ARTICLE IN PRESS
JID: ECOSTA [m3Gsc;February 28, 2022;13:22]

Econometrics and Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Econometrics and Statistics

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

Sparse simulation-based estimator built on quantiles

Paola Stolfi[a,*], Mauro Bernardi[b], Lea Petrella[c]

[a] Institute for Applied Mathematics, National Research Council of Italy, Italy
[b] Department of Statistical Sciences, University of Padova, Italy
[c] MEMOTEF Department, Sapienza University of Rome, Italy

ARTICLE INFO

Article history:
Received 1 February 2021
Revised 20 January 2022
Accepted 20 January 2022
Available online xxx

Keywords:
directional quantiles
method of simulated quantiles
sparse regularisation
SCAD
Elliptical Stable distribution

ABSTRACT

The method of simulated quantiles is extended to a general multivariate framework and to provide sparse estimation of the scaling matrix. The method is based on the minimisation of a distance between appropriate statistics evaluated on the true and synthetic data simulated from the postulated model. Those statistics are functions of the quantiles providing an effective way to deal with distributions that do not admit moments of any order like the α–Stable or the Tukey lambda distribution. The lack of a natural ordering represents the major challenge for the extension of the method to the multivariate framework, which is addressed by considering the notion of projectional quantile. The SCAD ℓ_1–penalty is then introduced in order to achieve sparse estimation of the scaling matrix which is mostly responsible for the curse of dimensionality. The asymptotic properties of the proposed estimator have been discussed and the method is illustrated and tested on several synthetic datasets simulated from the Elliptical Stable distribution for which alternative methods are recognised to perform poorly.

1. Introduction

Model–based statistical inference primarily deals with parameters estimation, that usually 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 (GMM) of Hansen (1982) or its generalised versions (EMM) of 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) of McFadden (1989), the method of simulated maximum likelihood (SML) of Gouriéroux and Monfort (1996) and its nonparametric version introduced in Kristensen and Shin (2012) or

* Corresponding author.
E-mail address: p.stolfi@iac.cnr.it (P. Stolfi).

https://doi.org/10.1016/j.ecosta.2022.01.006

Please cite this article as: P. Stolfi, M. Bernardi and L. Petrella, Sparse simulation-based estimator built on quantiles, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2022.01.006

the indirect inference method (II) of Gouriéroux et al. (1993), are the only viable solutions to the inferential problem. Despite their appealing characteristics of only requiring to be able to simulate from the specified Data Generating Process (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 form of identification between the parameters of interests and those of the auxiliary model. The quantile–matching estimation method (QM) of Koenker (2005), exploits the same idea behind the method of moments without requiring any other condition. The QM approach estimates model parameters by matching the empirical quantiles with their theoretical counterparts thereby requiring only the existence of a closed form expression for the quantile function. Such difficulty is solved by the method of simulated quantiles (MSQ) recently proposed by Dominicy and Veredas (2013) as a simulation–based extension of the QM of Koenker (2005). As any other simulation–based method, 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 data generated from the postulated theoretical model. The MSQ, relying on quantiles, can be applied only on univariate distributions. This work introduces the multivariate method of simulated quantiles (MMSQ) that is a possible extension of the MSQ to deal with multivariate data. The extension of the MSQ to a multivariate framework is not trivial because it requires the definition 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), as a particular version of directional quantile. An important methodological contribution of the paper concerns the choice of the relevant directions to project data in order to summarise the information for the parameters of interest. A general solution is provided for Elliptical distributions and for those Skew–Elliptical distributions that are closed under linear combinations. The asymptotic theory of the proposed MMSQ estimator has been derived under standard conditions on the underlying true DGP.

As any other simulation–based method the MMSQ does 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. Since we are dealing with multivariate distributions, the right identification of the sparsity patterns becomes crucial because it reduces the number of parameters to be estimated, and therefore, it reduces the complexity of the model. Several works related to sparse estimation of the covariance or precision matrix are available in literature, both methodological and applied showing the growing need of sparse estimator to handle modern statistical issues and to analyse modern datasets; Gao and Massam (2015) for instance, construct a gene expression network by proposing a sparse estimator of the precision matrix to handle symmetry–constrained graphical models; Bien and Tibshirani (2011) reduce the complexity of a covariance graph by proposing a sparse estimator of the covariance matrix; Friedman et al. (2008) propose a fast algorithm to achieve sparse estimator of the precision matrix; 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. These are just some examples that, on the one hand, endorse the importance of introducing sparse estimators that automatically shrink to zero for some parameters, such as, for example, the off–diagonal elements of the variance–covariance matrix, while on the other hand highlight the fact that these methods have been developed only within the Gaussian framework. The second contribution of this work fills this gap by handling the curse of dimensionality within a high–dimensional non–Gaussian framework. Specifically, the proposed approach penalises the objective function of the MMSQ by adding a SCAD ℓ_1–penalisation term that shrinks to zero the off–diagonal elements of the scale matrix of the postulated distribution. The asymptotic properties of penalised MMSQ (S–MMSQ) is discussed by showing how it inherits the Oracle properties of the SCAD estimator within the likelihood framework and, since the chosen penalty is concave, a fast and efficient algorithm to solve the optimisation problem has been given.

The effectiveness of the MMSQ and S–MMSQ methods have been tested on several synthetic datasets simulated from the Elliptical Stable distribution previously considered by Lombardi and Veredas (2009). 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, while a recent overview of multivariate Stable distributions can be found in Nolan (2008). The proposed methods can be effectively used to make inference on the parameters of many others large–dimensional heavy–tailed distributions such as, Stable, Skew–Elliptical Stable (Branco and Dey (2001)), Copula (Oh and Patton (2013)), multivariate Gamma (Mathai and Moschopoulos (1992)) and Tempered Stable (Koponen (1995)). These distributions are widely used in econometric applications, indeed they allow for infinite variance, skewness and heavy–tails that exhibit power decay allowing extreme events to have higher probability mass than in Gaussian model (see for instance Nolan (2003), Bianchi et al. (2010), Cherubini et al. (2004), Semenikhine et al. (2018)). However, their use is mainly limited to univariate applications because parameters estimation in multivariate frameworks becomes unfeasible. The MMSQ and S–MMSQ allow to use these distribution even in high–dimensional framework. In Stolfi et al. (2018) the authors apply the S–MMSQ to solve a portfolio optimisation problem under value-at-risk constraints where the joint returns follow a multivariate skew-elliptical stable distribution while in Bernardi and Stolfi (2020) the authors apply the S–MMSQ to estimates the parameters distributions of financial returns which are assumed to be Elliptically Stable, the estimates are then used to perform a dominance test that allows to determine whether or not a financial institution can be classified as being more systemically important. Here is the structure of the paper. In Section 2 the multivariate Method of Simulated Quantiles is introduced, and the basic asymptotic properties are discussed. Section 3 deals with the curse of dimensionality by introducing the S–MMSQ estimator. The effectiveness of the method is tested in Section 4, where several synthetic datasets from the Elliptical Stable distribution are considered. Section 5 concludes. Technical proofs of the theorems are deferred to Supplementary material.

2. Multivariate method of simulated quantiles

In this Section we first define the notation that will be used throughout the paper. Then the basic concepts on projectional quantiles are recalled, in order to introduce the multivariate method of simulated quantiles. Finally results about the consistency and asymptotic properties of the estimator are discussed.

2.1. Notations

Throughout the paper, unless specified differently, the following notation is used. $\mathbf{Y} \in \mathbb{R}^m$ denotes a random vector with probability density function $f_{\mathbf{Y}}\left(\cdot, \boldsymbol{\vartheta}\right)$ and distribution function $F_{\mathbf{Y}}\left(\cdot, \boldsymbol{\vartheta}\right)$ both depending on the set of parameters $\boldsymbol{\vartheta} \in \boldsymbol{\Theta} \subset \mathbb{R}^k$; $\{\mathbf{y}_i\}_{i=1}^n$ denotes a sample of observations from $\mathbf{Y}$; $\mathbf{u} \in \mathbb{S}^{m-1}$ is a vector in the unit sphere $\mathbb{S}^{m-1} = \left\{\mathbf{u} \in \mathbb{R}^m : \mathbf{u}'\mathbf{u} = 1\right\}$; $\tau \in (0, 1)$ is the quantile level; $\mathbb{E}[\cdot]$ stands for the expectation.

2.2. Projectional quantiles

The MMSQ requires the prior definition of the concept of multivariate quantile, a notion still vague until recently, because of the lack of a natural ordering in dimension greater than one. Here, we relies on the definition of projectional quantiles introduced by Hallin et al. (2010), Paindaveine and Šiman (2011) and Kong and Mizera (2012), reported below.

Definition 2.1. The $\tau\mathbf{u}$ projectional quantile of $\mathbf{Y}$ is defined as

$$q_{\tau\mathbf{u}} \in \left\{\arg\min_{q\in\mathbb{R}} \Psi_{\tau\mathbf{u}}(q)\right\}, \tag{1}$$

where

$$\Psi_{\tau\mathbf{u}}(q) = \mathbb{E}\left[\rho_\tau\left(\mathbf{u}'\mathbf{Y} - q\right)\right], \tag{2}$$

and $\rho_\tau(z) = z(\tau - \mathbf{1}_{(-\infty,0)}(z))$ denotes the quantile loss function evaluated at $z \in \mathbb{R}$.

Clearly the $\tau\mathbf{u}$–projectional quantile is the τ–quantile of the univariate random variable $Z = \mathbf{u}'\mathbf{Y}$ defined as the projection of $\mathbf{Y}$ over the direction given by $\mathbf{u}$. 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 definition of univariate quantile which preserves the ordering.

The empirical counterpart of the $\tau\mathbf{u}$–projectional quantile is defined as

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

where $\Psi^n_{\tau\mathbf{u}}(q) = \frac{1}{n}\sum_{i=1}^n[\rho_\tau(\mathbf{u}'\mathbf{y}_i - q)]$ denotes the empirical version of the loss function defined in equation (2).

2.3. The multivariate method of simulated quantiles

The MSQ introduced by Dominicy and Veredas (2013) is a likelihood–free simulation–based inferential procedure based on matching quantile–based measures, it is particularly useful in situations where either the density function is not analytically available and/or moments do not exist, and 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, is that empirical quantiles are robust ordered statistics being able to achieve high protection against bias induced by the presence of outlier contamination. To introduce MMSQ estimator let us define the following notations.

Let $\boldsymbol{\Phi}_{\boldsymbol{\vartheta}}$ be a $b \times 1$ vector of projectional quantile functions assumed to be continuously differentiable with respect to $\boldsymbol{\vartheta}$ for all $\mathbf{Y}$ and measurable for $\mathbf{Y}$ and for all $\boldsymbol{\vartheta} \subset \boldsymbol{\Theta}$. Let us assume also that $\boldsymbol{\Phi}_{\boldsymbol{\vartheta},}$ cannot be computed analytically but it can be empirically estimated. Let $\hat{\boldsymbol{\Phi}}_n$ be the vector of projectional quantile functions computed over the sample, that is measurable of $\mathbf{Y}$; let $\tilde{\boldsymbol{\Phi}}_\vartheta$ be its counterpart computed over simulated data, the subscript $\boldsymbol{\vartheta}$ emphasises that it depends on the parameters used to simulate the data.

At each iteration $j = 1, 2, \ldots$, the MMSQ estimates $\tilde{\boldsymbol{\Phi}}_{\vartheta^{(j)}}$ on R samples simulated from $\mathbf{y}^*_{r,j} \sim F_{\mathbf{Y}}\left(\cdot, \boldsymbol{\vartheta}^{(j)}\right)$, for $r = 1, 2, \ldots, R$, that is $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}^{(j)}} = \frac{1}{R}\sum_{r=1}^R \tilde{\boldsymbol{\Phi}}^r_{\boldsymbol{\vartheta}^{(j)}}$, where the superscript r identifies the simulated path on which the vector of projectional quantiles has been computed. The parameters are subsequently updated by minimising the distance between the two vectors of projectional quantiles functions $\hat{\boldsymbol{\Phi}}_{n,}$ and $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}^{(j)}}$ as follows

$$\hat{\boldsymbol{\vartheta}}_n = \arg\min_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}} Q_n\left(\boldsymbol{\vartheta}\right), \tag{3}$$

$$Q_n\left(\boldsymbol{\vartheta}\right) = \left(\hat{\boldsymbol{\Phi}}_n - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}\right)'\mathbf{W}\left(\hat{\boldsymbol{\Phi}}_n - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}\right), \tag{4}$$

where $\mathbf{W}$ is a $b \times b$ symmetric positive definite weighting matrix. Asymptotic results presented in section 2.4 provide information regarding the choice of the optimal weighting matrix $\mathbf{W}$.

The vector of projectional quantile functions $\boldsymbol{\Phi}_{\boldsymbol{\vartheta}}$ 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) 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)

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

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 depends either on the nature of the 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 scaling matrix of multivariate elliptical distributions by means of a measure of co–dispersion which consists in the interquartile range of the standardised variables projected along the bisector. The MMSQ based on projectional quantiles is more flexible and it allows to deal with more general distributions than elliptically contoured one because it relies on the construction of quantile based measures of variables projected along a set of optimal directions which depend upon the considered distribution. The selection of the relevant direction is deferred to Section 4.

2.4. Asymptotic theory

In this section we recall the asymptotic theory that holds for simulation–based estimators, the proofs which require some algebra are available in supplementary materials, the other can be derived by Dominicy and Veredas (2013) and Gouriéroux et al. (1993).

2.4.1. Preliminary notations and hypothesis

Remark 2.1. It is worth noting that the minimisation problem in equation 1, admits a unique solution if the distribution of the random vector $\mathbf{Y}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m$, with finite first order moment, and if $f_{\mathbf{Y}}$ has connected support (see Kong and Mizera (2012) and Paindaveine and Šiman (2011) for further details). The only assumption that we need to worry about is the finiteness of the first order moment, while the remaining two conditions are always satisfied because we restrict our attention to absolutely continuous random variables in this work. Thus, hereafter, we will assume that the random variable $\mathbf{Y}$ has finite first order moment.

For the sake of clarity, let us first define the following notations:

- Let $\boldsymbol{Y} \in \mathbb{R}^m$ be a random vector with cumulative distribution function $F_{\boldsymbol{Y}}$ and variance–covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Y}}$
- Let $\{\boldsymbol{y}_i\}_{i=1}^n$ be a sample of iid observations from $F_{\boldsymbol{Y}}$
- Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_K \in \mathbb{S}^{m-1}$ and $Z_k = \boldsymbol{u}_k'\boldsymbol{Y}$ be the projected random variable along $\boldsymbol{u}_k$ with density function $f_{Z_k}(\cdot)$ and cumulative distribution function $F_{Z_k}(\cdot)$, for $k = 1, 2, \dots, K$
- Let $q_{\tau\boldsymbol{u}_k}$ be the projectional quantile along the direction $\boldsymbol{u}_k$, namely $q_{\tau\boldsymbol{u}_k}$ is the quantile of random variable Z_k defined before, then $\hat{q}_{\tau\boldsymbol{u}_k}$ is the empirical quantile computed over the random sample $\{\boldsymbol{y}_i\}_{i=1}^n$
- Let $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_s)$ where $\tau_j \in (0, 1)$, then $\boldsymbol{q}_{\boldsymbol{\tau},\boldsymbol{u}_k} = \left(q_{\tau_1\boldsymbol{u}_k}, q_{\tau_2\boldsymbol{u}_k}, \dots, q_{\tau_s\boldsymbol{u}_k}\right)$ is the vector of directional quantiles along the direction $\boldsymbol{u}_k$ and $\hat{\boldsymbol{q}}_{\boldsymbol{\tau},\boldsymbol{u}_k} = \left(\hat{q}_{\tau_1\boldsymbol{u}_k}, \hat{q}_{\tau_2\boldsymbol{u}_k}, \dots, \hat{q}_{\tau_s\boldsymbol{u}_k}\right)$ is the vector of empirical directional quantiles computed over the random sample $\{\boldsymbol{y}_i\}_{i=1}^n$.

The next theorem establish the asymptotic properties of projectional quantiles which are needed to built the asymptotic theory of the MMSQ.

Theorem 2.1. *Considering the notations introduced above, let us assume* $Var(Z_k) < \infty$, *for* $k = 1, 2, \dots, K$, F_{Z_k} *differentiable in* $q_{\tau_j\boldsymbol{u}_k}$ *and* $F'_{Z_k}\left(q_{\tau_j\boldsymbol{u}_k}\right) = f_{Z_k}\left(q_{\tau_j\boldsymbol{u}_k}\right) > 0$, *for* $k = 1, 2, \dots, K$ *and* $j = 1, 2, \dots, s$. *Then*

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

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

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

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

for $r, c = 1, 2, \dots, s$, $\tau_r \wedge \tau_c$ *stands for the minimum between* τ_r *and* τ_c*;*

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

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

as $n \to \infty$, *where* $\boldsymbol{q}_{\tau_j} = \left(q_{\tau_j \boldsymbol{u}_1}, \dots, q_{\tau_j \boldsymbol{u}_K}\right)$,

$$\upsilon_{r,c} = \begin{cases} -\dfrac{\tau_j^2}{f_{Z_r}\left(q_{\tau_j \boldsymbol{u}_r}\right) f_{Z_c}\left(q_{\tau_j \boldsymbol{u}_c}\right)} + \dfrac{F_{Z_r,Z_c}\left(q_{\tau_j \boldsymbol{u}_r}, q_{\tau_j \boldsymbol{u}_c}\right)}{f_{Z_r}\left(q_{\tau_j \boldsymbol{u}_r}\right) f_{Z_c}\left(q_{\tau_j \boldsymbol{u}_c}\right)} & \text{for} \quad r \neq c \\ \dfrac{\tau_j(1-\tau_j)}{f_{Z_r}\left(q_{\tau_j \boldsymbol{u}_r}\right)^2} & \text{for} \quad r = c, \end{cases}$$

$F_{Z_r,Z_c}(\cdot)$ *is the joint cumulative function of variables* (Z_r, Z_c) *for* $r, c = 1, 2, \dots, K$*;*

(iii) given τ_j *and* τ_l *with* $j, l = 1, 2, \dots, s$ *and* $j \neq l$ *and given* $\boldsymbol{u}_k$ *and* $\boldsymbol{u}_t$ *with* $k, t = 1, 2, \dots, K$ *and* $k \neq t$, *it holds*

$$\sqrt{n}\left(\hat{q}_{\tau_j \boldsymbol{u}_k} - q_{\tau_j \boldsymbol{u}_k}, \hat{q}_{\tau_l \boldsymbol{u}_t} - q_{\tau_l \boldsymbol{u}_t}\right) \xrightarrow{d} \mathcal{N}(0, \rho),$$

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

$$\rho = -\frac{\tau_j \tau_l}{f_{Z_k}\left(q_{\tau_j \boldsymbol{u}_k}\right) f_{Z_t}\left(q_{\tau_l \boldsymbol{u}_t}\right)} + \frac{F_{Z_k,Z_t}\left(q_{\tau_j \boldsymbol{u}_k}, q_{\tau_l \boldsymbol{u}_t}\right)}{f_{Z_k}\left(q_{\tau_j \boldsymbol{u}_k}\right) f_{Z_t}\left(q_{\tau_l \boldsymbol{u}_t}\right)}. \tag{5}$$

The proof of this theorem which is quite technical can be found in supplementary materials.

To establish the asymptotic properties of the MMSQ estimates we need the following set of assumptions, which are standard assumptions for simulation–based estimators.

Assumption 2.1. There exists a unique/unknown value $\boldsymbol{\vartheta}_0 \subset \boldsymbol{\Theta}$ such that the sample functions of projectional quantiles equal the theoretical one, provided that for each parameter of interest there is at least one projectional quantiles statistic depending on that parameter. That is $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0 \;\Leftrightarrow\; \hat{\boldsymbol{\Phi}}_n = \boldsymbol{\Phi}_{\boldsymbol{\vartheta}_0}$, where $\boldsymbol{\Phi}_{\boldsymbol{\vartheta}_0}$ and $\hat{\boldsymbol{\Phi}}_n$ are respectively the vector of projectional quantile functions and the empirical counterpart defined in section 2.3.

Assumption 2.2. $\hat{\boldsymbol{\vartheta}}$ is the unique minimiser of $\left(\hat{\boldsymbol{\Phi}}_n - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}\right)' \mathbf{W} \left(\hat{\boldsymbol{\Phi}}_n - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}\right)$.

Assumption 2.3. Let $\hat{\boldsymbol{\Omega}}_n$ be the sample variance–covariance matrix of $\hat{\boldsymbol{\Phi}}_n$ and $\boldsymbol{\Omega}_{\boldsymbol{\vartheta}_0}$ be the non–singular variance–covariance matrix of $\boldsymbol{\Phi}_{\boldsymbol{\vartheta}_0}$, then $\hat{\boldsymbol{\Omega}}_n$ converges to $\boldsymbol{\Omega}_{\boldsymbol{\vartheta}_0}$ as n goes to infinity.

Assumption 2.4. The matrix $\left(\frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}}{\partial \boldsymbol{\vartheta}'} \mathbf{W} \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}}}{\partial \boldsymbol{\vartheta}}\right)$ is non–singular.

It is worth to note that Assumption 2.1 requires that quantile–based summary statistics together with the corresponding directions have to be chosen accurately, namely for each parameter there must be at least one directional quantile–based statistic which is a function of that parameter. This requirement is needed to exclude free parameters which would lead to infinite solution of the optimisation problem defined in Equations (3) and (4). Assumption 2.2 holds true from the previous assumption and from the fact that the objective function is a quadratic function. Assumption 2.3 comes from the law of large numbers and from a proper choice of the scoring function $\boldsymbol{\Phi}$. Assumption 2.4 holds true since we are considering absolutely continue random variables with cumulative distribution function being differentiable. Although most of the assumptions holds since they are standard assumptions in simulation–based estimation (Gouriéroux et al. (1993)), for the sake of completeness we found worthwhile to list them.

2.4.2. Main asymptotic results

Theorem 2.2. *Under the hypothesis of Theorem 2.1 and assumptions 2.1–2.3, the following properties hold:*

$$\sqrt{n}\left(\hat{\boldsymbol{\Phi}}_n - \boldsymbol{\Phi}_{\boldsymbol{\vartheta}_0}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\vartheta}_0}\right),$$

$$\sqrt{n}\left(\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}} - \boldsymbol{\Phi}_{\boldsymbol{\vartheta}_0}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\vartheta}_0}\right),$$

as $n \to \infty$, *where* $\boldsymbol{\Omega}_{\boldsymbol{\vartheta}_0} = \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}_0}}{\partial \mathbf{q}'} \boldsymbol{\Lambda} \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}_0}}{\partial \mathbf{q}}$, $\boldsymbol{q} = \left(\boldsymbol{q}_{\tau_1,\boldsymbol{u}_1}, \boldsymbol{q}_{\tau_2,\boldsymbol{u}_2}, \dots, \boldsymbol{q}_{\tau_K,\boldsymbol{u}_K}\right)'$, $\boldsymbol{\Lambda}$ *is the variance–covariance matrix of the projectional quantiles whose elements are defined in Theorem 2.1 according to each couple of entry and* $\frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}_0}}{\partial \boldsymbol{q}} = \operatorname{diag}\left\{\frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}_0}}{\partial \mathbf{q}_{\tau_1,\boldsymbol{u}_1}}, \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}_0}}{\partial \mathbf{q}_{\tau_2,\boldsymbol{u}_2}}, \dots, \frac{\partial \boldsymbol{\Phi}_{\boldsymbol{\vartheta}_0}}{\partial \mathbf{q}_{\tau_K,\boldsymbol{u}_K}}\right\}$.

Next theorem shows the asymptotic properties of the MMSQ estimator.

Theorem 2.3. *Under the hypothesis of Theorem 2.1 and assumptions 2.1 – 2.4, we have*

$$\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\left(1+\frac{1}{R}\right)\boldsymbol{D}_{\boldsymbol{\vartheta}_0}\mathbf{W}\boldsymbol{\Omega}_{\boldsymbol{\vartheta}_0}\mathbf{W}'\boldsymbol{D}'_{\boldsymbol{\vartheta}_0}\right),$$

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

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

Corollary 2.1. *Under the hypothesis of Theorem 2.1 and assumptions 2.1 – 2.4, the optimal weighting matrix is* $\mathbf{W}^*=\boldsymbol{\Omega}^{-1}_{\boldsymbol{\vartheta}_0}$. *Therefore, the efficient method of simulated quantiles estimator E–MMSQ, as* $n\to\infty$, *has the following asymptotic distribution*

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

These results can be easily derived from previous works, see Dominicy and Veredas (2013) and Gouriéroux et al. (1993).

3. Handling sparsity

In this section the MMSQ estimator is extended in order to achieve sparse estimation of the scaling matrix. The curse of dimensionality is a well known issue in multivariate and high–dimensional analysis. Indeed, in such settings the number of parameters to be estimated grows at least quadratically with the dimension of the model. Sparse estimators became therefore very powerful instruments to face this problem, indeed they correctly identify sparsity pattern and, as consequence, reduce the complexity of the model together with the computational effort required for the estimation. Although large literature is devoted to sparse estimators, they are only related to likelihood–based setting, therefore the need to introduce the S–MMSQ estimator. The smoothly clipped absolute deviation (SCAD) ℓ_1–penalty of Fan and Li (2001) is introduced into the MMSQ objective function. Formally, let $\boldsymbol{\Sigma}$ be the scaling matrix of $\boldsymbol{Y}$, whose elements are denoted as σ_{ij} for $i,j=1,\ldots,m$, 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 concave penalty function defined as:

$$p_\lambda(|\gamma|)=\begin{cases}\lambda|\gamma| & \text{if } |\gamma|\le\lambda\\ \frac{1}{a-1}\left(a\lambda|\gamma|-\frac{\gamma^2}{2}\right)-\frac{\lambda^2}{2(a-1)} & \text{if } \lambda<\gamma\le a\lambda\\ \frac{\lambda^2(a+1)}{2} & \text{if } a\lambda<|\gamma|,\end{cases} \tag{6}$$

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 derivative equal to 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 penalised MMSQ estimator minimises the penalised MMSQ objective function, defined as follows

$$\hat{\boldsymbol{\vartheta}}=\arg\min_{\boldsymbol{\vartheta}}\mathcal{Q}^\star\left(\boldsymbol{\vartheta}\right), \tag{7}$$

$$\mathcal{Q}^\star\left(\boldsymbol{\vartheta}\right)=\left(\hat{\boldsymbol{\Phi}}_n-\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}\right)'\mathbf{W}\left(\hat{\boldsymbol{\Phi}}_n-\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}\right)+n\sum_{i<j}p_\lambda\left(|\sigma_{ij}|\right), \tag{8}$$

where $\mathcal{Q}^\star\left(\boldsymbol{\vartheta}\right)$ is the penalised distance between $\hat{\boldsymbol{\Phi}}_n$ and $\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}$. 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. Given the smoothness of the MMSQ objective function, it turns out that the S-MMSQ is an Oracle estimator as detailed in the following theorem.

Theorem 3.1. *Let* $\boldsymbol{\vartheta}_0=\left(\boldsymbol{\vartheta}_0^1,\boldsymbol{\vartheta}_0^0\right)$ *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* $\boldsymbol{\vartheta}_0^0=\mathbf{0}\in\mathbb{R}^{k-s}$ *and let* $\mathcal{A}=\left\{(i,j):i<j,\sigma_{ij,0}\in\boldsymbol{\vartheta}_0^1\right\}$. *Given the SCAD penalty function* $p_\lambda\left(|\sigma_{ij}|\right)$, *for a sequence of* λ_n *such that* $\lambda_n\to0$, *and* $\sqrt{n}\lambda_n\to\infty$, *as* $n\to\infty$

i) *there exists a local minimiser* $\hat{\boldsymbol{\vartheta}}$ *of* $\mathcal{Q}^\star\left(\boldsymbol{\vartheta}\right)$ *in (7) with* $\|\hat{\boldsymbol{\vartheta}}-\boldsymbol{\vartheta}_0\|=\mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. *Furthermore, we have*

$$\lim_{n\to\infty}\mathbb{P}\left(\hat{\boldsymbol{\vartheta}}^0=0\right)=1; \tag{9}$$

ii) *the non–zero parameters* $\hat{\boldsymbol{\vartheta}}^1$ *has the following asymptotic distribution:*

$$\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}^1 - \boldsymbol{\vartheta}_0^1\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \left(1+\frac{1}{R}\right)\boldsymbol{D}_{\boldsymbol{\vartheta}_0^1}\mathbf{W}^1\boldsymbol{\Omega}_{\boldsymbol{\vartheta}_0^1}\mathbf{W}^{1\prime}\boldsymbol{D}'_{\boldsymbol{\vartheta}_0^1}\right), \tag{10}$$

as $n\to\infty$, *where* $\boldsymbol{D}_{\boldsymbol{\vartheta}^1} = \left(\frac{\partial\boldsymbol{\Phi}_{\boldsymbol{\vartheta}^1}}{\partial\boldsymbol{\vartheta}^{1\prime}}\mathbf{W}^1\frac{\partial\boldsymbol{\Phi}_{\boldsymbol{\vartheta}^1}}{\partial\boldsymbol{\vartheta}^1}\right)^{-1}\frac{\partial\boldsymbol{\Phi}_{\boldsymbol{\vartheta}^1}}{\partial\boldsymbol{\vartheta}^1}$ *and* $\mathbf{W}^1$ *is the weighting matrix related to the non-zero parameters.*

3.1. 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. Therefore, we consider an implementation that is similar to the Graphical Lasso algorithm introduced by Friedman et al. (2008) combined with the Maximisation–Minimisation algorithm exploiting a local quadratic approximation of the SCAD proposed by Fan and Li (2001) and Hunter and Li (2005). We outline the steps of the algorithm below.

The objective function $\mathcal{Q}^\star$ in equation (7) can be locally approximated, except for a constant term by

$$\mathcal{Q}^\star(\boldsymbol{\vartheta}) \approx (\hat{\boldsymbol{\Phi}}_n - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0})'\mathbf{W}(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}) - \frac{\partial\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}}{\partial\boldsymbol{\vartheta}}\mathbf{W}(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0})(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) + \frac{1}{2}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)'\frac{\partial\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}}{\partial\boldsymbol{\vartheta}}\mathbf{W}\frac{\partial\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}}{\partial\boldsymbol{\vartheta}}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) + \frac{n}{2}\boldsymbol{\vartheta}'\boldsymbol{S}_\lambda(\boldsymbol{\vartheta}_0)\boldsymbol{\vartheta},$$

where $\boldsymbol{S}_\lambda(\boldsymbol{\vartheta}_0) = \operatorname{diag}\left\{\mathbf{0}, \frac{p'_\lambda(|\sigma_{ij,0}|)}{|\sigma_{ij,0}|}; i>j, \sigma_{ij,0}\in\boldsymbol{\vartheta}_0^1\right\}$. By applying the first order condition we get the following formula

$$\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0 - \left[\frac{\partial\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}}{\partial\boldsymbol{\vartheta}'}\mathbf{W}\frac{\partial\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}}{\partial\boldsymbol{\vartheta}} + n\boldsymbol{S}_\lambda(\boldsymbol{\vartheta}_0)\right]^{-1} \times \left[-\frac{\partial\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}}{\partial\boldsymbol{\vartheta}}\mathbf{W}(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}_0}) + n\boldsymbol{S}_\lambda(\boldsymbol{\vartheta}_0)\boldsymbol{\vartheta}_0\right], \tag{11}$$

which allows to find the optimal solution iteratively.

Now, in order to get a positive definite matrix at each iteration, let $\boldsymbol{\Sigma}$ be a correlation matrix of dimension $n\times n$ and let us consider the following partition

$$\boldsymbol{\Sigma} = \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\sigma}_{12}\\ \boldsymbol{\sigma}'_{12} & 1\end{bmatrix},$$

where $\boldsymbol{\Sigma}_{11}$ is a matrix of dimension $(n-1)\times(n-1)$ and $\boldsymbol{\sigma}_{12}$ is a vector of dimension $n-1$. At each iteration $j=1,2,\dots$ of the optimisation algorithm we apply a step of the Newton–Raphson algorithm to $\boldsymbol{\sigma}_{12}$, using the formula in equation (11), as follows

$$\hat{\boldsymbol{\sigma}}_{12}^{(j)} = \hat{\boldsymbol{\sigma}}_{12}^{(j-1)} - \left[\left.\frac{\partial\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}}{\partial\boldsymbol{\sigma}_{12}'}\right|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}^{(j-1)}}\mathbf{W}\left.\frac{\partial\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}}{\partial\boldsymbol{\sigma}_{12}}\right|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}^{(j-1)}} + n\boldsymbol{S}_\lambda\left(\hat{\boldsymbol{\sigma}}_{12}^{(j-1)}\right)\right]^{-1} \times\left[-\left.\frac{\partial\tilde{\boldsymbol{\Phi}}_{\boldsymbol{\vartheta}}}{\partial\boldsymbol{\sigma}_{12}'}\right|_{\boldsymbol{\vartheta}=\hat{\boldsymbol{\vartheta}}^{(j-1)}}\mathbf{W}\left(\hat{\boldsymbol{\Phi}}_n - \tilde{\boldsymbol{\Phi}}_{\hat{\boldsymbol{\vartheta}}^{(j-1)}}\right) + n\boldsymbol{S}_\lambda\left(\hat{\boldsymbol{\sigma}}_{12}^{(j-1)}\right)\hat{\boldsymbol{\sigma}}_{12}^{(j-1)}\right], \tag{12}$$

where $\hat{\boldsymbol{\vartheta}}^{(j-1)}$ is the estimate of $\boldsymbol{\vartheta}_0$ at iteration $j-1$. Then we consider the transformation

$$\boldsymbol{\sigma}_{12}^{\star(j)} \to \frac{\hat{\boldsymbol{\sigma}}_{12}^{(j)}}{1+\hat{\boldsymbol{\sigma}}_{12}^{(j)\prime}\left[\hat{\boldsymbol{\Sigma}}_{11}^{(j-1)}\right]^{-1}\hat{\boldsymbol{\sigma}}_{12}^{(j)}},$$

and update the scaling matrix as

$$\hat{\boldsymbol{\Sigma}}^{(j)} = \begin{bmatrix}\hat{\boldsymbol{\Sigma}}_{11}^{(j-1)} & \boldsymbol{\sigma}_{12}^{\star(j)}\\ \boldsymbol{\sigma}_{12}^{\star(j)} & 1\end{bmatrix}.$$

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.

Regarding the parameter R, it helps in controlling the variability of the simulation paths. We considered several values of R, namely from 5 to 50 and it turned out that for $R\geq 10$ the quantile–based statistics did not change significantly, therefore we choose $R=10$, as also suggested in Dominicy and Veredas (2013).

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.

4. Synthetic data examples

As mentioned in the introduction the Stable distribution plays an interesting role in modelling multivariate data. Its peculiarity of heaving heavy tailed properties and its closeness under summation make it appealing in the financial contest. Despite its characteristics, estimation of parameters has been always challenging and this feature greatly limited its use in applied works requiring simulation–based methods. In this section we briefly introduce the multivariate Elliptical Stable distribution (ESD) previously considered by Lombardi and Veredas (2009).

4.1. Multivariate Elliptical Stable distribution

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

$$\mathbf{Y} =^d \boldsymbol{\xi} + \mathcal{R}\boldsymbol{\Gamma}\mathbf{U}, \tag{13}$$

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 $\mathbb{S}^{m-1}$ 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, i.e., $\mathbf{Y} \sim \mathcal{ESD}_m\big(\alpha, \boldsymbol{\xi}, \boldsymbol{\Omega}\big)$. 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), Embrechts et al. (2005) and 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 $\mathbf{Y} \sim \mathcal{ESD}_m\big(\alpha, \boldsymbol{\xi}, \boldsymbol{\Omega}\big)$ if and only if $\mathbf{Y}$ has the following stochastic representation as a scale mixture of Gaussian distributions

$$\mathbf{Y} = \boldsymbol{\xi} + \zeta^{\frac{1}{2}}\boldsymbol{X}, \tag{14}$$

where $\zeta \sim \mathcal{S}_{\frac{\alpha}{2}}(0, \bar{\omega}_\alpha, 1)$ and $\boldsymbol{X} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ independent of ζ. 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.

4.2. How to choose optimal directions

Before we turn to illustrate our simulation framework, we should solve an important issue related to the application of the MMSQ that concerns the choice of the directions. Indeed, the easiest solution is to choose an equally spaced grid of directions, an approach that would be computational expensive. Therefore, we choose optimal directions $\mathbf{u}^*$ according to the following definition 4.1 which allows to maximise the information contained in the chosen measure.

Definition 4.1. Let us consider a given parameter of interest $\vartheta^\star \subset \Theta_k \in \mathbb{R}^k$ 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 $\vartheta^\star$, 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

$$\mathbf{u}_{\max} \in \left\{\arg\max_{\mathbf{u}\in\mathbb{S}^{h-1}} q_{\tau\mathbf{u}}\right\}. \tag{15}$$

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 (15) 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) or numerically. It is worth to note that, when finding the optimal direction, the choice of h only depends on the parameter of interest and on the marginalisation properties of the distribution regardless the dimension m, as described in section 4.4. If more than one optimal direction is found for a given parameter, the corresponding projectional quantile based function will be computed along all the directions found. As said in the beginning, the best would be to find just one optimal direction for parameter because it makes the optimisation problem easier to solve, thus adding direction has only an impact of the efficiency of the minimisation problem.

4.3. Initialisation

Let us fix $m \geq 2$. Since each variables Y_i have univariate Elliptical Stable distribution, then marginals' parameters can be estimated using the approach of McCulloch (1986). The off–diagonal parameter of the scale matrix is estimated using the following procedure. For each couple of variables $\boldsymbol{Y}_{ij} = (Y_i, Y_j)'$ it holds $\boldsymbol{Y}_{ij} \sim \mathcal{ESD}_2(\alpha, \boldsymbol{\xi}_{ij}, \boldsymbol{\Omega}_{ij})$, where $\boldsymbol{\xi}_{ij} = (\xi_i, \xi_j)'$ and $\boldsymbol{\Omega}_{ij} = [\begin{smallmatrix}\omega_i^2 & \omega_{ij} \\ \omega_{ij} & \omega_j^2\end{smallmatrix}]$. Let us consider the standardised variables $\boldsymbol{X}_{ij} = (X_i, X_j)'$ where $(X_i, X_j) = (\frac{Y_i - \xi_i}{\omega_{ii}}, \frac{Y_j - \xi_j}{\omega_{jj}})$, then $\boldsymbol{X}_{ij} \sim \mathcal{ESD}_2(\alpha, \mathbf{0}, \bar{\boldsymbol{\Omega}}_{ij})$ where $\bar{\boldsymbol{\Omega}}_{ij} = [\begin{smallmatrix}1 & \rho_{ij} \\ \rho_{ij} & 1\end{smallmatrix}]$. Using the Definition 4.1, it turns out that the optimal direction for ρ_{ij} is $\boldsymbol{u} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)'$. Therefore, we project $\boldsymbol{X}_{ij}$ along $\boldsymbol{u}$ and we obtain the variable $X_{\boldsymbol{u}} = \boldsymbol{u}'\boldsymbol{X}_{ij}$ such that $X_{\boldsymbol{u}} \sim \mathcal{ESD}_1\left(\alpha, 0, 1 + \rho_{ij}\right)$. Now, since $X_{\boldsymbol{u}}$ is a univariate random variable we can apply the method of McCulloch (1986) to initialise the scale of a univariate ESD.

The proposed initialisation provides initial conditions in a neighbourhood of the global minimiser allowing a fast convergence of the optimisation problem to the optimal solution. It is worth to note that random initial conditions may not lead to the global minimiser.

4.4. Simulation results

In this Section we consider simulation examples for the ESD distribution $\mathbf{Y} \sim \mathcal{ESD}_m\left(\alpha, \boldsymbol{\xi}, \boldsymbol{\Omega}\right)$ as defined in section 4.1. We first need to select the quantile–based measures which are informative for each parameter, thus, we select for α a measure $\kappa_{\mathbf{u}}$ related to the kurtosis of the distribution, for the locations the median $m_{\mathbf{u}}$ and for the elements of the scaling matrix we opt for a measure of dispersion $\varsigma_{\mathbf{u}}$, and all the measures will be calculated along appropriately chosen directions, as it will be discussed later in this section. More in detail, we choose

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

where $\mathbf{u} \in \mathcal{S}^{m-1}$ defines a relevant direction. Of course more and/or different quantile-based statistics can be used, we tested several measures and select those providing better results. 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 $i = 1, 2, \ldots, m$, where ω_{ii} is the i–th element of the main diagonal of the matrix $\boldsymbol{\Omega}$. By exploiting this property, 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 $i = 1, 2, \ldots, m$ are the canonical directions. It still remains to consider the optimal directions for the off–diagonal elements of the scale matrix ω_{ij}, with $i, j = 1, 2, \ldots, m$ and $i \neq j$. Again we exploit the closure with respect to marginalisation. Specifically, let $\boldsymbol{Y}_{ij} = \left(Y_i, Y_j\right)$, then $\boldsymbol{Y}_{ij} \sim \mathcal{ESD}_2\left(\alpha, \boldsymbol{\xi}_{ij}, \boldsymbol{\Omega}_{ij}\right)$ as already pointed out in paragraph 4.3. Moreover, let $Y_{ij,\mathbf{u}} = \mathbf{u}'\boldsymbol{Y}_{ij}$ be the projection of $\boldsymbol{Y}_{ij}$ along $\mathbf{u}$, then $Y_{ij,\mathbf{u}} \sim \mathcal{ESD}_1\left(\alpha, \mathbf{u}'\boldsymbol{\xi}_{ij}, \mathbf{u}'\boldsymbol{\Omega}_{ij}\mathbf{u}\right)$, (see Embrechts et al. (2005)), from which we have the following representation of the projected ESD random variable

$$Y_{ij,\mathbf{u}} = \mathbf{u}'\boldsymbol{\xi}_{ij} + \sqrt{\mathbf{u}'\boldsymbol{\Omega}_{ij}\mathbf{u}}Z, \tag{16}$$

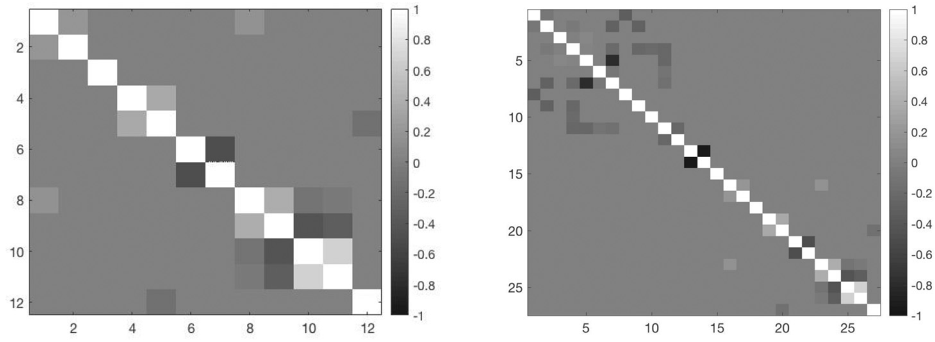
where $Z \sim \mathcal{ESD}_1(\alpha, 0, 1)$. Thus, following Definition 4.1, to find the optimal directions we solve

$$\mathbf{u}_{\max} = \arg\max_{\mathbf{u}\in\mathbb{S}^1} \mathbf{u}'\boldsymbol{\xi}_{ij} + \sqrt{\mathbf{u}'\boldsymbol{\Omega}_{ij}\mathbf{u}}, \tag{17}$$

which is a quadratic optimisation problem that can be solved using the method of Lagrangian multiplier plugging in the value obtained from initialisation procedure detailed in section 4.3.

The optimal direction $\mathbf{u}_{\max}$ is then plugged into $\mathbf{u}^* = \left(0, \ldots, u_{1,\max}, \ldots, u_{2,\max}, \ldots, 0\right)$. In these simulation examples we use as weighting matrix the identity matrix.

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 those considered in Lombardi and Veredas (2009). We consider two different sample sizes $n = 500, 2000$ and we fix



(a) $\mathbf{\Sigma}_{12}^{s}$ (b) $\mathbf{\Sigma}_{27}^{s}$

Fig. 1. Band structure of the scale matrices considered in the two simulation examples to test the performances the S–MMSQ method.

Table 1
Frobenius norm, F1–Score and Kullbach–Leibler information between the true scale matrix 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	1.70	1.90	1.95	2.00
Frobenius norm	*Dimension 12*				*Dimension 27*			
GLasso	1.595	1.0392	0.81693	0.59958	4.5938	2.6358	1.8644	0.76058
	(0.5232)	*(0.47659)*	*(0.34976)*	*(0.074508)*	*(3.0661)*	*(2.071)*	*(1.5321)*	*(0.045129)*
SCAD	1.5043	0.94056	0.7378	0.58001	4.5847	2.5318	1.7361	0.56438
	(0.56176)	*(0.51448)*	*(0.39827)*	*(0.11528)*	*(3.3211)*	*(2.1012)*	*(1.5937)*	*(0.063328)*
Adaptive Lasso	1.4486	0.90578	0.69566	0.50441	4.084	2.2872	1.7195	0.65269
	(0.5416)	*(0.46181)*	*(0.34298)*	*(0.087179)*	*(3.2848)*	*(1.6282)*	*(1.5373)*	*(0.047185)*
S–MMSQ	1.6618	1.4111	1.293	1.2417	2.6987	2.449	2.3426	2.1677
	(0.21718)	*(0.22563)*	*(0.22635)*	*(0.22013)*	*(0.2791)*	*(0.26377)*	*(0.28864)*	*(0.24305)*
F_1–score	*Dimension 12*				*Dimension 27*			
GLasso	0.1313	0.012143	0.019025	0	0.037952	0.007118	0.0036548	0
	(0.23919)	*(0.079663)*	*(0.1103)*	*(0)*	*(0.10075)*	*(0.07118)*	*(0.036548)*	*(0)*
SCAD	0.26295	0.17153	0.15148	0.23174	0.033123	0.0085093	0.0036548	0.0015072
	(0.27865)	*(0.22789)*	*(0.21994)*	*(0.23612)*	*(0.095177)*	*(0.072389)*	*(0.036548)*	*(0.015072)*
Adaptive Lasso	0.2431	0.080443	0.057361	0.037187	0.13042	0.040525	0.0075655	0
	(0.33484)	*(0.17254)*	*(0.1628)*	*(0.10126)*	*(0.23048)*	*(0.15814)*	*(0.075655)*	*(0)*
S–MMSQ	0.40246	0.55827	0.62059	0.69567	0.83754	0.75499	0.71847	0.66897
	(0.17051)	*(0.14057)*	*(0.13682)*	*(0.089005)*	*(0.097355)*	*(0.086734)*	*(0.079755)*	*(0.048205)*
KL	*Dimension 12*				*Dimension 27*			
GLasso	0.68981	0.29197	0.18876	0.10059	6.6643	2.3116	0.98558	0.17044
	(0.36107)	*(0.25353)*	*(0.17321)*	*(0.024998)*	*(8.8661)*	*(4.2476)*	*(1.7369)*	*(0.021347)*
SCAD	0.63751	0.24506	0.16588	0.09049	6.8927	2.2701	0.92517	0.095768
	(0.39392)	*(0.24673)*	*(0.19988)*	*(0.03358)*	*(8.9943)*	*(4.3379)*	*(1.8981)*	*(0.018791)*
Adaptive Lasso	0.58807	0.2294	0.14527	0.0735	6.627	2.3228	0.96203	0.13577
	(0.34298)	*(0.2109)*	*(0.15541)*	*(0.022154)*	*(8.9975)*	*(4.6305)*	*(2.0405)*	*(0.020124)*
S–MMSQ	0.96549	0.77602	0.67512	0.64992	58.4657	53.6626	51.8209	48.6645
	(0.20521)	*(0.22501)*	*(0.21598)*	*(0.21598)*	*(7.8325)*	*(9.2006)*	*(8.7965)*	*(8.7965)*

$R = 10$. In supplementary materials we report estimation results obtained over 100 replications. Our results show that the MMSQ estimator is unbiased and that the empirical coverages are 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.

To illustrate the performance of the S–MMSQ method two simulation examples are provided. The first considers a sample of $n = 500$ observations from a ESD of dimension $m = 12$, with locations at zero, four different values of the characteristic exponent $\alpha = (1.70, 1.90, 1.95, 2.00)$ and scale matrix $\mathbf{\Sigma}_{12}^{s}$ equal to that considered in Wang (2015). The band structure of the scale matrix $\mathbf{\Sigma}_{12}^{s}$ is reported in Figure 1 panel (a). The second example considers a sample of $n = 800$ observations from the ESD of dimension 27 with location and characteristic exponent chosen as before and block–diagonal scale matrix $\mathbf{\Sigma}_{27}^{s} = \text{diag}\left\{\mathbf{\Sigma}_{12}^{s}, \mathbf{\Sigma}_{15}^{s}\right\}$ and $\mathbf{\Sigma}_{15}^{s}$ is the covariance matrix in Section 4.2 of Wang (2010). The band structure of the scale matrix $\mathbf{\Sigma}_{27}^{s}$ is reported in Figure 1 panel (b).

We compare the S–MMSQ with three alternative algorithms: the graphical LASSO (GLASSO) of Friedman et al. (2008), the graphical LASSO with SCAD penalty (SCAD), the graphical adaptive Lasso (Adaptive Lasso) of Fan et al. (2009). The aim of the proposed simulation examples is to compare the performance of the proposed algorithm for different levels of deviations from the Gaussian assumption which represents the benchmark assumption for the competing algorithms. Results are reported in Table 1 in terms of average Frobenius norm, F1–Score and Kullback–Leibler (KL) divergence over 100 replications and their standard deviations. The S–MMSQ method performs very well with respect to the alternatives in terms of F_1-score for all the considered values of the characteristic exponent α. This means that the method correctly identifies the sparse structure of the matrices regardless the amount of the deviation from the Gaussian assumption. This results is confirmed by visual inspection of the figures included in supplementary materials reporting the band structure of the true and estimated matrices averaged across the 100 replications. The S–MMSQ method does a good job also in terms of Frobenius norm but only in dimension $m = 27$. The worst results are reported by the S–MMSQ in terms of KL divergence. A possible explanation for those results would be that maximum likelihood methods essentially minimise the KL divergence, therefore reported values for the alternative methods are the minimum obtainable.

5. Conclusion

In this paper we present an extension of the method of simulated quantiles proposed in Dominicy and Veredas (2013) to a multivariate framework. The method is useful when either the density function does not have an analytical expression or/and moments do not exits, provided that it can be easily simulated. This is the case of many distributions widely used in quantitative finance, for instance Stable, Elliptical Stable, Tempered Stable distributions allowing for skewness and heavy tails which are features that characterise financial data. Such distributions are mainly used in univariate setting because their estimation in multivariate framework lead to challenging numerical integration or to the need of misspecified models. The main challenge in extending the method of simulated quantiles to a multivariate framework is the choice of quantiles, which is not obvious out of univariate setting. Projectional quantiles along optimal directions are introduced in order to carry the information over the parameters of interest in an efficient way. The asymptotic theory of the MMSQ estimator comes from the well known asymptotic properties of the simulation–based methods if the direction are fixed. We also introduce a sparse version of the MMSQ using the SCAD ℓ_1–penalty into the MMSQ objective function in order to achieve sparse estimation of the scaling matrix. The need to introduce sparsity when dealing with modern datasets are clear from the recent huge literature around sparse estimators. Although this work places within this literature, it deeply differs from most of the works since it is not related to maximum likelihood estimation.

The smoothness of the MMSQ optimisation problem allows to extend the well known oracle properties of the SCAD estimator in the likelihood–based context to the MMSQ estimator. The method is illustrated using several synthetic datasets from the Elliptical Stable distribution for which alternative methods are recognised to perform poorly.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Paola Stolfi: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Supervision. **Mauro Bernardi:** Conceptualization, Methodology, Writing – original draft. **Lea Petrella:** Writing – original draft.

Aknowledgement

We sincerely thank Dott. Davide Vergni for the insightful discussions we had when working on this project and for the critical reading of this paper and Dott. Filippo Castiglione for its helpful advices and support during the revision process.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ecosta.2022.01.006

References

Bernardi, M., Stolfi, P., 2020. A dominance test for measuring financial connectedness. The European Journal of Finance 26 (2-3), 119–141.
Bianchi, M.L., Rachev, S.T., Kim, Y.S., Fabozzi, F.J., 2010. Tempered stable distributions and processes in finance: numerical analysis. In: Mathematical and statistical methods for actuarial sciences and finance. Springer, pp. 33–42.
Bien, J., Tibshirani, R.J., 2011. Sparse estimation of a covariance matrix. Biometrika 98 (4), 807–820. doi:10.1093/biomet/asr054.
Branco, M.D., Dey, D.K., 2001. A general class of multivariate skew-elliptical distributions. J. Multivariate Anal. 79 (1), 99–113. doi:10.1006/jmva.2000.1960.
Cherubini, U., Luciano, E., Vecchiato, W., 2004. Copula methods in finance. John Wiley & Sons.

Dominicy, Y., Ogata, H., Veredas, D., 2013. Inference for vast dimensional elliptical distributions. Comput. Statist. 28 (4), 1853–1880. doi:10.1007/s00180-012-0384-3.

Dominicy, Y., Veredas, D., 2013. The method of simulated quantiles. J. Econometrics 172 (2), 235–247. doi:10.1016/j.jeconom.2012.08.010.

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

Fan, J., Feng, Y., Wu, Y., 2009. Network exploration via the adaptive lasso and SCAD penalties. Ann. Appl. Stat. 3 (2), 521–541. doi:10.1214/08-AOAS215.

Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist. Assoc. 96 (456), 1348–1360. doi:10.1198/016214501753382273.

Fang, K.T., Kotz, S., Ng, K.W., 1990. Symmetric multivariate and related distributions. Monographs on Statistics and Applied Probability, 36. Chapman and Hall, Ltd., London doi:10.1007/978-1-4899-2937-2.

Friedman, J., Hastie, T., Tibshirani, R., 2008. Sparse inverse covariance estimation with the graphical lasso. Biostatistics (Oxford, England) 9 (3), 432–441. doi:10.1093/biostatistics/kxm045.

Gallant, A.R., Tauchen, G., 1996. Which moments to match? Econometric Theory 12 (4), 657–681. doi:10.1017/S0266466600006976.

Gao, X., Massam, H., 2015. Estimation of symmetry-constrained Gaussian graphical models: application to clustered dense networks. J. Comput. Graph. Statist. 24 (4), 909–929. doi:10.1080/10618600.2014.937811.

Gouriéroux, C., Monfort, A., 1996. Simulation-based econometric methods. Oxford University Press.

Gouriéroux, C., Monfort, A., Renault, E., 1993. Indirect inference. Journal of Applied Econometrics 8 (S1), S85–S118. doi:10.1002/jae.3950080507.

Hallin, M., Paindaveine, D., Šiman, M., 2010. Multivariate quantiles and multiple-output regression quantiles: from L_1 optimization to halfspace depth. Ann. Statist. 38 (2), 635–669. doi:10.1214/09-AOS723.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50 (4), 1029–1054. doi:10.2307/1912775.

Hunter, D.R., Li, R., 2005. Variable selection using mm algorithms. Ann. Statist. 33 (4), 1617–1642. doi:10.1214/009053605000000200.

Kim, T.-H., White, H., 2004. On more robust estimation of skewness and kurtosis. Finance Research Letters 1 (1), 56–73. doi:10.1016/S1544-6123(03)00003-5.

Koenker, R., 2005. Quantile regression. Econometric Society Monographs, 38. Cambridge University Press, Cambridge doi:10.1017/CBO9780511754098.

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

Koponen, I., 1995. Analytic approach to the problem of convergence of truncated lévy flights towards the gaussian stochastic process. Phys. Rev. E 52, 1197–1199. doi:10.1103/PhysRevE.52.1197.

Kristensen, D., Shin, Y., 2012. Estimation of dynamic models with nonparametric simulated maximum likelihood. J. Econometrics 167 (1), 76–94. doi:10.1016/j.jeconom.2011.09.042.

Lombardi, M.J., Veredas, D., 2009. Indirect estimation of elliptical stable distributions. Comput. Statist. Data Anal. 53 (6), 2309–2324. doi:10.1016/j.csda.2008.04.035.

Mathai, A.M., Moschopoulos, P.G., 1992. A form of multivariate gamma distribution. Ann. Inst. Statist. Math. 44 (1), 97–106. doi:10.1007/BF00048672.

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. doi:10.2307/1913621.

McNeil, A.J., Frey, R., Embrechts, P., 2015. Quantitative risk management, revised Princeton University Press, Princeton, NJ. Concepts, techniques and tools

Meinshausen, N., Bühlmann, P., 2006. High-dimensional graphs and variable selection with the lasso. Ann. Statist. 34 (3), 1436–1462. doi:10.1214/009053606000000281.

Nolan, J.P., 2003. Modeling financial data with stable distributions. In: Rachev, S.T. (Ed.), Handbook of Heavy Tailed Distributions in Finance. In: Handbooks in Finance, 10.1016/B978-044450896-6.50005-4, 1. North-Holland, Amsterdam, pp. 105–130.

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. doi:10.1007/s00180-013-0396-7.

Oh, D.H., Patton, A.J., 2013. Simulated method of moments estimation for copula-based multivariate models. Journal of the American Statistical Association 108 (502), 689–700.

Paindaveine, D., Šiman, M., 2011. On directional multiple-output quantile regression. J. Multivariate Anal. 102 (2), 193–212. doi:10.1016/j.jmva.2010.08.004.

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

Semenikhine, V., Furman, E., Su, J., 2018. On a multiplicative multivariate gamma distribution with applications in insurance. Risks 6 (3), 79.

Serfling, R., 2002. Quantile functions for multivariate analysis: approaches and applications. Statist. Neerlandica 56 (2), 214–232. doi:10.1111/1467-9574.00195. Special issue: Frontier research in theoretical statistics, 2000 (Eindhoven)

Stolfi, P., Bernardi, M., Petrella, L., 2018. The sparse method of simulated quantiles: An application to portfolio optimization. Statistica Neerlandica 72 (3), 375–398.

Wang, H., 2010. Sparse seemingly unrelated regression modelling: applications in finance and econometrics. Comput. Statist. Data Anal. 54 (11), 2866–2877. doi:10.1016/j.csda.2010.03.028.

Wang, H., 2015. Scaling it up: Stochastic search structure learning in graphical models. Bayesian Anal. 10 (2), 351–377. doi:10.1214/14-BA916.

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