Big[\rho_\tau \left(\mathbf{b}' \mathbf{Y} - q \right) \Big], \tag{3.3}$$

and $\rho_\tau(z) = z \left(\tau - \mathbb{1}_{(-\infty, 0)}(z) \right)$ *denotes the quantile loss function evaluated at* $z \in \mathbb{R}$*,* $q \in \mathbb{R}$*,* $\mathbf{b} \in \mathbb{R}^m$ *and* $\mathbb{E}(\cdot)$ *denotes the expectation operator.*

The term directional is due to the fact that the multivariate quantile defined above is associated to a unit vector $\mathbf{u} \in \mathbb{S}^{m-1}$.

Assumption 11. *The distribution of the random vector* $\boldsymbol{Y}$ *is absolutely continuous with respect to the Lebesgue measure on* $\mathbb{R}^m$*, with finite first order moment, having density* $f_{\mathbf{Y}}$ *that has connected support.*

Under assumption 11, for any $\tau \in (0, 1)$ the minimisation problem defined in equation (3.2) admits a unique solution $(a^{\tau \mathbf{u}}, \mathbf{b}^{\tau \mathbf{u}})$, which uniquely identifies one hyperplane $\pi^{\tau \mathbf{u}} \in \Pi^{\tau \mathbf{u}}$.

A special case of directional quantile is obtained by setting $\mathbf{b} = \mathbf{u}$; in that case the directional quantile $(a^{\tau \mathbf{u}}, \mathbf{u})$ becomes a scalar value and it inherits all the properties of the usual univariate quantile. This particular case of directional quantile is called projectional quantile, whose formal definition reported below is due to Kong and Mizera (2012) and Paindaveine and Šiman (2011).

Definition 12. *Let* $\mathbf{Y} = (Y_1, Y_2, \dots, Y_m)$ *be a* m*–dimensional random vector in* $\mathbb{R}^m$*,* $\mathbf{u} \in \mathbb{S}^{m-1}$ *be a vector in the unit sphere* $\mathbb{S}^{m-1}$*, and* $\tau \in (0, 1)$*.*

The $\tau\mathbf{u}$ projectional quantile of $\mathbf{Y}$ is defined as follows.

$$q^{\tau\mathbf{u}} \in \left\{ \arg\min_{q\in\mathbb{R}} \Psi^{\tau\mathbf{u}}\left(q\right) \right\}, \tag{3.4}$$

where $\Psi^{\tau\mathbf{u}}\left(q\right) = \Psi^{\tau\mathbf{u}}\left(q, \mathbf{u}\right)$ in equation (3.3).

Clearly the $\tau\mathbf{u}$–projectional quantile is the τ–quantile of the univariate random variable $\mathbf{u}'\mathbf{Y}$. This feature makes the definition of projectional quantile particularly appealing in order to extend the MSQ to a multivariate setting because, once the direction is properly chosen, it reduces to the usual univariate quantiles. Given a sample of observations $\{\mathbf{y}_i\}_{i=1}^n$ from $\mathbf{Y}$, the empirical version of the projectional quantile is defined as

$$q_n^{\tau\mathbf{u}} \in \left\{ \arg\min_{q} \Psi_n^{\tau\mathbf{u}}\left(q\right) \right\},$$

where $\Psi_n^{\tau\mathbf{u}}\left(q\right) = \frac{1}{n}\sum_{i=1}^n \left[\rho_\tau\left(\mathbf{u}'\mathbf{y}_i - q\right)\right]$ denotes the empirical version of the loss function defined in equation (3.3).

3.2.2 The method of simulated quantiles

The MSQ introduced by Dominicy and Veredas (2013) is likelihood–free simulation–based inferential procedure based on matching quantile–based measures, that is particularly useful in situations where either the density function does is not analytically available and/or moments do not exist. Since it is essentially a simulation–based method it can be applied to all those random variables that can be easily simulated. In the contest of MSQ, parameter are estimated by minimising the distance between an appropriately chosen vector of functions of empirical quantiles and their simulated counterparts based on the postulated parametric model. An appealing characteristic of the MSQ that makes it a valid alternative to other likelihood–free methods, such as the indirect inference of Gouriéroux et al. (1993), is that the MSQ does not rely on a necessarily misspecified auxiliary model. Furthermore, empirical quantiles are robust ordered statistics being able to achieve high protection against bias induced by the presence of outlier contamination.

Here we introduce the MMSQ using the notion of projectional quantiles defined in Section 3.2.1. Let $\mathbf{Y}$ be a d–dimensional random variable with distribution function $F_{\mathbf{Y}}(\cdot, \boldsymbol{\vartheta})$, which depends on a vector of unknown parameters $\boldsymbol{\vartheta} \subset \Theta \in \mathbb{R}^k$, and $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)'$ be a vector of n independent realisations of $\mathbf{Y}$. Let $\mathbf{u}_1, \dots, \mathbf{u}_K \in \mathbb{S}^{m-1}$ be a set of directions and let $\mathbf{q}_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau_k}, \mathbf{u}_k} = \left(q_{\boldsymbol{\vartheta}}^{\tau_1 \mathbf{u}_k}, q_{\boldsymbol{\vartheta}}^{\tau_2 \mathbf{u}_k}, \dots, q_{\boldsymbol{\vartheta}}^{\tau_{s_k} \mathbf{u}_k}\right)$ be a $s \times 1$ vector of projectional quantiles at given confidence levels $\tau_i \in (0, 1)$ with $i = 1, 2, \dots, s_k$, and $k = 1, \dots, K$. Let $\boldsymbol{\Phi}_{\boldsymbol{\vartheta}} = \boldsymbol{\Phi}\left(\mathbf{q}_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau_1}, \mathbf{u}_1}, \dots, \mathbf{q}_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau_K}, \mathbf{u}_K}\right)$ be a $b \times 1$ vector of quantile functions assumed to be continuously differentiable with respect to $\boldsymbol{\vartheta}$ for all $\mathbf{Y}$ and measurable for $\mathbf{Y}$ and for all $\boldsymbol{\vartheta} \subset \Theta$. Let us assume also that $\boldsymbol{\Phi}_{\boldsymbol{\vartheta}}$ cannot be computed analytically but it can be empirically estimated on simulated data; denote those quantities by $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^{r}$. Let $\hat{\mathbf{q}}^{\boldsymbol{\tau_k}, \mathbf{u_k}} = \left(\hat{q}^{\tau_1 \mathbf{u}_k}, \hat{q}^{\tau_2 \mathbf{u}_k}, \dots, \hat{q}^{\tau_{s_k} \mathbf{u}_k}\right)$ be a $s \times 1$ vector of sample projectional quantiles and let $\hat{\boldsymbol{\Phi}} = \boldsymbol{\Phi}\left(\hat{\mathbf{q}}^{\boldsymbol{\tau_1}, \mathbf{u}_1}, \dots, \hat{\mathbf{q}}^{\boldsymbol{\tau_K}, \mathbf{u}_K}\right)$ be a $b \times 1$ vector of functions of sample projectional quantiles.

The MMSQ at each iteration $j = 1, 2, \dots$ estimate $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}$ on a sample of R replication simulated from $\mathbf{y}_{r,j}^{*} \sim F_{\mathbf{Y}}\left(\cdot, \boldsymbol{\vartheta}^{(j)}\right)$, for $r = 1, 2, \dots, R$, as $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_j}^{R} = \frac{1}{R} \sum_{r=1}^{R} \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_j}^{r}$, where $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_j}^{r}$ is the function $\boldsymbol{\Phi}_{\boldsymbol{\vartheta}}$ computed at the r–th simulation path. The parameters are subsequently updated by minimising the distance between the vector of quantile measures calculated on the true observations $\hat{\boldsymbol{\Phi}}$ and that calculated on simulated realisations $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_j}^{R}$ as follows

$$\hat{\boldsymbol{\vartheta}} = \arg \min_{\boldsymbol{\vartheta} \in \Theta} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^{R}\right)' \mathbf{W}_{\boldsymbol{\vartheta}} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^{R}\right), \tag{3.5}$$

where $\mathbf{W}_{\boldsymbol{\vartheta}}$ is a $b \times b$ symmetric positive definite weighting matrix. The method of simulated quantiles of Dominicy and Veredas (2013) reduces to the selection of the first canonical direction $\mathbf{u}_1 = (1, 0, \dots, 0)$ as relevant direction in the projectional quantile.

The vector of functions of projectional quantiles $\boldsymbol{\Phi}_{\boldsymbol{\vartheta}}$ should be carefully selected in order to be as informative as possible for the vector of parameters of interest. In their applications, Dominicy and Veredas (2013) only propose to use the MSQ to estimate the parameters of univariate Stable law. Toward this end they consider the following vector of quantile–based statistics, as in McCulloch (1986) and Kim and White (2004)

$$\boldsymbol{\Phi}_{\boldsymbol{\vartheta}} = \left(\frac{q_{0.95} + q_{0.05} - 2q_{0.5}}{q_{0.95} - q_{0.05}}, \frac{q_{0.95} - q_{0.05}}{q_{0.75} - q_{0.25}}, q_{0.75} - q_{0.25}, q_{0.5} \right)'.$$

where the first element of the vector is a measure of skewness, the second one is a measure of kurtosis and the last two measures refer to scale and location, respectively. Of course, the selection of the quantile–based summary statistics depend either on the kind of parameter and on the assumed distribution. The MMSQ generalises also the MSQ proposed by Dominicy et al. (2013) where they estimate the elements of the variance–covariance matrix of multivariate elliptical distributions by means of a measure of co–dispersion which consists in the in interquartile range of the standardised variables projected along the bisector. The MMSQ based on projectional quantiles is more flexible and it allows us to deal with more general distributions than elliptically contoured distributions because it relies on the construction of quantile based measures on the variables projected along an optimal directions which depend upon the distribution considered.

3.2.3 Asymptotic theory

In this section we establish consistency and asymptotic normality of the proposed MMSQ estimator. The next theorem establish the asymptotic properties of projectional quantiles.

Theorem 13. *Let* $\mathbf{Y} \in \mathbb{R}^m$ *be a random vector with cumulative distribution function* $F_{\mathbf{Y}}$ *and variance–covariance matrix* $\boldsymbol{\Sigma}_{\mathbf{Y}}$. *Let* $\{\mathbf{y}_i\}_{i=1}^n$ *be a sample of iid observations from* $F_{\mathbf{Y}}$. *Let* $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_K \in \mathbb{S}^{m-1}$ *and* $Z_k = \mathbf{u}_k'\mathbf{Y}$ *be the projected random variable along* $\mathbf{u}_k$ *with cumulative distribution function* F_{Z_k}, *for* $k = 1, 2, \ldots, K$. *Let* $\boldsymbol{\tau}_k = (\tau_{1,k}, \tau_{2,k}, \ldots, \tau_{s,k})$ *where* $\tau_{j,k} \in (0, 1)$, $\mathbf{q}^{\boldsymbol{\tau}_k, \mathbf{u}_k} = (q^{\tau_{1,k}\mathbf{u}_k}, q^{\tau_{2,k}\mathbf{u}_k}, \ldots, q^{\tau_{s,k}\mathbf{u}_k})$ *be the vector of directional quantiles along the direction* $\mathbf{u}_k$ *and suppose* $Var\,(Z_k) < \infty$, *for*

$k = 1, 2, \ldots, K$. *Let us assume that* F_{Z_k} *is differentiable in* $q^{\tau_{j,k}\mathbf{u}_k}$ *and* $F'_{Z_k}\left(q^{\tau_{j,k}\mathbf{u}_k}\right) = f_{Z_k}\left(q^{\tau_{j,k}\mathbf{u}_k}\right) > 0$, *for* $k = 1, 2, \ldots, K$ *and* $j = 1, 2, \ldots, s$. *Then*

(i) for a given direction $\mathbf{u}_k$, *with* $k = 1, 2, \ldots, K$, *it holds*

$$\sqrt{n}\left(\hat{\mathbf{q}}^{\boldsymbol{\tau}_k\mathbf{u}_k} - \mathbf{q}^{\boldsymbol{\tau}_k\mathbf{u}_k}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\eta}\right),$$

as $n \to \infty$, *where* $\boldsymbol{\eta}$ *denotes a* $(K \times K)$ *symmetric matrix whose generic* (r, c) *entry is*

$$\eta_{r,c} = \frac{\tau_r \wedge \tau_c - \tau_r\tau_c}{f_{Z_k}\left(\mathbf{q}^{\tau_r,\mathbf{u}_k}\right) f_{Z_k}\left(\mathbf{q}^{\tau_c,\mathbf{u}_k}\right)},$$

for $r, c = 1, \ldots, K$;

(ii) for a given level τ_j, *with* $j = 1, 2, \ldots, s$, *it holds*

$$\sqrt{n}\left(\hat{\mathbf{q}}^{\tau_j} - \mathbf{q}^{\tau_j}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\eta}\right),$$

as $n \to \infty$, *where* $\mathbf{q}^{\tau_j} = \left(q^{\tau_j\mathbf{u}_1}, \ldots, q^{\tau_j\mathbf{u}_K}\right)$,

$$\eta_{r,c} = \begin{cases} -\frac{\tau_j^2}{f_{Z_r}\left(q^{\tau_j\mathbf{u}_r}\right)f_{Z_c}\left(q^{\tau_j\mathbf{u}_c}\right)} + \frac{F_{Z_r,Z_c}\left(\mathbf{q}^{\tau_j,r,c},\boldsymbol{\Sigma}_{Z_r,Z_c}\right)}{f_{Z_r}\left(q^{\tau_j\mathbf{u}_r}\right)f_{Z_c}\left(q^{\tau_j\mathbf{u}_c}\right)}, & \text{for} \quad r \neq c \\ \frac{\tau_j(1-\tau_j)}{f_{Z_r}\left(q^{\tau_j\mathbf{u}_r}\right)^2}, & \text{for} \quad r = c, \end{cases}$$

and $\boldsymbol{\Sigma}_{Z_r,Z_c}$ *denotes the variance–covariance matrix of the random variables* Z_r *and* Z_c *and* $\mathbf{q}^{\tau_j,r,c} = \left(q^{\tau_j\mathbf{u}_r}, q^{\tau_j\mathbf{u}_c}\right)$, *for* $r, c = 1, 2, \ldots, K$;

(iii) given τ_j *and* τ_l *with* $j, l = 1, 2, \ldots, s$ *and* $j \neq l$ *and given* $\mathbf{u}_s$ *and* $\mathbf{u}_t$ *with* $s, t = 1, 2, \ldots, K$ *and* $s \neq t$, *it holds*

$$\sqrt{n}\left(\hat{q}^{\tau_j\mathbf{u}_s} - q^{\tau_j\mathbf{u}_s}, \hat{q}^{\tau_l\mathbf{u}_t} - q^{\tau_l\mathbf{u}_t}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\eta}\right),$$

as $n \to \infty$*, where*

$$\eta_{r,c} = -\frac{\tau_j \tau_l}{f_{Z_s}\left(q^{\tau_j}\right) f_{Z_t}\left(q^{\tau_l}\right)} + \frac{F_{Z_s,Z_t}\left(\left(q^{\tau_j \mathbf{u}_s}, q^{\tau_l \mathbf{u}_t}\right), \boldsymbol{\Sigma}_{Z_s,Z_t}\right)}{f_{Z_s}\left(q^{\tau_j}\right) f_{Z_t}\left(q^{\tau_l}\right)}, \quad for \quad r \neq c. \tag{3.6}$$

Proof. See Appendix A. □

To establish the asymptotic properties of the MMSQ estimates we need the following set of assumptions.

Assumption 14. *There exists a unique/unknown true value* $\boldsymbol{\vartheta}_0 \subset \Theta$ *such that the sample function of projectional quantiles equal the theoretical one, provided that each quantile–based summary statistic is computed along a direction that is informative for the parameter of interest. That is* $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0 \;\Leftrightarrow\; \hat{\boldsymbol{\Phi}} = \boldsymbol{\Phi}_{\boldsymbol{\vartheta}_0}$.

Assumption 15. $\boldsymbol{\vartheta}_0$ *is the unique minimiser of* $\left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^{R}\right)' \mathbf{W}_{\boldsymbol{\vartheta}} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^{R}\right)$.

Assumption 16. *Let* $\hat{\boldsymbol{\Omega}}$ *be the sample variance–covariance matrix of* $\hat{\boldsymbol{\Phi}}$ *and* $\boldsymbol{\Omega}_{\boldsymbol{\vartheta}}$ *be the variance–covariance matrix of* $\boldsymbol{\Phi}_{\boldsymbol{\vartheta}}$*, then* $\hat{\boldsymbol{\Omega}}$ *converges to* $\boldsymbol{\Omega}_{\boldsymbol{\vartheta}}$.

Assumption 17. *The matrix* $\boldsymbol{\Omega}_{\boldsymbol{\vartheta}}$ *is non–singular.*

Assumption 18. *The matrix* $\left(\frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}}{\partial \boldsymbol{\vartheta}'} \mathbf{W}_{\boldsymbol{\vartheta}} \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}}{\partial \boldsymbol{\vartheta}}\right)$ *is non–singular.*

Under these assumptions we show the asymptotic properties of functions of quantiles.

Theorem 19. *Under the hypothesis of Theorem 13 and assumptions 6–9, we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\Phi}} - \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\vartheta}}\right)$$
$$\sqrt{n}\left(\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^{R} - \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\vartheta}}\right),$$

as $n \to \infty$*, where* $\boldsymbol{\Omega}_{\boldsymbol{\vartheta}} = \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}}{\partial \mathbf{q}'} \boldsymbol{\eta} \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}}{\partial \mathbf{q}}$, $\mathbf{q} = \left(\mathbf{q}^{\boldsymbol{\tau}_1,\mathbf{u}_1}, \mathbf{q}^{\boldsymbol{\tau}_2,\mathbf{u}_2}, \dots, \mathbf{q}^{\boldsymbol{\tau}_K,\mathbf{u}_K}\right)'$, $\boldsymbol{\eta}$ *is the variance–covariance matrix of the projectional quantiles* $\mathbf{q}$ *defined in Theorem 13 and* $\frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}}{\partial \mathbf{q}} = diag\left\{\frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}}{\partial \mathbf{q}^{\boldsymbol{\tau}_1,\mathbf{u}_1}}, \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}}{\partial \mathbf{q}^{\boldsymbol{\tau}_2,\mathbf{u}_2}}, \dots, \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}}{\partial \mathbf{q}^{\boldsymbol{\tau}_K,\mathbf{u}_K}}\right\}$.

Proof. See Appendix A. □

Next theorem shows the asymptotic properties of the MMSQ estimator.

Theorem 20. *Under the hypothesis of Theorem 13 and assumptions 6–10, we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\left(1+\frac{1}{R}\right)\mathbf{D}_{\boldsymbol{\vartheta}}\mathbf{W}_{\boldsymbol{\vartheta}}\boldsymbol{\Omega}_{\boldsymbol{\vartheta}}\mathbf{W}_{\boldsymbol{\vartheta}}'\mathbf{D}_{\boldsymbol{\vartheta}}'\right),$$

as $n\to\infty$, *where* $\mathbf{D}_{\boldsymbol{\vartheta}}=\left(\frac{\partial\boldsymbol{\Phi}_{\boldsymbol{\vartheta}}}{\partial\boldsymbol{\vartheta}'}\mathbf{W}_{\boldsymbol{\vartheta}}\frac{\partial\boldsymbol{\Phi}_{\boldsymbol{\vartheta}}}{\partial\boldsymbol{\vartheta}}\right)^{-1}\frac{\partial\boldsymbol{\Phi}_{\boldsymbol{\vartheta}}}{\partial\boldsymbol{\vartheta}}$.

Proof. See Appendix A. □

The next corollary provides the optimal weighting matrix $\mathbf{W}_{\boldsymbol{\vartheta}}$.

Corollary 21. *Under the hypothesis of Theorem 13 and assumptions 6–10, the optimal weighting matrix is*

$$\mathbf{W}_{\boldsymbol{\vartheta}}^{*}=\boldsymbol{\Omega}_{\boldsymbol{\vartheta}}^{-1}.$$

Therefore, the efficient method of simulated quantiles estimator E–MMSQ has the following asymptotic distribution

$$\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\left(1+\frac{1}{R}\right)\left(\frac{\partial\boldsymbol{\Phi}_{\boldsymbol{\vartheta}}}{\partial\boldsymbol{\vartheta}'}\boldsymbol{\Omega}_{\boldsymbol{\vartheta}}^{-1}\frac{\partial\boldsymbol{\Phi}_{\boldsymbol{\vartheta}}}{\partial\boldsymbol{\vartheta}}\right)^{-1}\right),$$

as $n\to\infty$.

3.3 Handling sparsity

3.3.1 Sparse method of simulated quantiles

In this section the MMSQ estimator is extended in order to achieve sparse estimation of the scaling matrix. Specifically, the smoothly clipped absolute deviation (SCAD) ℓ_1–penalty of Fan and Li (2001) is introduced into the MMSQ objective function. Formally, let $\mathbf{Y}\in\mathbb{R}^m$ be a random vector and

$\boldsymbol{\Sigma} = (\sigma_{i,j})_{i,j=1}^{n}$ be its scale matrix and we are interested in providing a sparse estimation of $\boldsymbol{\Sigma}$. To achieve this target we adopt a modified version of the MMSQ objective function obtained by adding the SCAD penalty to the off–diagonal elements of the covariance matrix in line with Bien and Tibshirani (2011). The SCAD function is a non convex penalty function with the following form

$$p_\lambda\left(|\gamma|\right) = \begin{cases} \lambda|\gamma| & \text{if } |\gamma| \leq \lambda \\ \frac{1}{a-1}\left(a\lambda|\gamma| - \frac{\gamma^2}{2}\right) - \frac{\lambda^2}{2(a-1)} & \text{if } \lambda < \gamma \leq a\lambda \\ \frac{\lambda^2(a+1)}{2} & \text{if } a\lambda < |\gamma|, \end{cases} \tag{3.7}$$

which corresponds to quadratic spline function with knots at λ and $a\lambda$. The SCAD penalty is continuously differentiable on $(-\infty; 0) \cup (0; \infty)$ but singular at 0 with its derivatives zero outside the range $[-a\lambda; a\lambda]$. This results in small coefficients being set to zero, a few other coefficients being shrunk towards zero while retaining the large coefficients as they are. The sparse MMSQ estimator (S–MMSQ) minimises the penalised MMSQ objective function, defined as follows

$$\hat{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta}} \mathcal{Q}^\star\left(\boldsymbol{\vartheta}\right), \tag{3.8}$$

where

$$\mathcal{Q}^\star\left(\boldsymbol{\vartheta}\right) = \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^R\right)' \mathbf{W}_{\boldsymbol{\vartheta}} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^R\right) + n \sum_{i<j} p_\lambda\left(|\sigma_{ij}|\right) \tag{3.9}$$

is the penalised distance between $\hat{\boldsymbol{\Phi}}$ and $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^R$ defined in Section 3.2.2. As shown in Fan and Li (2001), the SCAD estimator, with appropriate choice of the regularisation (tuning) parameter, possesses a sparsity property, i.e., it estimates zero components of the true parameter vector exactly as zero with probability approaching one as sample size increases while still being consistent for the non–zero components. An immediate consequence of the sparsity property of the SCAD estimator is that the asymptotic distribution of the estimator remains the same whether or not the correct zero restrictions are imposed in the course of the SCAD estimation procedure. They call them the oracle properties.

Let $\boldsymbol{\vartheta}_0 = (\boldsymbol{\vartheta}_0^1, \boldsymbol{\vartheta}_0^0)$ be the true value of the unknown parameter $\boldsymbol{\vartheta}$, where $\boldsymbol{\vartheta}_0^1 \in \mathbb{R}^s$ is the subset of non–zero parameters and $\boldsymbol{\vartheta}_0^0 = 0 \in \mathbb{R}^{k-s}$ and let $\mathcal{A} = \{(i,j) : i < j, \sigma_{ij,0} \in \boldsymbol{\vartheta}_0^1\}$. The following definition of oracle estimator has been formalised in Zou (2006).

Definition 22. *An oracle estimator $\hat{\boldsymbol{\vartheta}}_{\text{oracle}}$ has the following properties:*

(i) consistent variable selection: $\lim_{n\to\infty} \mathbb{P}(\mathcal{A}_n = \mathcal{A}) = 1$, *where* $\mathcal{A}_n = \left\{(i,j) : i < j, \hat{\sigma}_{ij} \in \hat{\boldsymbol{\vartheta}}^1_{\text{oracle}}\right\}$;

(ii) asymptotic normality: $\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}^1_{\text{oracle}} - \boldsymbol{\vartheta}_0^1\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, *as* $n \to \infty$, *where* $\boldsymbol{\Sigma}$ *is the variance covariance matrix of* $\boldsymbol{\vartheta}_0^1$.

Following Fan and Li (2001), in the remaining of this section we establish the oracle properties of the S–MMSQ estimator. We first prove the sparsity property.

Theorem 23. *Given the SCAD penalty function* $p_\lambda(|\sigma_{ij}|)$, *for a sequence of* λ_n *such that* $\lambda_n \to 0$, *and* $\sqrt{n}\lambda_n \to \infty$, *as* $n \to \infty$, *there exists a local minimiser* $\hat{\boldsymbol{\vartheta}}$ *of* $\mathcal{Q}^\star(\boldsymbol{\vartheta})$ *in (3.8) with* $\|\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. *Furthermore, we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\hat{\boldsymbol{\vartheta}}^0 = 0\right) = 1. \tag{3.10}$$

Proof. See Appendix A. □

The following theorem establishes the asymptotic normality of the S–MMSQ estimator; we denote by $\boldsymbol{\vartheta}^1$ the subvector of $\boldsymbol{\vartheta}$ that does not contain zero off–diagonal elements of the variance covariance matrix and by $\hat{\boldsymbol{\vartheta}}^1$ the corresponding penalised MMSQ estimator.

Theorem 24. *Given the SCAD penalty function $p_\lambda(|\sigma_{ij}|)$, for a sequence $\lambda_n \to 0$ and $\sqrt{n}\lambda_n \to \infty$ as $n \to \infty$, then $\hat{\boldsymbol{\vartheta}}^1$ has the following asymptotic distribution:*

$$\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}^1 - \boldsymbol{\vartheta}_0^1\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \left(1+\frac{1}{R}\right)\left(\frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}}{\partial \boldsymbol{\vartheta}^{1\prime}} \boldsymbol{\Omega}_{\boldsymbol{\vartheta}_0^1}^{-1} \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}}{\partial \boldsymbol{\vartheta}^1}\right)^{-1}\right), \tag{3.11}$$

as $n \to \infty$.

Proof. See Appendix A. □

3.3.2 Algorithm

The objective function of the sparse estimator is the sum of a convex function and a non convex function which complicates the minimisation procedure. Here, we adapt the algorithms proposed by Fan and Li (2001) and Hunter and Li (2005) to our objective function in order to allow a fast procedure for the minimisation problem.

The first derivative of the penalty function can be approximated as follows

$$\left[p_\lambda(|\sigma_{ij}|)\right]' = p_\lambda'(|\sigma_{ij}|)\,\mathrm{sgn}(\sigma_{ij}) \approx \frac{p_\lambda'(|\sigma_{ij,0}|)}{|\sigma_{ij,0}|}\sigma_{ij}, \tag{3.12}$$

when $\sigma_{ij} \neq 0$. We use it in the first order Taylor expansion of the penalty to get

$$p_\lambda(|\sigma_{ij}|) \approx p_\lambda(|\sigma_{ij,0}|) + \frac{1}{2}\frac{p_\lambda'(|\sigma_{ij,0}|)}{|\sigma_{ij,0}|}\left(\sigma_{ij}^2 - \sigma_{ij,0}^2\right), \tag{3.13}$$

for $\sigma_{ij} \approx \sigma_{ij,0}$. The objective function $\mathcal{Q}^\star$ in equation (3.8) can be locally approximated, except for a constant term by

$$\mathcal{Q}^\star(\boldsymbol{\vartheta}) \approx \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}\right)' \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}\right) - \frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}\right)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)$$
$$+ \frac{1}{2}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)' \frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) + \frac{n}{2} \boldsymbol{\vartheta}' \mathbf{P}_\lambda(\boldsymbol{\vartheta}_0) \boldsymbol{\vartheta}, \tag{3.14}$$

where $\mathbf{P}_\lambda(\boldsymbol{\vartheta}_0) = \mathrm{diag}\left\{\mathbf{0}, \frac{p'_\lambda(|\sigma_{ij,0}|)}{|\sigma_{ij,0}|}; i > j, \sigma_{ij,0} \in \boldsymbol{\vartheta}^1_0\right\}$, for which the first order condition becomes

$$\begin{aligned}
\frac{\partial \mathcal{Q}^\star(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} &\approx -\frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}\right) + \frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}'} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) + n\mathbf{P}_\lambda(\boldsymbol{\vartheta}_0)\boldsymbol{\vartheta} \\
&= -\frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}\right) + \frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}'} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) \\
&\quad + n\mathbf{P}_\lambda(\boldsymbol{\vartheta}_0)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) + n\mathbf{P}_\lambda(\boldsymbol{\vartheta}_0)\boldsymbol{\vartheta}_0 \\
&= (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)' \left[\frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}'} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} + \boldsymbol{\Sigma}_\lambda(\boldsymbol{\vartheta}_0)\right] - \frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}\right) \\
&\quad + \mathbf{P}_\lambda(\boldsymbol{\vartheta}_0)\boldsymbol{\vartheta}_0 \\
&= 0,
\end{aligned} \tag{3.15}$$

and therefore

$$\begin{aligned}
\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0 &- \left[\frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}'} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} + n\mathbf{P}_\lambda(\boldsymbol{\vartheta}_0)\right]^{-1} \\
&\times \left[-\frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}\right) + n\mathbf{P}_\lambda(\boldsymbol{\vartheta}_0)\boldsymbol{\vartheta}_0\right].
\end{aligned} \tag{3.16}$$

The optimal solution can be find iteratively, as follows

$$\boldsymbol{\vartheta}^{(k+1)} = \boldsymbol{\vartheta}^{(k)} - \left[\frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}^{(k)}}}{\partial \boldsymbol{\vartheta}'} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}^{(k)}}}{\partial \boldsymbol{\vartheta}} + n \mathbf{P}_\lambda \left(\boldsymbol{\vartheta}^{(k)} \right) \right]^{-1} \times \left[-\frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}^{(k)}}}{\partial \boldsymbol{\vartheta}} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}^{(k)}} \right) + n \mathbf{P}_\lambda \left(\boldsymbol{\vartheta}^{(k)} \right) \boldsymbol{\vartheta}^{(k)} \right], \tag{3.17}$$

and if $\vartheta_j^{(k+1)} \approx 0$, then $\vartheta_j^{(k+1)}$ is set equal zero. When the algorithm converges the estimator satisfies the following equation

$$-\frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0} \right) + n \mathbf{P}_\lambda \left(\boldsymbol{\vartheta}_0 \right) \boldsymbol{\vartheta}_0 = 0, \tag{3.18}$$

that is the first order condition of the minimisation problem of the S–MMSQ estimator.

The algorithm used above and introduced by Fan and Li (2001) is called local quadratic approximation (LQA). Hunter and Li (2005) showed that LQA applied to penalised maximum likelihood is an MM algorithm. Indeed, we define

$$\Psi_{|\sigma_{ij,0}|} \left(|\sigma_{ij}| \right) = p_\lambda \left(|\sigma_{ij,0}| \right) + \frac{1}{2} \frac{p'_\lambda \left(|\sigma_{ij,0}| \right)}{|\sigma_{ij,0}|} \left(\sigma_{ij}^2 - \sigma_{ij,0}^2 \right), \tag{3.19}$$

since the SCAD penalty is concave it holds

$$\Psi_{|\sigma_{ij,0}|} \left(|\sigma_{ij}| \right) \geq p_\lambda \left(|\sigma_{ij}| \right), \quad \forall |\sigma_{ij}|, \tag{3.20}$$

and equality holds when $|\sigma_{ij}| = |\sigma_{ij,0}|$. Then $\Psi_{|\sigma_{ij,0}|} \left(|\sigma_{ij}| \right)$ majorise $p_\lambda \left(|\sigma_{ij}| \right)$, and it holds

$$\Psi_{|\sigma_{ij,0}|} \left(|\sigma_{ij}| \right) < \Psi_{|\sigma_{ij,0}|} \left(|\sigma_{ij,0}| \right) \Rightarrow p_\lambda \left(|\sigma_{ij}| \right) < p_\lambda \left(|\sigma_{ij,0}| \right), \tag{3.21}$$

that is called descendent property. This feature allows us to construct an MM algorithm: at each iteration k we construct $\Psi_{|\sigma_{ij}^{(k)}|} \left(|\sigma_{ij}| \right)$ and then minimize it to get $\sigma_{ij}^{(k+1)}$, that satisfies $p_\lambda \left(|\sigma_{ij}^{(k+1)}| \right) < p_\lambda \left(|\sigma_{ij}^{(k)}| \right)$. Let us consider the following

$$S_k(\boldsymbol{\vartheta}) = \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^R\right)' \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^R\right) + n \sum_{i>j} \Psi_{|\sigma_{ij}^{(k)}|}\left(|\sigma_{ij}|\right), \tag{3.22}$$

then $S_k(\boldsymbol{\vartheta})$ majorise $\mathcal{Q}^\star(\boldsymbol{\vartheta})$; thus we only need to minimise $S_k(\boldsymbol{\vartheta})$, that can be done as explained above. Hunter and Li (2005) proposed an improved version of LQA for penalised maximum likelihood, aimed at avoiding to zero out the parameters too early during the iterative procedure. We present their method applied to S–MMSQ as follows

$$\begin{aligned}
p_{\lambda,\epsilon}\left(|\sigma_{ij}|\right) &= p_\lambda\left(|\sigma_{ij}|\right) - \epsilon \int_0^{|\sigma_{ij}|} \frac{p'_\lambda\left(|\sigma_{ij,0}|\right)}{\epsilon + t} dt \\
\mathcal{Q}^\star_\epsilon(\boldsymbol{\vartheta}) &= \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^R\right)' \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^R\right) + n \sum_{i>j} p_{\lambda,\epsilon}\left(|\sigma_{ij}|\right) \\
\Psi_{|\sigma_{ij,0}|,\epsilon}\left(|\sigma_{ij}|\right) &= p_{\lambda,\epsilon}\left(|\sigma_{ij,0}|\right) + \frac{p'_\lambda\left(|\sigma_{ij,0}|\right)}{2\left(\epsilon + |\sigma_{ij,0}|\right)}\left(\sigma_{ij}^2 - \sigma_{ij,0}^2\right) \\
S_{k,\epsilon}(\boldsymbol{\vartheta}) &= \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^R\right)' \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^R\right) + n \sum_{i>j} \Psi_{|\sigma_{ij}^{(k)}|,\epsilon}\left(|\sigma_{ij}|\right),
\end{aligned}$$

where $\bar{\boldsymbol{\vartheta}}$ is a consistent estimator of $\boldsymbol{\vartheta}$. They proved that as $\epsilon \downarrow 0$ the perturbed objective function $\mathcal{Q}^\star_\epsilon(\boldsymbol{\vartheta})$ converges uniformly to the not perturbed one $\mathcal{Q}^\star(\boldsymbol{\vartheta})$ and that if $\hat{\boldsymbol{\vartheta}}_\epsilon$ is a minimiser of $\mathcal{Q}^\star_\epsilon(\boldsymbol{\vartheta})$ then any limit point of the sequence $\left\{\hat{\boldsymbol{\vartheta}}_\epsilon\right\}_{\epsilon\downarrow 0}$ is a minimiser of $\mathcal{Q}^\star(\boldsymbol{\vartheta})$. This construction allows to define $\Psi_{|\sigma_{i,j}^{(k)}|,\epsilon}\left(|\sigma_{i,j}|\right)$ even when $\sigma_{i,j}^{(k)} \approx 0$. The authors also provided a way to choose the value of the perturbation ϵ and suggested the following

$$\epsilon = \frac{\tau}{2np'_\lambda(0)} \min\left\{|\sigma_{i,j}^{(0)}| : \sigma_{i,j}^{(0)} \neq 0\right\}, \tag{3.23}$$

with the following tuning constant $\tau = 10^{-8}$.

3.3.3 Tuning paramenter selection

The SCAD penalty requires the selection of two tuning parameters (a, λ). The first tuning parameter is fixed at $a = 3.7$ as suggested in Fan and Li

(2001), while the parameter λ is selected using K–fold cross validation, in which the original sample is divided in K subgroups T_k, called folds. The validation function is

$$CV(\lambda) = \sum_{k=1}^{K} \frac{1}{n_k} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}^{R}_{\hat{\boldsymbol{\vartheta}}_{\lambda,k}} \right) \mathbf{W}_{\hat{\boldsymbol{\vartheta}}_{\lambda,k}} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}^{R}_{\hat{\boldsymbol{\vartheta}}_{\lambda,k}} \right), \tag{3.24}$$

where $\hat{\boldsymbol{\vartheta}}_{\lambda,k}$ denotes the parameters estimate on the sample $\left(\cup_{i=1}^{K} T_k\right) \setminus T_k$ with λ as tuning parameter. Then the optimal value is chosen as $\lambda^* = \arg\min_{\lambda} CV(\lambda)$; again the minimisation is performed over a grid of values for λ.

3.3.4 Implementation

The symmetric and positive definiteness properties of the variance–covariance matrix should be preserved at each step of the optimisation process. Preserving those properties is a difficult task since the constraints that ensure the definite positiveness of a matrix are non linear. Here we propose two alternative solutions. The first solution relies on the approach of Levina et al. (2008) that induces sparse estimation of the elements of the Cholesky factor of the correlation matrix. The second solution instead considers a sequential column–wise factorisation of the correlation matrix that naturally preserves the positive definiteness of the matrix at each stage of the optimisation procedure.

Let now briefly recall the hyper–spherical parametrisation of the Cholesky factor introduced by Pourahmadi and Wang (2015) that maps the pairwise correlations onto the space of angles. Let $\mathbf{R}$ be a correlation matrix and $\mathbf{L}$

the corresponding Cholesky factor, i.e., $\mathbf{R} = \mathbf{L}\mathbf{L}'$, then

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ \cos\theta_{2,1} & \sin\theta_{2,1} & 0 & 0 & \cdots & 0 \\ \cos\theta_{3,1} & \cos\theta_{3,2}\sin\theta_{3,1} & \sin\theta_{3,2}\sin\theta_{3,1} & 0 & \cdots & 0 \\ \cos\theta_{4,1} & \cos\theta_{4,2}\sin\theta_{4,1} & \cos\theta_{4,3}\sin\theta_{4,1}\sin\theta_{4,2} & \prod_{i=1}^{3}\sin\theta_{4,i} & \cdots & 0 \\ \vdots & & & & & \\ \cos\theta_{n,1} & \cos\theta_{n,2}\sin\theta_{n,1} & \cos\theta_{n,3}\sin\theta_{n,1}\sin\theta_{n,2} & \cos\theta_{n,4}\prod_{i=1}^{3}\sin\theta_{n,i} & \cdots & \prod_{i=1}^{3}\sin\theta_{n,i} \end{bmatrix} \tag{3.25}$$

where $\theta_{i,j} \in (0, \pi)$. Unfortunately, the angles domain in equation (3.25) does not include zero and, moreover, we should introduce a one–to–one correspondence between zeros of the two parametrisations. Therefore, we should introduce the following translation:

$$\tilde{\theta}_{i,j} = \theta_{i,j} - \frac{\pi}{2}, \tag{3.26}$$

so that $\tilde{\theta}_{i,j} \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and $\tilde{\theta}_{i,j} = 0 \leftrightarrow \mathbf{L}_{ij} = 0$. The previous approach sparsifies the Cholesky factor and, in general, does not induce a sparse correlation matrix unless we are in the particular case of a decomposable graphs.

Let us recall some basic definitions from graph theory in order to define such graphs. A graph is defined as a couple (V, E) where V is a set of vertices and $E \subset V \times V$ be a set of edges. We assume that V is finite. The vertices u, v are adjacent if $(u, v) \in E$. If all the vertices are adjacent to each other then the graph is complete. Any path that begins and ends at the same vertex is called a cycle.

Definition 25 (Zareifard et al. 2016, Lauritzen 1996). *An undirected graph is said to be decomposable if any induced subgraph does not contain cordless cycle of length greater than or equal to four.*

Lemma 26 (Zareifard et al. 2016, Paulsen et al. 1989). *Let $\boldsymbol{\Omega}$ be an arbitrary positive definite matrix with zero restrictions according to decomposable graph $G = (V, E)$, i.e., $\boldsymbol{\Omega}_{ij} = 0$ if $(i, j) \notin E$. Then there exists an* *ordering of the vertices such that if* $\mathbf{\Omega} = \mathbf{LL}'$ *is the Cholesky decomposition corresponding to this ordering, then for* $i < j$

$$\mathbf{L}_{ij} = 0 \leftrightarrow (V, E) \notin E. \tag{3.27}$$

Hence the zero in $\mathbf{\Omega}$ *are preserved in the lower triangle of the corresponding matrix* $\mathbf{L}$ *obtained from the Cholesky decomposition.*

The previous Lemma tell us that if the optimisation returns a decomposable graph then we are also providing a sparse estimator of the correlation matrix. However, if the graph is not decomposable we can consider a different implementation in the same spirit of the column–wise update of the Graphical Lasso algorithm of Friedman et al. (2008).

We outline the steps of the algorithm below. Let $\mathbf{\Omega}$ be a correlation matrix of dimension $n \times n$ and partition $\mathbf{\Omega}$ as follows

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{\Omega}_{11} & \boldsymbol{\omega}_{12} \\ \boldsymbol{\omega}'_{12} & 1 \end{bmatrix}, \tag{3.28}$$

where $\mathbf{\Omega}_{11}$ is a matrix of dimension $(n-1) \times (n-1)$ and $\boldsymbol{\omega}_{12}$ is a vector of dimension $n-1$, and consider the transformation consider the transformation $\boldsymbol{\omega}^{\star}_{12} \to \frac{\hat{\boldsymbol{\omega}}_{12}}{1+\hat{\boldsymbol{\omega}}'_{12}\mathbf{\Omega}_{11}^{-1}\hat{\boldsymbol{\omega}}_{12}}$ where $\hat{\boldsymbol{\omega}}_{12}$ is obtained by applying a step of the Newton–Raphson algorithm to $\boldsymbol{\omega}_{12}$ as follows

$$\begin{aligned} \hat{\boldsymbol{\omega}}_{12} = \boldsymbol{\omega}_{12} - &\left[\frac{\partial \tilde{\mathbf{\Phi}}^R_{\boldsymbol{\omega}_{12}}}{\partial \boldsymbol{\omega}_{12}{}'} \mathbf{W}_{\bar{\boldsymbol{\omega}}_{12}} \frac{\partial \tilde{\mathbf{\Phi}}^R_{\boldsymbol{\omega}_{12}}}{\partial \boldsymbol{\omega}_{12}} + n\mathbf{\Sigma}_{\lambda}\left(\boldsymbol{\omega}_{12}\right) \right]^{-1} \\ &\times \left[-\frac{\partial \tilde{\mathbf{\Phi}}^R_{\sigma_{12}}}{\partial \sigma_{12}} \mathbf{W}_{\bar{\sigma}_{12}} \left(\hat{\mathbf{\Phi}} - \tilde{\mathbf{\Phi}}^R_{\sigma_{12}} \right) + n\mathbf{\Sigma}_{\lambda}\left(\boldsymbol{\omega}_{12}\right)\boldsymbol{\omega}_{12} \right]. \end{aligned} \tag{3.29}$$

Once we update the last column, we shift the next to the last at the end and repeat the steps described above. We repeat this procedure until convergence.

3.4 Synthetic data examples

In this Section we illustrate the performance of our methodology and compare it with three alternative methods on concrete simulation examples. We consider the ESD data generating process in Section 3.4.2 data generating process and the SESD 3.4.3 of different dimensions $m = \{2, 5, 12\}$. For each model, we consider several sample sizes: $T = \{500, 1000\}$ for the method of simulated quantiles and $T = 200$ for the sparse method of simulated quantiles. For each generated sample, we estimate the parameters using the MMSQ (S–MMSQ) and three different alternatives: the Graphical Lasso of Friedman et al. (2008) (GLasso), the graphical model with SCAD penalty (SCAD) and the graphical model with adaptive Lasso of Fan et al. (2009) (Adaptive Lasso).

To assess the performance of the MMSQ (S–MMSQ) we compute the Frobenius norm, the F_1–score measure and the Kullback–Leibler divergence, which are defined as follows

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{m} |a_{i,j}|^2} \tag{3.30}$$

$$F_1 - score = \frac{2TP}{2TP + FP + FN} \in [0, 1] \tag{3.31}$$

$$KL = \frac{1}{2}\left[\mathrm{tr}\left(\Omega_{true}^{-1}\hat{\Omega}\right) - m - \log\left(\frac{|\hat{\Omega}|}{|\Omega_{true}|}\right)\right], \tag{3.32}$$

where TP, FP and FN are the number of true positives, false positives and false negatives, respectively, see

Before presenting our results we illustrate the method we employ to select the relevant directions.

3.4.1 How to choose directions

An important issue related to the application of the MMSQ is the choice of the directions which are relevant for the parameters of interest. Indeed, a trivial approach would be to consider a grid of directions uniformly spaced on the unit circle. However, this solution would be computational expensive in large dimensional settings thus we consider a different and more effective approach. Specifically, we choose optimal directions $\mathbf{u}^*$ according to Definition 27 which allows to maximise the information contained in the selected measures.

Definition 27. *Let us consider a given parameter of interest $\boldsymbol{\vartheta}^\star \subset \Theta_k \in \mathbb{R}^k$ and consider the subset $\mathbf{Y}^\star = (Y_1^\star, \ldots, Y_l^\star, \ldots, Y_h^\star)$ of h variables of $\mathbf{Y} \in \mathbb{R}^m$ assumed to be informative for the parameter $\boldsymbol{\vartheta}^\star$, and the projectional quantile $q^{\tau\mathbf{u}}$ of $\mathbf{Y}^\star$ at a given τ, with $\mathbf{u} \in \mathbb{S}^{h-1}$. An optimal direction $\mathbf{u}^* \in \mathbb{S}^{m-1}$ for $\mathbf{Y}^\star$ is defined as the vector whose i–th coordinate is*

$$u_i^* = \begin{cases} u_{\max,l} & if\ Y_i = Y_l^\star \\ 0 & otherwise, \end{cases}$$

where $u_{\max,l}$ is the l–th coordinate of the vector

$$\mathbf{u}_{\max} \in \left\{ \arg \max_{\mathbf{u} \in \mathbb{S}^{h-1}} q^{\tau\mathbf{u}} \right\}. \tag{3.33}$$

If for example, $h = 2$, then the optimal direction is

$$\mathbf{u}^* = (0, \ldots, u_{\max,1}, 0, \ldots, 0, u_{\max,2}, \ldots, 0),$$

where $u_{\max,1}$ and $u_{\max,2}$ are the i–th and j–th coordinate respectively, which is informative for the covariances between Y_i and Y_j. The optimal solutions defined in (3.33) are computed using the Lagrangian function as follows

$$\mathcal{L}(\mathbf{u}, \lambda) = q^{\tau\mathbf{u}} - \lambda (\|\mathbf{u}\| - 1),$$

by solving

$$\nabla \mathcal{L}(\mathbf{u}, \lambda) = 0,$$

where ∇ stands for the gradient. This equation can be solved analytically, for instance when $m = h = 2$ for ESD distribution as shown in section

3.4.2, or numerically.

Let $\mathbf{U}^*$ be a set of optimal solutions $\mathbf{u}_j^*$ and let

$$\begin{aligned}\boldsymbol{\Phi}_{\boldsymbol{\vartheta}} &= \left(\boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}_1,\mathbf{u}_1^*}, \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}_2,\mathbf{u}_2^*}, \dots, \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}_K,\mathbf{u}_K^*}\right)' \in \mathbb{R}^B \\ \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^R &= \left(\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}_1,\mathbf{u}_1^*,R}, \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}_2,\mathbf{u}_2^*,R}, \dots, \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}_K,\mathbf{u}_K^*,R}\right)' \in \mathbb{R}^B \\ \hat{\boldsymbol{\Phi}} &= \left(\hat{\boldsymbol{\Phi}}^{\boldsymbol{\tau}_1,\mathbf{u}_1^*}, \hat{\boldsymbol{\Phi}}^{\boldsymbol{\tau}_2,\mathbf{u}_2^*}, \dots, \hat{\boldsymbol{\Phi}}^{\boldsymbol{\tau}_K,\mathbf{u}_K^*}\right)' \in \mathbb{R}^B,\end{aligned}$$

where K is the cardinality of $\mathbf{U}^*$, $B = \sum_{i=1}^K b_i$ and b_i is the dimension of $\boldsymbol{\Phi}_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}_i,\mathbf{u}_i^*}$ for $i = 1, 2, \dots, K$, then the MMSQ minimises the square distance defined in equation (3.5) between $\hat{\boldsymbol{\Phi}}$ and $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^R$ along the optimal directions $\mathbf{U}^*$.

3.4.2 Elliptical Stable distribution

In this Section we consider simulation examples for the ESD distribution $\mathbf{Y} \sim \mathcal{ESD}_m(\alpha, \boldsymbol{\xi}, \boldsymbol{\Omega})$ as defined in section 2.2. In order to apply the MMSQ, we first need to select the quantile–based measures which are informative for each of the parameters of interest $(\alpha, \boldsymbol{\xi}, \boldsymbol{\Omega})$ where the shape parameter $\alpha \in (0, 2)$ controls for the tail behaviour of the distribution, while $\boldsymbol{\xi} \in \mathcal{R}^m$ and $\boldsymbol{\Omega}$ denote the location parameter and the positive definite $m \times m$ scaling matrix, respectively. Since the quantile–based measures should be informative for the correspondent parameter, we select for α a measure related to the kurtosis of the distribution, for the locations the median and for the elements of the scaling matrix we opt for a measure of dispersion, and all the measures will be calculated along appropriately chosen directions, as it will be discussed later in this section. Summarising, for kurtosis, location and scale parameters we choose respectively

$$\kappa_{\mathbf{u}} = \frac{q_{0.95,\mathbf{u}} - q_{0.05,\mathbf{u}}}{q_{0.75,\mathbf{u}} - q_{0.25\mathbf{u}}}$$
$$m_{\mathbf{u}} = q_{0.5,\mathbf{u}}$$
$$\varsigma_{\mathbf{u}} = q_{0.75,\mathbf{u}} - q_{0.25\mathbf{u}},$$

where $\mathbf{u} \in \mathcal{S}^{m-1}$ defines a relevant direction. Next, we need to identify the optimal directions. To this end we can consider the relevant properties of the ESD. Specifically, as shown for example by Embrechts et al. (2005), the ESD is closed under marginalisation, i.e., $Y_i \sim \mathcal{ESD}_1(\alpha, \xi_i, \omega_{ii})$, for $i = 1, 2, \dots, m$, where ω_{ii} is the i–th element of the main diagonal of the matrix $\mathbf{\Omega}$. By exploiting the closure with respect to marginalisation, from definition 27 we conclude that the optimal directions for the shape parameter α, for the locations ξ_i and for the diagonal elements of the scale matrix ω_{ii}, for $i = 1, 2, \dots, m$ are the canonical directions. It still remains to consider the optimal directions for the off–diagonal elements of the scale matrix ω_{ij}, with $i, j = 1, 2, \dots, m$ and $i \neq j$. Again we exploit the closure with respect to marginalisation. Specifically, let $\mathbf{Z}_{ij} = (Y_i, Y_j)$, then $\mathbf{Z}_{ij} \sim \mathcal{ESD}_2\left(\alpha, \boldsymbol{\xi}_{ij}, \mathbf{\Omega}_{ij}\right)$, where

$$\boldsymbol{\xi}_{ij} = (\xi_i, \xi_j)', \qquad \mathbf{\Omega}_{ij} = \begin{pmatrix} \omega_{ii} & \omega_{ij} \\ \omega_{ij} & \omega_{jj} \end{pmatrix}.$$

Moreover, let $\mathbf{u} \in \mathbb{S}^1$ and $Z_{ij,\mathbf{u}} = \mathbf{u}'\mathbf{Z}_{ij}$ be the projection of $\mathbf{Z}_{ij}$ along $\mathbf{u}$, then $Z_{ij,\mathbf{u}} \sim \mathcal{ESD}_1\left(\alpha, \mathbf{u}'\boldsymbol{\xi}_{ij}, \mathbf{u}'\mathbf{\Omega}_{ij}\mathbf{u}\right)$, (see Embrechts et al. 2005), from which we have the following representation of the projected ESD random variable

$$Z_{ij,\mathbf{u}} = \mathbf{u}'\boldsymbol{\xi}_{ij} + \sqrt{\mathbf{u}'\mathbf{\Omega}_{ij}\mathbf{u}}Z, \tag{3.34}$$

where $Z \sim \mathcal{ESD}_1(\alpha, 0, 1)$. Following Definition 27, in order to find the optimal directions we need to compute

$$\mathbf{u}_{\max} = \arg\max_{\mathbf{u}\in\mathbb{S}^1} q^{\tau\mathbf{u}}\left(\mathbf{Z}_{ij}\right), \tag{3.35}$$

where $q^{\tau\mathbf{u}}\left(\mathbf{Z}_{ij}\right)$ is the projectional quantile of $\mathbf{Z}_{ij}$, i.e., the τ–th level quantile of the random variable $Z_{ij,\mathbf{u}}$. Exploiting representation (3.34), it holds

$$\mathbf{u}_{\max} = \arg\max_{\mathbf{u}\in\mathbb{S}^1} \mathbf{u}'\boldsymbol{\xi}_{ij} + \sqrt{\mathbf{u}'\boldsymbol{\Omega}_{ij}\mathbf{u}}, \tag{3.36}$$

which is a quadratic optimisation problem that can be solved using the method of Lagrangian multiplier, as follows

$$\mathcal{L}\left(\mathbf{u}, \lambda\right) = \mathbf{u}'\boldsymbol{\xi}_{ij} + \sqrt{\mathbf{u}'\boldsymbol{\Omega}_{ij}\mathbf{u}} - \lambda\left(\|\mathbf{u}\| - 1\right). \tag{3.37}$$

The solution requires to set to zero the gradient of the Lagrangian $\nabla\mathcal{L}\left(\mathbf{u}, \lambda\right) = 0$, that is

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial u_1} &= \frac{\left(\omega_{ii}^2 u_1 + \omega_{ij} u_2\right)}{\sqrt{\omega_{ii}^2 u_1^2 + \omega_{jj}^2 u_2^2 + 2\omega_{ij} u_1 u_2}} - 2\lambda u_1 = 0 \\
\frac{\partial\mathcal{L}}{\partial u_2} &= \frac{\left(\omega_{jj}^2 u_2 + \omega_{ij} u_1\right)}{\sqrt{\omega_{ii}^2 u_1^2 + \omega_{jj}^2 u_2^2 + 2\omega_{ij} u_1 u_2}} - 2\lambda u_2 = 0 \\
\frac{\partial\mathcal{L}}{\partial\lambda} &= u_1^2 + u_2^2 - 1 = 0,
\end{aligned} \tag{3.38}$$

and from the first two equations, we obtain

$$\begin{aligned}
& u_2\left(\sigma_1^2 u_1 + \omega_{ij} u_2\right) - u_1\left(\omega_{jj}^2 u_2 + \omega_{ij} u_1\right) = 0 \\
& u_2^2 + u_2 u_1 \frac{\omega_{ii}^2 - \omega_{jj}^2}{\omega_{ij}} - u_1^2 = 0 \\
& u_2 = \frac{u_1}{2}\left(-\frac{\omega_{ii}^2 - \omega_{jj}^2}{\omega_{ij}} \pm \sqrt{\left(\frac{\omega_{ii}^2 - \omega_{jj}^2}{\omega_{ij}}\right)^2 + 4}\right).
\end{aligned}$$

By inserting the previous expression for u_2 into equation (3.38), we solve for u_1

$$
\begin{aligned}
&u_1^2 + \frac{u_1^2}{4}\left(-\frac{\omega_{ii}^2-\omega_{jj}^2}{\omega_{ij}} \pm \sqrt{\left(\frac{\omega_{ii}^2-\omega_{jj}^2}{\omega_{ij}}\right)^2+4}\right)^2 = 1\\
&u_1^2\left[1+\frac{1}{4}\left(-\frac{\omega_{ii}^2-\omega_{jj}^2}{\omega_{ij}} \pm \sqrt{\left(\frac{\omega_{ii}^2-\omega_{jj}^2}{\omega_{ij}}\right)^2+4}\right)^2\right] = 1\\
&u_1 = \pm\frac{1}{\sqrt{\left[1+\frac{1}{4}\left(-\frac{\omega_{ii}^2-\omega_{jj}^2}{\omega_{ij}} \pm \sqrt{\left(\frac{\omega_{ii}^2-\omega_{jj}^2}{\omega_{ij}}\right)^2+4}\right)^2\right]}},
\end{aligned}
\tag{3.39}
$$

where the sign of u_1 depends on the sign of ω_{ij}. The optimal direction $\mathbf{u}_{\max}$ is then plugged into $\mathbf{u}^* = (0,\ldots,u_{1,\max},\ldots,u_{2,\max},\ldots,0)$ as explained in Definition 27.

To illustrate the effectiveness of the MMSQ we replicate the simulation study considered in Lombardi and Veredas (2009). Specifically, we consider two dimensions of the random vector $\mathbf{Y}$, $m = 2, 5$ and, for each dimension, we consider three values of the shape parameters $\alpha = \{1.7, 1.9, 1.95\}$, while the location parameter $\boldsymbol{\xi}$ is always set to zero and the scale matrices are

$$
\boldsymbol{\Omega}_2^s = \begin{pmatrix} 0.5 & 0.9 \\ 0.9 & 2 \end{pmatrix}, \tag{3.40}
$$

for $m = 2$, and

$$
\boldsymbol{\Omega}_5^s = \begin{pmatrix} 0.25 & 0.25 & 0.4 & 0 & 0 \\ 0.25 & 0.5 & 0.4 & 0 & 0 \\ 0.4 & 0.4 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2.55 \\ 0 & 0 & 0 & 2.55 & 4 \end{pmatrix}, \tag{3.41}
$$

for $m = 5$. We also consider two different sample sizes $n = 500, 2000$ and we fix $R = 5$. We also consider a simulation example of dimension $m = 12$, with $n = 500$ and $R = 5$ where the location parameters are equal to zero, as in previous examples, while the scale matrix $\mathbf{\Omega}_{12}^s$ is that considered in Wang (2015) and reported below

$$\begin{pmatrix}
0.239 & 0.117 & 0 & 0 & 0 & 0 & 0 & 0.031 & 0 & 0 & 0 & 0 \\
0.117 & 1.554 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.362 & 0.002 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.002 & 0.199 & 0.094 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.094 & 0.349 & 0 & 0 & 0 & 0 & 0 & 0 & -0.036 \\
0 & 0 & 0 & 0 & 0 & 0.295 & -0.229 & 0.002 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -0.229 & 0.715 & 0 & 0 & 0 & 0 & 0 \\
0.031 & 0 & 0 & 0 & 0 & 0.002 & 0 & 0.164 & 0.112 & -0.028 & -0.008 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.112 & 0.518 & -0.193 & -0.09 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.028 & -0.193 & 0.379 & 0.167 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.008 & -0.09 & 0.167 & 0.159 & 0 \\
0 & 0 & 0 & 0 & -0.036 & 0 & 0 & 0 & 0 & 0 & 0 & 0.207
\end{pmatrix}. \tag{3.42}$$

In order to show the results we introduce the following notation: let

$$\mathbf{R}_m^s = (\rho_{ij})_{i,j}^m \tag{3.43}$$

the m–dimensional matrix defined as

$$\mathbf{\Omega}_m^s = \mathbf{D}_m^{-1}\mathbf{R}_m^s\mathbf{D}_m^{-1} \tag{3.44}$$

where

$$\mathbf{D}_m = \operatorname{diag}\left\{\sqrt{\omega}_{11}, \dots, \sqrt{\omega}_{mm}\right\}. \tag{3.45}$$

In Table D.1, we report estimation results obtained over 1000 replications for $m = 2$, for all the values of α with $n = 500, 2000$ while in Tables D.2, D.3, D.4, we report results for $m = 5$, for the different values of the shape parameter $\alpha = \{1.7, 1.9, 1.95\}$. Specifically, each table reports the bias (BIAS), the standard error (SSD) and the empirical coverage probability (ECP) of the estimated parameters. Our results show that the MMSQ estimator is always unbiased, indeed the BIAS is always less than 0.25 in dimension $m = 2$ and less that 0.15 in dimension $m = 5$. The SSDs are always small, in particular for $n = 500$ it is always less then 0.5. The empirical coverages are always in line with their expected values for all but the diagonal elements of the scale matrix $\sqrt{\omega_{ii}}$ for $i = 1, 2, \dots, m$ for which they display lower values than expected, which means that in those cases the asymptotic standard errors are underestimated.

In Tables D.5, D.6 and D.7 we report the estimation results over $1,000$ replications for $m = 12$ and $\alpha = 1.7$, in Tables D.8, D.9, D.10 for $m = 12$ and $\alpha = 1.9$ and in Tables D.11, D.12, D.13 for $m = 12$ and $\alpha = 1.95$. The columns contains the same informations listed before, that are BIAS, SSD and ECP. The results show that the bias is always less then 0.2 in absolute value and the sample standard error is always less then 0.3. The empirical coverage probability is in general lower then the expected value, due to the small sample size, and, as observed in the previous examples, the lowest values correspond to the parameters $\sqrt{\omega_{ii}}$ for $i = 1, 2, \dots, 12$.

3.4.3 Skew Elliptical Stable distribution

In this Section we consider simulation examples for the SESD distribution $\mathbf{Y} \sim \mathcal{SESD}_m(\alpha, \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\delta})$ as defined in section 2.3. Specifically, we replicate the simulation study considered in section 3.4.2. Again, let us start by defining the quantile–based measures for the parameters of interest. For the shape parameter α, the locations ξ_i and scale parameters ω_{ii}, $i = 1, 2, \dots, m$ we consider the same measures as before, i.e., $(\kappa_{\mathbf{u}}, m_{\mathbf{u}}, \varsigma_{\mathbf{u}})$, while for the skewness parameters δ_i we consider the following measure

$$\gamma_{\mathbf{u}} = \frac{q_{0.99,\mathbf{u}} + q_{0.01,\mathbf{u}} - 2q_{0.5,\mathbf{u}}}{q_{0.99,\mathbf{u}} - q_{0.01,\mathbf{u}}}.$$

Then, as before, we need to identify the optimal directions for each parameter. Let us start by the locations. Due to the skewness, the median computed along the canonical direction is not anymore a good measure for the locations. Therefore, we consider a transformation of the data in order to remove the skewness. By the properties of the skewing mechanism, $\mathbf{Y}^- = -\mathbf{Y} \sim \mathcal{SESD}_m(\alpha, \boldsymbol{\xi}, \boldsymbol{\Omega}, -\boldsymbol{\delta})$ independent of $\mathbf{Y}$, and by applying

Theorem 9, it holds

$$\mathbf{Z} = \frac{\mathbf{Y} + \mathbf{Y}^{-}}{\sqrt{2}} \sim \mathcal{SESD}_m \left(\alpha, \sqrt{2}\boldsymbol{\xi}, \boldsymbol{\Omega}, \mathbf{0}\right), \tag{3.46}$$

which means that the variable $\mathbf{Z}$ is symmetric and, up to a constant, it has the same location parameter of $\mathbf{Y}$. Therefore, we choose, as informative measure for the locations, the median of the transformed variable $\mathbf{Z}$ in equation (3.46). In order to estimate the remaining parameters, we first consider that, for Theorem 8, univariate marginal variables Y_i, for $i = 1, 2, \dots, m$ have Skew Elliptical stable distribution, i.e., $Y_i \sim \mathcal{SESD}_1 (\alpha, \xi_i, \omega_{ii}, \delta_i)$. Therefore, the quantile–based measures for the shape, skewness and for the diagonal elements of the scale matrix are computed along the canonical directions.

Now we need to identify the optimal directions for the off–diagonal elements of the scale matrix. To this end, as before, we consider the bivariate marginal variables $\mathbf{Z}_{ij} = (Y_i, Y_j)'$ for $1 \leq i < j \leq m$. From Theorem 8 it holds $\mathbf{Z}_{ij} \sim \mathcal{SESD}_2 (\alpha, \xi_{ij}, \boldsymbol{\Omega}_{ij}, \boldsymbol{\delta}_{ij})$ where $\boldsymbol{\xi}_{ij}$ and $\boldsymbol{\Omega}_{ij}$ are defined as before, while $\boldsymbol{\delta}_{ij} = (\delta_i, \delta_j)'$. Moreover, let $\mathbf{Y}_{ij}^{-} \sim \mathcal{SESD}_2 (\alpha, \xi_{ij}, \boldsymbol{\Omega}_{ij}, -\boldsymbol{\delta}_{ij})$ independent of $\mathbf{Y}_{ij}$ and let us consider the same construction introduced for the locations, that is the random variable $\mathbf{Z}_{ij} = \frac{\mathbf{Y}_{ij} + \mathbf{Y}_{ij}^{-}}{\sqrt{2}}$, having distribution $\mathbf{Z}_{ij} \sim \mathcal{SESD}_2 \left(\alpha, \sqrt{2}\boldsymbol{\xi}_{ij}, \boldsymbol{\Omega}_{ij}, \mathbf{0}\right)$. Since $\mathbf{Z}_{ij}$ is a symmetric variable we can apply the same approach detailed in Section 3.4.2. Therefore, we choose the optimal direction $\mathbf{u}^* \in \mathbb{S}^1$ such that

$$\mathbf{u}^* = \arg\max_{\mathbf{u} \in \mathbb{S}^1} \sqrt{\mathbf{u}' \boldsymbol{\Omega}_{ij} \mathbf{u}}. \tag{3.47}$$

As in Section 3.4.2 we consider three values of the kurtosis parameter, $\alpha = \{1.7, 1.9, 1.95\}$ and set the locations to zero. We set the skewness $\boldsymbol{\delta} = (0.9, 0.9)$ and consider the scale matrix in (3.40) for the simulation study in dimension $m = 2$. In the second example in dimension $m = 5$ we set the skewness $\boldsymbol{\delta} = (0, 0, 0, 0.9, 0.9)$ and consider the scale matrix in equation (3.41). The third example is in dimension $m = 12$ and we set the skewness $\boldsymbol{\delta} = (0, 0, 0, 6, 0, 0, 0, 0, 0, 0, 0.6, 0.6, 0)$ and consider the scale matrix in equation (3.42).

In Tables D.14 we report the estimation results over $1,000$ replications for $m = 2$, in Tables D.15, D.16 and D.17 we report the estimation results for $m = 5$ and il Tables D.18, D.19, D.20, D.21, D.22 and D.23 we report the estimation results for $m = 12$. The columns contains the same informations listed before, that are BIAS, SSD and ECP.

α	1.70	1.90	1.95	2.00
Frobenius norm	*Dimension 12*			
GLasso	12.7379	3.9739	3.0410	2.6243
	(126.4986)	(2.7219)	(0.1621)	(0.0034)
SCAD	12.7135	3.9622	3.0349	2.6385
	(125.8675)	(2.6617)	(0.1580)	(0.0033)
Adaptive Lasso	13.0491	3.9439	3.0348	2.6347
	(147.5606)	(2.6404)	(0.1631)	(0.0034)
S–MMSQ	1.3755	1.5232	1.5903	1.7177
	(0.1930)	(0.2925)	(0.3687)	(0.3078)
F_1–score	*Dimension 12*			
GLasso	0.1555	0.0000	0.0000	0.0000
	(0.0673)	(0.0000)	(0.0000)	(0.0000)
SCAD	0.3360	0.1705	0.1453	0.2634
	(0.0701)	(0.0584)	(0.0384)	(0.0800)
Adaptive Lasso	0.2644	0.0526	0.0250	0.0284
	(0.1281)	(0.0153)	(0.0075)	(0.0085)
S–MMSQ	0.9058	0.7489	0.7314	0.6626
	(0.0019)	(0.0129)	(0.0105)	(0.0166)
KL	*Dimension 12*			
GLasso	11.7517	1.9744	1.6304	2.9770
	(248.9867)	(0.5865)	(0.0321)	(0.0262)
SCAD	11.7324	2.1747	1.7942	3.2730
	(242.2986)	(0.6994)	(0.0371)	(0.0196)
Adaptive Lasso	12.7734	2.0221	1.6785	3.0782
	(361.2139)	(0.5995)	(0.0361)	(0.0314)
S–MMSQ	0.8365	0.9974	1.0868	1.2585
	(0.2147)	(0.4067)	(0.5500)	(0.4935)

Table 3.1: Frobenius norm, F1–Score and Kullback–Leibler information between the true scale matrix in equation (3.42) of the Elliptical Stable distribution and the matrices estimated by alternative methods: the Graphical Lasso of Friedman et al. (2008) (GLasso), the graphical model with SCAD penalty (SCAD), the graphical model with adaptive Lasso of Fan et al. (2009) (Adaptive Lasso) and the S–MMSQ. The measures are evaluated over 100 replications, we report the mean and the variances in brackets.

α	1.70	1.90	1.95	2.00
Frobenius norm		*Dimension 12*		
GLasso	80.8238 ($5.9815{*}10^4$)	10.5276 ($1.0072{*}10^3$)	4.5713 (52.0067)	2.6035 ($0.6127{*}10^{-3}$)
SCAD	80.7741 ($5.9823{*}10^4$)	10.5748 ($1.0073{*}10^3$)	4.5493 (51.8079)	2.6064 ($0.6048{*}10^{-3}$)
Adaptive Lasso	78.2717 ($5.4857{*}10^4$)	10.0179 ($0.8584{*}10^3$)	4.4925 (47.1308)	2.6073 ($0.6168{*}10^{-3}$)
S–MMSQ	1.2175 (0.0306)	1.2196 (0.0303)	1.2184 (0.0293)	1.2232 (0.0309)
F_1–score		*Dimension 12*		
GLasso	0.2572 (0.1077)	0.0609 (0.0446)	0.0278 (0.0193)	0.0000 (0.0000)
SCAD	0.3702 (0.0946)	0.1320 (0.0593)	0.1097 (0.0377)	0.0750 (0.0186)
Adaptive Lasso	0.3953 (0.1765)	0.1067 (0.0628)	0.0457 (0.0321)	0.0000 (0.0000)
S–MMSQ	0.8396 (0.0014)	0.8398 (0.0014)	0.8395 (0.0014)	0.8391 (0.0014)
KL		*Dimension 12*		
GLasso	120.2312 ($1.6762{*}10^5$)	10.3107 ($1.6178{*}10^3$)	3.5421 (87.2424)	3.0728 (0.0044)
SCAD	120.3742 ($1.6758{*}10^5$)	10.5506 ($1.6171{*}10^3$)	3.5556 (81.9598)	3.1799 (0.0041)
Adaptive Lasso	119.4011 ($1.6013{*}10^5$)	9.3621 ($1.2379{*}10^3$)	3.6616 (95.5155)	3.1236 (0.0040)
S–MMSQ	0.6607 (0.0246)	0.6640 (0.0241)	0.6616 (0.0239)	0.6656 (0.0245)

Table 3.2: Frobenius norm, F1–Score and Kullbach–Leibler information between the true scale matrix in equation (3.42) of the Skew – Elliptical Stable distribution and the matrices estimated by alternative methods: the Graphical Lasso of Friedman et al. (2008) (GLasso), the graphical model with SCAD penalty (SCAD), the graphical model with adaptive Lasso of Fan et al. (2009) (Adaptive Lasso) and the S–MMSQ. The measures are evaluated over 100 replications, we report the mean and the variances in brackets.

Chapter 4

Sparse Indirect Inference

4.1 Introduction

Indirect inference (II) methods (Gouriéroux et al. 1993 and Gallant and Tauchen 1996) are likelihood–free alternatives to maximum likelihood or moment–based estimation methods for parametric inference which are particularly valuable when a closed–form expression for the density is not analytically available. Following the approach of Gouriéroux et al. (1993), throughout the chapter we consider the following dynamic model

$$y_t = r\left(y_{t-1}, \mathbf{x}_t, u_t, \boldsymbol{\vartheta}\right) \tag{4.1}$$

$$u_t = \phi\left(u_{t-1}, \epsilon_t, \boldsymbol{\vartheta}\right), \qquad \forall t = 1, 2, \ldots, T, \tag{4.2}$$

where $\mathbf{x}_t$ are exogenous variables whereas u_t and ϵ_t are latent variables. Concerning the process defined in equations (4.1)–(4.2), we assume that: *(i)* $\mathbf{x}_t$ is an homogeneous Markov process with transition distribution F_0 independent of ϵ_t and u_t; *(ii)* the process ϵ_t is a white noise whose distribution G_0 is known, and *(iii)* the process $\{y_t, \mathbf{x}_t\}$ is weekly stationary. We further assume that the joint density function of the observations $\{y_t, \mathbf{x}_t\}_{t=1}^{T}$ is not known analytically. The II method replaces the maximum likelihood estimator of the parameter $\boldsymbol{\vartheta}$ in equations (4.1)–(4.2) with a quasi–maximum likelihood estimator which relies on an alternative auxiliary model and then corrects for misspecification inconsistency by simulating from the original model. Specifically, let $Q(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta})$ the auxiliary criterion function, which depends on the observations $\{y_t, \mathbf{x}_t\}_{t=1}^T$ and on the auxiliary parameter $\boldsymbol{\beta} \in \mathbf{B} \subset \mathbb{R}^q$, such that $\lim_{T\to\infty} Q_T(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}) = Q_\infty(F_0, G_0, \boldsymbol{\vartheta}_0, \boldsymbol{\beta})$, a.s., where $\boldsymbol{\vartheta}_0$ is the true parameter of interest, then

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}\in\mathbf{B}} Q(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}). \tag{4.3}$$

Under the additional assumptions that the limit criterion is continuous in $\boldsymbol{\beta}$ and has a unique maximum $\boldsymbol{\beta}_0$, then the estimator $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}_0$, that is unknown since it depends on F_0 and $\boldsymbol{\vartheta}_0$ that are unknown. To overcome this problem, the II method simulates, for each value of $\boldsymbol{\vartheta}$, H paths $\tilde{\mathbf{y}}^h$ for $h = 1, 2, \dots, H$ and computes the QML estimate $\tilde{\boldsymbol{\beta}}^h$ for the auxiliary model in equation (4.3) and subsequently minimises the following objective function

$$\hat{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta}} \left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H} \tilde{\boldsymbol{\beta}}^h\right)' \hat{\boldsymbol{\Omega}} \left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H} \tilde{\boldsymbol{\beta}}^h\right), \tag{4.4}$$

for an appropriately chosen positive–definite square symmetric matrix $\hat{\boldsymbol{\Omega}}$. Moreover, indirect estimators are consistent and asymptotically Normally distributed under mild regularity conditions, see Gouriéroux et al. (1993). Among those assumptions, the most important refer to the bijectivity of the binding function

$$b(F, G, \boldsymbol{\vartheta}) = \arg\max_{\boldsymbol{\beta}\in\boldsymbol{B}} Q(F, G, \boldsymbol{\vartheta}, \boldsymbol{\beta}), \tag{4.5}$$

that maps the parameter space of the auxiliary model onto the parameter space of the true model and the full–column rank of the matrix $\frac{\partial b}{\partial \boldsymbol{\vartheta}}(F_0, G_0, \cdot)$.

The remaining of this chapter is organised as follows. Section 4.2 introduces the Sparse Indirect Inference (S–II) estimator. Asymptotic theory of the proposed S–II estimator is presented in Section 4.3, while Section 4.4 details the algorithm. Section 4.5 concludes by applying the methodology to a synthetic dataset from a linear regression model with Stable innovations.

4.2 Sparse method of indirect inference

In order to achieve sparse estimation of the parameter $\boldsymbol{\vartheta}$, as for the method of simulated quantiles considered in the previous chapter, we introduce the Smoothly Clipped Absolute Deviation (SCAD) ℓ_1–penalty of Fan and Li (2001) into the indirect inference objective function. The SCAD function is a non–convex penalty function defined as

$$p_\lambda\left(|\gamma|\right) = \begin{cases} \lambda|\gamma| & \text{if } |\gamma| \leq \lambda \\ \frac{1}{a-1}\left(a\lambda|\gamma| - \frac{\gamma^2}{2}\right) - \frac{\lambda^2}{2(a-1)} & \text{if } \lambda < \gamma \leq a\lambda \\ \frac{\lambda^2(a+1)}{2} & \text{if } a\lambda < |\gamma|, \end{cases} \tag{4.6}$$

which corresponds to quadratic spline function with knots at λ and $a\lambda$. The SCAD penalty is continuously differentiable on $(-\infty; 0) \cup (0; \infty)$ but singular at 0 with its derivatives being zero outside the range $[-a\lambda; a\lambda]$. This results in small coefficients being set to zero, a few other coefficients being shrunk towards zero while retaining the large coefficients as they are. The S–II estimator minimises the penalised II objective function, as follows

$$\hat{\boldsymbol{\vartheta}}^* = \arg\min_{\boldsymbol{\vartheta}} \mathcal{D}\left(\boldsymbol{\vartheta}\right), \tag{4.7}$$

where

$$\mathcal{D}\left(\boldsymbol{\vartheta}\right) = \left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}^h\right)' \hat{\boldsymbol{\Omega}} \left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}^h\right) + n\sum_i p_\lambda\left(|\boldsymbol{\vartheta}_i|\right), \tag{4.8}$$

where $\hat{\boldsymbol{\Omega}}$ is a positive–definite square symmetric matrix. A similar approach in a different context has been recently proposed by Blasques and Duplinskiy (2015).

4.3 Asymptotic theory

As shown in Fan and Li (2001), the SCAD estimator, with appropriate choice of the regularisation (tuning) parameter, possesses a sparsity property, i.e., it estimates zero components of the true parameter vector exactly as zero with probability approaching one as sample size increases while still being consistent for the non–zero components. An immediate consequence of the sparsity property of the SCAD estimator is the, so called, oracle property, i.e., the asymptotic distribution of the estimator remains the same whether or not the correct zero restrictions are imposed in the course of the SCAD estimation procedure. More specifically, let $\boldsymbol{\vartheta}_0 = (\boldsymbol{\vartheta}_0^1, \boldsymbol{\vartheta}_0^0)$ be the true value of the unknown parameter $\boldsymbol{\vartheta}$, where $\boldsymbol{\vartheta}_0^1 \in \mathbb{R}^s$ is the subset of non–zero parameters and $\boldsymbol{\vartheta}_0^0 = 0 \in \mathbb{R}^{k-s}$ and let $A = \{i : \boldsymbol{\vartheta}_i \in \boldsymbol{\vartheta}_0^1\}$, we consider the definition of oracle estimator that has been formalised in Zou (2006).

Definition 28. *An oracle estimator $\hat{\boldsymbol{\vartheta}}_{\text{oracle}}$ has the following properties:*

(i) consistent variable selection:

$$\lim_{n\to\infty} \mathbb{P}\left(A_n = A\right) = 1, \tag{4.9}$$

where $A_n = \left\{i : \hat{\boldsymbol{\vartheta}}_i \in \hat{\boldsymbol{\vartheta}}_{\text{oracle}}^1\right\}$;

(ii) asymptotic normality:

$$\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}_{\text{oracle}}^1 - \boldsymbol{\vartheta}_0^1\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right), \tag{4.10}$$

as $n \to \infty$, where $\boldsymbol{\Sigma}$ is the variance covariance matrix of $\boldsymbol{\vartheta}_0^1$.

In the remainder the Section we establish the oracle properties of the S–II estimator. To this end, the following set of assumptions are needed:

(i)

$$\xi_T = \sqrt{T}\left(\frac{\partial Q}{\partial \boldsymbol{\beta}}\left(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}_0\right) - \frac{1}{H}\sum_{h=1}^{H}\frac{\partial Q}{\partial \boldsymbol{\beta}}\left(\tilde{\mathbf{y}}^h, \mathbf{X}, \boldsymbol{\beta}_0\right)\right), \tag{4.11}$$

is asymptotically Normal with mean zero, and asymptotic variance–covariance matrix given by $W = \lim_{T\to\infty} V(\xi_T)$;

(ii)

$$\lim_{T\to\infty} V\left(\sqrt{T}\frac{\partial Q}{\partial \boldsymbol{\beta}}\left(\tilde{\mathbf{y}}^h, \mathbf{X}, \boldsymbol{\beta}_0\right)\right) = \mathbf{I}_0, \tag{4.12}$$

and the limit is independent of the initial values z_0^h, for $h = 1, 2, \ldots, H$;

(iii)

$$\lim_{T\to\infty} \operatorname{Cov}\left(\sqrt{T}\frac{\partial Q}{\partial \boldsymbol{\beta}}\left(\tilde{\mathbf{y}}^h, \mathbf{X}, \boldsymbol{\beta}_0\right), \sqrt{T}\frac{\partial Q}{\partial \boldsymbol{\beta}}\left(\tilde{\mathbf{y}}^l, \mathbf{X}, \boldsymbol{\beta}_0\right)\right) = \mathbf{K}_0, \tag{4.13}$$

and the limit is independent of z_0^h and z_0^l for $h \neq l$;

(iv)

$$\operatorname{plim}_{T\to\infty} - \frac{\partial^2 Q}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\left(\tilde{\mathbf{y}}^h, \mathbf{X}, \boldsymbol{\beta}_0\right) = -\frac{\partial^2 Q_\infty}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\left(F_0, G_0, \boldsymbol{\vartheta}_0, \boldsymbol{\beta}_0\right) = \mathbf{J}_0, \tag{4.14}$$

and the limit is independent of z_0^h;

(v)

$$\lim_{T\to\infty} \hat{\boldsymbol{\Omega}} = \boldsymbol{\Omega}. \tag{4.15}$$

The next Theorem states that the estimator defined in equation (4.7) satisfies the sparsity property.

Theorem 29. *Given the SCAD penalty function $p_\lambda(\cdot)$, for a sequence of λ_n such that $\lambda_n \to 0$, and $\sqrt{n}\lambda_n \to \infty$, as $n \to \infty$, there exists a local minimiser $\hat{\boldsymbol{\vartheta}}$ of $\mathcal{D}(\boldsymbol{\vartheta})$ in (4.7) with $\|\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. Furthermore, we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\hat{\boldsymbol{\vartheta}}^0 = 0\right) = 1. \tag{4.16}$$

Proof. See Appendix A. □

The following theorem establishes the asymptotic normality of the penalised SCAD II estimator; we denote by $\boldsymbol{\vartheta}^1$ the subvector of $\boldsymbol{\vartheta}$ that does not contain zero elements and by $\hat{\boldsymbol{\vartheta}}^1$ the corresponding penalised II estimator.

Theorem 30. *Given the SCAD penalty function $p_\lambda\left(|\boldsymbol{\vartheta}_i|\right)$, for a sequence $\lambda_n \to 0$ and $\sqrt{n}\lambda_n \to \infty$ as $n \to \infty$, then $\hat{\boldsymbol{\vartheta}}^1$ has the following asymptotic distribution:*

$$\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}^1 - \boldsymbol{\vartheta}_0^1\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \left(1 + \frac{1}{H}\right)\mathbf{W}\right), \tag{4.17}$$

as $n \to \infty$, where

$$\mathbf{W} = \left(b'\left(F_0, G_0, \boldsymbol{\vartheta}_0\right)' \boldsymbol{\Omega} b'\left(F_0, G_0, \boldsymbol{\vartheta}_0\right)\right)^{-1} \mathbf{W}_1 \left(b'\left(F_0, G_0, \boldsymbol{\vartheta}_0\right)' \boldsymbol{\Omega} b'\left(F_0, G_0, \boldsymbol{\vartheta}_0\right)\right)^{-1},$$

and

$$\mathbf{W}_1 = b'\left(F_0, G_0, \boldsymbol{\vartheta}_0\right)' \boldsymbol{\Omega}\mathbf{J}_0^{-1}\left(\mathbf{I}_0 - \mathbf{K}_0\right)\mathbf{J}_0^{-1}\boldsymbol{\Omega} b'\left(F_0, G_0, \boldsymbol{\vartheta}_0\right), \tag{4.18}$$

where $b'\left(F_0, G_0, \boldsymbol{\vartheta}_0\right) = \frac{\partial b(F_0, G_0, \boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}^{1\prime}}$ is the first derivative of the binding function $b\left(F_0, G_0, \boldsymbol{\vartheta}_0\right)$.

Proof. See Appendix A. □

4.4 Algorithm

The objective function of the sparse estimator is the sum of a convex function and a non convex function which complicates the minimisation procedure. Here, we adapt the algorithms proposed by Fan and Li (2001) and Hunter and Li (2005) to our objective function in order to allow a fast procedure for the minimisation problem.

The first derivative of the penalty function can be approximated as follows

$$\left[p_\lambda\left(|\boldsymbol{\vartheta}_i|\right)\right]' = p'_\lambda\left(|\boldsymbol{\vartheta}_i|\right)\operatorname{sgn}\left(\boldsymbol{\vartheta}_i\right) \approx \frac{p'_\lambda\left(|\boldsymbol{\vartheta}_{i0}|\right)}{|\boldsymbol{\vartheta}_{i0}|}\boldsymbol{\vartheta}_i, \tag{4.19}$$

when $\boldsymbol{\vartheta}_i \neq 0$. We use it in the first order Taylor expansion of the penalty to get

$$p_\lambda\left(|\boldsymbol{\vartheta}_i|\right) \approx p_\lambda\left(|\boldsymbol{\vartheta}_{i0}|\right) + \frac{1}{2}\frac{p'_\lambda\left(|\boldsymbol{\vartheta}_{i0}|\right)}{|\boldsymbol{\vartheta}_{i0}|}\left(\boldsymbol{\vartheta}_i^2 - \boldsymbol{\vartheta}_{i0}^2\right), \tag{4.20}$$

for $\boldsymbol{\vartheta}_i \approx \boldsymbol{\vartheta}_{i0}$. The objective function $\mathcal{D}$ in equation (4.7) can be locally approximated, except for a constant term by

$$\begin{aligned}
\mathcal{D}\left(\boldsymbol{\vartheta}\right) \approx & \left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}\right)\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}\right) \\
& - \frac{1}{H}\sum_{h=1}^{H}\frac{\partial\tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial\boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}\right)\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\right) \\
& + \frac{1}{2}\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\right)'\frac{1}{H}\sum_{h=1}^{H}\frac{\partial\tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial\boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\frac{1}{H}\sum_{h=1}^{H}\frac{\partial\tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial\boldsymbol{\vartheta}}\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\right) \\
& + \frac{n}{2}\boldsymbol{\vartheta}'\bar{\mathbf{P}}_{\lambda_n}\boldsymbol{\vartheta},
\end{aligned} \tag{4.21}$$

where $\bar{\mathbf{P}}_{\lambda_n} = \operatorname{diag}\left\{\frac{p'_{\lambda_n}(|\boldsymbol{\vartheta}_{0i}|)}{|\boldsymbol{\vartheta}_{0i}|}; \boldsymbol{\vartheta}_{0i} \in \boldsymbol{\vartheta}_0^1\right\}$. Then the first order condition becomes

$$\begin{aligned}
\frac{\partial \mathcal{D}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} &\approx -\frac{1}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}\right) \\
&\quad + \frac{1}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\frac{1}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) + n\bar{\mathbf{P}}_{\lambda_n}\boldsymbol{\vartheta} \\
&= -\frac{1}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}\right) \\
&\quad + \frac{1}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\frac{1}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) \\
&\quad + n\bar{\mathbf{P}}_{\lambda_n}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) + n\bar{\mathbf{P}}_{\lambda_n}\boldsymbol{\vartheta}_0 \\
&= 0, \qquad (4.22)
\end{aligned}$$

therefore

$$\begin{aligned}
&\left[\frac{1}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\frac{1}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} + n\bar{\mathbf{P}}_{\lambda_n}\right](\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) = \\
&\qquad \frac{1}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}\right) - n\bar{\mathbf{P}}_{\lambda_n}\boldsymbol{\vartheta}_0 \qquad (4.23)
\end{aligned}$$

and

$$\begin{aligned}
\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0 &- \left[\frac{1}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\frac{1}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} + n\bar{\mathbf{P}}_{\lambda_n}\right]^{-1} \\
&\times \left[\frac{1}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}\right) - n\bar{\mathbf{P}}_{\lambda_n}\boldsymbol{\vartheta}_0\right]. \qquad (4.24)
\end{aligned}$$

The optimal solution can be find iteratively, as follows

$$\boldsymbol{\vartheta}^{(k+1)} = \boldsymbol{\vartheta}^{(k)} - \left[\frac{1}{H} \sum_{h=1}^{H} \frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}^{(k)}}}{\partial \boldsymbol{\vartheta}} \hat{\boldsymbol{\Omega}} \frac{1}{H} \sum_{h=1}^{H} \frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}^{(k)}}}{\partial \boldsymbol{\vartheta}} + n \bar{\mathbf{P}}^{(k)}_{\lambda_n} \right]^{-1}$$
$$\times \left[\frac{1}{H} \sum_{h=1}^{H} \frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}^{(k)}}}{\partial \boldsymbol{\vartheta}} \hat{\boldsymbol{\Omega}} \left(\hat{\boldsymbol{\beta}} - \frac{1}{H} \sum_{h=1}^{H} \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}^{(k)}} \right) - n \bar{\mathbf{P}}^{(k)}_{\lambda_n} \boldsymbol{\vartheta}^{(k)} \right], \tag{4.25}$$

where $\bar{\mathbf{P}}^{(k)}_{\lambda_n} = \operatorname{diag} \left\{ \frac{p'_{\lambda_n}\left(|\boldsymbol{\vartheta}^{(k)}_i|\right)}{|\boldsymbol{\vartheta}^{(k)}_i|} ; \boldsymbol{\vartheta}^{(k)}_i \neq 0 \right\}$ and if $\boldsymbol{\vartheta}^{(k+1)}_i \approx 0$, then $\boldsymbol{\vartheta}^{(k+1)}_i$ is set equal zero. When the algorithm converges the estimator satisfies the following equation

$$\frac{1}{H} \sum_{h=1}^{H} \frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} \hat{\boldsymbol{\Omega}} \left(\hat{\boldsymbol{\beta}} - \frac{1}{H} \sum_{h=1}^{H} \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0} \right) - n \bar{\mathbf{P}}_{\lambda_n} \boldsymbol{\vartheta}_0 = 0, \tag{4.26}$$

that is the first order condition of the minimisation problem of the SCAD II.

The algorithm used above and introduced by Fan and Li (2001) is called local quadratic approximation (LQA). Hunter and Li (2005) showed that LQA applied to penalised maximum likelihood is an MM algorithm. Indeed, we define

$$\Psi_{|\boldsymbol{\vartheta}_{0i}|} \left(|\boldsymbol{\vartheta}_i| \right) = p_\lambda \left(|\boldsymbol{\vartheta}_{0i}| \right) + \frac{1}{2} \frac{p'_\lambda \left(|\boldsymbol{\vartheta}_{0i}| \right)}{|\boldsymbol{\vartheta}_{0i}|} \left(\boldsymbol{\vartheta}_i^2 - \boldsymbol{\vartheta}_{0i}^2 \right), \tag{4.27}$$

since the SCAD penalty is concave it holds

$$\Psi_{|\boldsymbol{\vartheta}_{0i}|} \left(|\boldsymbol{\vartheta}_i| \right) \geq p_\lambda \left(|\boldsymbol{\vartheta}_i| \right), \quad \forall |\boldsymbol{\vartheta}_i|, \tag{4.28}$$

and equality holds when $|\boldsymbol{\vartheta}_i| = |\boldsymbol{\vartheta}_{0i}|$. Then $\Psi_{|\boldsymbol{\vartheta}_{0i}|} \left(|\boldsymbol{\vartheta}_i| \right)$ majorise $p_\lambda \left(|\boldsymbol{\vartheta}_i| \right)$, and it holds

$$\Psi_{|\boldsymbol{\vartheta}_{0i}|} \left(|\boldsymbol{\vartheta}_i| \right) < \Psi_{|\boldsymbol{\vartheta}_{0i}|} \left(|\boldsymbol{\vartheta}_{0i}| \right) \Rightarrow p_\lambda \left(|\boldsymbol{\vartheta}_i| \right) < p_\lambda \left(|\boldsymbol{\vartheta}_{0i}| \right), \tag{4.29}$$

that is called descendent property. This feature allows us to construct an MM algorithm: at each iteration k we construct $\Psi_{|\boldsymbol{\vartheta}_i^{(k)}|}(|\boldsymbol{\vartheta}_i|)$ and then minimize it to get $\boldsymbol{\vartheta}_i^{(k+1)}$, that satisfies $p_\lambda\left(|\boldsymbol{\vartheta}_i^{(k+1)}|\right) < p_\lambda\left(|\boldsymbol{\vartheta}_i^{(k)}|\right)$.

Let us consider the following

$$S_k(\boldsymbol{\vartheta}) = \left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}_{\boldsymbol{\vartheta}}^h\right)\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}_{\boldsymbol{\vartheta}}^h\right) + n\sum_i \Psi_{|\boldsymbol{\vartheta}_i^{(k)}|}(|\boldsymbol{\vartheta}_i|), \tag{4.30}$$

then $S_k(\boldsymbol{\vartheta})$ majorise $\mathcal{D}(\boldsymbol{\vartheta})$; thus we only need to minimise $S_k(\boldsymbol{\vartheta})$, that can be done as explained above.

Hunter and Li (2005) proposed an improved version of LQA for penalised maximum likelihood, aimed at avoiding to zero out the parameters too early during the iterative procedure. We present their method applied to SCAD II as follows

$$\begin{aligned}
p_{\lambda,\epsilon}(|\boldsymbol{\vartheta}_i|) &= p_\lambda(|\boldsymbol{\vartheta}_i|) - \epsilon\int_0^{|\boldsymbol{\vartheta}_i|}\frac{p'_\lambda(|\boldsymbol{\vartheta}_{0i}|)}{\epsilon + t}dt\\
\mathcal{D}_\epsilon(\boldsymbol{\vartheta}) &= \left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}_{\boldsymbol{\vartheta}}^h\right)\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}_{\boldsymbol{\vartheta}}^h\right) + n\sum_i p_{\lambda,\epsilon}(|\boldsymbol{\vartheta}_i|)\\
\Psi_{|\boldsymbol{\vartheta}_{0i}|,\epsilon}(|\boldsymbol{\vartheta}_i|) &= p_{\lambda,\epsilon}(|\boldsymbol{\vartheta}_{0i}|) + \frac{p'_\lambda(|\boldsymbol{\vartheta}_{0i}|)}{2(\epsilon + |\boldsymbol{\vartheta}_{0i}|)}\left(\boldsymbol{\vartheta}_i^2 - \boldsymbol{\vartheta}_{0i}^2\right)\\
S_{k,\epsilon}(\boldsymbol{\vartheta}) &= \left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}_{\boldsymbol{\vartheta}}^h\right)\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}_{\boldsymbol{\vartheta}}^h\right) + n\sum_i \Psi_{|\boldsymbol{\vartheta}_i^{(k)}|,\epsilon}(|\boldsymbol{\vartheta}_i|).
\end{aligned}$$

They proved that as $\epsilon \downarrow 0$ the perturbed objective function $\mathcal{D}_\epsilon(\boldsymbol{\vartheta})$ converges uniformly to the not perturbed one $\mathcal{D}(\boldsymbol{\vartheta})$ and that if $\hat{\boldsymbol{\vartheta}}_\epsilon$ is a minimiser of $\mathcal{D}_\epsilon(\boldsymbol{\vartheta})$ then any limit point of the sequence $\left\{\hat{\boldsymbol{\vartheta}}_\epsilon\right\}_{\epsilon\downarrow 0}$ is a minimiser of $\mathcal{D}(\boldsymbol{\vartheta})$. This construction allows to define $\Psi_{|\sigma_{i,j}^{(k)}|,\epsilon}(|\sigma_{i,j}|)$ even when $\sigma_{i,j}^{(k)} \approx 0$. The authors also provided a way to choose the value of the perturbation ϵ and suggested the following

$$\epsilon = \frac{\tau}{2np'_\lambda(0)}\min\left\{|\boldsymbol{\vartheta}_i^{(0)}| : \boldsymbol{\vartheta}_i^{(0)} \neq 0\right\}, \tag{4.31}$$

with the following tuning constant $\tau = 10^{-8}$.

4.4.1 Tuning paramenter selection

The SCAD penalty requires the selection of two tuning parameters (a, λ). The first tuning parameter is fixed at $a = 3.7$ as suggested in Fan and Li (2001), while the parameter λ is selected using K–fold cross validation, in which the original sample is divided in K subgroups T_k, called folds. The validation function is

$$CV(\lambda) = \sum_{k=1}^{K} \frac{1}{n_k} \left(\hat{\boldsymbol{\beta}} - \frac{1}{H} \sum_{h=1}^{H} \tilde{\boldsymbol{\beta}}^h_{\hat{\boldsymbol{\vartheta}}_{\lambda,k}} \right) \hat{\boldsymbol{\Omega}} \left(\hat{\boldsymbol{\beta}} - \frac{1}{H} \sum_{h=1}^{H} \tilde{\boldsymbol{\beta}}^h_{\hat{\boldsymbol{\vartheta}}_{\lambda,k}} \right), \tag{4.32}$$

where $\hat{\boldsymbol{\vartheta}}_{\lambda,k}$ denotes the parameters estimate on the sample $\left(\cup_{i=1}^{K} T_k\right) \setminus T_k$ with λ as tuning parameter. Then the optimal value is chosen as $\lambda^* = \arg\min_\lambda CV(\lambda)$; again the minimisation is performed over a grid of values for λ.

4.4.2 Alternative formulation

The indirect inference estimator can be formulated in several ways by changing the objective function defined in equation (4.4). Here we consider the formulation that involves the distance between scores, as follows

$$\hat{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta}} \left(\frac{1}{H} \sum_{i=1}^{H} \tilde{\nabla}^i_{\boldsymbol{\vartheta}} \right)' \hat{\boldsymbol{\Omega}} \left(\frac{1}{H} \sum_{i=1}^{H} \tilde{\nabla}^i_{\boldsymbol{\vartheta}} \right), \tag{4.33}$$

where $\nabla_{\boldsymbol{\vartheta}} = \frac{\partial \log p(\mathbf{y}|\mathbf{X},\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}$ is the score function of the auxiliary model and $\tilde{\nabla}^i_{\boldsymbol{\vartheta}} = \frac{\partial \log p\left(\tilde{\mathbf{y}}^i|\mathbf{X},\boldsymbol{\vartheta}\right)}{\partial \boldsymbol{\vartheta}}$ is the score function of the auxiliary model computed on the i–th simulation path. The asymptotic theory related to this formulation is similar to the one presented above. However, from a computational point of view, this formulation of the objective function of the indirect inference method is faster because it only requires the computation of the score of the auxiliary model.

The sparse version is obtained by adding the SCAD penalty:

$$\hat{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta}} \left(\frac{1}{H} \sum_{i=1}^{H} \tilde{\nabla}^i_{\boldsymbol{\vartheta}} \right)' \hat{\boldsymbol{\Omega}} \left(\frac{1}{H} \sum_{i=1}^{H} \tilde{\nabla}^i_{\boldsymbol{\vartheta}} \right) + n \sum_j p_{\lambda_n} \left(|\boldsymbol{\vartheta}_j| \right), \tag{4.34}$$

Now we want to find the iterative algorithm derived for the other formulation. To this end let $\mathcal{F}(\boldsymbol{\vartheta})$ be the objective function in (4.34) and apply the first order Taylor approximation

$$\begin{aligned} \mathcal{F}(\boldsymbol{\vartheta}) \approx & \left(\frac{1}{H} \sum_{i=1}^{H} \tilde{\nabla}^i_{\boldsymbol{\vartheta}_0} \right)' \hat{\boldsymbol{\Omega}} \left(\frac{1}{H} \sum_{i=1}^{H} \tilde{\nabla}^i_{\boldsymbol{\vartheta}_0} \right) \\ & + 2 \frac{1}{H} \sum_{i=1}^{H} \frac{\partial \tilde{\nabla}^i_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} \hat{\boldsymbol{\Omega}} \left(\frac{1}{H} \sum_{i=1}^{H} \tilde{\nabla}^i_{\boldsymbol{\vartheta}_0} \right) (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) \\ & + (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)' \frac{1}{H} \sum_{i=1}^{H} \frac{\partial \tilde{\nabla}_i^{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} \hat{\boldsymbol{\Omega}} \frac{1}{H} \sum_{i=1}^{H} \frac{\partial \tilde{\nabla}_i^{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) \\ & + \frac{n}{2} \boldsymbol{\vartheta}' \bar{\mathbf{P}}_{\lambda_n} (\boldsymbol{\vartheta}_0) \boldsymbol{\vartheta}, \end{aligned} \tag{4.35}$$

where $\bar{\mathbf{P}}_{\lambda_n}(\boldsymbol{\vartheta}_0) = \operatorname{diag}\left\{ \mathbf{0}; \frac{p'_{\lambda_n}(|\boldsymbol{\vartheta}_{0i}|)}{|\boldsymbol{\vartheta}_{0i}|}; \boldsymbol{\vartheta}_{0i} \in \boldsymbol{\vartheta}_0^1 \right\}$.

In order to solve the minimisation problem in (4.34) we compute the first derivative of the objective function

$$\begin{aligned} \frac{\partial \mathcal{F}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \approx & \frac{1}{H} \sum_{i=1}^{H} \frac{\partial \tilde{\nabla}^i_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}'} \hat{\boldsymbol{\Omega}} \left(\frac{1}{H} \sum_{i=1}^{H} \tilde{\nabla}^i_{\boldsymbol{\vartheta}_0} \right) \\ & + \frac{1}{H} \sum_{i=1}^{H} \frac{\partial \tilde{\nabla}^i_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}'} \hat{\boldsymbol{\Omega}} \frac{1}{H} \sum_{i=1}^{H} \frac{\partial \tilde{\nabla}^i_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) \\ & + \frac{n}{2} \bar{\mathbf{P}}_{\lambda_n} (\boldsymbol{\vartheta}_0) (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) + \frac{n}{2} \bar{\mathbf{P}}_{\lambda_n} (\boldsymbol{\vartheta}_0) \boldsymbol{\vartheta}_{\mathbf{0}} = 0. \end{aligned} \tag{4.36}$$

Therefore, we get

$$\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0 - \left[\frac{1}{H}\sum_{i=1}^{H}\frac{\partial\tilde{\nabla}^i_{\boldsymbol{\vartheta}_0}}{\partial\boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\frac{1}{H}\sum_{i=1}^{H}\frac{\partial\tilde{\nabla}^i_{\boldsymbol{\vartheta}_0}}{\partial\boldsymbol{\vartheta}} + \frac{n}{2}\bar{\mathbf{P}}_{\lambda_n}\left(\boldsymbol{\vartheta}_0\right)\right]^{-1} \times \left[\frac{1}{H}\sum_{i=1}^{H}\frac{\partial\tilde{\nabla}^i_{\boldsymbol{\vartheta}_0}}{\partial\boldsymbol{\vartheta}'}\hat{\boldsymbol{\Omega}}\left(\frac{1}{H}\sum_{i=1}^{H}\tilde{\nabla}^i_{\boldsymbol{\vartheta}_0}\right) + \frac{n}{2}\bar{\mathbf{P}}_{\lambda_n}\left(\boldsymbol{\vartheta}_0\right)\boldsymbol{\vartheta}_0\right]. \quad (4.37)$$

Then we obtain the following iterative algorithm

$$\boldsymbol{\vartheta}^{(k+1)} = \boldsymbol{\vartheta}^{(k)} - \left[\frac{1}{H}\sum_{i=1}^{H}\frac{\partial\tilde{\nabla}^i_{\boldsymbol{\vartheta}^{(k)}}}{\partial\boldsymbol{\vartheta}'}\hat{\boldsymbol{\Omega}}\frac{1}{H}\sum_{i=1}^{H}\frac{\partial\tilde{\nabla}^i_{\boldsymbol{\vartheta}^{(k)}}}{\partial\boldsymbol{\vartheta}} + \frac{n}{2}\bar{\mathbf{P}}_{\lambda_n}\left(\boldsymbol{\vartheta}^{(k)}\right)\right]^{-1} \times \left[\frac{1}{H}\sum_{i=1}^{H}\frac{\partial\tilde{\nabla}^i_{\boldsymbol{\vartheta}^{(k)}}}{\partial\boldsymbol{\vartheta}'}\hat{\boldsymbol{\Omega}}\left(\frac{1}{H}\sum_{i=1}^{H}\tilde{\nabla}^i_{\boldsymbol{\vartheta}^{(k)}}\right) + \frac{n}{2}\bar{\mathbf{P}}_{\lambda_n}\left(\boldsymbol{\vartheta}^{(k)}\right)\boldsymbol{\vartheta}^{(k)}\right], \quad (4.38)$$

that when it reaches the convergence, it holds

$$\frac{1}{H}\sum_{i=1}^{H}\frac{\partial\tilde{\nabla}^i_{\boldsymbol{\vartheta}^{(k)}}}{\partial\boldsymbol{\vartheta}'}\hat{\boldsymbol{\Omega}}\left(\frac{1}{H}\sum_{i=1}^{H}\tilde{\nabla}^i_{\boldsymbol{\vartheta}^{(k)}}\right) + \frac{n}{2}\bar{\mathbf{P}}_{\lambda_n}\left(\boldsymbol{\vartheta}^{(k)}\right)\boldsymbol{\vartheta}^{(k)} = 0, \quad (4.39)$$

that is the first order conditions of (4.34).

4.5 Sparse linear regression model with Stable innovations

Let $\mathbf{y} = (y_1, y_2, \ldots, y_T)'$ be the vector of observations on the scalar response variable Y, $\mathbf{X} = (\mathbf{x}_1', \mathbf{x}_2', \ldots, \mathbf{x}_T')'$ is the $(T \times p)$ matrix of observations on the p covariates, i.e., $\mathbf{x}_{j,t} = (x_{j,1}, x_{j,2}, \ldots, x_{j,p})$ and consider the following regression model

$$\mathbf{y} = \boldsymbol{\iota}_T\gamma + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (4.40)$$

where $\boldsymbol{\iota}_T$ is the $T \times 1$ vector of unit elements, $\gamma \in \mathbb{R}$ denotes the parameter related to the intercept of the model, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ is the $p \times 1$ vector of regression parameters. The innovation terms $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T)$ in equation (4.40) is assumed to follow a Stable distribution, i.e., $\varepsilon_i \sim \mathcal{S}_\alpha (\lambda, 0, \sigma)$, for $i = 1, 2, \ldots, n$ where $\mathcal{S}_\alpha (\lambda, 0, \sigma)$ denotes the α–Stable distribution centred at zero with characteristic exponent $\alpha \in (0, 2)$, shape parameter $\lambda \in (-1, 1)$ and scale $\sigma > 0$. We further assume that the elements of the vector of innovations ε_i are independent, i.e., $\varepsilon_j \perp\!\!\!\perp \varepsilon_k$, for any $j \neq k$ and they are independent of $\mathbf{x}_l$, for $l = 1, 2, \ldots, p$. As auxiliary distribution we consider the Student–t regression model defined as in equation (4.40), with the only difference that the error term follows a Student–t distribution $\boldsymbol{\varepsilon} \sim \mathcal{T} (0, \sigma^2, \nu)$. The metric we consider is the L_2 distance between the score of the auxiliary distribution evaluated at the true $\mathbf{y}$ and simulated $\tilde{\mathbf{y}}$ data. The next proposition provides the score and the hessian matrix of the auxiliary Student–t regression model.

Proposition 31. *The log–likelihood function of the auxiliary Student–t regression model is*

$$\ell_t (y_t, \gamma, \boldsymbol{\beta}, \sigma, \nu) = \log \Gamma \left(\frac{\nu + 1}{2} \right) - \frac{1}{2} \log \left(\sigma^2 \right) - \log (\nu \pi) - \log \Gamma \left(\frac{\nu}{2} \right) - \frac{\nu + 1}{2} \log \left(1 + \frac{\| y_t - \mu_t \|_2^2}{\sigma^2 \nu} \right), \tag{4.41}$$

where $\mu_t = \gamma + \mathbf{x}_t' \boldsymbol{\beta}$. The score of the auxiliary distribution with respect to the parameters $\vartheta = (\gamma, \boldsymbol{\beta}, \sigma, \nu)$ is $\nabla_\vartheta = (\nabla_\gamma, \nabla_{\boldsymbol{\beta}}, \nabla_\sigma, \nabla_\nu)$, where $\nabla_\theta =$

$\sum_{t=1}^{T} \nabla_t^{\theta}$, *for* $\theta = \{\gamma, \boldsymbol{\beta}, \sigma, \nu\}$ *and*

$$\nabla_t^{\gamma}\left(\mathbf{y}, \mathbf{X}\right) = \frac{\left(\nu + 1\right)\left(y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\right)}{\sigma^2\nu + \|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2} \tag{4.42}$$

$$\nabla_t^{\beta}\left(\mathbf{y}, \mathbf{X}\right) = \frac{\left(\nu + 1\right)\mathbf{x}_t'\left(y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\right)}{\sigma^2\nu + \|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2} \tag{4.43}$$

$$\nabla_t^{\sigma}\left(\mathbf{y}, \mathbf{X}\right) = -\frac{1}{\sigma} + \frac{\left(\nu + 1\right)\|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2}{\sigma^3\nu + \sigma\|y_t - \gamma - \mathbf{x}_t\boldsymbol{\beta}\|_2^2} \tag{4.44}$$

$$\begin{aligned}\nabla_t^{\nu}\left(\mathbf{y}, \mathbf{X}\right) = {} & \frac{1}{2}\psi\left(\frac{\nu+1}{2}\right) - \frac{1}{2}\psi\left(\frac{\nu}{2}\right) - \frac{1}{2\nu} \\ & - \frac{1}{2}\log\left(1 + \frac{\|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2}{\sigma^2\nu}\right) \\ & + \frac{\nu+1}{2\nu}\frac{\|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|^2}{\sigma^2\nu + \|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|^2},\end{aligned} \tag{4.45}$$

where $\psi\left(\cdot\right)$ *denotes the digamma function. Now we compute the Hessian matrix of the auxiliary univariate Student–t distribution. The Hessian matrix for observation* $t = 1, 2, \ldots, T$ *is*

$$\nabla_t^{\vartheta,\vartheta} = \begin{pmatrix} \nabla_t^{\gamma,\gamma} & \nabla_t^{\gamma,\boldsymbol{\beta}} & \nabla_t^{\gamma,\sigma} & \nabla_t^{\gamma,\nu} \\ \nabla_t^{\gamma,\boldsymbol{\beta}} & \nabla_t^{\boldsymbol{\beta},\boldsymbol{\beta}} & \nabla_t^{\boldsymbol{\beta},\sigma} & \nabla_t^{\boldsymbol{\beta},\nu} \\ \nabla_t^{\gamma,\sigma} & \nabla_t^{\boldsymbol{\beta},\sigma} & \nabla_t^{\sigma,\sigma} & \nabla_t^{\sigma,\nu} \\ \nabla_t^{\gamma,\nu} & \nabla_t^{\boldsymbol{\beta},\nu} & \nabla_t^{\sigma,\nu} & \nabla_t^{\nu,\nu} \end{pmatrix}, \tag{4.46}$$

where the diagonal elements are

$$
\begin{aligned}
\frac{\partial \nabla_t^{\gamma}(\mathbf{y}, \mathbf{X})}{\partial \gamma} &= (\nu+1)\frac{-\left(\sigma^2\nu + \|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right) + 2\|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2}{\left(\sigma^2\nu + \|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)^2} \\
&= (\nu+1)\frac{-\sigma^2\nu + \|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2}{\left(\sigma^2\nu + \|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)^2} \qquad (4.47) \\
\frac{\partial \nabla_t^{\beta}(\mathbf{y}, \mathbf{X})}{\partial \boldsymbol{\beta}} &= (\nu+1)\frac{-\mathbf{x}_t\mathbf{x}_t'\sigma^2\nu + \mathbf{x}_t\mathbf{x}_t'\|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2}{\left(\sigma^2\nu + \|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)^2} \qquad (4.48) \\
\frac{\partial \nabla_t^{\sigma}(\mathbf{y}, \mathbf{X})}{\partial \sigma} &= \frac{1}{\sigma^2} - \frac{(\nu+1)\left(3\sigma^2\nu + \|y_t - \gamma - \mathbf{x}_t\boldsymbol{\beta}\|_2^2\right)\|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2}{\left(\sigma^3\nu + \|y_t - \gamma - \mathbf{x}_t\boldsymbol{\beta}\|_2^2\right)^2} \qquad (4.49)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \nabla_t^{\nu}(\mathbf{y}, \mathbf{X})}{\partial \nu} &= \frac{1}{4}\psi'\left(\frac{\nu+1}{2}\right) - \frac{1}{4}\psi'\left(\frac{\nu}{2}\right) + \frac{1}{2\nu^2} \\
&\quad + \frac{1}{2\nu}\frac{\|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2}{\left(\sigma^2\nu + \|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)} \\
&\quad - \frac{1}{2\nu^2}\frac{\|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2}{\left(\sigma^2\nu + \|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)} \\
&\quad - \frac{\nu+1}{2\nu}\frac{\|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2\sigma^2}{\left(\sigma^2\nu + \|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)^2}. \qquad (4.50)
\end{aligned}
$$

Now we compute the off–diagonal elements with respect to the parameter γ:

$$
\begin{aligned}
\frac{\partial \nabla_t^{\gamma}(\mathbf{y}, \mathbf{X})}{\partial \boldsymbol{\beta}} &= -(\nu+1)\frac{\mathbf{x}_t'\left(\sigma^2\nu - \|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)}{\left(\sigma^2\nu + \|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)^2} \qquad (4.51) \\
\frac{\partial \nabla_t^{\gamma}(\mathbf{y}, \mathbf{X})}{\partial \sigma} &= -2\nu\sigma(\nu+1)\frac{\left(y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\right)}{\left(\sigma^2\nu + \|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)^2} \qquad (4.52) \\
\frac{\partial \nabla_t^{\gamma}(\mathbf{y}, \mathbf{X})}{\partial \nu} &= \frac{y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}}{\sigma^2\nu + \|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2} - \frac{\sigma^2(\nu+1)\|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2}{\left(\sigma^2\nu + \|y_t - \gamma - \mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)^2}. \qquad (4.53)
\end{aligned}
$$

Now we compute the off–diagonal elements with respect to the parameter

$\boldsymbol{\beta}$:

$$\frac{\partial \nabla_t^{\beta}(\mathbf{y},\mathbf{X})}{\partial \sigma} = -2\nu\sigma(\nu+1)\frac{\mathbf{x}_t'(y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta})}{\left(\sigma^2\nu+\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)^2} \tag{4.54}$$

$$\frac{\partial \nabla_t^{\beta}(\mathbf{y},\mathbf{X})}{\partial \nu} = \frac{\mathbf{x}_t'(y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta})\left[\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2-\sigma^2\right]}{\left(\sigma^2\nu+\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)^2}. \tag{4.55}$$

Now we compute the off–diagonal elements with respect to the parameter σ:

$$\begin{aligned}\frac{\partial \nabla_t^{\sigma}(\mathbf{y},\mathbf{X})}{\partial \boldsymbol{\beta}} &= -2(\nu+1)\frac{-\mathbf{x}_t'(y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta})\left(\sigma^3\nu+\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2\sigma\right)}{\sigma^2\left(\sigma^2\nu+\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)^2}\\ &\quad+2(\nu+1)\frac{\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2\mathbf{x}_t'(y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta})\sigma}{\sigma^2\left(\sigma^2\nu+\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)^2}\\ &= -2\sigma\nu(\nu+1)\frac{\mathbf{x}_t'(y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta})}{\left(\sigma^2\nu+\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)^2}\end{aligned} \tag{4.56}$$

$$\frac{\partial \nabla_t^{\sigma}(\mathbf{y},\mathbf{X})}{\partial \nu} = \frac{1}{\sigma}\frac{\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2}{\sigma^2\nu+\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2} - (\nu+1)\sigma\frac{\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2}{\left(\sigma^2\nu+\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)^2}, \tag{4.57}$$

and the off–diagonal elements with respect to ν:

$$\frac{\partial \nabla_t^{\nu}(\mathbf{y},\mathbf{X})}{\partial \beta} = \frac{(y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta})\mathbf{x}_t'}{\sigma^2\nu+\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2} - \frac{(\nu+1)\sigma^2(y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta})\mathbf{x}_t'}{\left(\sigma^2\nu+\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)^2} \tag{4.58}$$

$$\begin{aligned}\frac{\partial \nabla_t^{\nu}(\mathbf{y},\mathbf{X})}{\partial \sigma} &= \frac{1}{\sigma}\frac{\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2}{\sigma^2\nu+\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2}\\ &\quad-(\nu+1)\sigma\frac{\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2}{\left(\sigma^2\nu+\|y_t-\gamma-\mathbf{x}_t'\boldsymbol{\beta}\|_2^2\right)^2}.\end{aligned} \tag{4.59}$$

4.5.1 Simulation experiment

In Table 4.1, we report the empirical inclusion probabilities of the regression parameters obtained over $1,00$ replications of the α–Stable regression model defined in equation (4.40), for two values of $\alpha = (1.70, 1.95)$ with $n = 250$. The true parameters are defined in the column (Par.) of Table (4.40), while the scale parameter of the Stable distribution is held fixed at $\sigma = 0.05$. Our

Par.	True	# of zero	# of zero	Par.	True	# of zero	# of zero
		$\alpha = 1.70$	$\alpha = 1.95$			$\alpha = 1.70$	$\alpha = 1.95$
γ	1	0	0	β_{11}	0	0.7500	0.8378
β_1	1	0	0	β_{12}	0	0.8182	0.9459
β_2	2	0	0	β_{13}	0	0.7273	0.9189
β_3	3	0	0	β_{14}	0	0.7955	0.9730
β_4	1	0	0	β_{15}	0	0.7273	0.8378
β_5	2	0	0	β_{16}	0	0.7727	0.8919
β_6	3	0	0	β_{17}	0	0.8182	0.9189
β_7	1	0	0	β_{18}	0	0.8636	0.9459
β_8	2	0	0	β_{19}	0	0.8636	1.0000
β_9	3	0	0	β_{20}	0	0.8636	0.9459
β_{10}	0	0.6591	0.8919				

Table 4.1: # of zero represents the number of time the corresponding parameter is estimated as zero evaluated over 1000 replications for the regression parameters $(\gamma, \boldsymbol{\beta}')$ of the α–Stable regression model defined in equation (4.40).

simulation results confirm that the sparse Indirect estimator perform well in detecting zeros in linear non–Gaussian regression models.

Chapter 5

Real data application

5.1 Introduction

The aim of this Section is to introduce some interesting real data examples where the ESD and its skewed extension SESD can be successfully employed. The availability of a simple and effective likelihood–free method to deal with parameters estimation in those circumstances where the density is not analytically amenable gives the chance to overcome traditional deficiencies of standard approaches, that usually rely on the simplifying assumptions of Gaussian or Student–t distributions, in order to account for the presence of fat–tailed and skewed data. One of those examples regards the allocation of wealth among different risky assets, a problem known in mathematics and financial engineering as portfolio optimisation. The next Section 5.2 deals with the problem of picking investment opportunities form a basket of alternative risky assets in order to satisfy some optimality conditions. Optimality criteria trade expected returns off the riskiness profile of the alternative investment opportunities and deliver portfolios that are characterised by the best risk–return profile, a strategy that dates back to the seminal paper of Markowitz (1952). Markowitz (1952) portfolio optimisation theory relies on the untrustworthy assumptions that either asset returns be normally distributed or the investors' utility functions be quadratic in the portfolio weights. Both assumptions lead to the choice of the variance as the natural candidate risk measure in an optimisation framework where investors maximise portfolios expected utilities under the constraint of matching a fixed level of overall risk measuring the returns uncertainty. The choice of the variance as measure of the risk associated with the investments represents one of the main drawbacks of the Markowitz's mean–variance methodology that limit its reliability. Indeed, the variance equally weights under and over performances, leading to optimal investment portfolios usually outperformed by other portfolios. Moreover, the first two moments fully describe returns characteristics only within a Gaussian world, whereas most recent empirical works document the inadequacy of the Normal distribution to capture stylised facts frequently observed in financial time series, see, e.g., McNeil et al. (2015). Recent financial literature have provided evidence that skewness and fat–tails strongly characterise the probabilistic behaviuor of stock returns (see, e.g., Kraus and Litzenberger 1976, Friend and Westerfield 1980, Barone-Adesi 1985). Moreover, the third and forth moments of asset returns play a relevant role also in the portfolio selection process. To this end, Dittmar (2002) and Scott and Horvath (1980) found that rational investors often prefer assets with higher skewness and lower kurtosis and Jondeau and Rockinger (2009) have reported empirical evidence that the mean–variance criterion may fail to approximate the constant relative risk aversion expected utility when assets are characterised by highly asymmetric and fat–tailed distributions.

Throughout the chapter, we make three main contributions to the existing literature on portfolio optimisation. First, we consider the multivariate Elliptical Stable distribution as data generating process for the assets returns. See also Aas and Haff (2006) and Paolella (2007) for a discussion of the alternative set of distributions for modelling skewed and heavy–tailed data. The second contribution concerns the inferential procedure that allows for sparse estimation of the scale matrix of the returns distribution. The third contribution instead consists to replace the variance by a downside risk measure, the Value–at–Risk (VaR), as measure of portfolio uncertainty. The Value–at–Risk at a given confidence level $\lambda \in (0, 1)$ measures the maximum loss in value of a portfolio with probability $(1-\lambda)$ over a predetermined time horizon. The Value–at–Risk can be calculated as the $(1-\lambda)$–quantile of the portfolio distribution. This quantile–based shortfall risk measure is commonly used and accepted in the financial literature, see, e.g., Jorion (2006) and references therein.

5.2 Portfolio optimisation

Without claiming to be complete, here we first introduce and formalise mathematically the portfolio optimisation problem we consider throughout the section, then we briefly detail how to calculate the risk profile of the investment portfolio followed by a description of the data employed. We conclude the section by presenting our major empirical findings. For further details about portfolio optimisation procedures we refer to the recent books of Jondeau et al. (2007), Roncalli (2014) and McNeil et al. (2015).

5.2.1 Portfolio optimisation problem

At each time t in the evaluation period, the investor's portfolio decision is based on the minimisation of the following loss function

$$\begin{aligned} &\arg\min_{\mathbf{w}_t} -\mathbb{E}_t\left(\mathbf{w}_t'\mathbf{Y}_{t+1}\right) + \kappa\rho_t\left(\mathbf{w}_t'\mathbf{Y}_{t+1}\right) \\ &\text{s.t.} \qquad \mathbf{w}_t \geq 0, \qquad \mathbf{w}_t'\mathbf{1} = 1, \end{aligned} \tag{5.1}$$

where $\mathbf{Y}_t \in \mathbb{R}^m$ is the vector of returns $\mathbf{w}_t \in \mathbb{R}^m$ denotes the vector of portfolio's weights at time t held by the investor over the period $[t, t+1)$, $\mathbb{E}_t\left(\mathbf{w}_t'\mathbf{Y}_{t+1}\right)$ and $\rho_t\left(\mathbf{w}_t'\mathbf{Y}_{t+1}\right)$ are the one–step ahead portfolio's expected return and risk measure evaluated at time t, respectively, and $\kappa \geq 0$ is the investor's risk aversion parameter. The investor objective function in equation (5.1) is a weighted average of the portfolio's one–step ahead expected return $\mathbb{E}_t\left(\mathbf{w}_t'\mathbf{Y}_{t+1}\right)$ and the portfolio's risk measured by the one–step ahead predictive risk measure $\rho_t\left(\mathbf{w}_t'\mathbf{Y}_{t+1}\right)$. This specification is close to the one adopted by Ahn et al. (1999) and Berkelaar and Kouwenberg (2000), see also Rockafellar and Uryasev (2000) for an alternative nonparametric approach based on Conditional Value–at–Risk. The two constraints for the portfolio weights ensure a full investment of the available budget and exclude short selling. Relaxing the constraint on short–selling can be expected to enhance the opportunities over a long–only active equity portfolio. In the empirical part, the short selling constraints on the weights $\mathbf{w}_t$ will be relaxed giving investors more flexible investment opportunities.

5.2.2 Portfolio risk measure

The portfolio's risk measure we consider in our empirical application is the Value–at–Risk, i.e., $\rho_t\left(\mathbf{w}_t'\mathbf{Y}_{t+1}\right) = -VaR_t^{\lambda}\left(\mathbf{w}_t'\mathbf{Y}_{t+1}\right)$. The Value–at–Risk for a risky asset at a given confidence level λ is the $(1-\lambda)$ −quantile of the distribution of the asset returns, and measures the minimum loss that can occur in the $(1-\lambda)\,100\%$ of the worst cases, see Jorion (2006). The Value–at–Risk measure provides useful insights on the probability of observing large negative payoffs implied by the estimated observation distribution. More formally, let $Z_{t+1} = \mathbf{w}_t'\mathbf{Y}_{t+1}$ be the scalar random variable that characterises the portfolio's return at time $t+1$, the Value–at–Risk of Z_{t+1} at a given confidence level λ, $VaR_t^{\lambda}\left(Z_{t+1}\right)$, is defined as the smallest number $z_0 \in \mathbb{R}$ such that the probability that the portfolio's return is less than the threshold z_0 is not smaller than $1-\lambda$, and it corresponds to $(1-\lambda)$–level quantile of the distribution of Z_{t+1}, i.e.

$$
\begin{aligned}
VaR_t^{\lambda}\left(Z_{t+1}\right) &= \inf\left\{z_0 : \mathbb{P}\left(Z_{t+1} < z_0\right) \geq 1-\lambda\right\} \\
&= \inf\left\{z_0 \in \mathbb{R} : F_{Z_{t+1}}\left(z_0\right) \geq 1-\lambda\right\} \\
&= F_{Z_{t+1}}^{-1}\left(1-\lambda\right), \qquad (5.2)
\end{aligned}
$$

where $F_{Z_{t+1}}(\cdot)$ is the cdf of Z_{t+1}, $F_{Z_{t+1}}^{-1}(\cdot)$ is the inverse function of $F_{Z_{t+1}}(\cdot)$ provided one exists, and the last equality holds for continuous distributions. Here, we assume that the risky assets $\mathbf{Y}_t$ follow the ESD or the SESD distribution, so that the closure property of those distributions with respect to linear combination stated in Theorem 9 is helpful in order to derive the VaR of the predicted portfolio returns. Moreover, Proposition 2.15 provides an easy way to evaluate the cdf of the SESD involved in equation (5.2). Specifically, Proposition 2.15 states that the cumulative density function of the univariate Skew Elliptical distribution can be calculated as the cumulative density function of a bivariate Elliptical Stable distribution whose parameters are appropriately modified in order to account for the skewness parameter of the original distribution. Proposition 2.15 extends to the Skew Elliptical Stable distribution the result proposed by Azzalini and Capitanio (2003) for the Skew Student–t distribution. Unfortunately, even the cdf of the Elliptical Stable distribution is not analytically available and we have to resort to numerical methods to integrate the latent Stable factor using the fast Fourier transform, see, Paolella (2007).

Before moving on to the next Section concerning the empirical application and results, it is worth mentioning that, in general, the Value–at–Risk measure is not a coherent risk measure because of the lack of the sub–additivity property, and, as a consequence, it does not incentive diversification, see, e.g., Artzner et al. (1999) and Acerbi and Tasche (2002). The lack of the sub–additivity property should discourage the use of the VaR in portfolio allocation problems where diversification is the major concern, in favour of alternative risk measures such as the Tail Conditional Expectation, that preserves that property, see, e.g., McNeil et al. (2015). However, it can easily proven that the VaR is sub–additive for Elliptical distributions. As concerns the Skew Elliptical Stable distributions the next proposition states that the the sub–additivity property is satisfied.

Proposition 32. *Without loss of generality, let us consider a bivariate Skew Elliptical Stable distribution with location and scale matrix equal to* $\boldsymbol{\xi}_2 = (\xi_1, \xi_2)' = (0, 0)'$, $\boldsymbol{\Omega}_2 = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, *respectively, i.e.,*

$$\mathbf{Y}_2 \sim \mathcal{SESD}_2\left(\alpha, \boldsymbol{\xi}_2, \boldsymbol{\Omega}_2, \boldsymbol{\lambda}_2\right). \tag{5.3}$$

A sufficient condition for the Skew Elliptical distribution to preserve the sub–additivity property of the VaR is that

$$VaR_\tau\left(Y^\star\right) \leq VaR_\tau\left(Y_1\right) + VaR_\tau\left(Y_2\right), \tag{5.4}$$

where $\tau \in (0,1)$ *is the VaR confidence level and* $Y^\star \sim \mathcal{SESD}_1\left(\alpha, 0, 1, \delta^\star\right)$, $\delta^\star = \frac{\delta_1+\delta_2}{\sqrt{2(1+\rho)}}$, $Y_j \sim \mathcal{SESD}_1\left(\alpha, 0, 1, \delta_j\right)$, $\delta_j = \frac{\lambda_j}{\sqrt{1+\lambda_j^2}}$, *for* $j = 1, 2$. *If* $\delta_1 = \delta_2 = \bar{\delta}$, *then the VaR is sub–additive.*

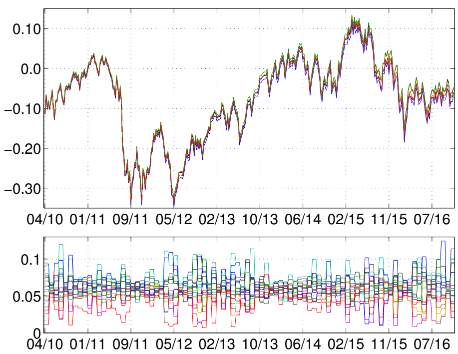
Proof. See Appendix A. □

5.2.3 Empirical application and results

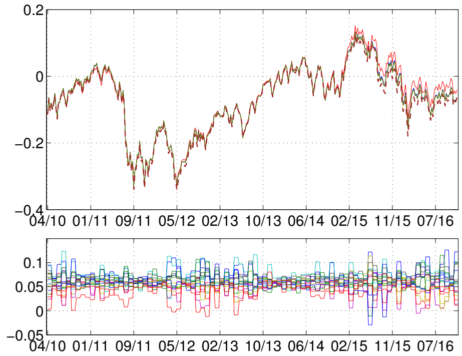
In our empirical application we consider a panel of MSCI European indexes. The basket consists on weekly returns of seventeen indexes, covering the period from January 6th, 1995 to November 25th, 2016. All the above mentioned return series have been downloaded from Datastream. Details about indexes' summary statistics over the whole sample are provided in Table F.1. Except for Portugal and Greece that experienced a dramatic decline during the global financial crisis and the subsequent European Sovereign debt crisis, annualised average returns, over the whole period, are positive and significant, ranging between 0.28% and 12.604% while annualised volatilities range between 21.174% and 35.733%. All the indexes are negatively skewed, suggesting that crashes occur more often than booms. Kurtosis measures are between 5.742 for Poland and 18.814 for Austria, a range that is not compatible with the Gaussian assumption. The dispersion of the Kurtosis measures across sectors suggests that European Stoxx Indexes are characterised by heterogeneous distribution patterns.The presence of heavy–tails is confirmed by the 1% weekly VaR. Moreover, the Jarque–Bera (JB) statistic confirms the departure from normality for all return series at the 1% level of significance. Regarding temporal dependence, as expected, we find no systematic evidence of serial correlation in market returns, but squared returns are strongly correlated, which suggests temporal variation in second moments. Turning to the multivariate characteristics of index returns, we notice that the correlation is the largest between France and Germany (0.890), while correlation is the lowest between Greece and Poland (0.359). From an unreported analysis, we also note that the correlations between the considered indexes have been much higher over the 2007–2009 financial crisis and much lower during the period of financial stability prior to 2007. We also note that, on average, correlations changes, on average, in a range of -31% to 25% between the two periods. Hence, the naive strategy is likely to overstate the diversification ability of the stock markets.

In order to perform our empirical analysis we estimated the ESD distribution using a rolling windows of $n = 800$ observations for the MMSQ and of $n = 200$ for the Sparse–MMSQ. The optimal tuning parameters of the SCAD penalty are selected by K–fold cross validation as described in Section 4.4.1, with $K = 5$. Parameters are re–estimated every four observations (about one month). As regards the portfolio allocation exercise, the forecasting horizon is set equal to $h = 1$, the VaR confidence level is fixed at $\lambda = 0.99$ and several levels of investors' risk aversion are considered $\kappa = \{0.10, 0.5, 1.0, 2.0, 5.0, 10.0\}$. For the sake of clarity results with larger or smaller levels of risk aversion are not reported but they are nearly indistinguishable from those corresponding to the corresponding nearest reported value.

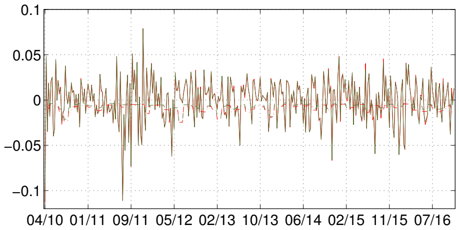
Now we turn to the description of our main findings about the empirical portfolio performance evaluation. To this end, we forecast the one–step ahead conditional returns' distribution over the whole sample period. The sequence of predictive distributions delivered by the competing models, are then used to build the mean–VaR optimal portfolios with an without the short selling constraint described in Section 5.2.1. The main objective of the portfolio application in neither to assess the validity of the ESD assumption for the distribution of the returns against alternatives, nor to gauge the effectiveness of the portfolio optimisation procedure that consider the Value–at–Risk as risk measure. The aim of this empirical exercise is instead that of comparing the results we obtain for the two estimation



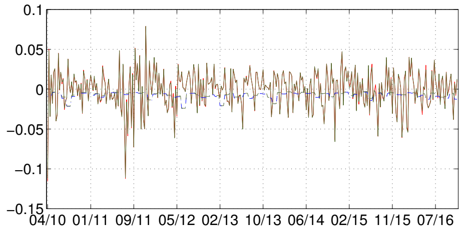
(a) Port. cum. ret. and weights (SS)



(b) Port. cum. ret. and weights



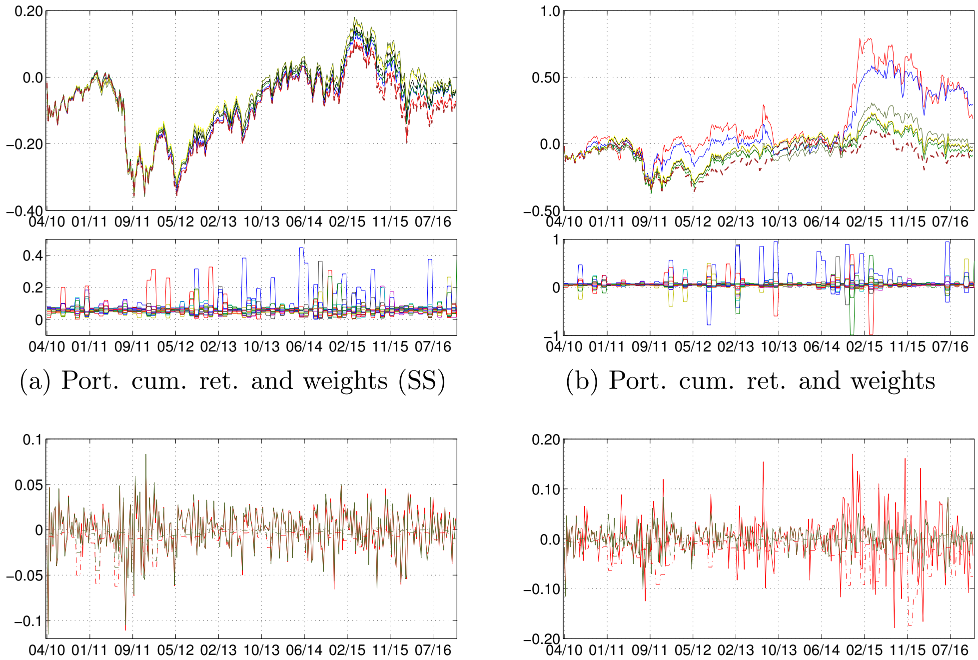
(c) Port. ret. (SS)



(d) Port. ret.

Figure 5.1: Mean–VaR$_{0.95}$ optimal portfolio results for the ESD distribution over the out–of–sample period from April 30th, 2010 to the end of the sampling period, November 25, 2016. ESD parameters are estimated by means of the MMSQ. Figures 5.1a and 5.1b plot the optimal portfolios cumulative returns with and without short selling constraints, respectively, for different values of the risk aversion parameter $\kappa = 0.1$ *(red)*, $\kappa = 0.5$ *(blue)*, $\kappa = 1$ *(green)*, $\kappa = 2$ *(black)*, $\kappa = 5$ *(yellow)*, $\kappa = 10$ *(olive)* in the top panel and the optimal portfolio weights for $\kappa = 10$ in the bottom panel. Figures 5.1c and 5.1d plot the optimal portfolio returns for $\kappa = 0.1$ *(red)*, and $\kappa = 10$ *(olive)*, with and without short selling constraints, respectively. The dotted thinned brown line represent the equally weighted portfolio cumulative returns which has been added for comparison.

methods, i.e., the MMSQ and the Sparse–MMSQ. To this end, Figures 5.1–5.2 depict the optimal portfolio decisions that are made over the last part of the sampling period form April 30th, 2010 to the end of the sample November 25, 2016. Figure 5.1 concerns the case where the model parameters are estimated by means of the MMSQ while Figure 5.2 presents results for its sparse counterpart. Let us consider the optimal portfolios with the short selling constraint, first. The most important evidence that emerges by inspecting Figures 5.1a–5.2a, reporting the cumulative returns



(a) Port. cum. ret. and weights (SS) (b) Port. cum. ret. and weights

(c) Port. ret. (SS) (d) Port. ret.

Figure 5.2: Mean–$\mathrm{VaR}_{0.95}$ optimal portfolio results for the ESD distribution over the out–of–sample period from April 30th, 2010 to the end of the sampling period, November 25, 2016. ESD parameters are estimated by means of the Sparse–MMSQ. Figures 5.1a and 5.1b plot the optimal portfolios cumulative returns with and without short selling constraints, respectively, for different values of the risk aversion parameter $\kappa = 0.1$ *(red)*, $\kappa = 0.5$ *(blue)*, $\kappa = 1$ *(green)*, $\kappa = 2$ *(black)*, $\kappa = 5$ *(yellow)*, $\kappa = 10$ *(olive)* in the top panel and the optimal portfolio weights for $\kappa = 10$ in the bottom panel. Figures 5.1c and 5.1d plot the optimal portfolio returns for $\kappa = 0.1$ *(red)*, and $\kappa = 10$ *(olive)*, with and without short selling constraints, respectively. The dotted thinned brown line represent the equally weighted portfolio cumulative returns which has been added for comparison.

for the MMSQ and the Sparse–MMSQ, is that, as the risk aversion coefficient κ increases from $\kappa = 0.1$ to $\kappa = 10$, cumulative returns increase as well. This evidence is stronger for the Sparse–MMSQ meaning that the shrinkage effect induced by the estimation method have a positive impact on the estimation of the scale matrix and, as a consequence, the portfolio results greatly benefited from a better estimate of the dependence structure among assets. The bottom panels of Figures 5.1a–5.2a plot the evolution over time of the optimal weights. Optimal weighs for the Sparse–MMSQ are characterised by a marked heterogenous behaviour while those implied by the MMSQ are flat and display lower levels of diversification. Figures 5.1c–5.2c plot the dynamic evolution of optimal portfolios returns for the highest $\kappa = 10$ and the lowest $\kappa = 0.1$ levels of risk aversion. Although not so evident, optimal portfolios returns that rely on parameters estimate from the Sparse–MMSQ have slightly lower variance than those built using the MMSQ. Indeed, visual inspection of Figures 5.1d–5.2d reveals that, when no short selling constraints are imposed the variability of optimal portfolios that rely on Sparse–MMSQ is incomparable lower than the optimal portfolio returns that rely on MMSQ, i.e., 0.34% versus 0.06%. Figure 5.2b concerning cumulative portfolio returns over the whole sample also confirms previous evidence.

Appendix A

Proofs of the main results

A.1 Chapter 2

Proof. Proposition 1.

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \frac{\int_{\mathbb{R}^+}\int_{\mathbb{R}} \phi_{d+1}\left((\mathbf{x}', y)', \mathbf{0}, \zeta \boldsymbol{\Omega}_\delta\right) h(\zeta)\, \mathbb{1}_{(0,\infty)}(y)\, dy d\zeta}{\int_{\mathbb{R}^+}\int_{\mathbb{R}^+} \phi(y, 0, \zeta)\, h(\zeta)\, dy d\zeta} \\
&= 2\int_0^{+\infty}\int_0^{+\infty} \frac{\zeta^{-\frac{d+1}{2}}}{(2\pi)^{\frac{d+1}{2}} |\boldsymbol{\Omega}_\delta|^{\frac{1}{2}}} \mathrm{e}^{-\frac{1}{2\zeta}(\mathbf{x}', y)\boldsymbol{\Omega}^{-1}(\mathbf{x}', y)'} h(\zeta)\, dy d\zeta \\
&= \frac{2}{(2\pi)^{\frac{d}{2}} |\bar{\boldsymbol{\Omega}}|^{\frac{1}{2}}} \int_0^{+\infty} \zeta^{-\frac{d}{2}} \mathrm{e}^{-\frac{\mathbf{x}'\bar{\boldsymbol{\Omega}}^{-1}\mathbf{x}}{2\zeta}} \\
&\quad \times \int_0^{+\infty} \sqrt{\frac{1}{2\pi\zeta\left(1-\boldsymbol{\delta}'\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\delta}\right)}} \mathrm{e}^{-\frac{\left(y-\boldsymbol{\delta}'\bar{\boldsymbol{\Omega}}^{-1}\mathbf{x}\right)^2}{2\zeta\left(1-\boldsymbol{\delta}'\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\delta}\right)}} h(\zeta)\, dy d\zeta. \qquad \text{(A.1)}
\end{aligned}
$$

Performing a simple change of variable, i.e., $t = \frac{\left(y - \boldsymbol{\delta}'\bar{\boldsymbol{\Omega}}^{-1}\mathbf{x}\right)}{\sqrt{\zeta\left(1-\boldsymbol{\delta}'\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\delta}\right)}}$, we get

$$\begin{aligned}
\int_0^{+\infty} \sqrt{\frac{1}{2\pi\zeta\left(1-\boldsymbol{\delta}'\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\delta}\right)}} \mathrm{e}^{-\frac{\left(y-\boldsymbol{\delta}'\bar{\boldsymbol{\Omega}}^{-1}\mathbf{x}\right)^2}{2\zeta\left(1-\boldsymbol{\delta}'\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\delta}\right)}} dy &= \int_{\frac{-\boldsymbol{\delta}'\bar{\boldsymbol{\Omega}}^{-1}\mathbf{x}}{\sqrt{\zeta\left(1-\boldsymbol{\delta}'\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\delta}\right)}}}^{+\infty} \phi\left(t\right) dt \\
&= \Phi\left(\frac{\boldsymbol{\delta}'\bar{\boldsymbol{\Omega}}^{-1}\mathbf{x}}{\sqrt{\zeta\left(1-\boldsymbol{\delta}'\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\delta}\right)}}\right) \\
&= \Phi\left(\frac{\boldsymbol{\lambda}'\mathbf{x}}{\sqrt{\zeta}}\right), \qquad \text{(A.2)}
\end{aligned}$$

where the last equality follows by applying equation (2.10). Substituting back equation (A.2) into (A.1), we obtain

$$f_{\mathbf{X}}\left(\mathbf{x}\right) = \frac{2}{\left(2\pi\right)^{\frac{d}{2}}\left|\bar{\boldsymbol{\Omega}}\right|^{\frac{1}{2}}} \int_0^{+\infty} \zeta^{-\frac{d}{2}} \mathrm{e}^{-\frac{\mathbf{x}'\bar{\boldsymbol{\Omega}}^{-1}\mathbf{x}}{2\zeta}} \Phi\left(\frac{\boldsymbol{\lambda}'\mathbf{x}}{\sqrt{\zeta}}\right) h\left(\zeta\right) d\zeta, \qquad \text{(A.3)}$$

which completes the proof. □

Proof. Lemma 7. The proof is similar to that in Branco and Dey (2001). Without loss of generality, assume $\boldsymbol{\xi} = \mathbf{0}$, then the moment generating function of the random variable $\mathbf{Y}$ is equal to

$$\begin{aligned}
M_{\mathbf{Y}}\left(\mathbf{t}\right) &= \mathbb{E}\left[e^{\mathbf{t}'\mathbf{Y}}\right] \\
&= 2\int_{\mathbb{R}^d} \exp\left\{\mathbf{t}'\mathbf{y}\right\} \int_0^{+\infty} \phi_d\left(\mathbf{y}, \mathbf{0}, \zeta\boldsymbol{\Omega}\right) \Phi_1\left(\frac{\boldsymbol{\lambda}'\boldsymbol{\omega}^{-1}\mathbf{y}}{\sqrt{\zeta}}\right) h\left(\zeta\right) d\zeta d\mathbf{y} \\
&= 2\int_{\mathbb{R}^d}\int_0^{+\infty} \frac{\exp\left\{-\frac{1}{2\zeta}\left[\mathbf{y}'\boldsymbol{\Omega}^{-1}\mathbf{y} - 2\zeta\mathbf{t}'\mathbf{y}\right]\right\}}{\left(2\pi\right)^{\frac{d}{2}}\left|\boldsymbol{\Omega}\right|^{\frac{1}{2}}\zeta^{\frac{d}{2}}} \\
&\quad \times \Phi_1\left(\frac{\boldsymbol{\lambda}'\boldsymbol{\omega}^{-1}\mathbf{y}}{\sqrt{\zeta}}\right) h\left(\zeta\right) d\zeta d\mathbf{y}.
\end{aligned}$$

Noting that

$$\left(\mathbf{y} - \zeta\boldsymbol{\Omega}\mathbf{t}\right)'\boldsymbol{\Omega}^{-1}\left(\mathbf{y} - \zeta\boldsymbol{\Omega}\mathbf{t}\right) = \mathbf{y}'\boldsymbol{\Omega}^{-1}\mathbf{y} + \zeta^2\mathbf{t}'\boldsymbol{\Omega}\mathbf{t} - 2\zeta\mathbf{t}'\mathbf{y}, \qquad \text{(A.4)}$$

then, after some algebraic manipulations and applying the Leibniz's rule for differentiation under the integral sign, the moment generating function can be written as

$$\begin{aligned}
M_{\mathbf{Y}}(\mathbf{t}) &= 2\int_0^{+\infty} \exp\left\{\frac{\zeta\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}}{2}\right\} \int_{\mathbb{R}^d} \frac{\Phi_1\left(\frac{\boldsymbol{\lambda}'\boldsymbol{\omega}^{-1}\mathbf{y}}{\sqrt{\zeta}}\right)}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Omega}|^{\frac{1}{2}}\zeta^{\frac{d}{2}}} \\
&\quad \times \exp\left\{-\frac{1}{2\zeta}(\mathbf{y}-\zeta\boldsymbol{\Omega}\mathbf{t})'\boldsymbol{\Omega}^{-1}(\mathbf{y}-\zeta\boldsymbol{\Omega}\mathbf{t})\right\} d\mathbf{y}d\zeta \\
&= 2\int_0^{+\infty} \exp\left\{\frac{\zeta\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}}{2}\right\} \int_{\mathbb{R}^d} \frac{\Phi_1\left(\frac{\boldsymbol{\lambda}'\boldsymbol{\omega}^{-1}(\mathbf{z}+\zeta\boldsymbol{\Omega}\mathbf{t})}{\sqrt{\zeta}}\right)}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Omega}|^{\frac{1}{2}}\zeta^{\frac{d}{2}}} \\
&\quad \times \exp\left\{-\frac{\mathbf{z}'\boldsymbol{\Omega}^{-1}\mathbf{z}}{2\zeta}\right\} d\mathbf{z}d\zeta \\
&= 2\int_0^{+\infty} \exp\left\{\frac{\zeta\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}}{2}\right\} \Phi_1\left(\frac{\sqrt{\zeta}\boldsymbol{\lambda}'\boldsymbol{\omega}^{-1}\boldsymbol{\Omega}\mathbf{t}}{\sqrt{1+\boldsymbol{\lambda}'\boldsymbol{\omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\omega}^{-1}\boldsymbol{\lambda}'}}\right) h(\zeta)\,d\zeta \qquad \text{(A.5)} \\
&= 2\int_0^{+\infty} \exp\left\{\frac{\zeta\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}}{2}\right\} \Phi_1\left(\frac{\sqrt{\zeta}\boldsymbol{\lambda}'\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}\mathbf{t}}{\sqrt{1+\boldsymbol{\lambda}'\bar{\boldsymbol{\Omega}}\boldsymbol{\lambda}'}}\right) h(\zeta)\,d\zeta \\
&= 2\int_0^{+\infty} \exp\left\{\frac{\zeta\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}}{2}\right\} \Phi_1\left(\sqrt{\zeta}\boldsymbol{\delta}'\boldsymbol{\omega}\mathbf{t}\right) h(\zeta)\,d\zeta, \qquad \text{(A.6)}
\end{aligned}$$

which completes the proof. Equation (A.5) has been obtained by applying Lemma 5.2 of Azzalini (2013). □

Proof. Theorem 8. Evaluating the moment generating function of $\mathbf{Y}$ in equation (A.7) at $\mathbf{t} = (\mathbf{s}', \mathbf{0}')'$ gives the moment generating function of $\mathbf{Y}_1$ as

$$M_{\mathbf{Y}_1}(\mathbf{s}) = \int_0^{\infty} M_{\mathbf{X}_1}\left(\sqrt{\zeta}\mathbf{s}\right) h(\zeta)\,d\zeta, \qquad \text{(A.7)}$$

where $M_{\mathbf{X}_1}\left(\sqrt{\zeta}\mathbf{s}\right)$ denotes the moment generating function of the Skew Normal distribution of the marginal vector $\mathbf{X}_1$ on partitioning $\mathbf{X}$ as $\mathbf{Y}$, and $\mathbf{X}_1 \mid \zeta \sim \mathcal{SN}_{d_1}(\boldsymbol{\xi}_1, \boldsymbol{\Omega}_{11}, \boldsymbol{\lambda}_1^*)$ with $\boldsymbol{\lambda}_1^*$ defined in equation (2.21), see Azzalini (2013). □

Proof. Theorem 9. The moment generating function of the random variable $\mathbf{Z} = \mathbf{d} + \mathbf{C}\mathbf{X}$ is equal to

$$\begin{aligned} M_{\mathbf{Z}}(\mathbf{t}) &= \exp\{\mathbf{t}'\mathbf{d}\}\,\mathbb{E}\left[e^{\mathbf{t}'(\mathbf{C}\mathbf{x})}\right] \\ &= 2\exp\{\mathbf{t}'\mathbf{d}\}\int_{\mathbb{R}^d}\exp\{\mathbf{t}'(\mathbf{C}\mathbf{x})\} \\ &\quad\times\int_0^{+\infty}\phi_d(\mathbf{x},\boldsymbol{\xi},\zeta\boldsymbol{\Omega})\,\Phi_1\left(\frac{\boldsymbol{\lambda}'\boldsymbol{\omega}^{-1}(\mathbf{x}-\boldsymbol{\xi})}{\sqrt{\zeta}}\right)h(\zeta)\,d\zeta d\mathbf{x}. \end{aligned} \tag{A.8}$$

Noting that

$$\begin{aligned} (\mathbf{y}-\boldsymbol{\xi}-\zeta\boldsymbol{\Omega}\mathbf{C}'\mathbf{t})'\,\boldsymbol{\Omega}^{-1}(\mathbf{y}-\boldsymbol{\xi}-\zeta\boldsymbol{\Omega}\mathbf{C}'\mathbf{t}) &= (\mathbf{y}-\boldsymbol{\xi})'\,\boldsymbol{\Omega}^{-1}(\mathbf{y}-\boldsymbol{\xi}) \\ &\quad+\zeta^2\mathbf{t}'\mathbf{C}\boldsymbol{\Omega}\mathbf{C}'\mathbf{t}-2\zeta\mathbf{t}'\mathbf{C}(\mathbf{y}-\boldsymbol{\xi}), \end{aligned} \tag{A.9}$$

then, after some algebraic manipulations and applying the Leibniz's rule for differentiation under the integral sign, the moment generating function can be written as

$$\begin{aligned} M_{\mathbf{Z}}(\mathbf{t}) = 2\exp\left\{\mathbf{t}'(\mathbf{d}+\mathbf{C}\boldsymbol{\xi})+\frac{\zeta\mathbf{t}'\mathbf{C}\boldsymbol{\Omega}\mathbf{C}'\mathbf{t}}{2}\right\}\int_0^{+\infty}\int_{\mathbb{R}^d}\frac{\Phi_1\left(\frac{\boldsymbol{\lambda}'\boldsymbol{\omega}^{-1}(\mathbf{x}-\boldsymbol{\xi})}{\sqrt{\zeta}}\right)}{(2\pi)^{\frac{d}{2}}\,|\boldsymbol{\Omega}|^{\frac{1}{2}}\zeta^{\frac{d}{2}}} \\ \times\exp\left\{-\frac{1}{2\zeta}(\mathbf{y}-\boldsymbol{\xi}-\zeta\boldsymbol{\Omega}\mathbf{C}'\mathbf{t})'\,\boldsymbol{\Omega}^{-1}(\mathbf{y}-\boldsymbol{\xi}-\zeta\boldsymbol{\Omega}\mathbf{C}'\mathbf{t})\right\}d\mathbf{y}h(\zeta)\,d\zeta. \end{aligned} \tag{A.10}$$

Rearranging terms, and considering the change of variable

$$\mathbf{u} = \boldsymbol{\Omega}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\xi}-\zeta\boldsymbol{\Omega}\mathbf{C}'\mathbf{t}),$$

we obtain

$$
\begin{aligned}
M_{\mathbf{Z}}(\mathbf{t}) &= 2\exp\left\{\mathbf{t}'(\mathbf{d}+\mathbf{C}\boldsymbol{\xi})+\frac{\zeta\mathbf{t}'\mathbf{C}\boldsymbol{\Omega}\mathbf{C}'\mathbf{t}}{2}\right\} \\
&\quad\times\int_0^{+\infty}\int_{\mathbb{R}^d}\Phi_1\left(\frac{\boldsymbol{\lambda}'\boldsymbol{\omega}^{-1}\left(\boldsymbol{\Omega}^{\frac{1}{2}}\mathbf{u}+\zeta\boldsymbol{\Omega}\mathbf{C}'\mathbf{t}\right)}{\sqrt{\zeta}}\right) \\
&\quad\times\frac{\exp\left\{-\frac{1}{2\zeta}\mathbf{u}'\mathbf{u}\right\}}{(2\pi\zeta)^{\frac{d}{2}}}d\mathbf{u}h(\zeta)\,d\zeta \\
&= 2\int_0^{+\infty}\exp\left\{\mathbf{t}'(\mathbf{d}+\mathbf{C}\boldsymbol{\xi})+\frac{\zeta\mathbf{t}'\mathbf{C}\boldsymbol{\Omega}\mathbf{C}'\mathbf{t}}{2}\right\} \\
&\quad\times\Phi_1\left(\frac{\boldsymbol{\lambda}'\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}\mathbf{C}'\sqrt{\zeta}\mathbf{t}}{\sqrt{1+\boldsymbol{\lambda}'\bar{\boldsymbol{\Omega}}\boldsymbol{\lambda}}}\right)h(\zeta)\,d\zeta \qquad\text{(A.11)} \\
&= 2\int_0^{+\infty}\exp\left\{\mathbf{t}'(\mathbf{d}+\mathbf{C}\boldsymbol{\xi})+\frac{\zeta\mathbf{t}'\mathbf{C}\boldsymbol{\Omega}\mathbf{C}'\mathbf{t}}{2}\right\}\Phi_1\left(\boldsymbol{\delta}'\boldsymbol{\omega}\mathbf{C}'\sqrt{\zeta}\mathbf{t}\right)h(\zeta)\,d\zeta, \qquad\text{(A.12)}
\end{aligned}
$$

which corresponds to the moment generating function of a SESD of dimension k with $\boldsymbol{\delta}_Z=\boldsymbol{\omega}_Z^{-1}\mathbf{C}\boldsymbol{\omega}\boldsymbol{\delta}$, where $\boldsymbol{\omega}_Z=(\boldsymbol{\Omega}_Z\odot I_h)^{\frac{1}{2}}$ and by applying equation (2.10) we get the expression for $\boldsymbol{\lambda}_Z$ in equation (2.25). Equation (A.11) has been obtained by applying Lemma 5.2 of Azzalini (2013). □

A.2 Chapter 3

Proof. Theorem 13.

(i) The proof of this result can be found in Cramér (1946).

(ii) We prove this part in a more general framework. We consider two confidence levels τ_1,τ_2 and two random variables Z_1,Z_2 and compute the asymptotic distribution of the vector of sample quantiles $\left(\hat{q}^{\tau_1,Z_1},\hat{q}^{\tau_2,Z_2}\right)$. If we choose $\tau_1=\tau_2$, $Z_1=\mathbf{u}_1'\mathbf{Y}$ and $Z_2=\mathbf{u}_2'\mathbf{Y}$ we get the result of the theorem.

Under the hypothesis of the theorem, the sample quantiles $\hat{q}^{\tau_1,Z_1}$ and

$\hat{q}^{\tau_2,Z_2}$ admit the Bahadur representation

$$\hat{q}^{\tau_j,Z_j} - q^{\tau_j,Z_j} = \frac{1}{n}\sum_{i=1}^{n} \frac{\tau_j - \mathbb{1}_{\left[z_{i,j} \le q^{\tau_j}\right]}}{f_{Z_j}\left(q^{\tau_j}\right)} + R_{n,j},$$

for $j = 1, 2$, where $R_{n,j} = o\left(\frac{1}{\sqrt{n}}\right)$. Let us start from the variance of $\hat{q}^{\tau_1,Z_1} - q^{\tau_1,Z_1}$.

$$\begin{aligned}
\operatorname{Var}\left(\hat{q}^{\tau_1,Z_1} - q^{\tau_1,Z_1}\right) &= \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\tau_1 - \mathbb{1}_{[z_{i,1} \le q^{\tau_1}]}}{f_{Z_1}\left(q^{\tau_1}\right)} + R_{n,1}\right) \\
&= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\tau_1 - \mathbb{1}_{[z_{i,1} \le q^{\tau_1}]}}{f_{Z_1}\left(q^{\tau_1}\right)} + R_{n,1}\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\tau_1 - \mathbb{1}_{[z_{i,1} \le q^{\tau_1}]}}{f_{Z_1}\left(q^{\tau_1}\right)}\right)^2\right. \\
&\qquad \left. +2R_{n,1}\frac{1}{n}\sum_{i=1}^{n} \frac{\tau_1 - \mathbb{1}_{[z_{i,1} \le q^{\tau_1}]}}{f_{Z_1}\left(q^{\tau_1}\right)} + R_{n,1}^2\right] \\
&= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\tau_1 - \mathbb{1}_{[z_{i,1} \le q^{\tau_1}]}}{f_{Z_1}\left(q^{\tau_1}\right)}\right)^2\right] \\
&\qquad + 2R_{n,1}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \frac{\tau_1 - \mathbb{1}_{[z_{i,1} \le q^{\tau_1}]}}{f_{Z_1}\left(q^{\tau_1}\right)}\right] + R_{n,1}^2 \\
&= \frac{1}{n^2 f_{Z_1}\left(q^{\tau_1}\right)^2}\mathbb{E}\left[\left(\sum_{i=1}^{n} \tau_1 - \mathbb{1}_{[z_{i,1} \le q^{\tau_1}]}\right)^2\right] \\
&\qquad + \frac{2R_{n,1}}{n f_{Z_1}\left(q^{\tau_1}\right)}\mathbb{E}\left[\sum_{i=1}^{n} \tau_1 - \mathbb{1}_{[z_{i,1} \le q^{\tau_1}]}\right] + R_{n,1}^2 \\
&= \frac{1}{n^2 f_{Z_1}\left(q^{\tau_1}\right)^2}\operatorname{Var}\left(\sum_{i=1}^{n} \tau_1 - \mathbb{1}_{[z_{i,1} \le q^{\tau_1}]}\right) + R_{n,1}^2 \\
&= \frac{\tau_1\left(1-\tau_1\right)}{f_{Z_1}\left(q^{\tau_1}\right)^2} + R_{n,1}^2,
\end{aligned}$$

where $R_{n,1}^2 = o\left(\frac{1}{n}\right)$. The same holds for the variance of $\hat{q}^{\tau_2,Z_2} - q^{\tau_2,Z_2}$.

Let us consider the covariance.

$$
\begin{aligned}
&\operatorname{Cov}\left(\hat{q}^{\tau_1,Z_1}-q^{\tau_1,Z_1},\hat{q}^{\tau_2,Z_2}-q^{\tau_2,Z_2}\right)\\
&=\operatorname{Cov}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_1-\mathbb{1}_{[z_{i,1}\leq q^{\tau_1}]}}{f\left(q^{\tau_1}\right)}+R_{n,1},\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_2-\mathbb{1}_{[z_{i,2}\leq q^{\tau_2}]}}{f\left(q^{\tau_2}\right)}R_{n,2}\right)\\
&=\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_1-\mathbb{1}_{[z_{i,1}\leq q^{\tau_1}]}}{f\left(q^{\tau_1}\right)}+R_{n,1}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_2-\mathbb{1}_{[z_{i,2}\leq q^{\tau_2}]}}{f\left(q^{\tau_2}\right)}+R_{n,2}\right)\right]\\
&=\frac{1}{n^2}\mathbb{E}\left[\sum_{i=1}^{n}\frac{\tau_1-\mathbb{1}_{[z_{i,1}\leq q^{\tau_1}]}}{f\left(q^{\tau_1}\right)}\sum_{i=1}^{n}\frac{\tau_2-\mathbb{1}_{[z_{i,2}\leq q^{\tau_2}]}}{f\left(q^{\tau_2}\right)}\right]\\
&\quad+\mathbb{E}\left[R_{n,1}\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_2-\mathbb{1}_{[z_{i,2}\leq q^{\tau_2}]}}{f\left(q^{\tau_2}\right)}\right]+\mathbb{E}\left[R_{n,2}\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_1-\mathbb{1}_{[z_{i,1}\leq q^{\tau_1}]}}{f\left(q^{\tau_1}\right)}\right]\\
&\quad+R_{n,1}R_{n,2}\\
&=\frac{1}{n^2}\mathbb{E}\left[\left(\frac{n\tau_1}{f\left(q^{\tau_1}\right)}-\sum_{i=1}^{n}\frac{\mathbb{1}_{[z_{i,1}\leq q^{\tau_1}]}}{f\left(q^{\tau_1}\right)}\right)\left(\frac{n\tau_2}{f\left(q^{\tau_2}\right)}-\sum_{i=1}^{n}\frac{\mathbb{1}_{[z_{i,2}\leq q^{\tau_2}]}}{f\left(q^{\tau_2}\right)}\right)\right]\\
&\quad+\frac{R_{n,1}}{nf\left(q^{\tau_2}\right)}\mathbb{E}\left[\sum_{i=1}^{n}\tau_2-\mathbb{1}_{[z_{i,2}\leq q^{\tau_2}]}\right]+\frac{R_{n,2}}{nf\left(q^{\tau_1}\right)}\mathbb{E}\left[\sum_{i=1}^{n}\tau_1-\mathbb{1}_{[z_{i,1}\leq q^{\tau_1}]}\right]\\
&\quad+R_{n,1}R_{n,2}\\
&=\frac{1}{n^2}\mathbb{E}\left[\frac{n\tau_1}{f\left(q^{\tau_1}\right)}\frac{n\tau_2}{f\left(q^{\tau_2}\right)}\right]-\frac{1}{n^2}\mathbb{E}\left[\frac{n\tau_1}{f\left(q^{\tau_1}\right)}\sum_{i=1}^{n}\frac{\mathbb{1}_{[z_{i,2}\leq q^{\tau_2}]}}{f\left(q^{\tau_2}\right)}\right]\\
&\quad-\frac{1}{n^2}\mathbb{E}\left[\frac{n\tau_2}{f\left(q^{\tau_2}\right)}\sum_{i=1}^{n}\frac{\mathbb{1}_{[z_{i,1}\leq q^{\tau_1}]}}{f\left(q^{\tau_1}\right)}\right]+\frac{1}{n^2}\mathbb{E}\left[\sum_{i=1}^{n}\frac{\mathbb{1}_{[z_{i,1}\leq q^{\tau_1}]}}{f\left(q^{\tau_1}\right)}\sum_{i=1}^{n}\frac{\mathbb{1}_{[z_{i,2}\leq q^{\tau_2}]}}{f\left(q^{\tau_2}\right)}\right]\\
&\quad+R_{n,1}R_{n,2}\\
&=\frac{\tau_1\tau_2}{f\left(q^{\tau_1}\right)f\left(q^{\tau_2}\right)}-\frac{\tau_1}{nf\left(q^{\tau_1}\right)f\left(q^{\tau_2}\right)}\mathbb{E}\left[\sum_{i=1}^{n}\mathbb{1}_{[z_{i,2}\leq q^{\tau_2}]}\right]\\
&\quad-\frac{\tau_2}{nf\left(q^{\tau_1}\right)f\left(q^{\tau_2}\right)}\mathbb{E}\left[\sum_{i=1}^{n}\mathbb{1}_{[z_{i,1}\leq q^{\tau_1}]}\right]\\
&\quad+\frac{1}{n^2f\left(q^{\tau_1}\right)f\left(q^{\tau_2}\right)}\mathbb{E}\left[\sum_{i=1}^{n}\mathbb{1}_{[z_{i,1}\leq q^{\tau_1}]}\sum_{i=1}^{n}\mathbb{1}_{[z_{i,2}\leq q^{\tau_2}]}\right]+R_{n,1}R_{n,2}
\end{aligned}
\tag{A.13}
$$

$$
\begin{aligned}
&= \frac{\tau_1\tau_2}{f(q^{\tau_1})f(q^{\tau_2})} - \frac{\tau_1}{f(q^{\tau_1})f(q^{\tau_2})}\mathbb{E}\left[\mathbb{1}_{[z_2\leq q^{\tau_2}]}\right] \\
&\quad - \frac{\tau_2}{f(q^{\tau_1})f(q^{\tau_2})}\mathbb{E}\left[\mathbb{1}_{[z_1\leq q^{\tau_1}]}\right] \\
&\quad + \frac{1}{f(q^{\tau_1})f(q^{\tau_2})}\mathbb{E}\left[\mathbb{1}_{[z_1\leq q^{\tau_1}]}\mathbb{1}_{[z_2\leq q^{\tau_2}]}\right] + R_{n,1}R_{n,2} \\
&= \frac{\tau_1\tau_2}{f(q^{\tau_1})f(q^{\tau_2})} - 2\frac{\tau_1\tau_2}{f(q^{\tau_1})f(q^{\tau_2})} + \frac{F_{Z_1,Z_2}\left(\mathbf{q}^{\tau},\boldsymbol{\Sigma}_{Z_1,Z_2}\right)}{f(q^{\tau_1})f(q^{\tau_2})} + R_{n,1}R_{n,2} \\
&= -\frac{\tau_1\tau_2}{f(q^{\tau_1})f(q^{\tau_2})} + \frac{F_{Z_1,Z_2}\left(\mathbf{q}^{\tau},\boldsymbol{\Sigma}_{Z_1,Z_2}\right)}{f(q^{\tau_1})f(q^{\tau_2})} + R_{n,1}R_{n,2},
\end{aligned}
$$

where $\mathbf{q}^{\tau} = (q^{\tau_1}, q^{\tau_2})'$ and $R_{n,1}R_{n,2} = o\left(\frac{1}{n}\right)$.

(iii) Using the Bahadur representation

$$
\begin{aligned}
&\operatorname{Cov}\left(\hat{q}^{\tau_k\mathbf{u}} - q^{\tau_k\mathbf{u}}, \hat{q}^{\tau_j\mathbf{u}} - q^{\tau_j\mathbf{u}}\right) \\
&= \operatorname{Cov}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_i - \mathbb{1}_{[z_i\leq q^{\tau_k\mathbf{u}}]}}{f(q^{\tau_k\mathbf{u}})} + R_{n,1}, \frac{1}{n}\sum_{i=1}^{n}\frac{\tau_j - \mathbb{1}_{\left[z_i\leq q^{\tau_j\mathbf{u}}\right]}}{f(q^{\tau_j\mathbf{u}})} + R_{n,2}\right) \\
&= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_k - \mathbb{1}_{[z_i\leq q^{\tau_k\mathbf{u}}]}}{f(q^{\tau_i\mathbf{u}})} + R_{n,1}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_j - \mathbb{1}_{\left[z_i\leq q^{\tau_j\mathbf{u}}\right]}}{f(q^{\tau_j\mathbf{u}})} + R_{n,2}\right)\right] \\
&= \frac{1}{n^2}\mathbb{E}\left[\sum_{i=1}^{n}\frac{\tau_k - \mathbb{1}_{[z_i\leq q^{\tau_k\mathbf{u}}]}}{f(q^{\tau_k\mathbf{u}})}\sum_{i=1}^{n}\frac{\tau_j - \mathbb{1}_{\left[z_i\leq q^{\tau_j\mathbf{u}}\right]}}{f(q^{\tau_j\mathbf{u}})}\right] \\
&\quad + R_{n,1}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_j - \mathbb{1}_{\left[z_i\leq q^{\tau_j\mathbf{u}}\right]}}{f(q^{\tau_j\mathbf{u}})}\right] + R_{n,2}\mathbb{E}\left[\sum_{i=1}^{n}\frac{\tau_k - \mathbb{1}_{[z_i\leq q^{\tau_k\mathbf{u}}]}}{f(q^{\tau_k\mathbf{u}})}\right] \\
&\quad + R_{n,1}R_{n,2} \\
&= \frac{1}{f(q^{\tau_k\mathbf{u}})f(q^{\tau_j\mathbf{u}})}\mathbb{E}\left[\left(\tau_k - \mathbb{1}_{[z_i\leq q^{\tau_k\mathbf{u}}]}\right)\left(\tau_j - \mathbb{1}_{\left[z_i\leq q^{\tau_j\mathbf{u}}\right]}\right)\right] + R_{n,1}R_{n,2} \\
&= \frac{1}{f(q^{\tau_k\mathbf{u}})f(q^{\tau_j\mathbf{u}})}\left(\tau_k\tau_j - \tau_k\mathbb{E}\left[\mathbb{1}_{\left[z_i\leq q^{\tau_j\mathbf{u}}\right]}\right] - \tau_j\mathbb{E}\left[\mathbb{1}_{[z_i\leq q^{\tau_k\mathbf{u}}]}\right]\right) \\
&\quad + \frac{1}{f(q^{\tau_k\mathbf{u}})f(q^{\tau_j\mathbf{u}})}\left(\mathbb{E}\left[\mathbb{1}_{[z_i\leq q^{\tau_k\mathbf{u}}]}\mathbb{1}_{\left[z_i\leq q^{\tau_j\mathbf{u}}\right]}\right]\right) + R_{n,1}R_{n,2} \\
&= \frac{\tau_k\wedge\tau j - \tau_k\tau_j}{f(q^{\tau_k\mathbf{u}})f(q^{\tau_j\mathbf{u}})} + R_{n,1}R_{n,2}.
\end{aligned}
$$

□

Proof. Theorem 19. The function $\hat{\mathbf{\Phi}}$ is assumed to be continuously differentiable, so Delta method applies

$$\hat{\mathbf{\Phi}} \approx \mathbf{\Phi}_{\boldsymbol{\vartheta}} + \frac{\partial \mathbf{\Phi}_{\boldsymbol{\vartheta}}}{\partial \mathbf{q}} \left(\hat{\mathbf{q}} - \mathbf{q}_{\boldsymbol{\vartheta}} \right), \tag{A.14}$$

then

$$\begin{aligned} \mathrm{Var}\left(\hat{\mathbf{\Phi}} \right) &\approx \mathrm{Var}\left(\frac{\partial \mathbf{\Phi}_{\boldsymbol{\vartheta}}}{\partial \mathbf{q}} \hat{\mathbf{q}} \right) \\ &= \frac{\partial \mathbf{\Phi}_{\boldsymbol{\vartheta}}}{\partial \mathbf{q}'} \mathrm{Cov}\left(\hat{\mathbf{q}} \right) \frac{\partial \mathbf{\Phi}_{\boldsymbol{\vartheta}}}{\partial \mathbf{q}}, \end{aligned} \tag{A.15}$$

where $\hat{\mathbf{q}} = (\hat{\mathbf{q}}^{\boldsymbol{\tau}_1 \mathbf{u}_1}, \ldots, \hat{\mathbf{q}}^{\boldsymbol{\tau}_K \mathbf{u}_K})$ and $\mathbf{q}_{\boldsymbol{\vartheta}} = (\mathbf{q}_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}_1 \mathbf{u}_1}, \ldots, \mathbf{q}_{\boldsymbol{\vartheta}}^{\boldsymbol{\tau}_K \mathbf{u}_K})$. □

Proof. Theorem 20. The first order condition of (3.5) is

$$\frac{1}{R} \sum_{r=1}^{R} \frac{\partial \tilde{\mathbf{\Phi}}_{\boldsymbol{\vartheta}}^{r}}{\partial \boldsymbol{\vartheta}} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \left(\hat{\mathbf{\Phi}}_{\boldsymbol{\vartheta}} - \frac{1}{R} \sum_{r=1}^{R} \tilde{\mathbf{\Phi}}_{\boldsymbol{\vartheta}}^{r} \right) = 0, \tag{A.16}$$

where $\bar{\boldsymbol{\vartheta}}$ is a consistent estimate of $\boldsymbol{\vartheta}$. Let us consider the first order Taylor expansion around the true parameter $\boldsymbol{\vartheta}_0$

$$\begin{aligned} &\frac{1}{R} \sum_{r=1}^{R} \frac{\partial \tilde{\mathbf{\Phi}}_{\boldsymbol{\vartheta}_0}^{r}}{\partial \boldsymbol{\vartheta}} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \left(\hat{\mathbf{\Phi}}_{\boldsymbol{\vartheta}} - \frac{1}{R} \sum_{r=1}^{R} \tilde{\mathbf{\Phi}}_{\boldsymbol{\vartheta}_0}^{r} \right) \\ &\quad - \frac{1}{R} \sum_{r=1}^{R} \frac{\partial \tilde{\mathbf{\Phi}}_{\boldsymbol{\vartheta}_0}^{r}}{\partial \boldsymbol{\vartheta}} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \frac{1}{R} \sum_{r=1}^{R} \frac{\partial \tilde{\mathbf{\Phi}}_{\boldsymbol{\vartheta}_0}^{r}}{\partial \boldsymbol{\vartheta}} \left(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0 \right) = o_p\left(1\right). \end{aligned} \tag{A.17}$$

From this equation we get

$$\sqrt{n} \left(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0 \right) \approx \left(\frac{\partial \tilde{\mathbf{\Phi}}_{\boldsymbol{\vartheta}}'}{\partial \boldsymbol{\vartheta}} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \frac{\partial \tilde{\mathbf{\Phi}}_{\boldsymbol{\vartheta}}}{\partial \boldsymbol{\vartheta}} \right)^{-1} \frac{\partial \tilde{\mathbf{\Phi}}_{\boldsymbol{\vartheta}}}{\partial \boldsymbol{\vartheta}} \mathbf{W}_{\bar{\boldsymbol{\vartheta}}} \sqrt{n} \left(\hat{\mathbf{\Phi}} - \frac{1}{R} \sum_{r=1}^{R} \tilde{\mathbf{\Phi}}_{\boldsymbol{\vartheta}_0}^{r} \right), \tag{A.18}$$

as $n \to \infty$. From Theorem 19

$$\sqrt{n}\left(\hat{\mathbf{\Phi}} - \frac{1}{R}\sum_{r=1}^{R}\tilde{\mathbf{\Phi}}^{r}_{\boldsymbol{\vartheta}_0}\right) \to^{d} \mathcal{N}\left(\mathbf{0}, \left(1+\frac{1}{R}\right)\mathbf{\Omega}_{\boldsymbol{\vartheta}}\right), \tag{A.19}$$

as $n \to \infty$, and $\tilde{\mathbf{\Phi}}^{r}_{\boldsymbol{\vartheta}_0}$ converges to $\mathbf{\Phi}_{\boldsymbol{\vartheta}}$. Moreover since $\bar{\boldsymbol{\vartheta}}$ is consistent the matrix $\mathbf{W}_{\bar{\boldsymbol{\vartheta}}}$ converges to $\mathbf{W}_{\boldsymbol{\vartheta}}$. From these results we get

$$\mathrm{Var}\left(\sqrt{n}\left(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}\right)\right) \to \left(1+\frac{1}{R}\right)\left[\mathbf{H}_{\vartheta}^{-1}\frac{\partial \mathbf{\Phi}_{\boldsymbol{\vartheta}}}{\partial \boldsymbol{\vartheta}}\right]\mathbf{W}_{\boldsymbol{\vartheta}}\mathbf{\Omega}_{\boldsymbol{\vartheta}}\mathbf{W}_{\boldsymbol{\vartheta}}'\left[\mathbf{H}_{\boldsymbol{\vartheta}}^{-1}\frac{\partial \mathbf{\Phi}_{\boldsymbol{\vartheta}}}{\partial \boldsymbol{\vartheta}}\right]', \tag{A.20}$$

as $n \to \infty$, where $\mathbf{H}_{\boldsymbol{\vartheta}} = \frac{\partial \mathbf{\Phi}_{\boldsymbol{\vartheta}}}{\partial \boldsymbol{\vartheta}'}\mathbf{W}_{\boldsymbol{\vartheta}}\frac{\partial \mathbf{\Phi}_{\boldsymbol{\vartheta}}}{\partial \boldsymbol{\vartheta}}$. □

Proof. Theorem 23. We prove this theorem following Fan and Li (2001) and Gao and Massam (2015). In the following we denote by σ_{ij}^0 and σ_{ij} respectively the zero and non zero off–diagonal elements of the variance covariance matrix.
Let us consider a ball $\|\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\| \leq Mn^{-\frac{1}{2}}$ for some finite constant M. In order to prove the result in equation (A.25), let us consider the first order condition of equation (3.8) and its first order taylor expansion

$$\begin{aligned}\frac{\partial \mathcal{Q}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} &= -2\frac{\partial \tilde{\mathbf{\Phi}}^{R}_{\boldsymbol{\vartheta}}}{\partial \boldsymbol{\vartheta}}\mathbf{W}_{\boldsymbol{\vartheta}}\left(\hat{\mathbf{\Phi}} - \tilde{\mathbf{\Phi}}^{R}_{\boldsymbol{\vartheta}}\right) + n\mathbf{v} \\ &\approx -2\frac{\partial \tilde{\mathbf{\Phi}}^{R}_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}\mathbf{W}_{\boldsymbol{\vartheta}}\left(\hat{\mathbf{\Phi}} - \tilde{\mathbf{\Phi}}^{R}_{\boldsymbol{\vartheta}_0}\right) + 2\frac{\partial \tilde{\mathbf{\Phi}}^{R}_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}'}\mathbf{W}_{\boldsymbol{\vartheta}}\frac{\partial \tilde{\mathbf{\Phi}}^{R}_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\right) + n\mathbf{v},\end{aligned} \tag{A.21}$$

where $\mathbf{v} = \left(\mathbf{0}; p'_{\lambda_n}\left(|\sigma_{ij}|\right)\mathrm{sgn}\left(\sigma_{ij}\right), i<j\right)$. The first two terms are $\mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. Regarding the penalisation term, let us first consider the zero off–diagonal element σ_{ij}^0. For a given λ_n, the first derivative $p'_{\lambda_n}\left(|\sigma_{ij}|\right)$ with respect to

$|\sigma_{ij}|$ is given by

$$p'_{\lambda_n}(|\sigma_{ij}|) = \begin{cases} \lambda_n & \text{if } |\sigma_{ij}| \leq \lambda_n \\ \frac{(a\lambda_n - |\sigma_{ij}|)}{a-1} & \text{if } \lambda_n < |\sigma_{ij}| \leq a\lambda_n \\ 0 & \text{if } a\lambda_n < |\sigma_{ij}|, \end{cases} \tag{A.22}$$

and it holds

$$\lim_{|\sigma_{ij}|\to 0} \frac{p'_{\lambda_n}(|\sigma_{ij}|)}{\lambda_n} = 1. \tag{A.23}$$

Then, for a generic σ^0_{ij}, the corresponding element in $n\mathbf{v}$ can be written as

$$n\lambda_n \mathrm{sgn}(\sigma_{ij}) \frac{p'_{\lambda_n}(|\sigma_{ij}|)}{\lambda_n} = n\lambda_n \mathrm{sgn}(\sigma_{ij}). \tag{A.24}$$

We rewrite (A.21) as follows

$$\frac{\partial \mathcal{Q}(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = n\lambda_n \left\{ \lambda_n^{-1}\mathbf{v} - \mathcal{O}_p\left(n^{-\frac{n}{2}}\lambda_n^{-1}\right), \right\} \tag{A.25}$$

Since $\liminf_{n\to\infty} \liminf_{|\sigma_{ij}|\to 0} \frac{p'_{\lambda_n}(|\sigma_{ij}|)}{\lambda_n} > 0$ and $\sqrt{n}\lambda_n \to \infty$, the term $n\mathbf{v}$ has asymptotic order higher that $\mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$ and dominates the equation (A.25). This means that the sign of $\frac{\partial \mathcal{Q}(\boldsymbol{\vartheta})}{\partial \sigma_{ij}}$ is determined by the sign of σ_{ij}, i.e. for any local minimiser it holds $\hat{\sigma}_{i,j} = 0$ with probability 1. Now consider the case in which σ_{ij} is not a zero element, then using the Taylor approximation we can calculate the following

$$\begin{aligned}
\mathcal{Q}(\boldsymbol{\vartheta}_0) - \mathcal{Q}(\boldsymbol{\vartheta}) &= \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}\right)' \mathbf{W}_{\boldsymbol{\vartheta}_0} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}\right) - \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}}\right)' \mathbf{W}_{\boldsymbol{\vartheta}} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}}\right) \\
&\quad + n \sum_{i<j} \left[p_\lambda\left(|\sigma^0_{ij}|\right) - p_\lambda\left(|\sigma_{ij}|\right)\right] \\
&\approx 2 \frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} \mathbf{W}_{\boldsymbol{\vartheta}_0} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}\right) (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) \\
&\quad + (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)' \left[-2 \frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}'} \mathbf{W}_{\boldsymbol{\vartheta}_0} \frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}\right] (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) \\
&\quad - n \sum_{i<j} p'_{\lambda_n}\left(|\sigma_{ij}|\right) \operatorname{sgn}\left(\sigma_{ij}\right) \left(\sigma_{ij} - \sigma^0_{ij}\right) \\
&\quad - n \sum_{i<j} p''_{\lambda_n}\left(|\sigma_{ij}|\right) \left(\sigma_{ij} - \sigma^0_{ij}\right)^2,
\end{aligned}$$

where $p''_{\lambda_n}(|\sigma_{ij}|)$ stands for the second derivative. For n large enough the summation term in equation (A.26) is negligible since $\sigma_{ij} \neq 0$ and

$$\begin{aligned}
&\lim_{n\to\infty} p'_{\lambda_n}\left(|\sigma_{ij}|\right) = 0 \\
&\lim_{n\to\infty} p''_{\lambda_n}\left(|\sigma_{ij}|\right) = 0.
\end{aligned} \tag{A.26}$$

The same holds for the fist term. The matrix

$$-2 \frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}'} \mathbf{W}_{\boldsymbol{\vartheta}_0} \frac{\partial \tilde{\boldsymbol{\Phi}}^R_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}, \tag{A.27}$$

is negative definite and for n large it dominates the other terms, therefore $\mathcal{Q}(\boldsymbol{\vartheta}_0) - \mathcal{Q}(\boldsymbol{\vartheta}) \leq 0$. This implies that there exist a local minimizer $\hat{\boldsymbol{\vartheta}}$ such that $\|\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. □

Proof. Theorem 24. Let us consider the first order Taylor expansion with respect to $\boldsymbol{\vartheta}_0^1$ of the first order condition computed in equation (A.26)

$$\begin{aligned}
\frac{\partial Q(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}^1} &= -2\frac{\partial \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^R}{\partial \boldsymbol{\vartheta}^1}\mathbf{W}_{\boldsymbol{\vartheta}^1}\left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}^R\right) + n\mathbf{v} \\
&= -2\frac{\partial \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}^R}{\partial \boldsymbol{\vartheta}^1}\mathbf{W}_{\boldsymbol{\vartheta}_0^1}\left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}^R\right) + 2\left(\frac{\partial \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}^R}{\partial \boldsymbol{\vartheta}^{1\prime}}\mathbf{W}_{\boldsymbol{\vartheta}_0^1}\frac{\partial \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}^R}{\partial \boldsymbol{\vartheta}^1}\right)\left(\boldsymbol{\vartheta}^1 - \boldsymbol{\vartheta}_0^1\right) \\
&\quad + n\mathbf{v}_0 + n\mathbf{P}_0\left(\boldsymbol{\vartheta}^1 - \boldsymbol{\vartheta}_0^1\right) = 0,
\end{aligned} \tag{A.28}$$

where $\mathbf{v} = \left(\mathbf{0}; p'_{\lambda_n}(|\sigma_{ij}|)\,\mathrm{sgn}(\sigma_{ij}), i < j\right)$ and $\mathbf{v}_0$ is $\mathbf{v}$ computed at the true value of the variance covariance matrix; $\mathbf{P} = \mathrm{diag}\left\{\mathbf{0}, p''_{\lambda_n}(|\sigma_{ij}|), i < j\right\}$ and $\mathbf{P}_0$ is $\mathbf{P}$ computed at the true parameter of the variance covariance matrix.

$$\begin{aligned}
&2\left(\frac{\partial \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}^R}{\partial \boldsymbol{\vartheta}^{1\prime}}\mathbf{W}_{\boldsymbol{\vartheta}_0^1}\frac{\partial \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}^R}{\partial \boldsymbol{\vartheta}^1}\right)\left(\boldsymbol{\vartheta}^1 - \boldsymbol{\vartheta}_0^1\right) + n\mathbf{v} + n\mathbf{P}\left(\boldsymbol{\vartheta}^1 - \boldsymbol{\vartheta}_0^1\right) \\
&= 2\frac{\partial \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}^R}{\partial \boldsymbol{\vartheta}^1}\mathbf{W}_{\boldsymbol{\vartheta}_0^1}\left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}^R\right)\sqrt{n}\left[2\left(\frac{\partial \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}^R}{\partial \boldsymbol{\vartheta}^{1\prime}}\mathbf{W}_{\boldsymbol{\vartheta}_0^1}\frac{\partial \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}^R}{\partial \boldsymbol{\vartheta}^1}\right) + n\mathbf{P}_0\right] \\
&\quad \times \left\{\boldsymbol{\vartheta}^1 - \boldsymbol{\vartheta}_0^1 + \left[2\left(\frac{\partial \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}^R}{\partial \boldsymbol{\vartheta}^{1\prime}}\mathbf{W}_{\boldsymbol{\vartheta}_0^1}\frac{\partial \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}^R}{\partial \boldsymbol{\vartheta}^1}\right) + n\mathbf{P}_0\right]^{-1} n\mathbf{v}_0\right\} \\
&= 2\frac{\partial \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}^R}{\partial \boldsymbol{\vartheta}^1}\mathbf{W}_{\boldsymbol{\vartheta}_0^1}\sqrt{n}\left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}^R\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}^1}\mathbf{W}_{\boldsymbol{\vartheta}_0^1}\boldsymbol{\Omega}_{\boldsymbol{\vartheta}_0}\mathbf{W}'_{\boldsymbol{\vartheta}^1}\frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}^{1\prime}}\right).
\end{aligned} \tag{A.29}$$

Since $\mathbf{v}_0$ and $\mathbf{P}_0$ vanish asymptotically, we apply the same argument of Theorem 20 to complete the proof. □

A.3 Chapter 4

Proof. Theorem 29. Let us consider a ball $\|\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$ where $m < \infty$, and let us compute the first order condition of (4.7)

$$\frac{\partial \mathcal{D}}{\partial \boldsymbol{\vartheta}} = -\frac{1}{H}\sum_{i=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}_{\boldsymbol{\vartheta}}^{h}}{\partial \boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{i=1}^{H}\tilde{\boldsymbol{\beta}}_{\boldsymbol{\vartheta}}^{h}\right) + n\boldsymbol{v} \tag{A.30}$$

where $\boldsymbol{v}_i = \frac{\partial p_\lambda}{\boldsymbol{\vartheta}_i}\left(|\boldsymbol{\vartheta}_i|\right)$. Now we consider the taylor expansion

$$\begin{aligned}\frac{\partial \mathcal{D}}{\partial \boldsymbol{\vartheta}} = &-\frac{1}{H}\sum_{i=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}_{\boldsymbol{\vartheta}_0}^{h}}{\partial \boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{i=1}^{H}\tilde{\boldsymbol{\beta}}_{\boldsymbol{\vartheta}_0}^{h}\right) + \\ &+ \frac{1}{H}\sum_{i=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}_{\boldsymbol{\vartheta}_0}^{h}}{\partial \boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\frac{1}{H}\sum_{i=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}_{\boldsymbol{\vartheta}_0}^{h}}{\partial \boldsymbol{\vartheta}}\left(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0\right) + n\boldsymbol{v}_0\end{aligned} \tag{A.31}$$

where $\boldsymbol{v}_{i0} = \frac{\partial p_\lambda}{\boldsymbol{\vartheta}_{i0}}\left(|\boldsymbol{\vartheta}_i|\right)$. From the asymptotic properties of the indirect inference estimator, it holds that the first two terms are of order $\mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. Looking at the penalisation term, let us first consider the zero elements $\vartheta_i^0 \in \boldsymbol{\vartheta}^0$. For a given λ_n the first derivative $p'_{\lambda_n}\left(|\vartheta_i|\right)$ is given by

$$p'_{\lambda_n}\left(|\vartheta_i|\right) = \begin{cases}\lambda_n & \text{if } |\vartheta_i| \le \lambda_n\\ \frac{(a\lambda_n - |\vartheta_i|)}{a-1} & \text{if } \lambda_n < |\vartheta_i| \le a\lambda_n\\ 0 & \text{if } a\lambda_n < |\vartheta_i|,\end{cases} \tag{A.32}$$

and it holds

$$\lim_{|\vartheta_i|\to 0}\frac{p'_{\lambda_n}\left(|\vartheta_i|\right)}{\lambda_n} = 1. \tag{A.33}$$

Then for a generic $\vartheta_i^0 \in \boldsymbol{\vartheta}^0$ the corresponding element in $n\boldsymbol{v}$ can be written as

$$n\lambda_n \mathrm{sgn}\left(\vartheta_i\right)\frac{p'_{\lambda_n}\left(|\vartheta_i|\right)}{\lambda_n} = n\lambda_n \mathrm{sgn}\left(\vartheta_i\right) \tag{A.34}$$

Then we can write

$$\frac{\partial \mathcal{D}}{\partial \boldsymbol{\vartheta}} = n\lambda_n \left(\lambda_n^{-1} \boldsymbol{v}_0 - \mathcal{O}_p \left(n^{-\frac{1}{2}} \lambda_n^{-1} \right) \right), \tag{A.35}$$

since the $\liminf_{n\to\infty} \liminf_{|\vartheta_i|\to 0} \frac{p'_{\lambda_n}(|\vartheta_i|)}{\lambda_n} > 0$ and $\sqrt{n}\lambda_n \to \infty$ the term $n\boldsymbol{v}_0$ has asymptotic order higher than $\mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$ and dominates the equation. This means that the sign of $\frac{\partial \mathcal{D}}{\partial \boldsymbol{\vartheta}}$ is determined by the sign of ϑ_i, then for any local minimizer it holds that $\hat{\vartheta}_i = 0$ with probability one.
Now let us consider the case in which $\vartheta_i \in \boldsymbol{\vartheta}^1$, that is ϑ_i is not a zero element. Then, by using the taylor approximation, it holds

$$\begin{aligned}
\mathcal{D}(\boldsymbol{\vartheta}_0) - \mathcal{D}(\boldsymbol{\vartheta}) &= \left(\hat{\boldsymbol{\beta}} - \frac{1}{H} \sum_{i=1}^{H} \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0} \right) \hat{\boldsymbol{\Omega}} \left(\hat{\boldsymbol{\beta}} - \frac{1}{H} \sum_{i=1}^{H} \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0} \right) - \\
&- \left(\hat{\boldsymbol{\beta}} - \frac{1}{H} \sum_{i=1}^{H} \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}} \right) \hat{\boldsymbol{\Omega}} \left(\hat{\boldsymbol{\beta}} - \frac{1}{H} \sum_{i=1}^{H} \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}} \right) \\
&+ n \sum_i \left(p_{\lambda_n}(|\vartheta_{i0}|) - p_{\lambda_n}(|\vartheta_i|) \right) = \\
&= \frac{2}{H} \sum_{h=1}^{H} \frac{\partial \tilde{\boldsymbol{\beta}}_{\boldsymbol{\vartheta}_0^h}}{\partial \boldsymbol{\vartheta}} \hat{\boldsymbol{\Omega}} \left(\hat{\boldsymbol{\beta}} - \frac{1}{H} \sum_{i=1}^{H} \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0} \right) (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) - \\
&- (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)' \frac{2}{H} \sum_{h=1}^{H} \frac{\partial \tilde{\boldsymbol{\beta}}_{\boldsymbol{\vartheta}_0^h}}{\partial \boldsymbol{\vartheta}} \hat{\boldsymbol{\Omega}} \frac{2}{H} \sum_{h=1}^{H} \frac{\partial \tilde{\boldsymbol{\beta}}_{\boldsymbol{\vartheta}_0^h}}{\partial \boldsymbol{\vartheta}} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) - \\
&- n \sum_i p'_{\lambda_n}(|\vartheta_i|) \operatorname{sgn}(\vartheta_i) (\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{i0}) \\
&- n \sum_i p''_{\lambda_n}(|\vartheta_i|) (\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{i0})^2,
\end{aligned} \tag{A.36}$$

for $n >> 0$ the last term is negligible because $||\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0|| \leq Mn^{-\frac{1}{2}}$ and the same holds for the first term. In the second term there is the following matrix

$$-\frac{2}{H} \sum_{h=1}^{H} \frac{\partial \tilde{\boldsymbol{\beta}}_{\boldsymbol{\vartheta}_0^h}}{\partial \boldsymbol{\vartheta}} \hat{\boldsymbol{\Omega}} \frac{2}{H} \sum_{h=1}^{H} \frac{\partial \tilde{\boldsymbol{\beta}}_{\boldsymbol{\vartheta}_0^h}}{\partial \boldsymbol{\vartheta}} \tag{A.37}$$

that is negative definite and for $T >> 0$, it dominates the other terms therefore $\mathcal{D}(\boldsymbol{\vartheta}_0) - \mathcal{D}(\boldsymbol{\vartheta}) \leq 0$. This implies that there is a local minimizer $\hat{\boldsymbol{\vartheta}}$ such that $||\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0|| \leq Mn^{-\frac{1}{2}}$. □

Proof. 30 Let us consider the second Taylor expansion of the first order condition of 4.7

$$\begin{aligned}\frac{\partial \mathcal{D}}{\partial \boldsymbol{\vartheta}} = & -\frac{2}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}\right) + \\ & + \frac{2}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\frac{1}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) + n\mathbf{v}_0 + n\mathbf{P}_0(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) = 0\end{aligned} \tag{A.38}$$

where $\mathbf{v}_{0i} = p_{\lambda_n}(|\vartheta_{0i}|)$ and $\mathbf{P}_{0i} = p'_{\lambda_n}(|\vartheta_{0i}|)\,\mathrm{sgn}(\vartheta_i)$. Let us rewrite the above equation

$$\begin{aligned}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)\left[\frac{2}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\frac{1}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} + n\mathbf{P}_0\right] = \\ \frac{2}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}\right) + n\mathbf{v}_0\end{aligned} \tag{A.39}$$

$$\begin{aligned}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) = & \left[\frac{2}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\frac{1}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}} + n\mathbf{P}_0\right]^{-1} \times \\ & \left[\frac{2}{H}\sum_{h=1}^{H}\frac{\partial \tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{\vartheta}}\hat{\boldsymbol{\Omega}}\left(\hat{\boldsymbol{\beta}} - \frac{1}{H}\sum_{h=1}^{H}\tilde{\boldsymbol{\beta}}^h_{\boldsymbol{\vartheta}_0}\right) - n\mathbf{v}_0\right]\end{aligned} \tag{A.40}$$

Since $\mathbf{v}_0$ and $\mathbf{P}_0$ vanish asymptotically the theorem comes from the asymptotic theory of indirect inference estimator developed in Gouriéroux et al. (1993). □

A.4 Chapter 5

Proof. In general, the VaR is sub–additive if and only if $VaR_\tau(Y_1+Y_2) \leq VaR_\tau(Y_1) + VaR_\tau(Y_2)$. Then, by applying Proposition (8) and Proposition (9), we get the marginal distribution of $Y_j \sim \mathcal{SESD}_1(\alpha, 0, 1, \delta_j)$ with $\delta_j = \frac{\lambda_j}{\sqrt{1+\lambda_j^2}}$ for $j = 1, 2$ and the distribution of the sum $Y^\star = Y_1 + Y_2 \sim \mathcal{SESD}_1(\alpha, 0, 2(1+\rho), \delta^\star)$ with $\delta^\star = \frac{\delta_1+\delta_2}{\sqrt{2(1+\rho)}}$ and

$$VaR_\tau(Y^\star) = \sqrt{2(1+\rho)}VaR_\tau\left(\alpha, \frac{\delta_1+\delta_2}{\sqrt{2(1+\rho)}}\right) \tag{A.41}$$

$$VaR_\tau(Y_1) + VaR_\tau(Y_2) = VaR_\tau(\alpha, \delta_1) + VaR_\tau(\alpha, \delta_2), \tag{A.42}$$

where $VaR_\tau(\alpha, \delta_1)$ denotes the VaR at confidence level τ of a standardised univariate Skew Elliptical distribution with shape parameter δ and characteristic exponent α. Assume marginals are equal, i.e., $\delta_1 = \delta_2 = \delta$, then, rearranging previous equations, the sub–additivity property is preserved if and only if

$$\frac{VaR_\tau\left(\alpha, \frac{\sqrt{2}\delta}{\sqrt{1+\rho}}\right)}{VaR_\tau(\alpha, \delta)} \leq \frac{\sqrt{2}\delta}{\sqrt{1+\rho}}. \tag{A.43}$$

Since the SESD is a continuous distribution with strictly increasing cdf, then $VaR_\tau(\alpha, \delta) = F_X^{-1}(\tau, \alpha, \delta)$, where $F_X^{-1}(\tau, \alpha, \delta)$ denotes the inverse of the cdf of a univariate random variable $X \sim \mathcal{SESD}_1(\alpha, 0, 1, \delta)$. Let $Z_1 \sim \mathcal{SESD}_1\left(\alpha, 0, 1, \frac{\sqrt{2}\delta}{\sqrt{1+\rho}}\right)$ and $Z_2 \sim \mathcal{SESD}_1(\alpha, 0, 1, \delta)$, then equation (A.43) becomes

$$F_{Z_1}^{-1}\left(\tau, \alpha, \frac{\sqrt{2}\delta}{\sqrt{1+\rho}}\right) \leq \frac{\sqrt{2}\delta}{\sqrt{1+\rho}} F_{Z_2}^{-1}(\tau, \alpha, \delta), \tag{A.44}$$

and after a simple change of variable, we get

$$\tau \leq F_{Z_1}\left(\frac{\sqrt{2}\delta}{\sqrt{1+\rho}}F_{Z_2}^{-1}(\alpha,\delta), \alpha, \frac{\sqrt{2}\delta}{\sqrt{1+\rho}}\right) = \frac{\sqrt{1+\rho}}{\sqrt{2}\delta}\tau^{\star\star}_{\rho,\delta}, \tag{A.45}$$

where $\tau^{\star\star}_{\rho,\delta} = F_{Z_1}\left(F_{Z_2}^{-1}(\alpha,\delta), \frac{\sqrt{2}\delta}{\sqrt{1+\rho}}\right)$. Observe also that if $\tau < 0.5$ and $\delta > 0$, then $\tau^{\star\star}_{\rho,\delta} \leq \tau$. This means that if $\delta > 0$, then $\sqrt{1+\rho}\tau^{\star\star}_{\rho,\delta} \geq \sqrt{2}\delta\tau$ and

$$\frac{\sqrt{1+\rho}}{\sqrt{2}\delta} \geq \frac{\tau}{\tau^{\star\star}} > 1, \tag{A.46}$$

To complete the proof it is sufficient to prove that the SESD is well defined if and only if the matrix

$$\bar{\mathbf{\Omega}} = \begin{bmatrix} 1 & \delta & \delta \\ \delta & 1 & \rho \\ \delta & \rho & 1 \end{bmatrix}, \tag{A.47}$$

is a proper positive definite correlation matrix. Therefore, $1 - \boldsymbol{\delta}'\mathbf{C}\boldsymbol{\delta} > 0$ which implies $2\delta^2 < 1 + \rho$. $\square$

Appendix B

SESD parameters initialisation

The parameters of the Skew Elliptical Stable distribution can be initialised by using the quantile–based measures introduced in Section 3.2.2. In this appendix we construct such initial estimates by following the same approach of McCulloch (1986).

Let $\mathbf{Y} \sim \mathcal{SESD}_m\left(\alpha, \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\delta}\right)$, we first assume $m = 1$, and then we consider the case $m \geq 2$ as an extension. Let us consider the quantile–based measure for the skewness and kurtosis parameters defined as

$$\nu_\delta\left(\alpha, \delta\right) = \frac{q_{0.95} + q_{0.05} - 2q_{0.5}}{q_{0.95} - q_{0.05}} \tag{B.1}$$

$$\nu_\alpha\left(\alpha, \delta\right) = \frac{q_{0.95} - q_{0.05}}{q_{0.75} - q_{0.25}}, \tag{B.2}$$

which only depend on the parameters of the SESD governing the skewness and the kurtosis of the distribution. In order to initialise the parameters (α, δ) of the SESD we consider a tabulation of the measure defined in equations (B.1)–(B.2), over a grid of values of the parameters. Tabulated values are reported in Tables B.3 and B.4. The empirical counterpart of equations (B.1)–(B.2) defined as

$$\hat{\nu}_{\alpha}(\alpha, \delta) = \frac{\hat{q}_{0.95} - \hat{q}_{0.05}}{\hat{q}_{0.75} - \hat{q}_{0.25}} \tag{B.3}$$

$$\hat{\nu}_{\delta}(\alpha, \delta) = \frac{\hat{q}_{0.95} + \hat{q}_{0.05} - 2\hat{q}_{0.5}}{\hat{q}_{0.95} - \hat{q}_{0.05}}, \tag{B.4}$$

are consistent estimators of the corresponding theoretical measures. Moreover, $\nu_{\alpha}(\alpha, \delta)$ is a strictly decreasing function of α meaning that it identifies the kurtosis parameter, while $\nu_{\beta}(\alpha, \delta)$ is strictly increasing in δ for each α, meaning that it identifies the skewness parameter given the information on α. Therefore, we can invert equations (B.1)–(B.2) in order to express the parameters of interest as a function of the quantile–based measures, as follows

$$\alpha = \psi_1(\nu_{\alpha}, \nu_{\delta}) \tag{B.5}$$

$$\delta = \psi_2(\nu_{\alpha}, \nu_{\delta}). \tag{B.6}$$

Consistent estimates of the parameters of interest can be obtained by considering the empirical counterparts of equations (B.7)–(B.8), as follows

$$\hat{\alpha} = \psi_1(\hat{\nu}_{\alpha}, \hat{\nu}_{\delta}) \tag{B.7}$$

$$\hat{\delta} = \psi_2(\hat{\nu}_{\alpha}, \hat{\nu}_{\delta}). \tag{B.8}$$

Tabulated values of the parameters of interest α and δ over a grid of values of the corresponding quantile–based measures are reported in tables B.2 and B.1 so that it would be possible to retrieve consistent estimates by linear interpolation.

Concerning the scale parameter σ, we consider the following quantile–based measure of dispersion

$$\nu_{\sigma} = \frac{q_{0.75} - q_{0.25}}{\sigma}, \tag{B.9}$$

which depends on both the kurtosis and the skewness parameters, i.e.,

$\nu_\sigma = \phi_3(\alpha, \delta)$. Table B.5 tabulates ν_σ over a grid of skewness and shape parameters (α, δ). Plugging the previous estimates of the characteristic exponent and asymmetry parameters $\left(\hat{\alpha}, \hat{\delta}\right)$ and the sample quantiles $(\hat{q}_{0.25}, \hat{q}_{0.75})$, into the empirical counterpart of equation (B.9) we obtain the initial value of the scale parameter σ, i.e. As

$$\hat{\sigma} = \frac{\hat{q}_{0.75} - \hat{q}_{0.25}}{\phi_3\left(\hat{\alpha}, \hat{\delta}\right)}. \tag{B.10}$$

As regards the location parameter, an initial estimate is computed as follows. First, we see that, if

$$Y \sim \mathcal{SESD}(\alpha, \xi, \omega, \delta),$$

then

$$Y^- \sim \mathcal{SESD}(\alpha, \xi, \omega, -\delta),$$

and the transformation

$$Z = \frac{Y + Y^-}{\sqrt{2}} \sim \mathcal{SESD}\left(\alpha, \sqrt{2}\xi, \omega, 0\right)$$

is a symmetric random variable. Therefore, we can consider the usual median rescaled by a constant factor as efficient estimator of the location on the transformed data Z. The variable Z can be constructed by changing the sign of a bootstrap sample from the original data Y.

Now we consider the case $m \geq 2$. From Theorem 8, the marginal variables Y_i have univariate Elliptical Stable distribution, i.e.,

$$Y_i \sim \mathcal{SESD}_1(\alpha, \xi_i, \omega_i, \delta_i).$$

Therefore, the marginals parameters can be estimated using the procedure detailed before. As concerns the off–diagonal parameters of the scale matrix, they are estimated as follows. From theorem 8, for each couple of variables $\mathbf{Y}_{ij} = (Y_i, Y_j)'$ it holds

$$\mathbf{Y}_{ij} \sim \mathcal{SESD}_2 \left(\alpha, \boldsymbol{\xi}_{ij}, \boldsymbol{\Omega}_{ij}, \boldsymbol{\delta}_{ij}\right),$$

where $\boldsymbol{\xi}_{ij} = (\xi_i, \xi_j)'$ and $\boldsymbol{\Omega}_{ij} = \begin{bmatrix} \omega_i^2 & \omega_{ij} \\ \omega_{ij} & \omega_j^2 \end{bmatrix}$. Now, let

$$\mathbf{Y}_{ij}^{-} \sim \mathcal{SESD} \left(\alpha, \boldsymbol{\xi}_{ij}, \boldsymbol{\Omega}_{ij}, -\boldsymbol{\delta}_{ij}\right)$$

and construct the transformed random variable

$$\mathbf{Z}_{ij} = \frac{\mathbf{Y}_{ij} + \mathbf{Y}_{ij}^{-}}{\sqrt{2}}, \tag{B.11}$$

such that $\mathbf{Z}_{ij} \sim \mathcal{SESD} \left(\alpha, \sqrt{2}\boldsymbol{\xi}_{ij}, \Omega_{ij}, \mathbf{0}\right)$. Let us consider the standardised variables $\mathbf{X}_{ij} = (X_i, X_j)'$ where

$$(X_i, X_j) = \left(\frac{Z_i - \sqrt{2}\xi_i}{\omega_{ii}}, \frac{Z_j - \sqrt{2}\xi_j}{\omega_{jj}}\right), \tag{B.12}$$

then $\mathbf{X}_{ij} \sim \mathcal{SESD} \left(\alpha, \mathbf{0}, \bar{\boldsymbol{\Omega}}_{ij}, \mathbf{0}\right)$ where $\bar{\boldsymbol{\Omega}}_{ij} = \begin{bmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{bmatrix}$. Using the Definition 27, it turns out that the optimal direction for ρ_{ij} is $\mathbf{u} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)'$. Therefore, we project $\mathbf{X}_{ij}$ along $\mathbf{u}$ and we obtain the variable $X_{\mathbf{u}} = \mathbf{u}'\mathbf{X}_{ij}$ such that $X_{\mathbf{u}} \sim \mathcal{SESD} (\alpha, 0, 1 + \rho_{ij}, 0)$, by applying Theorem 9. Now, since $X_{\mathbf{u}}$ is a univariate random variable we can apply the previous method to initialise the scale of a univariate SESD. Therefore

$$\nu_\sigma = \frac{q_{0.75} - q_{0.25}}{1 + \rho_{ij}}, \tag{B.13}$$

and by using the consistent estimates computed before we get a initial estimate of the correlation parameter

$$\hat{\rho}_{ij} = \frac{\hat{q}_{0.75} - \hat{q}_{0.25}}{\phi_3 \left(\hat{\alpha}, 0\right)} - 1. \tag{B.14}$$

ν	ν_δ									
ν_α	**0.0**	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**
2.7	0.002	0.299	0.544	0.703	0.724	0.746	0.767	0.788	0.812	0.832
2.8	0.001	0.253	0.473	0.656	0.710	0.725	0.740	0.754	0.900	0.900
2.9	0.001	0.223	0.423	0.602	0.706	0.724	0.742	0.759	0.900	0.900
3.0	-0.001	0.200	0.387	0.555	0.699	0.725	0.750	0.775	0.800	0.850
3.2	-0.003	0.167	0.332	0.487	0.623	0.710	0.734	0.759	0.783	0.802
3.4	-0.003	0.155	0.313	0.459	0.593	0.703	0.722	0.741	0.760	0.900
3.6	-0.003	0.145	0.294	0.435	0.566	0.688	0.725	0.753	0.780	0.900
3.8	-0.001	0.137	0.279	0.414	0.543	0.661	0.718	0.746	0.774	0.818
4.0	-0.005	0.130	0.267	0.393	0.520	0.636	0.710	0.732	0.754	0.777
4.2	-0.006	0.123	0.254	0.379	0.500	0.612	0.706	0.735	0.764	0.793
4.4	-0.006	0.123	0.254	0.379	0.500	0.612	0.706	0.735	0.764	0.793
4.6	-0.008	0.118	0.243	0.363	0.478	0.590	0.699	0.725	0.750	0.775
4.8	-0.007	0.112	0.231	0.348	0.459	0.568	0.678	0.721	0.747	0.773
5.0	-0.007	0.112	0.231	0.348	0.459	0.568	0.678	0.721	0.747	0.773
6.0	-0.003	0.106	0.210	0.321	0.427	0.529	0.633	0.712	0.746	0.779
8.0	-0.001	0.092	0.184	0.282	0.380	0.474	0.574	0.675	0.729	0.768
10.0	-0.001	0.088	0.179	0.273	0.365	0.457	0.555	0.655	0.723	0.764
15.0	-0.002	0.075	0.151	0.232	0.319	0.403	0.489	0.585	0.689	0.749
25.0	0.002	0.067	0.136	0.206	0.286	0.362	0.442	0.533	0.638	0.732
35.0	0.000	0.065	0.130	0.195	0.266	0.340	0.417	0.502	0.608	0.719
45.0	0.003	0.060	0.120	0.183	0.252	0.322	0.393	0.475	0.575	0.699

Table B.1: Tabulation of the skewness parameter δ as function of the quantile–based measures ν_δ and ν_α.

α	ν_δ									
ν_α	**0.0**	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**
2.7	1.728	1.730	1.742	1.763	1.763	1.763	1.763	1.763	1.727	1.727
2.8	1.659	1.664	1.673	1.688	1.703	1.703	1.703	1.703	1.703	1.703
2.9	1.606	1.608	1.619	1.636	1.651	1.651	1.651	1.651	1.550	1.550
3.0	1.559	1.560	1.566	1.581	1.591	1.605	1.605	1.605	1.605	1.877
3.2	1.481	1.482	1.489	1.494	1.512	1.527	1.527	1.527	1.527	1.445
3.4	1.418	1.419	1.425	1.431	1.438	1.463	1.463	1.463	1.463	1.455
3.6	1.364	1.365	1.367	1.377	1.383	1.394	1.406	1.406	1.406	1.349
3.8	1.318	1.319	1.320	1.329	1.335	1.345	1.356	1.356	1.356	1.168
4.0	1.277	1.277	1.279	1.281	1.292	1.301	1.311	1.311	1.311	1.311
4.2	1.240	1.240	1.242	1.243	1.253	1.261	1.269	1.269	1.269	1.269
4.4	1.206	1.207	1.209	1.210	1.220	1.227	1.235	1.235	1.235	1.235
4.6	1.177	1.177	1.178	1.179	1.183	1.188	1.194	1.201	1.201	1.201
4.8	1.149	1.149	1.149	1.150	1.153	1.158	1.163	1.168	1.168	1.168
5.0	1.125	1.125	1.125	1.126	1.128	1.133	1.138	1.143	1.143	1.143
6.0	1.025	1.024	1.023	1.024	1.024	1.027	1.028	1.030	1.030	1.030
8.0	0.895	0.895	0.894	0.893	0.893	0.891	0.890	0.887	0.884	0.884
10.0	0.815	0.815	0.814	0.812	0.810	0.807	0.804	0.798	0.793	0.793
15.0	0.697	0.697	0.696	0.695	0.692	0.689	0.689	0.684	0.678	0.671
25.0	0.590	0.590	0.590	0.588	0.588	0.584	0.580	0.575	0.567	0.560
35.0	0.535	0.535	0.535	0.535	0.533	0.530	0.525	0.520	0.512	0.505
45.0	0.498	0.498	0.498	0.498	0.496	0.493	0.493	0.489	0.485	0.479

Table B.2: Tabulation of the kurtosis parameter α as function of the quantile–based measures ν_δ and ν_α.

ν_α				δ				
α	**0.0**	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**
1.95	2.475	2.475	2.474	2.475	2.477	2.480	2.483	2.492
1.90	2.516	2.515	2.515	2.515	2.519	2.520	2.525	2.539
1.85	2.561	2.561	2.560	2.560	2.564	2.567	2.573	2.590
1.80	2.611	2.613	2.612	2.610	2.620	2.624	2.632	2.651
1.75	2.671	2.672	2.671	2.674	2.680	2.689	2.699	2.722
1.70	2.735	2.739	2.742	2.745	2.754	2.763	2.778	2.806
1.65	2.815	2.819	2.823	2.829	2.839	2.853	2.871	2.901
1.60	2.912	2.913	2.916	2.922	2.939	2.954	2.976	3.013
1.55	3.020	3.022	3.026	3.036	3.052	3.072	3.102	3.140
1.50	3.145	3.149	3.155	3.166	3.182	3.207	3.238	3.285
1.45	3.292	3.294	3.303	3.315	3.336	3.359	3.393	3.445
1.40	3.461	3.464	3.472	3.485	3.509	3.532	3.574	3.625
1.35	3.654	3.660	3.667	3.683	3.705	3.729	3.777	3.828
1.30	3.880	3.883	3.892	3.901	3.930	3.961	4.004	4.057
1.25	4.139	4.138	4.150	4.159	4.188	4.219	4.264	4.313
1.20	4.438	4.440	4.452	4.458	4.480	4.517	4.559	4.610
1.15	4.792	4.791	4.796	4.802	4.823	4.860	4.905	4.943
1.10	5.206	5.205	5.202	5.211	5.225	5.260	5.302	5.331
1.05	5.702	5.695	5.686	5.696	5.703	5.737	5.760	5.783
1.00	6.292	6.288	6.272	6.275	6.282	6.299	6.304	6.314
0.95	7.010	6.993	6.981	6.977	6.973	6.972	6.959	6.949
0.90	7.886	7.876	7.858	7.845	7.818	7.792	7.752	7.706
0.85	8.998	8.978	8.949	8.915	8.863	8.811	8.731	8.654
0.80	10.414	10.391	10.336	10.265	10.185	10.102	9.950	9.831
0.75	12.280	12.246	12.156	12.076	11.948	11.759	11.564	11.366
0.70	14.789	14.730	14.623	14.482	14.271	14.003	13.717	13.406
0.65	18.312	18.215	18.107	17.856	17.539	17.154	16.645	16.182
0.60	23.391	23.358	23.153	22.697	22.196	21.583	20.831	20.166
0.55	31.201	31.190	30.795	30.116	29.265	28.298	27.222	26.253
0.50	44.112	44.009	43.373	42.034	40.624	39.353	37.532	35.955
0.45	67.257	67.095	65.432	63.380	60.721	58.418	55.649	53.461
0.40	114.237	113.281	110.339	105.497	100.657	95.941	90.804	87.166

Table B.3: Tabulation of the quantile–based measures ν_α as function of the skewness parameter δ and the kurtosis parameter α.

ν_δ				$\boldsymbol{\delta}$				
$\boldsymbol{\alpha}$	**0.0**	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**
1.95	0.000	0.004	0.009	0.016	0.026	0.040	0.058	0.086
1.90	0.000	0.010	0.020	0.033	0.048	0.068	0.093	0.128
1.85	0.000	0.018	0.036	0.054	0.076	0.104	0.137	0.179
1.80	0.000	0.026	0.052	0.077	0.109	0.144	0.183	0.235
1.75	0.000	0.033	0.067	0.100	0.138	0.180	0.225	0.286
1.70	0.000	0.040	0.079	0.118	0.165	0.213	0.264	0.329
1.65	0.000	0.045	0.090	0.135	0.188	0.240	0.299	0.367
1.60	0.000	0.049	0.100	0.152	0.207	0.266	0.328	0.401
1.55	0.000	0.055	0.110	0.166	0.225	0.287	0.356	0.434
1.50	0.000	0.060	0.120	0.180	0.243	0.309	0.382	0.461
1.45	0.000	0.065	0.129	0.191	0.260	0.328	0.405	0.487
1.40	0.000	0.069	0.137	0.204	0.275	0.347	0.427	0.510
1.35	0.000	0.074	0.145	0.215	0.289	0.365	0.447	0.535
1.30	0.000	0.078	0.152	0.224	0.305	0.383	0.468	0.557
1.25	0.000	0.082	0.159	0.235	0.317	0.400	0.489	0.580
1.20	0.000	0.086	0.165	0.246	0.331	0.420	0.509	0.601
1.15	0.000	0.090	0.174	0.258	0.346	0.437	0.529	0.620
1.10	0.000	0.090	0.182	0.267	0.359	0.454	0.549	0.642
1.05	0.000	0.094	0.191	0.280	0.374	0.472	0.568	0.664
1.00	0.000	0.098	0.198	0.292	0.389	0.490	0.586	0.685
0.95	0.000	0.103	0.207	0.306	0.405	0.508	0.607	0.704
0.90	0.000	0.109	0.217	0.318	0.421	0.528	0.625	0.725
0.85	0.000	0.114	0.223	0.329	0.438	0.546	0.645	0.746
0.80	0.000	0.117	0.235	0.345	0.455	0.566	0.667	0.767
0.75	0.000	0.129	0.250	0.361	0.474	0.588	0.690	0.789
0.70	0.000	0.133	0.263	0.378	0.497	0.613	0.715	0.810
0.65	0.000	0.141	0.275	0.395	0.524	0.638	0.740	0.833
0.60	0.000	0.149	0.292	0.418	0.551	0.667	0.766	0.855
0.55	0.000	0.153	0.309	0.447	0.580	0.698	0.793	0.877
0.50	0.000	0.169	0.326	0.469	0.610	0.729	0.823	0.901
0.45	0.000	0.178	0.348	0.504	0.647	0.768	0.854	0.922
0.40	0.000	0.197	0.383	0.545	0.692	0.806	0.884	0.942

Table B.4: Tabulation of the quantile–based measures ν_δ as function of the skewness parameter δ and the kurtosis parameter α.

ν_ω				$\boldsymbol{\delta}$				
$\boldsymbol{\alpha}$	**0.0**	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**
1.95	1.349	1.345	1.333	1.311	1.278	1.236	1.183	1.113
1.90	1.351	1.347	1.335	1.313	1.281	1.239	1.186	1.117
1.85	1.353	1.349	1.337	1.316	1.284	1.242	1.189	1.121
1.80	1.356	1.352	1.339	1.318	1.286	1.245	1.192	1.126
1.75	1.358	1.354	1.341	1.320	1.289	1.248	1.196	1.131
1.70	1.361	1.356	1.344	1.323	1.292	1.252	1.202	1.136
1.65	1.363	1.358	1.346	1.325	1.295	1.256	1.206	1.143
1.60	1.365	1.360	1.349	1.329	1.299	1.261	1.213	1.150
1.55	1.367	1.363	1.351	1.332	1.303	1.265	1.218	1.159
1.50	1.370	1.365	1.354	1.335	1.307	1.271	1.227	1.169
1.45	1.372	1.368	1.357	1.338	1.312	1.278	1.235	1.181
1.40	1.375	1.370	1.360	1.343	1.317	1.285	1.244	1.194
1.35	1.377	1.373	1.364	1.347	1.324	1.294	1.256	1.208
1.30	1.379	1.377	1.367	1.353	1.332	1.304	1.268	1.226
1.25	1.383	1.380	1.372	1.360	1.341	1.315	1.283	1.246
1.20	1.387	1.384	1.378	1.367	1.351	1.329	1.302	1.269
1.15	1.391	1.389	1.384	1.376	1.363	1.345	1.323	1.296
1.10	1.397	1.396	1.392	1.386	1.377	1.364	1.347	1.328
1.05	1.403	1.403	1.402	1.399	1.393	1.386	1.376	1.365
1.00	1.411	1.412	1.414	1.414	1.414	1.412	1.411	1.411
0.95	1.422	1.424	1.429	1.432	1.438	1.444	1.453	1.464
0.90	1.437	1.439	1.446	1.454	1.467	1.482	1.502	1.529
0.85	1.454	1.458	1.467	1.481	1.502	1.530	1.563	1.606
0.80	1.477	1.482	1.494	1.515	1.546	1.586	1.638	1.700
0.75	1.505	1.510	1.528	1.557	1.599	1.656	1.728	1.816
0.70	1.540	1.548	1.571	1.610	1.667	1.745	1.840	1.962
0.65	1.583	1.595	1.624	1.675	1.754	1.855	1.986	2.149
0.60	1.641	1.654	1.694	1.763	1.867	2.002	2.174	2.392
0.55	1.715	1.731	1.787	1.880	2.018	2.199	2.424	2.718
0.50	1.813	1.837	1.910	2.038	2.226	2.467	2.776	3.177
0.45	1.944	1.977	2.085	2.259	2.519	2.853	3.286	3.842
0.40	2.126	2.176	2.326	2.586	2.957	3.438	4.083	4.896

Table B.5: Tabulation of the quantile–based measures ν_ω as function of the skewness parameter δ and the kurtosis parameter α.

Appendix C

Numerical evaluation of the SESD cdf

In this appendix we detail the procedure used to evaluate the cumulative density function of the Skew Elliptical Stable distribution. From the basic properties of the characteristic function, we have

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} dF_X(x) = \int_{-\infty}^{\infty} e^{itx} f_X(x)\, dt, \tag{C.1}$$

and the density function can be obtained by inverting the characteristic function

$$f_X(x) = \int_{-\infty}^{\infty} e^{-itx} \phi_X(t)\, dt. \tag{C.2}$$

Therefor, the cumulative density function can be obtained by integrating che density function

$$F_X(s) = \int_{-\infty}^{s} f_X(x)\, dx = \int_{-\infty}^{s} \int_{-\infty}^{\infty} e^{-itx} \phi_X(t)\, dt dx. \tag{C.3}$$

We are going to use these simple formula in order to get an approximation of the density and cumulative function of the Skew Elliptical Stable Distribution. However this approach is quite general thus it can be adapted to all those distribution whose characteristic function is known but whose density function is not analytically tractable. To this end we use Fast Fourier Transform (FFT) that we briefly recall below. Let us start by introducing the Discrete Fourier Transform (DFT). Let g_t be a continuous function such that $\int_{-\infty}^{\infty} |g(t)| < \infty$, then the function

$$G(x) = \int_{-\infty}^{\infty} e^{-itx} g(t)\, dt, \tag{C.4}$$

is called the Fourier transform of $g(t)$. The discrete version is given by

$$G_n = \frac{1}{T} \sum_{i=1}^{T-1} g_t e^{-\frac{2\pi i t n}{T}}, \qquad n = 0, 1, \ldots, T-1, \tag{C.5}$$

where $g_t = g(t)$ that can be evaluated by using the inverse discrete Fourier transform

$$g_t = \sum_{n=0}^{T-1} G_n e^{\frac{2\pi i n t}{T}}, \qquad t = 0, 1, \ldots, T-1. \tag{C.6}$$

The FFT is the DFT computed in a smart way in order to reduce the computational time. Below we report the main passages that lead to computation of the FFT. Let $z_T = e^{-\frac{2\pi i}{T}}$, we rewrite the DFT of g_t as follows

$$\begin{aligned} G_n &= \frac{1}{T} \sum_{t=0}^{T-1} g_t \left(z_T^n\right)^t = \frac{1}{T} \sum_{t=0}^{\frac{T}{2}-1} g_{2t} \left(z_T^n\right)^{2t} + \frac{1}{T} \sum_{t=0}^{\frac{T}{2}-1} g_{2t+1} \left(z_T^n\right)^{2t+1} = \\ &= \frac{1}{T} \sum_{t=0}^{\frac{T}{2}-1} g_{2t} \left(z_T^{2n}\right)^t + \frac{\left(z_T^n\right)}{T} \sum_{t=0}^{\frac{T}{2}-1} g_{2t+1} \left(z_T^{2n}\right)^t = S_1 + S_2. \end{aligned} \tag{C.7}$$

Thus we divided the the series in odd and even indexes. We observe that

$$\begin{aligned} z_T^{\frac{T}{2}} &= \left(e^{-\frac{2\pi i}{T}}\right)^{\frac{T}{2}} = e^{-\pi i} = -1 \\ Z_T^T &= e^{-2\pi i} = 1, \end{aligned} \tag{C.8}$$

then if we consider the $n+\frac{T}{2}$–th element of the DFT we get

$$\begin{aligned} G_{n+\frac{T}{2}} &= \frac{1}{T}\sum_{t=0}^{\frac{T}{2}-1} g_{2t}\left(z_T^{n+\frac{T}{2}}\right)^{2t} + \frac{1}{T}\sum_{t=0}^{\frac{T}{2}-1} g_{2t+1}\left(z_T^{n+\frac{T}{2}}\right)^{2t+1} \\ &= \frac{1}{T}\sum_{t=0}^{\frac{T}{2}-1} g_{2t}\left(z_T^{2n} z_T^{T}\right)^{t} + \frac{\left(z_T^{n+\frac{T}{2}}\right)}{T}\sum_{t=0}^{\frac{T}{2}-1} g_{2t+1}\left(z_T^{2n}\right)^{t} \\ &= \frac{1}{T}\sum_{t=0}^{\frac{T}{2}-1} g_{2t}\left(z_T^{2n} z_T^{T}\right)^{t} - \frac{\left(z_T^{n}\right)}{T}\sum_{t=0}^{\frac{T}{2}-1} g_{2t+1}\left(z_T^{2n}\right)^{t} = S_1 - S_2. \end{aligned} \tag{C.9}$$

Thus S_1 and S_2 have to be computed until $\frac{T}{2}-1$ and this reduces considerably the computational time. We us this result to approximate the characteristic function, indeed as we have seen before, the characteristic function is the Fourier transform of the density function. Then

$$\phi_X(s) = \int_{-\infty}^{\infty} e^{isx} f_X(x)\,dx \approx \int_{l}^{u} e^{isx} f_X(x)\,dx \approx \sum_{n=0}^{T-1} e^{isx_n} P_n, \tag{C.10}$$

where $P_n = f_X(x_n)\,\Delta x$, $x+n = l+n\Delta x$ and $\Delta x = \frac{u-l}{T}$. Now we multiply the formula above by e^{-isl}

$$\begin{aligned} \phi_X(s)\,e^{-isl} \approx \sum_{n=0}^{T-1} e^{isx_n} e^{-isl} P_n = \sum_{n=0}^{T-1} e^{isx_n - l} P_n = \\ \sum_{n=0}^{T-1} e^{is\Delta x} P_n = \sum_{n=0}^{T-1} e^{2\pi int} P_n = g_t, \end{aligned} \tag{C.11}$$

where $s\Delta x = \frac{2\pi t}{T}$. We consider a T-length grid of values for s: $s_t = \frac{2\pi t}{T}$ with $t = -\frac{T}{2}, -\frac{T}{2}+1, \ldots, \frac{T}{2}-1, \frac{T}{2}$. As next step we apply the DFT to g_t in order to get P_n as follows

$$P_n = \frac{1}{T}\sum_{t=0}^{T-1} g_t e^{-\frac{2\pi itn}{T}} \quad n = 0, 1, \ldots, T-1. \tag{C.12}$$

Since $P_n = f_X(x_n)\,\Delta x$, we easily get the density function. Now we can get also an approximation of the cumulative distribution function, indeed we consider the following

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(t)\,dt \approx \int_{l}^{x} f_X(t)\,dt \\ &\approx \sum_{n=0}^{N} f_X(x_n)\,\Delta x = \sum_{n=0}^{N} P_n, \end{aligned} \tag{C.13}$$

where $x_0 = l$ and $x_N = x$. Finally we can get the quantile by numerically invert this formula.

Appendix D

Tables of simulated examples

D.1 Elliptical Stable

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.70**	-0.0075	0.0996	0.7970	-0.0041	0.0535	0.7650
ξ_1	0.00	0.0016	0.0443	0.9380	0.0013	0.0201	0.9500
ξ_2	0.00	0.0088	0.0841	0.9440	0.0021	0.0385	0.9590
ω_{11}	0.50	0.0112	0.2904	0.6030	-0.0046	0.0605	0.5330
ω_{22}	2.00	-0.0409	0.3599	0.6870	-0.0059	0.1439	0.6910
ω_{12}	0.90	-0.1044	0.2841	0.8090	-0.0369	0.1680	0.8280
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.90**	-0.0315	0.0876	0.8750	-0.0141	0.0626	0.8760
ξ_1	0.00	-0.0003	0.0444	0.9390	0.0010	0.0209	0.9440
ξ_2	0.00	0.0029	0.0891	0.9240	0.0005	0.0401	0.9510
ω_{11}	0.50	-0.0069	0.2040	0.6480	0.0045	0.4682	0.6120
ω_{22}	2.00	-0.0412	0.3563	0.7700	0.0002	0.4357	0.7380
ω_{12}	0.90	-0.1862	0.3717	0.7530	-0.1373	0.3110	0.7730
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.95**	-0.0628	0.0974	0.8580	-0.0310	0.0586	0.8580
ξ_1	0.00	0.0006	0.0436	0.9360	0.0047	0.1060	0.9490
ξ_2	0.00	0.0038	0.0862	0.9220	-0.0008	0.0645	0.9520
ω_{11}	0.50	0.0111	0.5107	0.6270	-0.0014	0.2008	0.6310
ω_{22}	2.00	-0.0688	0.4903	0.7650	-0.0272	0.3358	0.7580
ω_{12}	0.90	-0.2227	0.4907	0.6980	-0.2181	0.4444	0.7190

Table D.1: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the locations $\boldsymbol{\xi} = (\xi_1, \xi_2)$, scale matrix $\boldsymbol{\Omega} = \{\omega_{ij}\}$, with $i, j = 1, 2$ and $i \leq j$ and tail parameter α of the bivariate Elliptical Stable distribution. The results reported above are obtained using 1000 replications for three different values of $\alpha = \{1.70, 1.90, 1.95\}$.

		$n=500$			$n=2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.70**	-0.0055	0.0613	0.7958	-0.0001	0.0352	0.8006
ξ_1	0.00	-0.0008	0.0281	0.9409	0.0010	0.0470	0.9376
ξ_2	0.00	0.0011	0.0406	0.9479	-0.0001	0.0625	0.9315
ξ_3	0.00	-0.0024	0.0533	0.9550	0.0021	0.0984	0.9406
ξ_4	0.00	-0.0055	0.0785	0.9409	0.0177	0.5072	0.9527
ξ_5	0.00	0.0023	0.1149	0.9389	0.0278	1.0030	0.9436
$\sqrt{\omega}_{11}$	0.5000	-0.0047	0.0312	0.7688	-0.0015	0.0160	0.8187
$\sqrt{\omega}_{22}$	0.7071	0.0040	0.0393	0.7678	0.0019	0.0214	0.7795
$\sqrt{\omega}_{33}$	1.0000	-0.0058	0.0547	0.7247	-0.0033	0.0316	0.7402
$\sqrt{\omega}_{44}$	1.4142	0.0022	0.0801	0.7337	0.0040	0.0479	0.7422
$\sqrt{\omega}_{55}$	2.0000	-0.0091	0.1144	0.7047	0.0011	0.0681	0.7382
ρ_{12}	0.7071	-0.0171	0.1312	0.9650	-0.0080	0.0706	0.9778
ρ_{13}	0.8000	-0.0490	0.1764	0.9469	-0.0219	0.0983	0.9748
ρ_{14}	0.00	0.0124	0.1292	0.9269	0.0071	0.0657	0.9275
ρ_{15}	0.00	0.0178	0.1456	0.8859	0.0085	0.0724	0.8751
ρ_{23}	0.5657	-0.0167	0.1558	0.9289	0.0010	0.0841	0.9527
ρ_{24}	0.00	0.0103	0.1168	0.9109	0.0050	0.0708	0.8207
ρ_{25}	0.00	0.0252	0.1336	0.8749	0.0100	0.0723	0.8258
ρ_{34}	0.00	0.0101	0.1194	0.9600	0.0031	0.0648	0.9587
ρ_{35}	0.00	0.0119	0.1194	0.9660	0.0046	0.0619	0.9527
ρ_{45}	0.9016	-0.1466	0.2975	0.9489	-0.0253	0.0981	0.9778

Table D.2: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the locations $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_5)$, scale matrix $\mathbf{\Omega} = \{\omega_{ij}\}$, with $i, j = 1, 2, \ldots, 5$ and $i \leq j$, the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43 and tail parameter α of the Elliptical Stable distribution in dimension 5. The results reported above are obtained using 1000 replications for $\alpha = 1.70$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.90**	-0.0387	0.1539	0.9891	-0.0092	0.0571	0.7480
ξ_1	0.00	0.0030	0.0920	0.9659	-0.0007	0.0142	0.9540
ξ_2	0.00	0.0052	0.0768	0.9628	-0.0028	0.0192	0.9560
ξ_3	0.00	0.0123	0.3309	0.9566	-0.0012	0.0275	0.9580
ξ_4	0.00	0.0130	0.3594	0.9395	0.0029	0.1348	0.9570
ξ_5	0.00	0.0306	0.5551	0.9333	0.0000	0.2171	0.9410
$\sqrt{\omega_{11}}$	0.5000	-0.0062	0.0297	0.7628	-0.0024	0.0160	0.8150
$\sqrt{\omega_{22}}$	0.7071	-0.0021	0.0375	0.7736	0.0000	0.0180	0.7790
$\sqrt{\omega_{33}}$	1.0000	-0.0063	0.0499	0.7271	-0.0034	0.0249	0.7420
$\sqrt{\omega_{44}}$	1.4142	-0.0021	0.0779	0.7876	0.0017	0.0367	0.7520
$\sqrt{\omega_{55}}$	2.0000	-0.0123	0.1070	0.7426	0.0005	0.0523	0.7690
ρ_{12}	0.7071	-0.0260	0.1174	0.9643	-0.0043	0.0577	0.9890
ρ_{13}	0.8000	-0.0751	0.1486	0.9271	-0.0155	0.0703	0.9870
ρ_{14}	0.00	0.0123	0.1158	0.9519	0.0056	0.0615	0.9600
ρ_{15}	0.00	0.0266	0.1506	0.8992	0.0060	0.0669	0.8900
ρ_{23}	0.5657	-0.0318	0.1193	0.9287	0.0018	0.0631	0.9750
ρ_{24}	0.00	0.0116	0.1151	0.9442	0.0010	0.0613	0.8920
ρ_{25}	0.00	0.0182	0.1184	0.9240	0.0039	0.0634	0.8780
ρ_{34}	0.00	-0.0013	0.1107	0.9674	0.0014	0.0538	0.9750
ρ_{35}	0.00	0.0029	0.1103	0.9767	0.0014	0.0579	0.9740
ρ_{45}	0.9016	-0.1425	0.2389	0.9659	-0.0208	0.0903	0.9950

Table D.3: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the locations $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_5)$, scale matrix $\mathbf{\Omega} = \{\omega_{ij}\}$, with $i, j = 1, 2, \ldots, 5$ and $i \leq j$, the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43 and tail parameter α of the Elliptical Stable distribution in dimension 5. The results reported above are obtained using 1000 replications for $\alpha = 1.90$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.95**	-0.0662	0.2190	0.9854	-0.0278	0.1222	0.9910
ξ_1	0.00	-0.0011	0.0284	0.9610	-0.0010	0.0140	0.9550
ξ_2	0.00	0.0012	0.0415	0.9463	-0.0029	0.0193	0.9565
ξ_3	0.00	-0.0051	0.0572	0.9382	-0.0019	0.0274	0.9475
ξ_4	0.00	-0.0162	0.2342	0.9431	-0.0025	0.0400	0.9520
ξ_5	0.00	-0.0221	0.4252	0.9496	-0.0094	0.0596	0.9385
$\sqrt{\omega}_{11}$	0.5000	-0.0040	0.0293	0.7724	-0.0031	0.0136	0.8186
$\sqrt{\omega}_{22}$	0.7071	0.0004	0.0416	0.7333	-0.0012	0.0184	0.7676
$\sqrt{\omega}_{33}$	1.0000	-0.0005	0.0516	0.7561	-0.0033	0.0254	0.7481
$\sqrt{\omega}_{44}$	1.4142	-0.0010	0.0728	0.7577	0.0012	0.0409	0.7271
$\sqrt{\omega}_{55}$	2.0000	-0.0042	0.1085	0.7236	0.0021	0.0493	0.7661
ρ_{12}	0.7071	-0.0276	0.1349	0.9496	-0.0043	0.0602	0.9880
ρ_{13}	0.8000	-0.0783	0.1569	0.9154	-0.0231	0.0670	0.9835
ρ_{14}	0.00	0.0183	0.1326	0.9463	0.0044	0.0574	0.9550
ρ_{15}	0.00	0.0226	0.1373	0.8862	-0.0034	0.0548	0.9100
ρ_{23}	0.5657	-0.0339	0.1248	0.9171	-0.0064	0.0614	0.9805
ρ_{24}	0.00	0.0146	0.1123	0.9496	0.0019	0.0632	0.9220
ρ_{25}	0.00	0.0168	0.1233	0.9268	-0.0036	0.0684	0.9280
ρ_{34}	0.00	0.0068	0.1100	0.9659	-0.0006	0.0507	0.9805
ρ_{35}	0.00	0.0055	0.1053	0.9707	-0.0078	0.0524	0.9835
ρ_{45}	0.9016	-0.1245	0.1787	0.9447	-0.0416	0.0724	0.9910

Table D.4: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the locations $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_5)$, scale matrix $\boldsymbol{\Omega} = \{\omega_{ij}\}$, with $i, j = 1, 2, \ldots, 5$ and $i \leq j$, the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43 and tail parameter α of the Elliptical Stable distribution in dimension 5. The results reported above are obtained using 1000 replications for $\alpha = 1.95$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.70**	-0.0122	0.0377	0.9970	-0.0039	0.0222	1.0000
ξ_1	0.00	0.0024	0.0692	0.8989	-0.0010	0.0180	0.9577
ξ_2	0.00	0.0026	0.0771	0.9109	0.0044	0.0476	0.9296
ξ_3	0.00	0.0036	0.1834	0.9109	0.0032	0.0211	0.9859
ξ_4	0.00	0.0043	0.1726	0.9099	0.0002	0.0200	0.9155
ξ_5	0.00	0.0009	0.0584	0.9109	-0.0030	0.0213	0.9718
ξ_6	0.00	0.0043	0.0427	0.9149	0.0031	0.0208	0.9577
ξ_7	0.00	-0.0066	0.0939	0.9079	-0.0011	0.0311	0.9437
ξ_8	0.00	-0.0043	0.1197	0.9179	0.0004	0.0147	0.9859
ξ_9	0.00	-0.0057	0.2075	0.9199	-0.0032	0.0354	0.9014
ξ_{10}	0.00	-0.0058	0.0451	0.8969	-0.0038	0.0270	0.9155
ξ_{11}	0.00	0.0050	0.0270	0.8769	-0.0001	0.0182	0.9296
ξ_{12}	0.00	0.0052	0.1070	0.8989	0.0008	0.0182	0.9437
$\sqrt{\omega_{1,1}}$	0.4889	-0.0135	0.0284	0.5175	-0.0009	0.0140	0.7042
$\sqrt{\omega_{2,2}}$	1.2466	-0.0290	0.0607	0.5275	0.0010	0.0370	0.7324
$\sqrt{\omega_{3,3}}$	0.6017	-0.0030	0.0330	0.5706	0.0026	0.0177	0.7324
$\sqrt{\omega_{4,4}}$	0.4461	-0.0115	0.0258	0.5596	0.0006	0.0125	0.8028
$\sqrt{\omega_{5,5}}$	0.5908	-0.0063	0.0323	0.5475	-0.0020	0.0170	0.6761
$\sqrt{\omega_{6,6}}$	0.5431	-0.0130	0.0283	0.5586	-0.0018	0.0133	0.7887
$\sqrt{\omega_{7,7}}$	0.8456	-0.0207	0.0432	0.5135	-0.0055	0.0230	0.6901
$\sqrt{\omega_{8,8}}$	0.4050	-0.0153	0.0236	0.5235	-0.0037	0.0099	0.7606
$\sqrt{\omega_{9,9}}$	0.7197	-0.0170	0.0373	0.4985	-0.0005	0.0207	0.7324
$\sqrt{\omega_{10,10}}$	0.6156	-0.0144	0.0312	0.5305	-0.0023	0.0157	0.8169
$\sqrt{\omega_{11,11}}$	0.3987	-0.0057	0.0243	0.5736	-0.0026	0.0101	0.8169
$\sqrt{\omega_{12,12}}$	0.4550	-0.0072	0.0249	0.5886	0.0006	0.0115	0.8451

Table D.5: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for tail parameter α, the locations $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_12)$ and the scale matrix $\boldsymbol{\Omega} = \{\omega_{ij}\}$, with $i, j = 1, 2, \ldots, 12$ and $i \leq j$ of the 12–dimensional Elliptical Stable simulated experiment discussed in Section 3.4.2, for $\alpha = 1.7$ and sample size $n = 500$.

D.2 Skew Elliptical Stable

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
$\rho_{1,2}$	0.1920	-0.1107	0.1626	0.6907	-0.0511	0.1709	0.8732
$\rho_{1,3}$	0.00	0.0031	0.0980	0.9009	0.0231	0.1152	0.8451
$\rho_{1,4}$	0.00	0.0271	0.1001	0.7898	-0.0216	0.0856	0.8310
$\rho_{1,5}$	0.00	0.0177	0.1045	0.8739	-0.0006	0.0959	0.8873
$\rho_{1,6}$	0.00	0.0245	0.0955	0.8088	0.0092	0.1122	0.8310
$\rho_{1,7}$	0.00	0.0122	0.1155	0.8969	0.0527	0.2451	0.6197
$\rho_{1,8}$	0.1566	-0.0141	0.1181	0.8468	-0.0214	0.1040	0.8451
$\rho_{1,9}$	0.00	0.0125	0.1136	0.9049	0.0257	0.1460	0.7606
$\rho_{1,10}$	0.00	0.0285	0.1128	0.8569	0.0117	0.1741	0.8592
$\rho_{1,11}$	0.00	0.0335	0.1184	0.8328	0.0299	0.1249	0.8310
$\rho_{1,12}$	0.00	0.0240	0.0947	0.7558	-0.0030	0.0820	0.7606
$\rho_{2,3}$	0.00	-0.0235	0.1109	0.7858	-0.0090	0.1069	0.7606
$\rho_{2,4}$	0.00	-0.0178	0.1282	0.7728	-0.0357	0.1477	0.6761
$\rho_{2,5}$	0.00	-0.0059	0.1137	0.8038	0.0096	0.1473	0.7042
$\rho_{2,6}$	0.00	-0.0207	0.1232	0.7357	0.0050	0.1325	0.7042
$\rho_{2,7}$	0.00	-0.0267	0.1027	0.8408	-0.0035	0.1096	0.7606
$\rho_{2,8}$	0.00	-0.0057	0.1488	0.7708	-0.0305	0.1532	0.6056
$\rho_{2,9}$	0.00	-0.0179	0.1057	0.8048	0.0042	0.0958	0.8028
$\rho_{2,10}$	0.00	-0.0196	0.1093	0.8208	-0.0003	0.1087	0.6620
$\rho_{2,11}$	0.00	-0.0151	0.1396	0.7528	-0.0327	0.1422	0.6338
$\rho_{2,12}$	0.00	-0.0138	0.1342	0.7578	-0.0181	0.1279	0.7465
$\rho_{3,4}$	0.0075	0.0132	0.1142	0.9129	-0.0292	0.1324	0.8310
$\rho_{3,5}$	0.00	0.0037	0.0841	0.7918	-0.0221	0.0931	0.7183
$\rho_{3,6}$	0.00	0.0033	0.0962	0.8358	-0.0283	0.1299	0.7606
$\rho_{3,7}$	0.00	-0.0189	0.1030	0.8729	-0.0184	0.2103	0.6479
$\rho_{3,8}$	0.00	0.0218	0.1159	0.9309	-0.0092	0.1372	0.7887
$\rho_{3,9}$	0.00	-0.0092	0.1000	0.8188	0.0279	0.1697	0.7887
$\rho_{3,10}$	0.00	0.0111	0.0851	0.7648	0.0064	0.1293	0.6620
$\rho_{3,11}$	0.00	0.0088	0.1214	0.9219	0.0050	0.1942	0.7465
$\rho_{3,12}$	0.00	0.0078	0.1109	0.9169	0.0217	0.1306	0.8310
$\rho_{4,5}$	0.3567	-0.1125	0.1247	0.7387	-0.0364	0.0982	0.8451
$\rho_{4,6}$	0.00	0.0332	0.1111	0.8769	0.0014	0.2494	0.9296
$\rho_{4,7}$	0.00	0.0017	0.1188	0.9069	0.0758	0.2544	0.8169
$\rho_{4,8}$	0.00	0.0645	0.1047	0.8288	-0.0049	0.1108	0.8732
$\rho_{4,9}$	0.00	0.0188	0.1213	0.9259	0.0203	0.2027	0.7465
$\rho_{4,10}$	0.00	0.0388	0.1160	0.9129	0.0370	0.2248	0.8310
$\rho_{4,11}$	0.00	0.0481	0.1080	0.8258	0.0161	0.1380	0.9155
$\rho_{4,12}$	0.00	0.0215	0.0935	0.8338	0.0019	0.1162	0.7606

Table D.6: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43 of the 12–dimensional Elliptical Stable simulated experiment discussed in Section 3.4.2, for $\alpha = 1.7$ and sample size $n = 500$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
$\rho_{5,6}$	0.00	0.0033	0.0968	0.8388	0.0131	0.1406	0.8310
$\rho_{5,7}$	0.00	-0.0085	0.1079	0.8869	0.0558	0.2875	0.6338
$\rho_{5,8}$	0.00	0.0327	0.1162	0.9510	-0.0063	0.0982	0.8732
$\rho_{5,9}$	0.00	0.0019	0.1008	0.8559	0.0345	0.1175	0.8592
$\rho_{5,10}$	0.00	0.0115	0.0830	0.7818	0.0160	0.1255	0.8028
$\rho_{5,11}$	0.00	0.0250	0.1201	0.9359	-0.0151	0.1778	0.8873
$\rho_{5,12}$	-0.1339	0.0810	0.1064	0.8458	0.0038	0.1013	0.8310
$\rho_{6,7}$	-0.4896	0.1500	0.1010	0.8919	0.2298	0.2295	0.6620
$\rho_{6,8}$	0.0091	0.1500	0.1010	0.5075	0.0128	0.1550	0.8451
$\rho_{6,9}$	0.00	0.0518	0.1183	0.9199	0.0395	0.2633	0.7324
$\rho_{6,10}$	0.00	0.0016	0.1128	0.8949	-0.0026	0.1208	0.8028
$\rho_{6,11}$	0.00	0.0141	0.1024	0.8298	0.0130	0.1407	0.9577
$\rho_{6,12}$	0.00	0.0255	0.1184	0.9299	-0.0029	0.0980	0.8732
$\rho_{7,8}$	0.00	0.0238	0.1043	0.9029	0.0273	0.1519	0.9859
$\rho_{7,9}$	0.00	0.0213	0.1286	0.9760	0.0112	0.1849	0.9437
$\rho_{7,10}$	0.00	0.0059	0.0933	0.9489	0.0359	0.2611	0.9014
$\rho_{7,11}$	0.00	0.0088	0.0980	0.9590	-0.0110	0.1547	0.9718
$\rho_{7,12}$	0.00	0.0107	0.1226	0.9760	-0.0053	0.1277	0.9718
$\rho_{8,9}$	0.3843	0.0148	0.1171	0.9700	-0.1328	0.2150	0.8873
$\rho_{8,10}$	-0.1123	-0.1470	0.1576	0.8268	0.0595	0.1942	0.8732
$\rho_{8,11}$	-0.0495	0.1179	0.1158	0.9760	0.0792	0.2163	0.8451
$\rho_{8,12}$	0.00	0.0689	0.1148	0.8849	0.0220	0.1813	0.8310
$\rho_{9,10}$	-0.4356	0.1449	0.1112	0.6827	0.2272	0.2493	0.8592
$\rho_{9,11}$	-0.3136	0.1843	0.1120	0.7017	0.1211	0.2324	0.8732
$\rho_{9,12}$	0.00	0.0199	0.1193	0.9600	-0.0071	0.2032	0.8732
$\rho_{10,11}$	0.6803	-0.1749	0.1185	0.8268	-0.2771	0.2994	0.9296
$\rho_{10,12}$	0.00	0.0542	0.1274	0.9439	0.0185	0.2285	0.9437
$\rho_{11,12}$	0.00	0.0811	0.1263	0.9299	-0.0152	0.2901	1.0000

Table D.7: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43, of the 12–dimensional Elliptical Stable simulated experiment discussed in Section 3.4.2, for $\alpha = 1.7$ and sample size $n = 500$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.90**	-0.0347	0.1139	0.9940	-0.0018	0.0239	1.0000
ξ_1	0.00	0.0002	0.0306	0.9113	-0.0044	0.0187	0.9351
ξ_2	0.00	-0.0031	0.0756	0.9173	-0.0014	0.0496	0.9351
ξ_3	0.00	0.0104	0.0366	0.9002	0.0108	0.0239	0.9091
ξ_4	0.00	0.0011	0.0272	0.9093	0.0043	0.0186	0.9481
ξ_5	0.00	0.0007	0.0346	0.9264	0.0014	0.0222	0.9481
ξ_6	0.00	0.0017	0.0329	0.9062	0.0021	0.0219	0.9481
ξ_7	0.00	-0.0048	0.0515	0.9163	-0.0028	0.0328	0.9481
ξ_8	0.00	0.0004	0.0253	0.9022	-0.0017	0.0155	0.9221
ξ_9	0.00	0.0010	0.0433	0.9123	0.0013	0.0314	0.9351
ξ_{10}	0.00	-0.0076	0.0367	0.9062	0.0005	0.0259	0.9221
ξ_{11}	0.00	0.0028	0.0245	0.9042	0.0008	0.0168	0.9610
$\sqrt{\omega_{12}}$	0.00	0.0021	0.0275	0.9204	-0.0039	0.0173	0.9481
$\sqrt{\omega_{1,1}}$	0.4889	-0.0123	0.0259	0.5323	0.0018	0.0131	0.7792
$\sqrt{\omega_{2,2}}$	1.2466	-0.0144	0.0608	0.4980	0.0001	0.0285	0.7662
$\sqrt{\omega_{3,3}}$	0.6017	0.0011	0.0309	0.5756	0.0044	0.0130	0.7662
$\sqrt{\omega_{4,4}}$	0.4461	-0.0100	0.0261	0.5423	0.0014	0.0123	0.8052
$\sqrt{\omega_{5,5}}$	0.5908	-0.0064	0.0297	0.5655	-0.0008	0.0141	0.7922
$\sqrt{\omega_{6,6}}$	0.5431	-0.0092	0.0284	0.5454	-0.0003	0.0136	0.7532
$\sqrt{\omega_{7,7}}$	0.8456	-0.0117	0.0417	0.5333	-0.0021	0.0262	0.6364
$\sqrt{\omega_{8,8}}$	0.4050	-0.0125	0.0233	0.5242	-0.0006	0.0109	0.7662
$\sqrt{\omega_{9,9}}$	0.7197	-0.0158	0.0368	0.4980	0.0024	0.0162	0.8052
$\sqrt{\omega_{10,10}}$	0.6156	-0.0116	0.0293	0.5323	-0.0011	0.0163	0.7662
$\sqrt{\omega_{11,11}}$	0.3987	-0.0051	0.0224	0.5675	0.0003	0.0098	0.8442
$\sqrt{\omega_{12,12}}$	0.4550	-0.0053	0.0257	0.5665	0.0002	0.0126	0.7792

Table D.8: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the tail parameter α, the locations $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_1 2)$ and the scale matrix $\boldsymbol{\Omega} = \{\omega_{ij}\}$, with $i, j = 1, 2, \dots, 12$ and $i \leq j$ of the 12–dimensional Elliptical Stable simulated experiment discussed in Section 3.4.2, for $\alpha = 1.9$ and sample size $n = 500$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
$\rho_{1,2}$	0.1920	-0.0920	0.1656	0.7631	-0.0167	0.1454	0.8831
$\rho_{1,3}$	0.00	0.0010	0.1042	0.9032	0.0182	0.1415	0.7662
$\rho_{1,4}$	0.00	0.0316	0.1087	0.7510	-0.0087	0.0864	0.8571
$\rho_{1,5}$	0.00	0.0157	0.1077	0.8831	0.0057	0.1103	0.8312
$\rho_{1,6}$	0.00	0.0151	0.0934	0.8317	0.0048	0.1209	0.7792
$\rho_{1,7}$	0.00	0.0036	0.1195	0.8599	0.0507	0.1600	0.7403
$\rho_{1,8}$	0.1566	-0.0112	0.1193	0.8377	-0.0516	0.0924	0.7662
$\rho_{1,9}$	0.00	0.0112	0.1117	0.9113	0.0001	0.1207	0.7792
$\rho_{1,10}$	0.00	0.0189	0.1108	0.8841	-0.0041	0.1632	0.8961
$\rho_{1,11}$	0.00	0.0333	0.1258	0.8216	-0.0069	0.0870	0.9091
$\rho_{1,12}$	0.00	0.0274	0.0994	0.7782	-0.0004	0.0777	0.8312
$\rho_{2,3}$	0.00	-0.0293	0.1107	0.7782	-0.0438	0.1309	0.7532
$\rho_{2,4}$	0.00	-0.0215	0.1322	0.7500	-0.0038	0.1147	0.7532
$\rho_{2,5}$	0.00	-0.0212	0.1159	0.7742	-0.0138	0.0985	0.7792
$\rho_{2,6}$	0.00	-0.0155	0.1317	0.7298	0.0050	0.1140	0.7273
$\rho_{2,7}$	0.00	-0.0303	0.1011	0.8337	0.0026	0.0766	0.9091
$\rho_{2,8}$	0.00	-0.0072	0.1498	0.7692	0.0146	0.1239	0.8052
$\rho_{2,9}$	0.00	-0.0242	0.1007	0.8125	0.0054	0.0863	0.8052
$\rho_{2,10}$	0.00	-0.0228	0.1016	0.8034	0.0042	0.1153	0.7662
$\rho_{2,11}$	0.00	-0.0177	0.1428	0.7298	-0.0214	0.1469	0.7662
$\rho_{2,12}$	0.00	-0.0165	0.1351	0.7601	-0.0083	0.1033	0.8182
$\rho_{3,4}$	0.0075	0.0056	0.1162	0.9143	-0.0071	0.0855	0.9351
$\rho_{3,5}$	0.00	0.0043	0.0846	0.8024	-0.0059	0.0732	0.7792
$\rho_{3,6}$	0.00	-0.0007	0.0985	0.8165	0.0009	0.0830	0.8571
$\rho_{3,7}$	0.00	-0.0237	0.1070	0.8659	0.0452	0.1646	0.7792
$\rho_{3,8}$	0.00	0.0114	0.1236	0.9153	-0.0192	0.0939	0.9091
$\rho_{3,9}$	0.00	-0.0080	0.0991	0.8427	-0.0061	0.0981	0.8182
$\rho_{3,10}$	0.00	0.0091	0.0820	0.8024	0.0086	0.0652	0.9091
$\rho_{3,11}$	0.00	0.0038	0.1213	0.9345	-0.0241	0.0951	0.9091
$\rho_{3,12}$	0.00	0.0063	0.1169	0.8972	0.0016	0.0922	0.8831
$\rho_{4,5}$	0.3567	-0.0935	0.1276	0.7470	-0.0403	0.0873	0.8571
$\rho_{4,6}$	0.00	0.0320	0.1124	0.9093	0.0651	0.2488	0.7792
$\rho_{4,7}$	0.00	-0.0006	0.1221	0.8871	0.0348	0.2381	0.7273
$\rho_{4,8}$	0.00	0.0660	0.1133	0.7984	-0.0122	0.0818	0.9091
$\rho_{4,9}$	0.00	0.0157	0.1223	0.9224	0.0138	0.1548	0.8701
$\rho_{4,10}$	0.00	0.0389	0.1225	0.9143	0.0246	0.1724	0.7922
$\rho_{4,11}$	0.00	0.0602	0.1088	0.8357	-0.0234	0.0917	0.8571
$\rho_{4,12}$	0.00	0.0328	0.1045	0.8125	0.0051	0.0721	0.8831

Table D.9: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43 of the 12–dimensional Elliptical Stable simulated experiment discussed in Section 3.4.2, for $\alpha = 1.9$ and sample size $n = 500$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
$\rho_{5,6}$	0.00	0.0036	0.0986	0.8438	-0.0025	0.1236	0.8312
$\rho_{5,7}$	0.00	-0.0076	0.1112	0.8901	0.0199	0.1991	0.7532
$\rho_{5,8}$	0.00	0.0373	0.1272	0.9405	-0.0204	0.1208	0.8701
$\rho_{5,9}$	0.00	0.0065	0.1031	0.8669	0.0381	0.2042	0.7403
$\rho_{5,10}$	0.00	0.0187	0.0854	0.8266	0.0179	0.1157	0.8571
$\rho_{5,11}$	0.00	0.0322	0.1237	0.9355	0.0094	0.0875	0.9481
$\rho_{5,12}$	-0.1339	0.0763	0.1085	0.8468	0.0016	0.0973	0.8571
$\rho_{6,7}$	-0.4896	0.1313	0.0949	0.5948	0.1281	0.1835	0.7013
$\rho_{6,8}$	0.0091	0.0511	0.1225	0.9375	-0.0116	0.1063	0.9610
$\rho_{6,9}$	0.00	0.0117	0.1066	0.9284	0.0111	0.1607	0.8701
$\rho_{6,10}$	0.00	0.0172	0.0967	0.8629	0.0170	0.1836	0.7532
$\rho_{6,11}$	0.00	0.0324	0.1199	0.9315	-0.0168	0.0937	0.9610
$\rho_{6,12}$	0.00	0.0232	0.1083	0.8972	-0.0164	0.1106	0.9351
$\rho_{7,8}$	0.00	0.0152	0.1272	0.9657	0.0249	0.1489	0.9740
$\rho_{7,9}$	0.00	-0.0041	0.0978	0.9677	-0.0002	0.1557	0.9351
$\rho_{7,10}$	0.00	0.0059	0.1051	0.9597	-0.0174	0.1851	0.9091
$\rho_{7,11}$	0.00	0.0119	0.1298	0.9617	-0.0124	0.1221	0.9740
$\rho_{7,12}$	0.00	0.0031	0.1176	0.9677	0.0207	0.1586	0.9221
$\rho_{8,9}$	0.3843	-0.1229	0.1516	0.8619	-0.1415	0.1896	0.8571
$\rho_{8,10}$	-0.1123	0.1254	0.1269	0.8327	0.0303	0.2223	0.8442
$\rho_{8,11}$	-0.0495	0.1255	0.1164	0.7540	0.0506	0.1446	0.8571
$\rho_{8,12}$	0.00	0.0787	0.1233	0.8629	0.0692	0.1981	0.9091
$\rho_{9,10}$	-0.4356	0.1372	0.1027	0.7056	0.1518	0.2232	0.7922
$\rho_{9,11}$	-0.3136	0.1843	0.1157	0.6976	0.0917	0.1724	0.7922
$\rho_{9,12}$	0.00	0.0279	0.1220	0.9466	0.0002	0.1611	0.9351
$\rho_{10,11}$	0.6803	-0.1719	0.1204	0.8317	-0.1981	0.2618	0.9481
$\rho_{10,12}$	0.00	0.0569	0.1296	0.9567	0.0025	0.1785	0.9870
$\rho_{11,12}$	0.00	0.0886	0.1395	0.9365	0.0103	0.2176	0.9740

Table D.10: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43 of the 12–dimensional Elliptical Stable simulated experiment discussed in Section 3.4.2, for $\alpha = 1.9$ and sample size $n = 500$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.95**	-0.0805	0.2344	0.9837	0.0461	0.0263	1.0000
ξ_1	0.00	-0.0006	0.0309	0.9000	0.0003	0.0177	0.9730
ξ_2	0.00	-0.0055	0.0763	0.9173	0.0029	0.0491	0.9324
ξ_3	0.00	0.0085	0.0366	0.9153	0.0024	0.0247	0.9595
ξ_4	0.00	-0.0007	0.0267	0.9122	0.0040	0.0193	0.9324
ξ_5	0.00	0.0006	0.0355	0.9214	0.0005	0.0209	0.9730
ξ_6	0.00	0.0030	0.0328	0.9102	0.0032	0.0217	0.9459
ξ_7	0.00	-0.0054	0.0520	0.9000	-0.0011	0.0361	0.9324
ξ_8	0.00	0.0010	0.0253	0.9061	0.0011	0.0163	0.9595
ξ_9	0.00	0.0025	0.0429	0.9143	0.0014	0.0300	0.9324
ξ_{10}	0.00	-0.0059	0.0366	0.9071	-0.0015	0.0265	0.9189
ξ_{11}	0.00	0.0027	0.0244	0.9041	0.0021	0.0172	0.9324
ξ_{12}	0.00	0.0003	0.0277	0.9214	-0.0006	0.0194	0.9054
$\sqrt{\omega_{1,1}}$	0.4889	-0.0114	0.0259	0.5378	0.0033	0.0137	0.7027
$\sqrt{\omega_{2,2}}$	1.2466	-0.0135	0.0605	0.5408	-0.0018	0.0395	0.5541
$\sqrt{\omega_{3,3}}$	0.6017	0.0029	0.0320	0.5510	0.0013	0.0138	0.7973
$\sqrt{\omega_{4,4}}$	0.4461	-0.0110	0.0247	0.5408	0.0017	0.0101	0.8514
$\sqrt{\omega_{5,5}}$	0.5908	-0.0054	0.0296	0.5531	-0.0008	0.0169	0.7432
$\sqrt{\omega_{6,6}}$	0.5431	-0.0079	0.0274	0.5439	-0.0018	0.0144	0.6892
$\sqrt{\omega_{7,7}}$	0.8456	-0.0108	0.0414	0.5378	-0.0022	0.0222	0.7432
$\sqrt{\omega_{8,8}}$	0.4050	-0.0121	0.0230	0.5173	-0.0016	0.0119	0.7703
$\sqrt{\omega_{9,9}}$	0.7197	-0.0159	0.0355	0.4908	0.0003	0.0194	0.7297
$\sqrt{\omega_{10,10}}$	0.6156	-0.0117	0.0285	0.5367	-0.0014	0.0147	0.7432
$\sqrt{\omega_{11,11}}$	0.3987	-0.0049	0.0221	0.5806	-0.0014	0.0091	0.8919
$\sqrt{\omega_{12,12}}$	0.4550	-0.0045	0.0253	0.5622	0.0003	0.0110	0.8649

Table D.11: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the tail parameter α, the locations $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_12)$ and the scale matrix $\boldsymbol{\Omega} = \{\omega_{ij}\}$, with $i, j = 1, 2, \ldots, 12$ and $i \leq j$ of the 12–dimensional Elliptical Stable simulated experiment discussed in Section 3.4.2, for $\alpha = 1.95$ and sample size $n = 500$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
$\rho_{1,2}$	0.1920	-0.0909	0.1710	0.7602	-0.0190	0.1610	0.8784
$\rho_{1,3}$	0.00	-0.0064	0.1089	0.8827	0.0480	0.1898	0.6892
$\rho_{1,4}$	0.00	0.0283	0.1030	0.8031	-0.0245	0.0809	0.8108
$\rho_{1,5}$	0.00	0.0091	0.1067	0.8704	0.0071	0.1175	0.7432
$\rho_{1,6}$	0.00	0.0125	0.0934	0.8531	0.0441	0.1515	0.7027
$\rho_{1,7}$	0.00	0.0014	0.1186	0.8571	0.0241	0.2180	0.5946
$\rho_{1,8}$	0.1566	-0.0229	0.1209	0.8388	-0.0459	0.0872	0.8108
$\rho_{1,9}$	0.00	0.0137	0.1164	0.8888	0.0399	0.1791	0.7973
$\rho_{1,10}$	0.00	0.0219	0.1073	0.8908	-0.0033	0.1594	0.7027
$\rho_{1,11}$	0.00	0.0285	0.1260	0.8408	-0.0125	0.0674	0.9324
$\rho_{1,12}$	0.00	0.0165	0.0968	0.7867	-0.0009	0.0788	0.8243
$\rho_{2,3}$	0.00	-0.0310	0.1141	0.7827	0.0150	0.1124	0.8108
$\rho_{2,4}$	0.00	-0.0061	0.1423	0.7735	-0.0263	0.1328	0.6081
$\rho_{2,5}$	0.00	-0.0125	0.1189	0.7878	0.0019	0.1120	0.7162
$\rho_{2,6}$	0.00	-0.0043	0.1387	0.7347	0.0065	0.1412	0.7432
$\rho_{2,7}$	0.00	-0.0260	0.1069	0.8245	0.0080	0.0930	0.7162
$\rho_{2,8}$	0.00	0.0051	0.1555	0.7643	0.0139	0.1417	0.7297
$\rho_{2,9}$	0.00	-0.0172	0.1097	0.8306	0.0050	0.0936	0.7973
$\rho_{2,10}$	0.00	-0.0172	0.0998	0.8173	0.0162	0.1052	0.8108
$\rho_{2,11}$	0.00	-0.0093	0.1458	0.7612	0.0210	0.1235	0.7568
$\rho_{2,12}$	0.00	-0.0161	0.1356	0.7571	0.0213	0.1211	0.7568
$\rho_{3,4}$	0.0075	-0.0007	0.1175	0.9143	-0.0027	0.0928	0.8919
$\rho_{3,5}$	0.00	-0.0021	0.0824	0.8429	-0.0060	0.0658	0.8784
$\rho_{3,6}$	0.00	-0.0101	0.1002	0.8235	-0.0056	0.0957	0.8243
$\rho_{3,7}$	0.00	-0.0252	0.1098	0.8735	0.0279	0.1355	0.7568
$\rho_{3,8}$	0.00	0.0044	0.1214	0.9276	0.0094	0.0927	0.9324
$\rho_{3,9}$	0.00	-0.0161	0.1033	0.8367	0.0539	0.1940	0.7838
$\rho_{3,10}$	0.00	-0.0013	0.0840	0.7959	0.0011	0.0661	0.7297
$\rho_{3,11}$	0.00	-0.0039	0.1280	0.9000	0.0087	0.1048	0.8378
$\rho_{3,12}$	0.00	-0.0089	0.1174	0.9133	0.0130	0.0889	0.9324
$\rho_{4,5}$	0.3567	-0.1032	0.1305	0.7551	-0.0405	0.0710	0.8243
$\rho_{4,6}$	0.00	0.0330	0.1098	0.9143	0.0138	0.1905	0.7568
$\rho_{4,7}$	0.00	0.0084	0.1264	0.8857	0.0399	0.2155	0.7027
$\rho_{4,8}$	0.00	0.0674	0.1119	0.8245	-0.0204	0.0889	0.9189
$\rho_{4,9}$	0.00	0.0200	0.1215	0.9184	0.0164	0.1944	0.7973
$\rho_{4,10}$	0.00	0.0362	0.1217	0.9173	0.0147	0.1820	0.8243
$\rho_{4,11}$	0.00	0.0627	0.1136	0.8306	-0.0215	0.0803	0.9054
$\rho_{4,12}$	0.00	0.0328	0.0967	0.8408	0.0087	0.0720	0.9189

Table D.12: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43 of the 12–dimensional Elliptical Stable simulated experiment discussed in Section 3.4.2, for $\alpha = 1.95$ and sample size $n = 500$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
$\rho_{5,6}$	0.00	0.0007	0.0988	0.8378	0.0052	0.1081	0.8514
$\rho_{5,7}$	0.00	-0.0110	0.1131	0.8939	0.0444	0.1727	0.7297
$\rho_{5,8}$	0.00	0.0388	0.1291	0.9327	0.0139	0.1192	0.9189
$\rho_{5,9}$	0.00	0.0047	0.1044	0.8541	0.0410	0.1655	0.7568
$\rho_{5,10}$	0.00	0.0170	0.0880	0.8153	0.0014	0.0820	0.7973
$\rho_{5,11}$	0.00	0.0262	0.1196	0.9378	0.0106	0.1374	0.8649
$\rho_{5,12}$	-0.1339	0.0783	0.1133	0.8102	0.0321	0.0949	0.8108
$\rho_{6,7}$	-0.4896	0.1294	0.1007	0.6439	0.1291	0.1880	0.6757
$\rho_{6,8}$	0.0091	0.0481	0.1270	0.9184	-0.0002	0.1228	0.9054
$\rho_{6,9}$	0.00	0.0049	0.1095	0.9245	0.0316	0.2112	0.8108
$\rho_{6,10}$	0.00	0.0100	0.1017	0.8592	-0.0020	0.1551	0.8108
$\rho_{6,11}$	0.00	0.0242	0.1199	0.9480	0.0307	0.1140	0.9459
$\rho_{6,12}$	0.00	0.0135	0.1083	0.9214	-0.0042	0.1346	0.9459
$\rho_{7,8}$	0.00	0.0194	0.1363	0.9724	0.0087	0.1219	0.9865
$\rho_{7,9}$	0.00	0.0016	0.0942	0.9776	-0.0079	0.1735	0.9189
$\rho_{7,10}$	0.00	0.0096	0.1040	0.9704	0.0230	0.1575	0.9324
$\rho_{7,11}$	0.00	0.0178	0.1354	0.9684	0.0134	0.1583	0.9324
$\rho_{7,12}$	0.00	0.0099	0.1258	0.9735	-0.0087	0.1776	0.9459
$\rho_{8,9}$	0.3843	-0.1212	0.1638	0.8439	-0.1244	0.2053	0.8784
$\rho_{8,10}$	-0.1123	0.1288	0.1316	0.8510	0.0646	0.1509	0.8649
$\rho_{8,11}$	-0.0495	0.1280	0.1272	0.7531	0.0498	0.1478	0.8784
$\rho_{8,12}$	0.00	0.0774	0.1310	0.8735	0.0647	0.2082	0.8514
$\rho_{9,10}$	-0.4356	0.1453	0.1283	0.7286	0.1621	0.2234	0.8108
$\rho_{9,11}$	-0.3136	0.1822	0.1273	0.7306	0.0982	0.1659	0.7973
$\rho_{9,12}$	0.00	0.0317	0.1372	0.9582	0.0049	0.1566	0.9324
$\rho_{10,11}$	0.6803	-0.1694	0.1393	0.8367	-0.2226	0.2684	0.9054
$\rho_{10,12}$	0.00	0.0501	0.1396	0.9673	-0.0127	0.1506	1.0000
$\rho_{11,12}$	0.00	0.0970	0.1546	0.9459	-0.0266	0.2943	1.0000

Table D.13: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43 of the 12–dimensional Elliptical Stable simulated experiment discussed in Section 3.4.2, for $\alpha = 1.95$ and sample size $n = 500$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.70**	-0.0134	0.0922	0.8518	-0.0002	0.0457	0.8941
δ_1	0.90	-0.0192	0.0534	0.8998	-0.0175	0.0263	0.8971
δ_2	0.90	-0.0207	0.0539	0.9061	-0.0164	0.0270	0.9114
ξ_1	0.00	0.0007	0.0233	0.9885	0.0003	0.0120	0.9837
ξ_2	0.00	-0.0001	0.0478	0.9823	0.0014	0.0229	0.9919
ω_1	0.50	-0.0012	0.1047	0.9238	-0.0053	0.0496	0.9521
ω_2	2.00	-0.0059	0.4056	0.9405	-0.0172	0.2052	0.9450
ρ_{12}	0.90	-0.0849	0.0944	0.9885	-0.0731	0.0630	0.9827
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.90**	-0.0309	0.0784	0.9399	-0.0087	0.0464	0.9537
δ_1	0.90	-0.0326	0.0684	0.9695	-0.0177	0.0276	0.9698
δ_2	0.90	-0.0318	0.0627	0.9644	-0.0177	0.0288	0.9678
ξ_1	0.00	0.0008	0.0217	0.9868	0.0002	0.0112	0.9839
ξ_2	0.00	0.0009	0.0463	0.9786	0.0014	0.0223	0.9869
ω_1	0.50	-0.0200	0.0938	0.9389	-0.0110	0.0511	0.9547
ω_2	2.00	-0.0585	0.3761	0.9450	-0.0431	0.2047	0.9497
ρ_{12}	0.90	-0.0758	0.0811	0.9980	-0.0491	0.0503	0.9960
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.95**	-0.0502	0.0693	0.9458	-0.0246	0.0380	0.9629
δ_1	0.90	-0.0435	0.0779	0.9857	-0.0246	0.0274	0.9890
δ_2	0.90	-0.0446	0.0754	0.9847	-0.0246	0.0274	0.9809
ξ_1	0.00	0.0006	0.0214	0.9898	0.0001	0.0111	0.9869
ξ_2	0.00	0.0021	0.0457	0.9806	0.0013	0.0221	0.9880
ω_1	0.50	-0.0313	0.0924	0.9387	-0.0204	0.0476	0.9639
ω_2	2.00	-0.1138	0.3671	0.9479	-0.0820	0.1861	0.9578
ρ_{12}	0.90	-0.0835	0.0893	0.9939	-0.0514	0.0485	0.9980

Table D.14: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the locations $\boldsymbol{\xi} = (\xi_1, \xi_2)$, scale matrix $\boldsymbol{\Omega} = \{\omega_{ij}\}$, with $i, j = 1, 2$ and $i \leq j$, the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43, tail parameter α and skewness parameter $\boldsymbol{\delta} = (\delta_1, \delta_2)$ of the Skew Elliptical Stable distribution in dimension 2. The results reported above are obtained using 1000 replications for three values of $\alpha = \{1.70, 1.90, 1.95\}$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.70**	-0.0068	0.0690	0.9500	0.0013	0.0320	0.9500
δ_1	0.00	0.0048	0.0067	0.9400	0.0022	0.0010	0.1100
δ_2	0.00	0.0048	0.0063	0.9500	0.0082	0.0016	0.0100
δ_3	0.00	0.0040	0.0038	0.9100	0.0013	0.0005	0.0600
δ_4	0.90	-0.0116	0.1648	0.9700	0.0180	0.0185	0.8100
δ_5	0.90	-0.0179	0.1649	0.9700	0.0167	0.0234	0.9200
ξ_1	0.00	0.0016	0.0365	0.9600	0.0032	0.0218	0.9400
ξ_2	0.00	-0.0029	0.0534	0.9700	0.0023	0.0286	0.9400
ξ_3	0.00	0.0093	0.0757	0.9400	0.0065	0.0393	0.9500
ξ_4	0.00	-0.0051	0.0703	0.9700	0.0041	0.0356	0.9500
ξ_5	0.00	-0.0059	0.1089	0.9200	-0.0040	0.0618	0.9400
ω_{11}	0.2500	-0.0126	0.0259	0.9400	-0.0027	0.0140	0.9800
ω_{22}	0.5000	0.0184	0.0596	0.9200	0.0003	0.0261	0.9400
ω_{33}	1.0000	0.0038	0.0998	0.9700	0.0166	0.0538	0.9500
ω_{44}	2.0000	-0.1397	0.3571	0.9300	-0.1571	0.1700	0.8800
ω_{55}	4.0000	-0.4342	0.6637	0.9100	-0.1142	0.3980	0.9600
ρ_{12}	0.7071	-0.0438	0.1336	0.9400	-0.0345	0.1055	0.9100
ρ_{13}	0.8000	-0.1043	0.1487	0.9200	-0.0173	0.1050	0.9800
ρ_{14}	0.00	0.0075	0.0256	0.9300	0.0018	0.0148	0.9400
ρ_{15}	0.00	0.0085	0.0445	0.9700	0.0040	0.0170	0.9400
ρ_{23}	0.5657	-0.0851	0.1680	0.9300	-0.0323	0.1255	0.9700
ρ_{24}	0.00	0.0049	0.0306	0.9600	0.0032	0.0154	0.9200
ρ_{25}	0.00	0.0076	0.0414	0.9300	0.0053	0.0172	0.9600
ρ_{34}	0.00	0.0047	0.0277	0.9100	0.0022	0.0151	0.9300
ρ_{35}	0.00	0.0100	0.0332	0.9500	0.0032	0.0151	0.9300
ρ_{45}	0.9016	-0.0727	0.0785	0.8200	-0.0552	0.0573	0.9300

Table D.15: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the locations $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_5)$, scale matrix $\boldsymbol{\Omega} = \{\omega_{ij}\}$, with $i, j = 1, 2, \ldots, 5$ and $i \leq j$, the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43, tail parameter α and skewness parameter $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_5)$ of the Skew Elliptical Stable distribution in dimension 5. The results reported above are obtained using 1000 replications for $\alpha = 1.70$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.90**	-0.0194	0.0550	0.9100	0.0019	0.0307	0.9700
δ_1	0.00	0.0010	0.0053	0.9400	0.0001	0.0002	0.9500
δ_2	0.00	-0.0014	0.0080	0.9100	0.0004	0.0005	0.9200
δ_3	0.00	0.0007	0.0059	0.9400	0.0001	0.0001	0.9600
δ_4	0.90	-0.1184	0.3072	0.8700	-0.0029	0.0230	0.9400
δ_5	0.90	-0.1307	0.3039	0.8700	-0.0019	0.0248	0.9600
ξ_1	0.00	-0.0038	0.0397	0.9500	0.0009	0.0191	0.9700
ξ_2	0.00	-0.0047	0.0563	0.9600	0.0040	0.0254	0.9600
ξ_3	0.00	-0.0103	0.0771	0.9500	0.0007	0.0374	0.9400
ξ_4	0.00	-0.0001	0.0778	0.9500	0.0007	0.0386	0.9400
ξ_5	0.00	0.0029	0.1028	0.9200	0.0079	0.0589	0.9300
$\sqrt{\omega}_{11}$	0.5000	-0.0068	0.0293	0.9600	-0.0011	0.0156	0.9400
$\sqrt{\omega}_{22}$	0.7071	0.0002	0.0562	0.9700	0.0029	0.0393	0.9700
$\sqrt{\omega}_{33}$	1.0000	0.0657	0.1227	0.9200	-0.0097	0.0601	0.9600
$\sqrt{\omega}_{44}$	1.4142	-0.1053	0.5032	0.8900	-0.0003	0.1924	0.9400
$\sqrt{\omega}_{55}$	2.0000	-0.4325	0.9533	0.8700	0.1150	0.3819	0.9300
ρ_{12}	0.7071	0.0233	0.1612	0.9700	0.0015	0.1263	0.9200
ρ_{13}	0.8000	-0.0808	0.1602	0.9000	-0.0439	0.1816	0.9400
ρ_{14}	0.00	0.0164	0.0521	0.9500	0.0108	0.0373	0.9500
ρ_{15}	0.00	0.0103	0.0358	0.9400	0.0090	0.0315	0.9300
ρ_{23}	0.5657	-0.0445	0.1748	0.9300	-0.0301	0.1313	0.9400
ρ_{24}	0.00	0.0132	0.0525	0.9500	0.0081	0.0280	0.9200
ρ_{25}	0.00	0.0113	0.0389	0.9100	0.0107	0.0391	0.9500
ρ_{34}	0.00	0.0137	0.0502	0.9500	0.0097	0.0389	0.9600
ρ_{35}	0.00	0.0073	0.0352	0.9200	0.0108	0.0340	0.9400
ρ_{45}	0.9016	-0.0894	0.0802	0.8300	-0.0394	0.0773	0.9800

Table D.16: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the locations $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_5)$, scale matrix $\boldsymbol{\Omega} = \{\omega_{ij}\}$, with $i, j = 1, 2, \ldots, 5$ and $i \leq j$, the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43, tail parameter α and skewness parameter $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_5)$ of the Skew Elliptical Stable distribution in dimension 5. The results reported above are obtained using 1000 replications for $\alpha = 1.90$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.95**	-0.0569	0.0476	0.8000	-0.0216	0.0229	0.8600
δ_1	0.00	0.0007	0.0048	0.9000	0.0001	0.0001	0.9100
δ_2	0.00	-0.0007	0.0066	0.9300	0.0003	0.0009	0.9800
δ_3	0.00	-0.0000	0.0071	0.9400	-0.0001	0.0013	0.9900
δ_4	0.90	-0.1477	0.3327	0.8400	-0.0152	0.0923	0.9900
δ_5	0.90	-0.1681	0.3285	0.8400	-0.0172	0.0928	0.9900
ξ_1	0.00	0.0013	0.0356	0.9600	-0.0004	0.0197	0.9500
ξ_2	0.00	-0.0006	0.0507	0.9400	-0.0036	0.0266	0.9300
ξ_3	0.00	0.0121	0.0700	0.9400	0.0016	0.0395	0.9500
ξ_4	0.00	-0.0118	0.0733	0.9300	0.0033	0.0333	0.9600
ξ_5	0.00	-0.0162	0.1173	0.9400	-0.0040	0.0587	0.9600
$\sqrt{\omega}_{11}$	0.5000	-0.0109	0.0259	0.9500	-0.0012	0.0161	0.9400
$\sqrt{\omega}_{22}$	0.7071	-0.0026	0.0504	0.9500	0.0002	0.0340	0.9600
$\sqrt{\omega}_{33}$	1.0000	0.0584	0.1033	0.9300	-0.0244	0.0584	0.9300
$\sqrt{\omega}_{44}$	1.4142	-0.1518	0.5501	0.8800	-0.0699	0.2097	0.9600
$\sqrt{\omega}_{55}$	2.0000	-0.4465	1.0475	0.8500	-0.0296	0.3965	0.9800
ρ_{12}	0.7071	-0.0120	0.1683	0.9800	-0.0147	0.1015	0.9400
ρ_{13}	0.8000	-0.0961	0.1402	0.9100	-0.0058	0.1633	0.9700
ρ_{14}	0.00	0.0051	0.0364	0.9400	0.0081	0.0330	0.9500
ρ_{15}	0.00	0.0178	0.0601	0.9500	0.0095	0.0295	0.9500
ρ_{23}	0.5657	-0.0742	0.1622	0.9300	-0.0153	0.1309	0.9400
ρ_{24}	0.00	0.0062	0.0436	0.9400	0.0070	0.0329	0.9600
ρ_{25}	0.00	0.0177	0.0604	0.9500	0.0122	0.0317	0.9300
ρ_{34}	0.00	0.0022	0.0354	0.9400	0.0059	0.0317	0.9500
ρ_{35}	0.00	0.0129	0.0543	0.9700	0.0075	0.0253	0.9400
ρ_{45}	0.9016	-0.0984	0.0807	0.8400	-0.0564	0.0773	0.9900

Table D.17: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the locations $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_5)$, scale matrix $\boldsymbol{\Omega} = \{\omega_{ij}\}$, with $i, j = 1, 2, \ldots, 5$ and $i \leq j$, the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43, tail parameter α and skewness parameter $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_5)$ of the Skew Elliptical Stable distribution in dimension 5. The results reported above are obtained using 1000 replications for $\alpha = 1.95$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.70**	-0.0159	0.0503	0.9500	-0.0060	0.0214	0.9600
δ_1	0.00	0.0067	0.0582	0.9900	0.0001	0.0001	0.8200
δ_2	0.00	0.0073	0.0658	0.9900	0.0005	0.0009	0.9900
δ_3	0.60	-0.2108	0.3134	0.9900	-0.1617	0.2923	0.7000
δ_4	0.00	0.0189	0.1030	0.9700	0.0002	0.0004	0.9900
δ_5	0.00	0.0109	0.0713	0.9800	0.0003	0.0003	0.9500
δ_6	0.00	0.0122	0.0818	0.9800	0.0002	0.0004	0.9900
δ_7	0.00	0.0010	0.0028	0.9600	0.0002	0.0001	0.4600
δ_8	0.00	0.0144	0.0962	0.9800	0.0001	0.0001	0.8600
δ_9	0.00	0.0008	0.0035	0.9700	0.0002	0.0006	0.9800
δ_{10}	0.60	-0.0610	0.2530	0.8300	-0.0957	0.2707	0.7900
δ_{11}	0.60	-0.1717	0.3088	1.0000	-0.0913	0.2922	0.7600
δ_{12}	0.00	0.0072	0.0636	0.9900	0.0002	0.0003	0.9900
ξ_1	0.00	-0.0003	0.0380	0.9700	-0.0037	0.0195	0.9600
ξ_2	0.00	0.0003	0.1024	0.9400	-0.0017	0.0494	0.9200
ξ_3	0.00	-0.0023	0.0440	0.9600	0.0033	0.0226	0.9300
ξ_4	0.00	0.0036	0.0331	0.9400	0.0038	0.0193	0.9700
ξ_5	0.00	-0.0022	0.0466	0.9400	-0.0008	0.0253	0.9700
ξ_6	0.00	0.0023	0.0442	0.9200	0.0005	0.0214	0.9400
ξ_7	0.00	-0.0017	0.0645	0.9200	0.0065	0.0358	0.9500
ξ_8	0.00	-0.0020	0.0312	0.9400	-0.0029	0.0145	0.9300
ξ_9	0.00	-0.0110	0.0534	0.9500	-0.0005	0.0325	0.9500
ξ_{10}	0.00	0.0009	0.0411	0.9600	-0.0018	0.0256	0.9500
ξ_{11}	0.00	0.0018	0.0304	0.9100	-0.0024	0.0138	0.9700
ξ_{12}	0.00	0.0060	0.0364	0.9400	0.0018	0.0177	0.9400
$\sqrt{\omega_{1,1}}$	0.2390	0.0118	0.0258	0.9500	0.0004	0.0132	0.9600
$\sqrt{\omega_{2,2}}$	1.5540	0.0038	0.1600	0.9600	-0.0398	0.0835	0.9600
$\sqrt{\omega_{3,3}}$	0.3620	-0.0234	0.0700	0.9700	-0.0184	0.0470	0.9500
$\sqrt{\omega_{4,4}}$	0.1990	0.0014	0.0233	0.9400	-0.0031	0.0106	0.9700
$\sqrt{\omega_{5,5}}$	0.3490	-0.0132	0.0350	0.9200	0.0102	0.0187	0.9200
$\sqrt{\omega_{6,6}}$	0.2950	-0.0017	0.0304	0.9600	0.0030	0.0161	0.9400
$\sqrt{\omega_{7,7}}$	0.7150	0.0281	0.0738	0.9300	-0.0008	0.0361	0.9400
$\sqrt{\omega_{8,8}}$	0.1640	0.0113	0.0199	0.9400	-0.0011	0.0093	0.9500
$\sqrt{\omega_{9,9}}$	0.5180	0.0015	0.0559	0.9700	0.0058	0.0307	0.9500
$\sqrt{\omega_{10,10}}$	0.3790	0.0145	0.0800	0.9300	-0.0041	0.0506	0.9500
$\sqrt{\omega_{11,11}}$	0.1590	-0.0084	0.0276	0.9500	0.0003	0.0211	0.9800
$\sqrt{\omega_{12,12}}$	0.2070	0.0096	0.0257	0.9600	-0.0045	0.0116	0.9200
$\rho_{1,2}$	0.1920	-0.0266	0.2149	0.9400	0.0453	0.1629	0.9300
$\rho_{1,3}$	0.00	0.0079	0.1375	0.9400	0.0172	0.1068	0.9300
$\rho_{1,4}$	0.00	-0.0517	0.1815	0.9000	0.0004	0.0995	0.9500
$\rho_{1,5}$	0.00	0.0116	0.1596	0.9000	-0.0335	0.1749	0.9600
$\rho_{1,6}$	0.00	0.0421	0.1613	0.9400	-0.0148	0.1045	0.9600
$\rho_{1,7}$	0.00	-0.0649	0.1821	0.8900	0.0241	0.2593	0.9400
$\rho_{1,8}$	0.1566	-0.0539	0.1670	0.9400	0.0049	0.1011	0.9800
$\rho_{1,9}$	0.00	-0.0406	0.2244	0.9200	0.0009	0.2168	0.9300
$\rho_{1,10}$	0.00	0.0376	0.1915	0.9100	0.0014	0.1933	0.9600
$\rho_{1,11}$	0.00	-0.0382	0.2085	0.9500	-0.0218	0.1196	0.9500
$\rho_{1,12}$	0.00	-0.0246	0.1994	0.9200	-0.0132	0.0851	0.9500

Table D.18: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the locations $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_12)$, scale matrix $\mathbf{\Omega} = \{\omega_{ij}\}$, with $i, j = 1, 2, \ldots, 12$ and $i \leq j$, the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43, tail parameter α and skewness parameters $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_12)$ of the Skew Elliptical Stable distribution in dimension 12. The results reported above are obtained using 1000 replications for $\alpha = 1.70$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
$\rho_{2,3}$	0.00	-0.0105	0.1795	0.9100	0.0234	0.1599	0.9400
$\rho_{2,4}$	0.00	-0.0169	0.1946	0.9100	0.0542	0.1462	0.9300
$\rho_{2,5}$	0.00	0.0007	0.1957	0.9100	0.0419	0.1444	0.9300
$\rho_{2,6}$	0.00	-0.0165	0.1929	0.9300	0.0286	0.1115	0.9500
$\rho_{2,7}$	0.00	-0.0172	0.1815	0.9200	0.0152	0.1297	0.9400
$\rho_{2,8}$	0.00	0.0390	0.2026	0.9400	0.0600	0.1269	0.9100
$\rho_{2,9}$	0.00	0.0147	0.1834	0.9100	0.0170	0.1057	0.9400
$\rho_{2,10}$	0.00	0.0419	0.1403	0.9200	0.0468	0.0969	0.9000
$\rho_{2,11}$	0.00	-0.0288	0.2023	0.9200	0.0317	0.1141	0.9500
$\rho_{2,12}$	0.00	0.0354	0.1900	0.9400	0.0044	0.1443	0.9400
$\rho_{3,4}$	0.0075	-0.0010	0.1670	0.9100	0.0025	0.1236	0.9400
$\rho_{3,5}$	0.00	-0.0274	0.1688	0.9000	0.0156	0.1066	0.9100
$\rho_{3,6}$	0.00	-0.0237	0.1640	0.9300	0.0101	0.1124	0.9500
$\rho_{3,7}$	0.00	0.0024	0.1827	0.9400	-0.0398	0.2298	0.9500
$\rho_{3,8}$	0.00	-0.0050	0.1748	0.9200	0.0386	0.1506	0.9400
$\rho_{3,9}$	0.00	0.0180	0.1440	0.9000	0.0320	0.1783	0.9200
$\rho_{3,10}$	0.00	-0.0318	0.1876	0.9300	-0.0503	0.1701	0.9700
$\rho_{3,11}$	0.00	-0.0718	0.1811	0.9400	-0.0172	0.2102	0.9600
$\rho_{3,12}$	0.00	0.0207	0.1546	0.9400	0.0219	0.1212	0.9500
$\rho_{4,5}$	0.3567	-0.1032	0.1520	0.9300	-0.0624	0.0994	0.9200
$\rho_{4,6}$	0.00	0.0196	0.1824	0.9300	0.0531	0.1325	0.9500
$\rho_{4,7}$	0.00	-0.0489	0.1873	0.9400	-0.0270	0.3218	0.8900
$\rho_{4,8}$	0.00	0.0531	0.1754	0.9200	0.0233	0.2051	0.9400
$\rho_{4,9}$	0.00	-0.0043	0.1681	0.9600	-0.0077	0.2506	0.9100
$\rho_{4,10}$	0.00	0.0482	0.1740	0.9200	0.0323	0.2024	0.9400
$\rho_{4,11}$	0.00	0.0378	0.1811	0.9200	0.0163	0.1125	0.9200
$\rho_{4,12}$	0.00	0.0419	0.1402	0.9000	0.0440	0.0984	0.9300
$\rho_{5,6}$	0.00	0.0267	0.1981	0.9400	0.0360	0.1357	0.9700
$\rho_{5,7}$	0.00	-0.0269	0.1475	0.9400	-0.0229	0.2186	0.9300
$\rho_{5,8}$	0.00	0.0214	0.1678	0.9200	0.0315	0.0972	0.9300
$\rho_{5,9}$	0.00	-0.0103	0.1677	0.9200	0.0169	0.1878	0.9300
$\rho_{5,10}$	0.00	0.0341	0.1338	0.9000	-0.0102	0.0990	0.9300
$\rho_{5,11}$	0.00	0.0221	0.1602	0.9400	0.0285	0.1139	0.9300
$\rho_{5,12}$	-0.1339	0.0816	0.1486	0.9300	0.0691	0.1061	0.9200
$\rho_{6,7}$	-0.4896	0.2006	0.1852	0.8200	0.1638	0.2047	0.8500
$\rho_{6,8}$	0.0091	0.0595	0.1683	0.8800	-0.0068	0.1273	0.9500
$\rho_{6,9}$	0.00	0.0330	0.1702	0.9100	0.0073	0.1722	0.9200
$\rho_{6,10}$	0.00	0.0519	0.1300	0.9000	-0.0031	0.1453	0.9500
$\rho_{6,11}$	0.00	0.0631	0.1673	0.9300	-0.0160	0.1309	0.9500
$\rho_{6,12}$	0.00	0.0560	0.1740	0.9200	0.0176	0.1521	0.9400
$\rho_{7,8}$	0.00	-0.0517	0.2156	0.9000	0.0590	0.1520	0.9400
$\rho_{7,9}$	0.00	0.0255	0.1736	0.9200	-0.0013	0.1718	0.9300
$\rho_{7,10}$	0.00	-0.0258	0.1313	0.9500	0.0297	0.1596	0.9200
$\rho_{7,11}$	0.00	-0.0076	0.1327	0.9000	0.0537	0.1590	0.9200
$\rho_{7,12}$	0.00	-0.0555	0.1804	0.9600	0.0340	0.1239	0.9400
$\rho_{8,9}$	0.3843	-0.1052	0.1801	0.9300	-0.0929	0.1746	0.9000
$\rho_{8,10}$	-0.1123	0.0613	0.1531	0.9400	0.0361	0.1795	0.9300
$\rho_{8,11}$	-0.0495	0.0525	0.1673	0.9300	0.0496	0.1159	0.9400
$\rho_{8,12}$	0.00	0.0202	0.1745	0.9400	0.0427	0.1552	0.9300
$\rho_{9,10}$	-0.4356	0.2159	0.1607	0.8000	0.2032	0.2154	0.8000
$\rho_{9,11}$	-0.3136	0.1695	0.1696	0.8600	0.1573	0.1976	0.8900
$\rho_{9,12}$	0.00	0.0222	0.1477	0.9100	0.0536	0.1585	0.9400
$\rho_{10,11}$	0.6803	-0.1038	0.2065	0.9600	-0.1580	0.2559	0.8900
$\rho_{10,12}$	0.00	0.0490	0.1603	0.9300	-0.0005	0.1657	0.9200
$\rho_{11,12}$	0.00	0.0652	0.1588	0.9300	0.0381	0.1936	0.9200

Table D.19: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43 with $i, j = 1, 2, \ldots, 12$ and $i \leq j$ of the Skew Elliptical Stable distribution in dimension 12. The results reported above are obtained using 1000 replications for $\alpha = 1.70$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.90**	-0.0167	0.0385	0.9000	-0.0128	0.0223	0.9300
δ_1	0.00	0.0004	0.0006	0.9500	0.0000	0.0000	0.9800
δ_2	0.00	0.0005	0.0004	0.9000	0.0001	0.0000	0.0200
δ_3	0.60	-0.0116	0.0665	0.9600	-0.0070	0.0750	0.9700
δ_4	0.00	0.0010	0.0016	0.9500	0.0003	0.0000	0.0000
δ_5	0.00	0.0008	0.0008	0.8900	0.0000	0.0000	0.9800
δ_6	0.00	0.0014	0.0010	0.9300	0.0002	0.0000	0.0200
δ_7	0.00	0.0011	0.0012	0.9400	0.0004	0.0000	0.0000
δ_8	0.00	0.0003	0.0006	0.9200	0.0005	0.0000	0.0000
δ_9	0.00	0.0001	0.0002	0.9700	0.0003	0.0001	0.0000
δ_{10}	0.60	0.0184	0.0423	0.9300	0.0057	0.0654	0.9300
δ_{11}	0.60	-0.0257	0.0930	0.9600	-0.0099	0.1292	0.9800
δ_{12}	0.00	0.0008	0.0008	0.9100	0.0005	0.0000	0.0000
ξ_1	0.00	-0.0013	0.0375	0.9500	-0.0011	0.0189	0.9700
ξ_2	0.00	0.0051	0.0863	0.9600	0.0029	0.0529	0.9200
ξ_3	0.00	0.0072	0.0384	0.9400	0.0008	0.0229	0.9500
ξ_4	0.00	-0.0003	0.0340	0.9400	0.0001	0.0188	0.9700
ξ_5	0.00	0.0081	0.0469	0.9200	-0.0015	0.0235	0.9500
ξ_6	0.00	0.0043	0.0363	0.9300	-0.0015	0.0232	0.9500
ξ_7	0.00	0.0102	0.0610	0.9500	-0.0010	0.0324	0.9600
ξ_8	0.00	0.0050	0.0317	0.9300	0.0009	0.0164	0.9700
ξ_9	0.00	-0.0013	0.0492	0.9200	0.0043	0.0285	0.9500
ξ_{10}	0.00	-0.0017	0.0436	0.9700	0.0003	0.0206	0.9400
ξ_{11}	0.00	-0.0006	0.0308	0.9400	0.0015	0.0137	0.9600
ξ_{12}	0.00	0.0027	0.0363	0.9600	0.0027	0.0182	0.9400
$\sqrt{\omega}_{1,1}$	0.2390	0.0037	0.0265	0.9400	-0.0037	0.0123	0.9700
$\sqrt{\omega}_{2,2}$	1.5540	-0.0005	0.1499	0.9500	-0.0153	0.0907	0.9300
$\sqrt{\omega}_{3,3}$	0.3620	-0.0102	0.0430	0.9300	-0.0010	0.0202	0.9500
$\sqrt{\omega}_{4,4}$	0.1990	-0.0057	0.0211	0.9700	-0.0009	0.0110	0.9500
$\sqrt{\omega}_{5,5}$	0.3490	-0.0232	0.0342	0.9100	0.0100	0.0203	0.9400
$\sqrt{\omega}_{6,6}$	0.2950	0.0050	0.0338	0.9700	0.0054	0.0139	0.9400
$\sqrt{\omega}_{7,7}$	0.7150	0.0047	0.0662	0.9500	0.0016	0.0335	0.9600
$\sqrt{\omega}_{8,8}$	0.1640	0.0067	0.0153	0.9400	-0.0032	0.0093	0.9300
$\sqrt{\omega}_{9,9}$	0.5180	-0.0132	0.0543	0.9400	-0.0011	0.0286	0.9300
$\sqrt{\omega}_{10,10}$	0.3790	0.0083	0.0434	0.9600	0.0047	0.0211	0.9500
$\sqrt{\omega}_{11,11}$	0.1590	-0.0028	0.0156	0.9700	0.0083	0.0082	0.8200
$\sqrt{\omega}_{12,12}$	0.2070	0.0100	0.0215	0.9300	-0.0046	0.0095	0.9200
$\rho_{1,2}$	0.1920	0.0491	0.2273	0.9300	0.0556	0.1567	0.9200
$\rho_{1,3}$	0.00	0.0381	0.1821	0.9500	-0.0018	0.1902	0.9200
$\rho_{1,4}$	0.00	-0.0338	0.2357	0.9600	0.0024	0.0760	0.9500
$\rho_{1,5}$	0.00	0.0292	0.1585	0.9200	0.0161	0.1109	0.9700
$\rho_{1,6}$	0.00	0.0979	0.1790	0.9100	0.0014	0.1010	0.9600
$\rho_{1,7}$	0.00	0.0381	0.1667	0.9400	-0.0091	0.2719	0.9500
$\rho_{1,8}$	0.1566	0.0025	0.1662	0.9500	0.0273	0.0810	0.9600
$\rho_{1,9}$	0.00	0.1187	0.1851	0.9100	-0.0315	0.2580	0.9300
$\rho_{1,10}$	0.00	0.1147	0.1801	0.8800	0.0104	0.1979	0.9500
$\rho_{1,11}$	0.00	-0.0443	0.1771	0.9600	0.0095	0.0802	0.9500
$\rho_{1,12}$	0.00	0.0025	0.1896	0.9700	-0.0107	0.0713	0.9400

Table D.20: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the locations $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_12)$, scale matrix $\boldsymbol{\Omega} = \{\omega_{ij}\}$, with $i, j = 1, 2, \dots, 12$ and $i \leq j$, the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43, tail parameter α and skewness parameters $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_12)$ of the Skew Elliptical Stable distribution in dimension 12. The results reported above are obtained using 1000 replications for $\alpha = 1.90$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
$\rho_{2,3}$	0.00	0.0416	0.1970	0.9200	-0.0117	0.1702	0.9500
$\rho_{2,4}$	0.00	-0.0403	0.2556	0.9600	0.0225	0.1574	0.9500
$\rho_{2,5}$	0.00	0.0012	0.2214	0.9500	0.0637	0.1328	0.9200
$\rho_{2,6}$	0.00	0.0243	0.2270	0.9300	0.0322	0.1647	0.9600
$\rho_{2,7}$	0.00	-0.0029	0.1941	0.9400	-0.0125	0.1499	0.9800
$\rho_{2,8}$	0.00	0.1173	0.2349	0.9500	0.0712	0.1664	0.9300
$\rho_{2,9}$	0.00	0.0914	0.2185	0.9200	0.0141	0.1082	0.9300
$\rho_{2,10}$	0.00	0.0393	0.1975	0.9300	0.0474	0.1371	0.9300
$\rho_{2,11}$	0.00	-0.1021	0.2491	0.9500	0.0762	0.1437	0.9400
$\rho_{2,12}$	0.00	-0.0157	0.2119	0.9500	0.0286	0.1656	0.9500
$\rho_{3,4}$	0.0075	-0.0095	0.2200	0.9300	-0.0025	0.1437	0.9300
$\rho_{3,5}$	0.00	-0.0266	0.2072	0.9200	-0.0029	0.0861	0.9300
$\rho_{3,6}$	0.00	-0.0367	0.2193	0.9400	0.0381	0.1026	0.9300
$\rho_{3,7}$	0.00	0.0730	0.1555	0.9400	-0.0311	0.1458	0.9400
$\rho_{3,8}$	0.00	-0.0400	0.2173	0.9700	0.0326	0.1860	0.9200
$\rho_{3,9}$	0.00	0.1141	0.1603	0.8600	0.0446	0.1242	0.9400
$\rho_{3,10}$	0.00	0.0735	0.1786	0.9500	0.0717	0.1586	0.9300
$\rho_{3,11}$	0.00	0.0076	0.2129	0.9700	0.0655	0.1704	0.9400
$\rho_{3,12}$	0.00	0.0117	0.1910	0.9400	0.0205	0.1358	0.9500
$\rho_{4,5}$	0.3567	-0.0210	0.1710	0.9500	-0.0462	0.0796	0.8800
$\rho_{4,6}$	0.00	0.0756	0.1886	0.9400	0.0574	0.1265	0.9400
$\rho_{4,7}$	0.00	-0.0728	0.1932	0.9100	0.0265	0.2138	0.9200
$\rho_{4,8}$	0.00	0.0255	0.2168	0.9500	0.0512	0.1220	0.9600
$\rho_{4,9}$	0.00	0.0510	0.1959	0.9500	-0.0849	0.2375	0.9300
$\rho_{4,10}$	0.00	0.0509	0.1746	0.9400	0.0360	0.1955	0.9500
$\rho_{4,11}$	0.00	0.0156	0.1952	0.9700	0.0091	0.1342	0.9600
$\rho_{4,12}$	0.00	0.0573	0.1798	0.9300	0.0541	0.1169	0.9200
$\rho_{5,6}$	0.00	0.0729	0.1640	0.9400	0.0135	0.0749	0.9500
$\rho_{5,7}$	0.00	-0.0214	0.1955	0.9500	0.0071	0.1431	0.9600
$\rho_{5,8}$	0.00	0.0107	0.1951	0.9500	0.0241	0.1154	0.9500
$\rho_{5,9}$	0.00	0.0097	0.1980	0.9700	-0.0021	0.1650	0.9500
$\rho_{5,10}$	0.00	-0.0111	0.1608	0.9500	0.0032	0.1038	0.9300
$\rho_{5,11}$	0.00	0.0480	0.1877	0.9500	0.0085	0.1114	0.9700
$\rho_{5,12}$	-0.1339	0.0746	0.1775	0.9200	0.0366	0.0923	0.9300
$\rho_{6,7}$	-0.4896	0.1275	0.1809	0.8900	0.0830	0.1416	0.9100
$\rho_{6,8}$	0.0091	0.0572	0.1795	0.9500	0.0067	0.1173	0.9400
$\rho_{6,9}$	0.00	0.0917	0.1932	0.9400	-0.0239	0.1559	0.9300
$\rho_{6,10}$	0.00	0.0587	0.1545	0.9600	0.0048	0.1454	0.9200
$\rho_{6,11}$	0.00	0.0669	0.1667	0.9200	0.0018	0.1219	0.9600
$\rho_{6,12}$	0.00	0.0827	0.1644	0.9000	0.0070	0.1110	0.9500
$\rho_{7,8}$	0.00	-0.0414	0.2054	0.9400	0.0681	0.1293	0.9400
$\rho_{7,9}$	0.00	-0.0044	0.1862	0.9400	0.0163	0.1273	0.9400
$\rho_{7,10}$	0.00	0.0001	0.1624	0.9700	0.0501	0.1307	0.9600
$\rho_{7,11}$	0.00	0.0193	0.1764	0.9500	0.0684	0.1191	0.9100
$\rho_{7,12}$	0.00	-0.0877	0.1852	0.9300	0.0452	0.1154	0.9600
$\rho_{8,9}$	0.3843	-0.0223	0.1890	0.9800	-0.0470	0.1690	0.9500
$\rho_{8,10}$	-0.1123	0.0656	0.1750	0.9200	0.0516	0.2261	0.9600
$\rho_{8,11}$	-0.0495	0.0813	0.1708	0.9300	0.0701	0.1407	0.9000
$\rho_{8,12}$	0.00	0.0429	0.1665	0.9400	0.0759	0.1356	0.9300
$\rho_{9,10}$	-0.4356	0.2543	0.1548	0.6500	0.1887	0.2082	0.8700
$\rho_{9,11}$	-0.3136	0.1633	0.1591	0.8200	0.0937	0.1621	0.9200
$\rho_{9,12}$	0.00	0.0196	0.1888	0.9500	0.0506	0.1239	0.9300
$\rho_{10,11}$	0.6803	-0.0181	0.1712	0.9300	-0.0029	0.1761	0.9500
$\rho_{10,12}$	0.00	0.0557	0.1726	0.9500	0.0211	0.1425	0.9400
$\rho_{11,12}$	0.00	0.0755	0.1720	0.9500	0.0189	0.2116	0.9400

Table D.21: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43 with $i, j = 1, 2, \ldots, 12$ and $i \leq j$ of the Skew Elliptical Stable distribution in dimension 12. The results reported above are obtained using 1000 replications for $\alpha = 1.90$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.95**	-0.0282	0.0321	0.8900	-0.0249	0.0169	0.7600
δ_1	0.00	0.0004	0.0004	0.7400	0.0000	0.0000	0.0000
δ_2	0.00	0.0006	0.0004	0.9200	0.0001	0.0000	0.0100
δ_3	0.60	-0.0153	0.0446	0.9100	-0.0127	0.0633	0.9200
δ_4	0.00	0.0008	0.0011	0.9600	0.0003	0.0001	0.0100
δ_5	0.00	0.0009	0.0006	0.9000	0.0000	0.0000	0.0000
δ_6	0.00	0.0015	0.0009	0.7800	0.0002	0.0000	0.0300
δ_7	0.00	0.0008	0.0011	0.9600	0.0004	0.0000	0.0000
δ_8	0.00	0.0002	0.0003	0.9700	0.0005	0.0001	0.0000
δ_9	0.00	0.0001	0.0001	0.9400	0.0003	0.0001	0.0000
δ_{10}	0.60	0.0175	0.0301	0.9400	-0.0296	0.1203	0.9800
δ_{11}	0.60	-0.0356	0.0848	0.9600	-0.0117	0.0778	0.9100
δ_{12}	0.00	0.0007	0.0004	0.6500	0.0005	0.0001	0.0000
ξ_1	0.00	0.0028	0.0365	0.9400	-0.0011	0.0190	0.9500
ξ_2	0.00	0.0003	0.0824	0.9200	0.0016	0.0495	0.9300
ξ_3	0.00	-0.0022	0.0357	0.9500	0.0011	0.0208	0.9300
ξ_4	0.00	-0.0020	0.0292	0.9500	0.0029	0.0187	0.9600
ξ_5	0.00	0.0030	0.0423	0.9400	-0.0007	0.0258	0.9500
ξ_6	0.00	-0.0013	0.0378	0.9500	0.0005	0.0214	0.9400
ξ_7	0.00	0.0052	0.0589	0.9300	0.0075	0.0323	0.9400
ξ_8	0.00	-0.0041	0.0288	0.9600	-0.0031	0.0131	0.9400
ξ_9	0.00	-0.0102	0.0537	0.9400	-0.0014	0.0303	0.9200
ξ_{10}	0.00	0.0014	0.0347	0.9500	-0.0019	0.0226	0.9500
ξ_{11}	0.00	0.0026	0.0278	0.9400	-0.0019	0.0124	0.9300
ξ_{12}	0.00	0.0067	0.0309	0.9100	0.0018	0.0184	0.9200
$\sqrt{\omega_{1,1}}$	0.2390	0.0081	0.0266	0.9200	-0.0013	0.0126	0.9500
$\sqrt{\omega_{2,2}}$	1.5540	-0.0159	0.1501	0.9600	-0.0233	0.0900	0.9400
$\sqrt{\omega_{3,3}}$	0.3620	0.0032	0.0382	0.9500	0.0023	0.0204	0.9400
$\sqrt{\omega_{4,4}}$	0.1990	-0.0048	0.0207	0.9300	-0.0034	0.0104	0.9600
$\sqrt{\omega_{5,5}}$	0.3490	-0.0145	0.0345	0.9500	0.0102	0.0175	0.9200
$\sqrt{\omega_{6,6}}$	0.2950	0.0010	0.0310	0.9400	0.0030	0.0136	0.9400
$\sqrt{\omega_{7,7}}$	0.7150	-0.0034	0.0737	0.9600	0.0030	0.0331	0.9200
$\sqrt{\omega_{8,8}}$	0.1640	0.0082	0.0183	0.9300	-0.0008	0.0090	0.9500
$\sqrt{\omega_{9,9}}$	0.5180	-0.0123	0.0471	0.9100	0.0051	0.0267	0.9600
$\sqrt{\omega_{10,10}}$	0.3790	0.0029	0.0423	0.9600	-0.0019	0.0213	0.9500
$\sqrt{\omega_{11,11}}$	0.1590	-0.0041	0.0143	0.9200	0.0073	0.0078	0.8300
$\sqrt{\omega_{12,12}}$	0.2070	0.0093	0.0225	0.9400	-0.0052	0.0108	0.9300
$\rho_{1,2}$	0.1920	0.0007	0.2094	0.9600	0.0253	0.2257	0.9600
$\rho_{1,3}$	0.00	-0.0067	0.1638	0.9800	-0.0207	0.1922	0.9200
$\rho_{1,4}$	0.00	-0.0372	0.2198	0.9500	-0.0025	0.0827	0.9300
$\rho_{1,5}$	0.00	0.0362	0.1670	0.9400	0.0437	0.1509	0.9600
$\rho_{1,6}$	0.00	0.0958	0.1829	0.8900	0.0094	0.1358	0.9700
$\rho_{1,7}$	0.00	0.0878	0.2121	0.9400	-0.0080	0.2587	0.9300
$\rho_{1,8}$	0.1566	-0.0103	0.1730	0.9600	0.0358	0.0749	0.9500
$\rho_{1,9}$	0.00	0.1643	0.1978	0.8800	-0.0222	0.2573	0.9300
$\rho_{1,10}$	0.00	0.1239	0.1721	0.8900	-0.0082	0.2029	0.9400
$\rho_{1,11}$	0.00	-0.0534	0.1773	0.9300	0.0093	0.1046	0.9500
$\rho_{1,12}$	0.00	-0.0482	0.1884	0.9500	-0.0002	0.0769	0.9300

Table D.22: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the locations $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_12)$, scale matrix $\boldsymbol{\Omega} = \{\omega_{ij}\}$, with $i, j = 1, 2, \ldots, 12$ and $i \leq j$, the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43, tail parameter α and skewness parameters $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_12)$ of the Skew Elliptical Stable distribution in dimension 12. The results reported above are obtained using 1000 replications for $\alpha = 1.95$.

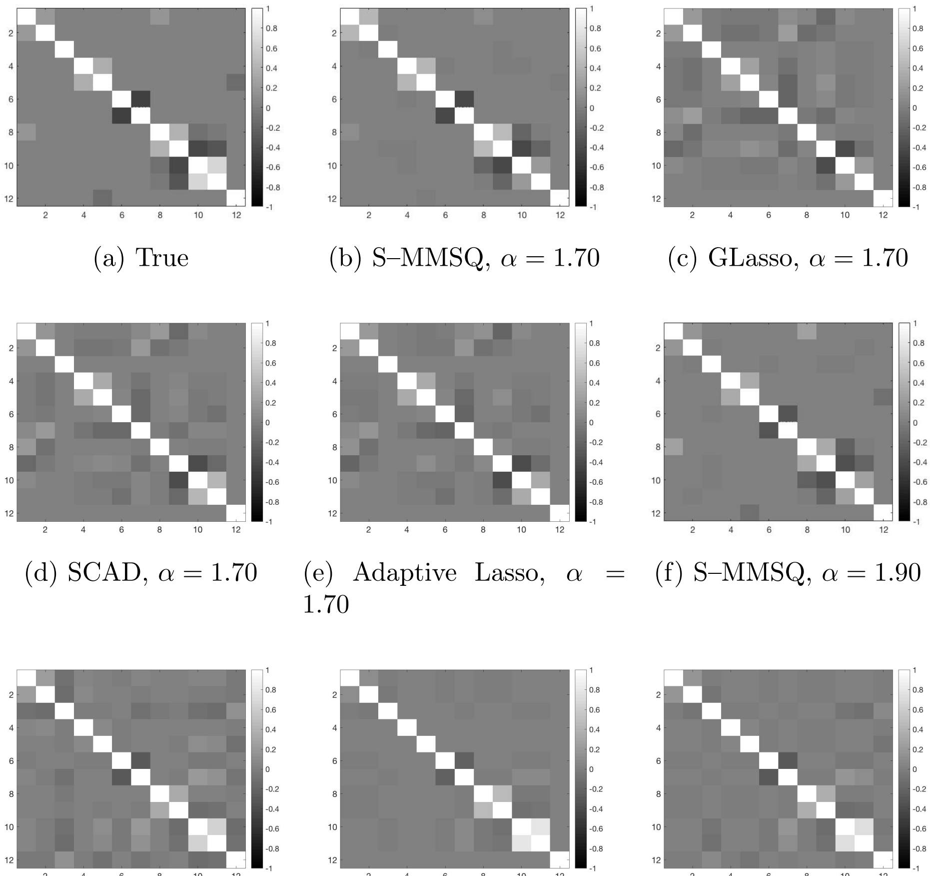
		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
$\rho_{2,3}$	0.00	0.0099	0.2368	0.9900	-0.0060	0.1218	0.9700
$\rho_{2,4}$	0.00	-0.0236	0.2403	0.9600	0.0653	0.1405	0.9300
$\rho_{2,5}$	0.00	0.0093	0.2666	0.9700	0.0614	0.1457	0.9600
$\rho_{2,6}$	0.00	0.0553	0.2437	0.9200	-0.0129	0.1614	0.9500
$\rho_{2,7}$	0.00	0.0099	0.2064	0.9600	0.0185	0.1217	0.9600
$\rho_{2,8}$	0.00	0.0541	0.2919	0.9600	0.0471	0.1644	0.9200
$\rho_{2,9}$	0.00	0.0982	0.2218	0.9300	0.0232	0.1360	0.9600
$\rho_{2,10}$	0.00	0.0335	0.2063	0.9600	0.0503	0.1258	0.9300
$\rho_{2,11}$	0.00	-0.0553	0.2535	0.9400	0.0562	0.1355	0.9600
$\rho_{2,12}$	0.00	-0.0392	0.2385	0.9600	0.0408	0.1475	0.9400
$\rho_{3,4}$	0.0075	-0.0088	0.1782	0.9200	-0.0372	0.1626	0.9300
$\rho_{3,5}$	0.00	0.0205	0.1989	0.9300	0.0163	0.1035	0.9600
$\rho_{3,6}$	0.00	-0.0674	0.2193	0.9300	0.0260	0.1160	0.9200
$\rho_{3,7}$	0.00	0.0958	0.1373	0.8600	-0.0481	0.1613	0.9700
$\rho_{3,8}$	0.00	-0.0639	0.2061	0.9500	-0.0255	0.1942	0.9100
$\rho_{3,9}$	0.00	0.0715	0.1683	0.9200	0.0321	0.1403	0.9500
$\rho_{3,10}$	0.00	0.0574	0.1536	0.9300	0.0453	0.1172	0.9100
$\rho_{3,11}$	0.00	-0.0292	0.1795	0.9400	0.0446	0.1944	0.9400
$\rho_{3,12}$	0.00	0.0093	0.1716	0.9500	-0.0098	0.1733	0.9300
$\rho_{4,5}$	0.3567	-0.0490	0.1461	0.9400	-0.0635	0.0916	0.8900
$\rho_{4,6}$	0.00	0.0644	0.1802	0.9400	0.0568	0.1324	0.9400
$\rho_{4,7}$	0.00	-0.0493	0.2022	0.9400	0.0199	0.2660	0.9300
$\rho_{4,8}$	0.00	0.0286	0.1529	0.9500	0.0473	0.1088	0.9100
$\rho_{4,9}$	0.00	0.0914	0.1821	0.9300	-0.0998	0.2579	0.9200
$\rho_{4,10}$	0.00	0.0596	0.1776	0.9400	-0.0138	0.1950	0.9300
$\rho_{4,11}$	0.00	0.0135	0.1632	0.9400	-0.0021	0.1379	0.9400
$\rho_{4,12}$	0.00	0.0357	0.1843	0.9500	0.0654	0.1059	0.9100
$\rho_{5,6}$	0.00	0.0141	0.1942	0.9500	0.0239	0.0857	0.9400
$\rho_{5,7}$	0.00	-0.0221	0.1929	0.9600	-0.0280	0.2037	0.9500
$\rho_{5,8}$	0.00	-0.0053	0.2052	0.9600	0.0131	0.1002	0.9500
$\rho_{5,9}$	0.00	-0.0248	0.1905	0.9700	-0.0048	0.1590	0.9600
$\rho_{5,10}$	0.00	0.0032	0.1572	0.9300	0.0042	0.0840	0.9300
$\rho_{5,11}$	0.00	-0.0140	0.1885	0.9300	-0.0016	0.1083	0.9500
$\rho_{5,12}$	-0.1339	0.0650	0.1542	0.9300	0.0590	0.1223	0.9300
$\rho_{6,7}$	-0.4896	0.1563	0.1890	0.8500	0.1076	0.1447	0.8800
$\rho_{6,8}$	0.0091	0.0973	0.1954	0.9200	0.0182	0.1292	0.9500
$\rho_{6,9}$	0.00	0.1200	0.1716	0.8900	-0.0565	0.1865	0.8900
$\rho_{6,10}$	0.00	0.1070	0.1657	0.8700	-0.0088	0.1514	0.9300
$\rho_{6,11}$	0.00	0.0392	0.1950	0.9500	0.0142	0.1166	0.9700
$\rho_{6,12}$	0.00	0.0880	0.2057	0.9200	0.0063	0.1285	0.9400
$\rho_{7,8}$	0.00	-0.0340	0.1872	0.9100	0.0723	0.1521	0.9300
$\rho_{7,9}$	0.00	0.0024	0.1840	0.9600	-0.0066	0.1401	0.9500
$\rho_{7,10}$	0.00	0.0100	0.1415	0.9700	0.0404	0.1464	0.9400
$\rho_{7,11}$	0.00	0.0363	0.1870	0.9300	0.0591	0.1249	0.9000
$\rho_{7,12}$	0.00	-0.1267	0.2059	0.9000	0.0294	0.1358	0.9600
$\rho_{8,9}$	0.3843	-0.0060	0.1654	0.9500	-0.1193	0.2174	0.9600
$\rho_{8,10}$	-0.1123	0.0923	0.1886	0.9100	0.0465	0.2441	0.9100
$\rho_{8,11}$	-0.0495	0.0834	0.1769	0.9300	0.0623	0.1629	0.9400
$\rho_{8,12}$	0.00	0.0433	0.1707	0.9500	0.0993	0.1734	0.9300
$\rho_{9,10}$	-0.4356	0.2419	0.1578	0.6800	0.2196	0.2023	0.8300
$\rho_{9,11}$	-0.3136	0.1518	0.1652	0.8200	0.0918	0.1629	0.9100
$\rho_{9,12}$	0.00	0.0228	0.1503	0.9500	0.0598	0.1404	0.9600
$\rho_{10,11}$	0.6803	0.0253	0.1527	0.9600	-0.0222	0.1926	0.9200
$\rho_{10,12}$	0.00	0.0852	0.1500	0.8800	0.0081	0.1722	0.9200
$\rho_{11,12}$	0.00	0.0911	0.1674	0.9100	-0.0183	0.2314	0.9300

Table D.23: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the off–diagonale elements ρ_{ij} of the matrix $\mathbf{R}$ defined in 3.43 with $i, j = 1, 2, \ldots, 12$ and $i \leq j$ of the Skew Elliptical Stable distribution in dimension 12. The results reported above are obtained using 1000 replications for $\alpha = 1.95$.

Appendix E

Figures of simulated examples

E.1 Elliptical Stable

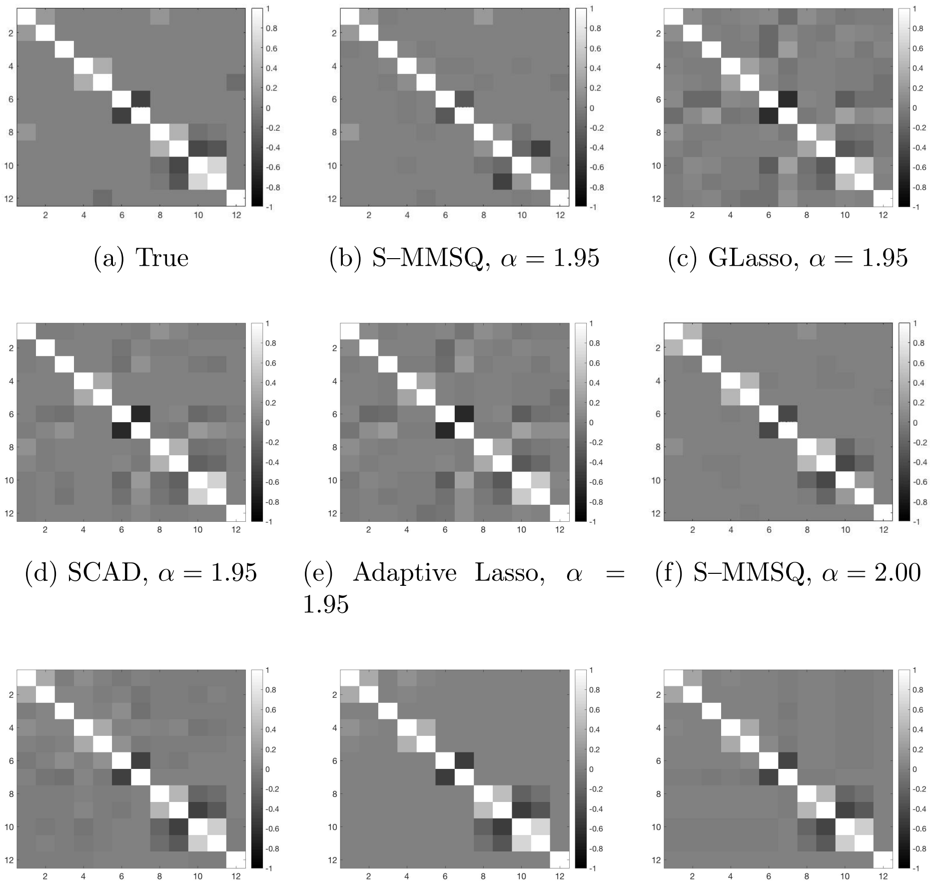


(a) True (b) S–MMSQ, $\alpha = 1.70$ (c) GLasso, $\alpha = 1.70$

(d) SCAD, $\alpha = 1.70$ (e) Adaptive Lasso, $\alpha = 1.70$ (f) S–MMSQ, $\alpha = 1.90$

(g) GLasso, $\alpha = 1.90$ (h) SCAD, $\alpha = 1.90$ (i) Adaptive Lasso, $\alpha = 1.90$

Figure E.1: Band structure of the true *(left)* and estimated *(right)* scale matrices through S–MMSQ of the 12–dimensional Elliptical Stable simulated experiment discussed in Section 3.4.2, for $\alpha = 1.70$ and sample size $n = 200$.



(a) True

(b) S–MMSQ, $\alpha = 1.95$

(c) GLasso, $\alpha = 1.95$

(d) SCAD, $\alpha = 1.95$

(e) Adaptive Lasso, $\alpha = 1.95$

(f) S–MMSQ, $\alpha = 2.00$

(g) GLasso, $\alpha = 2.00$

(h) SCAD, $\alpha = 2.00$

(i) Adaptive Lasso, $\alpha = 2.00$

Figure E.2: Band structure of the true and estimated scale matrices of the Elliptical Stable distribution experiment discussed in Section 3.4.2, for $\alpha = (1.95, 2.00)$ and sample size $n = 200$.

E.2 Skew Elliptical Stable

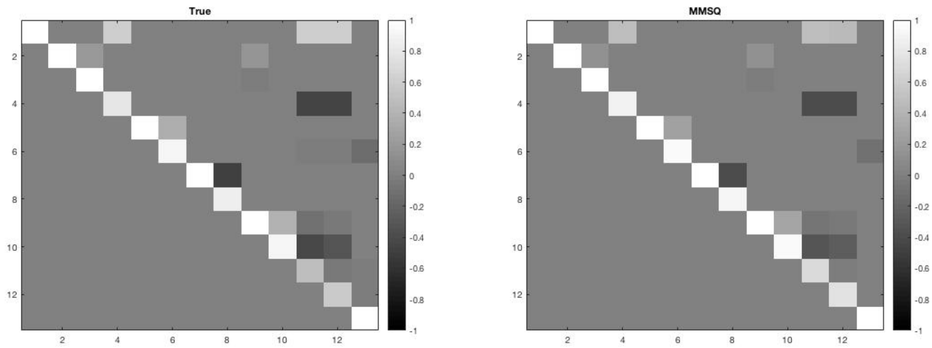
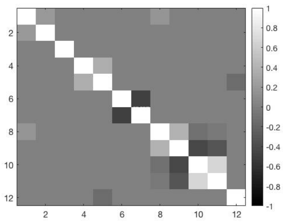
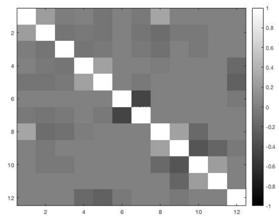


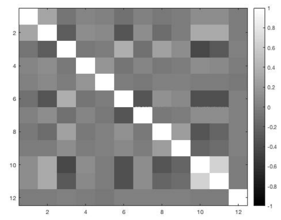
Figure E.3: Images displaying the band structure of the true *(left)* and estimated *(right)* matrices $\bar{\mathbf{\Omega}}$ through S–MMSQ of the Skew Elliptical Stable distribution in dimension 12 discussed in Section 3.4.3, for $\alpha = 1.70$ and sample size $n = 200$.



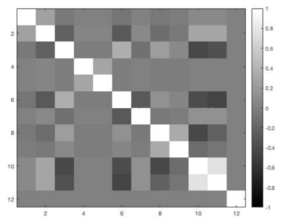
(a) True



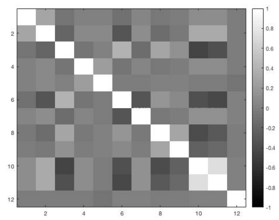
(b) S–MMSQ, $\alpha = 1.70$



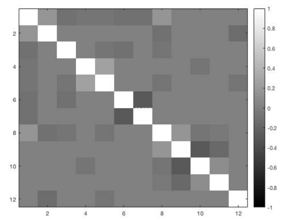
(c) GLasso, $\alpha = 1.70$



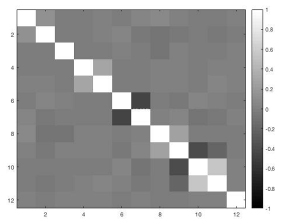
(d) SCAD, $\alpha = 1.70$



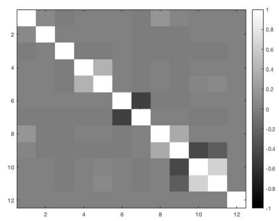
(e) Adaptive Lasso, $\alpha = 1.70$



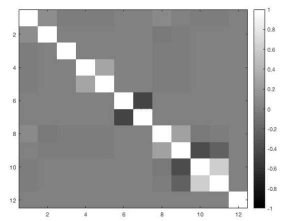
(f) S–MMSQ, $\alpha = 1.90$



(g) GLasso, $\alpha = 1.90$

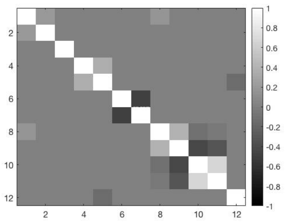


(h) SCAD, $\alpha = 1.90$

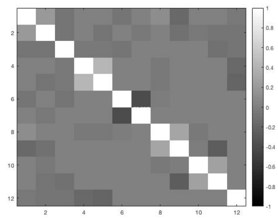


(i) Adaptive Lasso, $\alpha = 1.90$

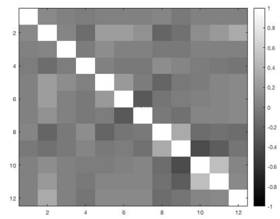
Figure E.4: Band structure of the true *(left)* and estimated *(right)* scale matrices through S–MMSQ of the 12–dimensional Skew – Elliptical Stable simulated experiment discussed in Section 3.4.3, for $\alpha = 1.70$ and sample size $n = 200$.



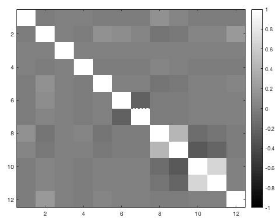
(a) True



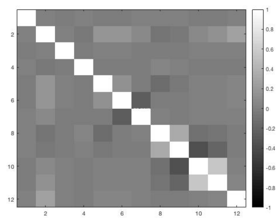
(b) S–MMSQ, $\alpha = 1.95$



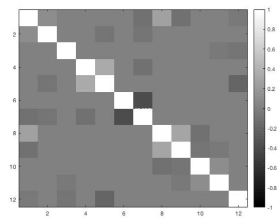
(c) GLasso, $\alpha = 1.95$



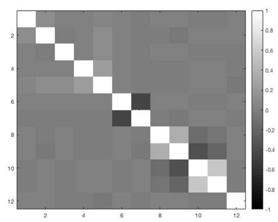
(d) SCAD, $\alpha = 1.95$



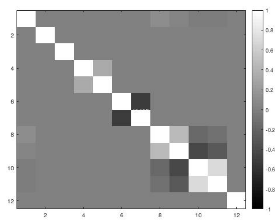
(e) Adaptive Lasso, $\alpha = 1.95$



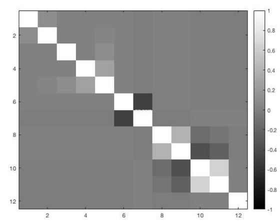
(f) S–MMSQ, $\alpha = 2.00$



(g) GLasso, $\alpha = 2.00$



(h) SCAD, $\alpha = 2.00$



(i) Adaptive Lasso, $\alpha = 2.00$

Figure E.5: Band structure of the true and estimated scale matrices of the Skew–Elliptical Stable distribution experiment discussed in Section 3.4.3, for $\alpha = (1.95, 2.00)$ and sample size $n = 200$.

Appendix F

Tables of real data examples

Summary Stat.	IT	FR	DE	PT	EL	NL	AT	BE	DK	FI	IE	PL	UK	CZ	ES	SE	HU
Min	-0.251	-0.245	-0.239	-0.190	-0.374	-0.297	-0.360	-0.217	-0.238	-0.268	-0.377	-0.204	-0.234	-0.275	-0.241	-0.238	-0.350
Max	0.184	0.123	0.156	0.118	0.224	0.154	0.174	0.118	0.128	0.198	0.171	0.161	0.125	0.183	0.135	0.174	0.185
Mean	0.280	4.751	5.014	-0.299	-8.513	5.052	0.769	4.211	10.231	6.441	0.517	3.605	3.444	3.484	5.567	8.298	12.604
Std Dev.	23.169	20.979	22.634	20.007	35.733	21.174	23.894	20.740	19.632	31.810	25.226	27.066	16.991	23.360	23.172	23.071	30.231
Skew	-0.758	-0.753	-0.672	-0.725	-0.743	-1.281	-1.752	-1.209	-1.127	-0.586	-1.606	-0.149	-0.990	-0.635	-0.707	-0.605	-0.954
Kurt	8.786	8.425	8.003	6.847	8.534	13.617	18.814	10.543	10.919	6.146	17.970	5.742	13.686	9.933	7.228	7.768	11.786
1% Str. Lev. (%)	-8.158	-7.404	-8.814	-9.011	-15.070	-8.899	-9.469	-8.841	-7.683	-13.220	-11.282	-10.459	-6.256	-8.718	-8.538	-8.671	-11.185
L–B Q–test	3.193	7.143	9.643	2.176	1.927	6.683	5.704	3.975	12.050	2.638	24.397	6.636	19.801	7.113	5.160	7.839	18.958
ACF of r_t	-0.009	-0.069	-0.020	0.031	0.006	-0.019	0.003	0.022	-0.036	-0.014	-0.089	0.021	-0.065	0.006	-0.051	-0.040	-0.001
ACF of r_t^2	0.197	0.116	0.179	0.246	0.126	0.062	0.159	0.404	0.240	0.155	0.136	0.223	0.089	0.103	0.133	0.097	0.087
JB	1723.441	1527.060	1292.796	814.234	1581.625	5745.001	12636.697	3021.599	3264.712	542.900	11291.079	366.526	5689.300	2392.750	957.389	1165.347	3893.189
Correlations	IT	FR	DE	PT	EL	NL	AT	BE	DK	FI	IE	PL	UK	CZ	ES	SE	HU
IT	1.000	0.835	0.787	0.640	0.477	0.783	0.632	0.658	0.622	0.576	0.592	0.411	0.751	0.408	0.818	0.704	0.515
FR	0.835	1.000	0.890	0.656	0.470	0.882	0.652	0.743	0.667	0.667	0.645	0.477	0.856	0.456	0.815	0.809	0.530
DE	0.787	0.890	1.000	0.633	0.452	0.851	0.631	0.716	0.671	0.649	0.630	0.506	0.815	0.458	0.784	0.809	0.533
PT	0.640	0.656	0.633	1.000	0.473	0.589	0.546	0.527	0.531	0.487	0.466	0.409	0.597	0.389	0.689	0.580	0.456
EL	0.477	0.470	0.452	0.473	1.000	0.441	0.445	0.426	0.408	0.327	0.408	0.359	0.434	0.386	0.491	0.413	0.390
NL	0.783	0.882	0.851	0.589	0.441	1.000	0.632	0.776	0.660	0.607	0.656	0.434	0.823	0.420	0.761	0.759	0.512
AT	0.632	0.652	0.631	0.546	0.445	0.632	1.000	0.594	0.553	0.427	0.566	0.428	0.653	0.439	0.642	0.585	0.539
BE	0.658	0.743	0.716	0.527	0.426	0.776	0.594	1.000	0.589	0.457	0.598	0.379	0.715	0.372	0.655	0.634	0.429
DK	0.622	0.667	0.671	0.531	0.408	0.660	0.553	0.589	1.000	0.507	0.560	0.375	0.673	0.427	0.589	0.664	0.440
FI	0.576	0.667	0.649	0.487	0.327	0.607	0.427	0.457	0.507	1.000	0.440	0.415	0.617	0.360	0.565	0.697	0.426
IE	0.592	0.645	0.630	0.466	0.408	0.656	0.566	0.598	0.560	0.440	1.000	0.368	0.638	0.384	0.571	0.571	0.442
PL	0.411	0.477	0.506	0.409	0.359	0.434	0.428	0.379	0.375	0.415	0.368	1.000	0.454	0.525	0.440	0.465	0.579
UK	0.751	0.856	0.815	0.597	0.434	0.823	0.653	0.715	0.673	0.617	0.638	0.454	1.000	0.432	0.730	0.758	0.517
CZ	0.408	0.456	0.458	0.389	0.386	0.420	0.439	0.372	0.427	0.360	0.384	0.525	0.432	1.000	0.439	0.441	0.568
ES	0.818	0.815	0.784	0.689	0.491	0.761	0.642	0.655	0.589	0.565	0.571	0.440	0.730	0.439	1.000	0.707	0.520
SE	0.704	0.809	0.809	0.580	0.413	0.759	0.585	0.634	0.664	0.697	0.571	0.465	0.758	0.441	0.707	1.000	0.517
HU	0.515	0.530	0.533	0.456	0.390	0.512	0.539	0.429	0.440	0.426	0.442	0.579	0.517	0.568	0.520	0.517	1.000

TABLE F.1: *(Top panel)*: summary statistics of the weekly returns of the MSCI european indexes, for the period form January 2nd, 1987 to March 3rd, 2015. The 7–th row, denoted by "1% Str. Lev." reports the 1% empirical quantile of the returns distribution, while the last row, denoted by "JB" reports the value of the Jarque–Berá test–statistics. *(Bottom panel)*: empirical correlations among the European indexes.

List of Figures

List of Tables

Bibliography

Aas, K. and Haff, I. (2006). The generalized hyperbolic skew student's t–distribution. *Journal of Financial Econometrics*, 4:275–309.

Abdul-Hamid, H. and Nolan, J. P. (1998). Multivariate stable densities as functions of one-dimensional projections. *J. Multivariate Anal.*, 67(1):80–89.

Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, 26:1487–1503.

Adler, R. J., Feldman, R. E., and Taqqu, M. S., editors (1998). *A practical guide to heavy tails*. Birkhäuser Boston, Inc., Boston, MA. Statistical techniques and applications, Papers from the workshop held in Santa Barbara, CA, December 1995.

Ahn, D.-H., Boudoukh, J., Richardson, M., and Whitelaw, R. F. (1999). Optimal risk management using options. *The Journal of Finance*, 54(1):359–375.

Artzner, P., Delbaen, F., Eber, J. M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9:203–228.

Azzalini, A. (2013). *The Skew-Normal and Related Families*. Cambridge University Press, Cambridge.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 65(2):367–389.

Barone-Adesi, G. (1985). Arbitrage equilibrium with skewed asset returns. *Journal of Financial and Quantitative Analysis*, 20:299–313.

Berkelaar, A. B. and Kouwenberg, R. (2000). Dynamic asset allocation and downside-risk aversion.

Bernardi, M. (2013). Risk measures for skew normal mixtures. *Statistics & Probability Letters*, 83:1819–1824.

Bien, J. and Tibshirani, R. J. (2011). Sparse estimation of a covariance matrix. *Biometrika*, 98(4):807–820.

Blasques, F. and Duplinskiy, A. (2015). Penalized indirect inference.

Bradley, B. O. and Taqqu, M. S. (2003). Financial risk and heavy tails. In Rachev, S. T., editor, *Handbook of Heavy Tailed Distributions in Finance*, volume 1 of *Handbooks in Finance*, pages 35 – 103. North-Holland, Amsterdam.

Branco, M. D. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *J. Multivariate Anal.*, 79(1):99–113.

Brenner, M. (1974). On the stability of the distribution of the market component in stock price changes. *Journal of Financial and Quantitative Analysis*, 9(6):945–961.

Buckle, D. J. (1995). Bayesian inference for stable distributions. *J. Amer. Statist. Assoc.*, 90(430):605–613.

Byczkowski, T., Nolan, J. P., and Rajput, B. (1993). Approximation of multidimensional stable densities. *J. Multivariate Anal.*, 46(1):13–31.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton Mathematical Series, vol. 9. Princeton University Press, Princeton, N. J.

Dittmar, R. F. (2002). Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *Journal of Finance*, 57:369–403.

Dominicy, Y., Ogata, H., and Veredas, D. (2013). Inference for vast dimensional elliptical distributions. *Comput. Statist.*, 28(4):1853–1880.

Dominicy, Y. and Veredas, D. (2013). The method of simulated quantiles. *J. Econometrics*, 172(2):235–247.

DuMouchel, W. H. (1973). On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *Ann. Statist.*, 1:948–957.

Embrechts, P., Frey, R., and McNeil, A. (2005). Quantitative risk management. *Princeton Series in Finance, Princeton*, 10.

Fama, E. F. (1965). Portfolio analysis in a stable paretian market. *Management Science*, 11(3):404–419.

Fama, E. F. (2012). The behavior of stock-market prices [reprint of J. Bus. **38** (1965), no. 1, 34–105]. In *Financial risk measurement and management*, volume 267 of *Internat. Lib. Crit. Writ. Econ.*, pages 256–327. Edward Elgar, Cheltenham.

Fama, E. F. and Roll, R. (1968). Some properties of symmetric stable distributions. *J. Amer. Statist. Assoc.*, 63:817–836.

Fan, J., Feng, Y., and Wu, Y. (2009). Network exploration via the adaptive lasso and SCAD penalties. *Ann. Appl. Stat.*, 3(2):521–541.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.*, 96(456):1348–1360.

Fang, K. T., Kotz, S., and Ng, K. W. (1990). *Symmetric multivariate and related distributions*, volume 36 of *Monographs on Statistics and Applied Probability.* Chapman and Hall, Ltd., London.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics (Oxford, England)*, 9(3):432–441.

Friend, I. and Westerfield, R. (1980). Co–skewness and capital asset pricing. *Journal of Finance*, 35:897–913.

Gallant, A. R. and Tauchen, G. (1996). Which moments to match? *Econometric Theory*, 12(4):657–681.

Gamba, A. (1999). Portfolio analysis with symmetric stable Paretian returns. In *Current topics in quantitative finance (Venice, 1997)*, Contrib. Manag. Sci., pages 48–69. Physica, Heidelberg.

Gao, X. and Massam, H. (2015). Estimation of symmetry-constrained Gaussian graphical models: application to clustered dense networks. *J. Comput. Graph. Statist.*, 24(4):909–929.

Genton, M. G., editor (2004). *Skew-elliptical distributions and their applications*. Chapman & Hall/CRC, Boca Raton, FL. A journey beyond normality.

Godsill, S. (1999). Mcmc and em-based methods for inference in heavy-tailed processes with alpha;-stable innovations. In *Higher-Order Statistics, 1999. Proceedings of the IEEE Signal Processing Workshop on*, pages 228–232.

Gouriéroux, C. and Monfort, A. (1996). *Simulation-based econometric methods*. CORE lectures. Oxford University Press.

Gouriéroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8(S1):S85–S118.

Hallin, M., Paindaveine, D., and Šiman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: from L_1 optimization to halfspace depth. *Ann. Statist.*, 38(2):635–669.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054.

Hunter, D. R. and Li, R. (2005). Variable selection using mm algorithms. *Ann. Statist.*, 33(4):1617–1642.

Jiang, W. and Turnbull, B. (2004). The indirect method: inference based on intermediate statistics—a synthesis and examples. *Statist. Sci.*, 19(2):239–263.

Jondeau, E., Poon, S.-h., and Rockinger, M. (2007). *Financial modeling under non-Gaussian distributions, Springer Finance*. Springer–Verlag London Ltd, London.

Jondeau, E. and Rockinger, M. (2009). The impact of shocks on higher moments. *Journal of Financial Econometrics*, 7:77–105.

Jorion, F. (2006). *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, The Inc, 3rd edition.

Kim, H.-M. and Genton, M. G. (2011). Characteristic functions of scale mixtures of multivariate skew-normal distributions. *J. Multivariate Anal.*, 102(7):1105–1117.

Kim, T.-H. and White, H. (2004). On more robust estimation of skewness and kurtosis. *Finance Research Letters*, 1(1):56–73.

Koenker, R. (2005). *Quantile regression*, volume 38 of *Econometric Society Monographs*. Cambridge University Press, Cambridge.

Kong, L. and Mizera, I. (2012). Quantile tomography: using quantiles with multivariate data. *Statist. Sinica*, 22(4):1589–1610.

Kraus, A. and Litzenberger, R. H. (1976). Skewness preference and the valuation of risk assets. *Journal of Finance*, 31:1085–1100.

Kristensen, D. and Shin, Y. (2012). Estimation of dynamic models with nonparametric simulated maximum likelihood. *J. Econometrics*, 167(1):76–94.

Lachos, V. H., Ghosh, P., and Arellano-Valle, R. B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statist. Sinica*, 20(1):303–322.

Lauritzen, S. L. (1996). *Graphical models*, volume 17. Clarendon Press.

Levina, E., Rothman, A., Zhu, J., et al. (2008). Sparse estimation of large covariance matrices via a nested lasso penalty. *The Annals of Applied Statistics*, 2(1):245–263.

Lombardi, M. J. and Veredas, D. (2009). Indirect estimation of elliptical stable distributions. *Comput. Statist. Data Anal.*, 53(6):2309–2324.

Mandelbrot, B. (2012). The variation of certain speculative prices [reprint of J. Bus. **36** (1963), no. 4, 394–419]. In *Financial risk measurement and management*, volume 267 of *Internat. Lib. Crit. Writ. Econ.*, pages 230–255. Edward Elgar, Cheltenham.

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance*, 7:77–91.

Matsui, M. and Takemura, A. (2009). Integral representations of one-dimensional projections for multivariate stable densities. *Journal of Multivariate Analysis*, 100(3):334 – 344.

McCulloch, J. H. (1986). Simple consistent estimators of stable distribution parameters. *Communications in Statistics-Simulation and Computation*, 15(4):1109–1136.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57(5):995–1026.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative risk management*. Princeton Series in Finance. Princeton University Press, Princeton, NJ, revised edition. Concepts, techniques and tools.

Meerschaert, M. M. and Scheffler, H.-P. (2003). Chapter 15 - portfolio modeling with heavy tailed random vectors. In Rachev, S. T., editor, *Handbook of Heavy Tailed Distributions in Finance*, volume 1 of *Handbooks in Finance*, pages 595 – 640. North-Holland, Amsterdam.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist.*, 34(3):1436–1462.

Mikosch, T. (2003). Modeling dependence and tails of financial time series. *Extreme Values in Finance, Telecommunications, and the Environment*, pages 185–286.

Mittnik, S., Rachev, S. T., and Paolella, M. S. (1998). Stable paretian modeling in finance: Some empirical and theoretical aspects. In Adler, R. J., Feldman, R. E., and Taqqu, M. S., editors, *A Practical Guide to Heavy Tails*, pages 79–110. Birkhauser Boston Inc., Cambridge, MA, USA.

Nolan, J. P. (2008). An overview of multivariate stable distributions. *Online: http://academic2. american. edu/ jpnolan/stable/overview. pdf*, 2(008).

Nolan, J. P. (2013). Multivariate elliptically contoured stable distributions: theory and estimation. *Comput. Statist.*, 28(5):2067–2089.

Ortobelli, S., Rachev, S., Huber, I., and Biglova, A. (2004). Optimal portfolio selection and risk management: a comparison between the stable Paretian approach and the Gaussian one. In *Handbook of computational and numerical methods in finance*, pages 197–252. Birkhäuser Boston, Boston, MA.

Paindaveine, D. and Šiman, M. (2011). On directional multiple-output quantile regression. *J. Multivariate Anal.*, 102(2):193–212.

Paolella, M. S. (2007). *Intermediate probability: a computational approach.* John Wiley & Sons, New York.

Paulsen, V. I., Power, S. C., and Smith, R. R. (1989). Schur products and matrix completions. *Journal of functional analysis*, 85(1):151–178.

Pourahmadi, M. and Wang, X. (2015). Distribution of random correlation matrices: Hyperspherical parameterization of the cholesky factor. *Statistics & Probability Letters*, 106:5–12.

Qiou, Z., Ravishanker, N., and Dey, D. K. (1999). Multivariate survival analysis with positive stable frailties. *Biometrics*, 55(2):637–644.

Rachev, S. and Mittnik, S. (2000). *Stable Paretian models in finance*. John Willey & Sons, New York.

Ravishanker, N. and Qiou, Z. (1998). Bayesian inference for time series with infinite variance stable innovations. In *A practical guide to heavy tails (Santa Barbara, CA, 1995)*, pages 259–280. Birkhäuser Boston, Boston, MA.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42.

Roncalli, T. (2014). *Introduction to risk parity and budgeting*. Chapman & Hall/CRC Financial Mathematics Series. CRC Press, Boca Raton, FL.

Salas-Gonzalez, D., Kuruoglu, E. E., and Ruiz, D. P. (2009). Finite mixture of α-stable distributions. *Digit. Signal Process.*, 19(2):250–264.

Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable non-Gaussian random processes*. Stochastic Modeling. Chapman & Hall, New York. Stochastic models with infinite variance.

Scott, R. C. and Horvath, P. A. (1980). On the cirection of preference for moments of higher order than the variance. *Journal of Finance*, 35(4):915–919.

Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications. *Statist. Neerlandica*, 56(2):214–232. Special issue: Frontier research in theoretical statistics, 2000 (Eindhoven).

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58(1):267–288.

Tsionas, E. G. (2013). Bayesian inference in multivariate stable distributions using copulae. *Available at SSRN 2214612*.

Tsionas, M. G. (2016). Bayesian analysis of multivariate stable distributions using one-dimensional projections. *J. Multivar. Anal.*, 143(C):185–193.

Wang, H. (2015). Scaling it up: Stochastic search structure learning in graphical models. *Bayesian Anal.*, 10(2):351–377.

Zareifard, H., Rue, H., Khaledi, M. J., and Lindgren, F. (2016). A skew gaussian decomposable graphical model. *Journal of Multivariate Analysis*, 145:58 – 72.

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.*, 38(2):894–942.

Zolotarev, V. M. (1964). On the representation of stable laws by integrals. *Trudy Mat. Inst. Steklov.*, 71:46–50.

Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.*, 101(476):1418–1429.