COLLANA DEL
DIPARTIMENTO DI ECONOMIA

MULTIVARIATE METHOD OF SIMULATED QUANTILES

Mauro Bernardi - Lea Petrella - Paola Stolfi

ISSN 2279-6916 Working papers
(Dipartimento di Economia Università degli studi Roma Tre) (online)

Working Paper n° 212, 2016

REDAZIONE:
Dipartimento di Economia
Università degli Studi Roma Tre
Via Silvio D'Amico, 77 - 00145 Roma
Tel. 0039-06-57335655 fax 0039-06-57335771
E-mail: dip_eco@uniroma3.it
http://dipeco.uniroma3.it

DIPARTIMENTO DI ECONOMIA

MULTIVARIATE METHOD OF SIMULATED QUANTILES

Mauro Bernardi - Lea Petrella - Paola Stolfi

Multivariate Method of Simulated Quantiles

Mauro Bernardi*[1], Lea Petrella†[2], and Paola Stolfi‡[3]

[1]Department of Statistical Sciences, University of Padova
[2]MEMOTEF Department, Sapienza University of Rome
[3]Department of Economics, Roma Tre University

Abstract

In this paper we extend the method of simulated quantiles (MSQ) of Dominicy and Veredas (2013) to a multivariate framework (MMSQ). The MSQ like alternative likelihood–free procedures is based on the minimisation of the distance between appropriate statistics based on the true and simulated data. 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 in the multivariate setting requires a careful definition of the concept of quantile. Here, we rely on the notion of projectional quantile recently introduced by Hallin et al. (2010) and Kong and Mizera (2012). We establish consistency and asymptotic normality of the proposed estimator. As a further improvement we introduce the smoothly clipped absolute deviation (SCAD) ℓ_1–penalty of Fan and Li (2001) into the MMSQ objective function in order to achieve sparse estimation of the scaling matrix which is the major responsible for the curse of dimensionality problem. We extend the asymptotic theory and we show that the sparse–MMSQ estimator enjoys the oracle properties under mild regularity conditions.

Keywords: directional quantiles, method of simulated quantiles, quantile matching, sparsity.

JEL Classification Code: C130, C150, C390.

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,

---

*Corresponding author: Via C. Battisti, 241, 35121 Padua, Italy. e-mail: mauro.bernardi@unipd.it, Tel.: +39.049.8274165.

†e-mail: lea.petrella@uniroma1.it.

‡e-mail: paola.stolfi@uniroma3.it.

in some particular situations the maximum likelihood estimator (MLE) is difficult to compute either because of the complexity of the model or because the density of the postulated model 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 estimate especially when the dimension of the data increases. In some other situations, instead, the dimension of the model prohibits the computation of the likelihood or, even its maximisation in a reasonable computational time. In all those situations 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 correctly specified moment conditions are available and provided that they identify the parameters of interest. 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 them 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–based estimation method (QM), Koenker (2005), instead exploits the idea of the MM by choosing as estimates those parameters that match the empirical percentiles with their theoretical counterpart. The QM method has several drawbacks the main being the knowledge of the quantile function in a closed form.

This paper focuses on the method of simulated quantiles recently proposed by Dominicy and Veredas (2013) as a simulation–based extension of the QM of Koenker (2005). 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, in this paper we further extend the method of simulated quantiles to deal with multivariate data, originating the multivariate method of simulated quantiles (MMSQ, hereafter). Of course, there are many similarities between MMSQ and II method. MMSQ 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. As its univariate counterpart, MMSQ minimise the distance between summary statistics based on quantiles of the observed and synthetic data. However, unlike II, MMSQ 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 MMSQ estimates is that they inherits the robust properties of quantiles while retaining levels of efficiency comparable with the II estimators.

The extension to multivariate data is not easy because a unique definition of multivariate quantile does not exist 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 directional and projectional quantiles of Hallin et al. (2010) and Kong and Mizera (2012). 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 quite interesting in order to relax the assumption of independence between variables. An important methodological contribution 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. Furthermore, the efficiency of the MMSQ decreases as the number of quantile functions increases. 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 that are closed under linear combinations.

Throughout the paper, 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 directional 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.

An issue frequently observed in high dimensions is the curse of dimensionality, i.e., the situation where the number of parameters grows quadratically or exponentially with the dimension of the problem. In those circumstances, the right identification of the sparsity pattern becomes crucial since 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. In this paper, we penalise the quadratic 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. 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.

The MMSQ is illustrated and its effectiveness is tested through examples where synthetic datasets are simulated from the Elliptical Stable distribution already considered by Lombardi and Veredas (2009).

The remainder of the paper is organised as follows. In Section 2 the basic concepts on directional and projectional quantiles are recalled. Section 3 introduces the method of simulated quantiles, while in Section 4 we establish consistency and asymptotic normality of the proposed estimator. The asymptotic variance of the estimator is necessary to select the optimal weighting matrix for the square distance in order to obtain the efficient MMSQ estimator. Section 5 deals with the curse of dimensionality problem, introduces the penalised MMSQ estimator that induces sparsity in the scale matrix using the SCAD ℓ_1–penalty and shows that the Sparse–MMSQ enjoys the oracle properties. Finally Section 6 illustrates the method on simulated datasets and Section 7 concludes.

2 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 1. *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* $\mathbb{S}^{m-1} = \{\mathbf{u} \in \mathbb{R}^m : \mathbf{u}'\mathbf{u} = 1\}$ *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*

$$\pi_{\tau\mathbf{u}} = \left\{\mathbf{Y} : \mathbf{b}'_{\tau\mathbf{u}}\mathbf{Y} - q_{\tau\mathbf{u}} = 0\right\}, \tag{1}$$

such that

$$(q_{\tau\mathbf{u}}, \mathbf{b}'_{\tau\mathbf{u}})' \in \left\{\arg\min_{(q,\mathbf{b})} \Psi_{\tau\mathbf{u}}(q, \mathbf{b}) \quad s.t. \quad \mathbf{b}'\mathbf{u} = 1\right\}, \tag{2}$$

where

$$\Psi_{\tau\mathbf{u}}(q, \mathbf{b}) = \mathbb{E}\Big[\rho_\tau(\mathbf{b}'\mathbf{Y} - q)\Big], \tag{3}$$

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 2. *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 2, for any $\tau \in (0,1)$ the minimisation problem defined in equation (2) admits a unique solution $(q_{\tau\mathbf{u}}, \mathbf{b}_{\tau\mathbf{u}})$, 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 $(q_{\tau\mathbf{u}}, \mathbf{u})$ 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 3. *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* $\mathbb{S}^{m-1}$*, and* $\tau \in (0, 1)$*. The* $\tau\mathbf{u}$ *projectional quantile of* $\mathbf{Y}$ *is defined as follows.*

$$q_{\tau\mathbf{u}} \in \left\{ \arg\min_{q\in\mathbb{R}} \Psi_{\tau\mathbf{u}}\left(q\right) \right\}, \tag{4}$$

where $\Psi_{\tau\mathbf{u}}(q) = \Psi_{\tau\mathbf{u}}(q, \mathbf{u})$ *in equation* (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

$$q_{\tau\mathbf{u}}^n \in \left\{ \arg\min_{q} \Psi_{\tau\mathbf{u}}^n\left(q\right) \right\}, \tag{5}$$

where $\Psi_{\tau\mathbf{u}}^n(q) = \frac{1}{n}\sum_{i=1}^n \Big[\rho_\tau\left(\mathbf{u}'\mathbf{y}_i - q\right)\Big]$ denotes the empirical version of the loss function defined in equation (3).

3 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 2. Let $\mathbf{Y}$ be a d–dimensional random variable with distribution function $F_{\mathbf{Y}}(\cdot, \vartheta)$, which depends on a vector of unknown parameters $\vartheta \subset \Theta \in \mathbb{R}^k$, and $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)'$ be a vector of n independent realisations of $\mathbf{Y}$. Moreover, let $\mathbf{q}_\vartheta^{\boldsymbol{\tau},\mathbf{u}} = (q_\vartheta^{\tau_1\mathbf{u}}, q_\vartheta^{\tau_2\mathbf{u},}, \ldots, q_\vartheta^{\tau_s\mathbf{u}})$ be a $s \times 1$ vector of projectional quantiles at given confidence levels $\tau_k \in (0, 1)$ with $k = 1, 2, \ldots, s$, and $\mathbf{u} \in \mathbb{S}^{m-1}$. Let $\boldsymbol{\Phi}_{\mathbf{u},\vartheta} = \boldsymbol{\Phi}\left(\mathbf{q}_\vartheta^{\boldsymbol{\tau},\mathbf{u}}\right)$ be a $b \times 1$ vector of quantile functionsassumed to be continuously differentiable with respect to ϑ for all $\mathbf{Y}$ and measurable for $\mathbf{Y}$ and for all $\vartheta \subset \Theta$. Let us assume also that $\boldsymbol{\Phi}_{\mathbf{u},\vartheta}$ cannot be computed analytically but it can be empirically estimated on simulated data; denote those quantities by $\tilde{\boldsymbol{\Phi}}^{r}_{\mathbf{u},\vartheta}$. Let $\hat{\mathbf{q}}^{\boldsymbol{\tau},\mathbf{u}} = \left(\hat{q}^{\tau_1\mathbf{u}}, \hat{q}^{\tau_2\mathbf{u}}, \dots, \hat{q}^{\tau_s\mathbf{u}}\right)$ be a $s \times 1$ vector of sample projectional quantiles with $\mathbf{u} \in \mathbb{S}^{m-1}$ and $\tau_k \in (0,1)$ with $k = 1, 2, \dots, s$, and let $\hat{\boldsymbol{\Phi}}_{\mathbf{u}} = \boldsymbol{\Phi}\left(\hat{\mathbf{q}}^{\boldsymbol{\tau},\mathbf{u}}\right)$ be a $b \times 1$ vector of functions of sample quantiles, that is measurable of $\mathbf{Y}$.

The MMSQ at each iteration $j = 1, 2, \dots$ estimate $\tilde{\boldsymbol{\Phi}}_{\mathbf{u},\vartheta}$ on a sample of R replication simulated from $\mathbf{y}^{*}_{r,j} \sim F_{\mathbf{Y}}\left(\cdot, \vartheta^{(j)}\right)$, for $r = 1, 2, \dots, R$, as $\tilde{\boldsymbol{\Phi}}^{R}_{\mathbf{u},\vartheta_j} = \frac{1}{R}\sum_{r=1}^{R} \tilde{\boldsymbol{\Phi}}^{r}_{\mathbf{u},\vartheta_j}$, where $\tilde{\boldsymbol{\Phi}}^{r}_{\mathbf{u},\vartheta_j}$ is the function $\boldsymbol{\Phi}_{\mathbf{u},\vartheta}$ 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 $\hat{\boldsymbol{\Phi}}_{\mathbf{u}}$ and that calculated on simulated realisations $\tilde{\boldsymbol{\Phi}}^{R}_{\mathbf{u},\vartheta_j}$ as follows

$$\hat{\vartheta} = \arg\min_{\vartheta \in \vartheta} \left(\hat{\boldsymbol{\Phi}}_{\mathbf{u}} - \tilde{\boldsymbol{\Phi}}^{R}_{\mathbf{u},\vartheta}\right)' \mathbf{W}_{\vartheta} \left(\hat{\boldsymbol{\Phi}}_{\mathbf{u}} - \tilde{\boldsymbol{\Phi}}^{R}_{\mathbf{u},\vartheta}\right), \tag{6}$$

where $\mathbf{W}_{\vartheta}$ 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 $\boldsymbol{\Phi}_{\mathbf{u},\vartheta}$ 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)

$$\boldsymbol{\Phi}_{\vartheta} = \left(\frac{q_{0.95} + q_{0.05} - 2q_{0.5}}{q_{0.95} - q_{0.05}}, \frac{q_{0.95} - q_{0.05}}{q_{0.75} - q_{0.25}}, q_{0.75} - q_{0.25}, q_{0.5}\right)'. \tag{7}$$

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. This can be done by choosing the optimal direction $\mathbf{u}^{*}$ according to the following Definition 4 which allows to maximise the information contained in the chosen measure.

Definition 4. *Let us consider a given parameter of interest $\vartheta^{\star} \subset \Theta_k \in \mathbb{R}^k$ and consider the subset $\mathbf{Y}^{\star} = (Y_1^{\star}, \dots, Y_l^{\star}, \dots, Y_h^{\star})$ 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*

$$u_i^{*} = \begin{cases} u_{\max,l} & \textit{if } Y_i = Y_l^{\star} \\ 0 & \textit{otherwise}, \end{cases} \tag{8}$$

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

$$\mathbf{u}_{\max} \in \left\{ \arg \max_{\mathbf{u} \in \mathbb{S}^{h-1}} q^{\tau \mathbf{u}} \right\}. \tag{9}$$

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

$$\mathbf{u}^* = \left(0, \ldots, u_{\max,1}, 0, \ldots, 0, u_{\max,2}, \ldots, 0\right), \tag{10}$$

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 (9) are computed using the Lagrangian function as follows

$$\mathcal{L}\left(\mathbf{u}, \lambda\right) = q_{\tau \mathbf{u}} - \lambda \left(\|\mathbf{u}\| - 1\right), \tag{11}$$

by solving

$$\nabla \mathcal{L}\left(\mathbf{u}, \lambda\right) = 0, \tag{12}$$

where ∇ stands for the gradient. This equation can be solved analytically, for instance when $m = h = 2$ for ESD distribution as shown in section 6.1, or numerically.

Let $\mathbf{U}^*$ collect all the optimal solutions $\mathbf{u}_j^*$ for $q^{\tau_j \mathbf{u}},\ j = 1, 2, \ldots, s$ and all the canonical directions and let

$$\boldsymbol{\Phi}_{\vartheta}^{\boldsymbol{\tau}, \mathbf{u}^*} = \left(\boldsymbol{\Phi}_{\vartheta}^{\mathbf{u}_1, \boldsymbol{\tau}}, \boldsymbol{\Phi}_{\vartheta}^{\mathbf{u}_2, \boldsymbol{\tau}}, \ldots, \boldsymbol{\Phi}_{\vartheta}^{\mathbf{u}_K, \boldsymbol{\tau}}\right)' \in \mathbb{R}^B \tag{13}$$

$$\tilde{\boldsymbol{\Phi}}_{\mathbf{u}^*, \vartheta}^{R} = \left(\tilde{\boldsymbol{\Phi}}_{\mathbf{u}_1^*, \vartheta}^{R}, \tilde{\boldsymbol{\Phi}}_{\mathbf{u}_2^*, \vartheta}^{R}, \ldots, \tilde{\boldsymbol{\Phi}}_{\mathbf{u}_K^*, \vartheta}^{R}\right)' \in \mathbb{R}^B \tag{14}$$

$$\hat{\boldsymbol{\Phi}}_{\mathbf{u}^*} = \left(\hat{\boldsymbol{\Phi}}_{\mathbf{u}_1^*}, \hat{\boldsymbol{\Phi}}_{\mathbf{u}_2^*}, \ldots, \hat{\boldsymbol{\Phi}}_{\mathbf{u}_K^*}\right)' \in \mathbb{R}^B, \tag{15}$$

where K is the cardinality of $\mathbf{U}^*$, $B = \sum_{i=1}^{K} b_i$ and b_i is the dimension of $\Phi_{\mathbf{u}_i, \vartheta}$ for $i = 1, 2, \ldots, K$, then the MMSQ minimises the square distance defined in equation (6) between $\hat{\boldsymbol{\Phi}}_{\mathbf{u}^*}$ and $\tilde{\boldsymbol{\Phi}}_{\mathbf{u}^*, \vartheta}^{R}$ along the optimal directions $\mathbf{U}^*$.

4 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 5. *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* $\boldsymbol{\tau}_k = \left(\tau_{1,k}, \tau_{2,k}, \ldots, \tau_{s,k}\right)$ *where* $\tau_{j,k} \in (0, 1)$, $\mathbf{q}_{\boldsymbol{\tau}_k, \mathbf{u}_k} = \left(q_{\tau_{1,k} \mathbf{u}_k}, q_{\tau_{2,k} \mathbf{u}_k}, \ldots, q_{\tau_{s,k} \mathbf{u}_k}\right)$ *be the vector of directional quantiles along the direction* $\mathbf{u}_k$ *and suppose* $Var\left(Z_k\right) < \infty$, *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*

$$\sqrt{n}\left(\hat{\mathbf{q}}_{\boldsymbol{\tau}_k,\mathbf{u}_k} - \mathbf{q}_{\boldsymbol{\tau}_k,\mathbf{u}_k}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\eta}\right), \tag{16}$$

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

$$\eta_{r,c} = \frac{\tau_r \wedge \tau_c - \tau_r \tau_c}{f_{Z_k}\left(q_{\tau_r \mathbf{u}_k}\right) f_{Z_k}\left(q_{\tau_c \mathbf{u}_k}\right)}, \tag{17}$$

for $r, c = 1, 2, \ldots, K$;

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

$$\sqrt{n}\left(\hat{\mathbf{q}}_{\tau_j} - \mathbf{q}_{\tau_j}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\eta}\right), \tag{18}$$

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

$$\eta_{r,c} = \begin{cases} -\frac{\tau_j^2}{f_{Z_r}\left(q_{\tau_j \mathbf{u}_r}\right) f_{Z_c}\left(q_{\tau_j \mathbf{u}_c}\right)} + \frac{F_{Z_r,Z_c}\left(\mathbf{q}_{\tau_j,r,c}, \boldsymbol{\Sigma}_{Z_r,Z_c}\right)}{f_{Z_r}\left(q_{\tau_j \mathbf{u}_r}\right) f_{Z_c}\left(q_{\tau_j \mathbf{u}_c}\right)}, & \textit{for} \quad r \neq c \\ \frac{\tau_j(1-\tau_j)}{f_{Z_r}\left(q_{\tau_j \mathbf{u}_r}\right)^2}, & \textit{for} \quad r = c, \end{cases} \tag{19}$$

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

$$\sqrt{n}\left(\hat{q}_{\tau_j \mathbf{u}_s} - q_{\tau_j \mathbf{u}_s}, \hat{q}_{\tau_l \mathbf{u}_t} - q_{\tau_l \mathbf{u}_t}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\eta}\right), \tag{20}$$

as $n \to \infty$, where

$$\eta_{r,c} = -\frac{\tau_j \tau_l}{f_{Z_s}\left(q_{\tau_j}\right) f_{Z_t}\left(q_{\tau_l}\right)} + \frac{F_{Z_s,Z_t}\left(\left(q_{\tau_j \mathbf{u}_s}, q_{\tau_l \mathbf{u}_t}\right), \boldsymbol{\Sigma}_{Z_s,Z_t}\right)}{f_{Z_s}\left(q_{\tau_j}\right) f_{Z_t}\left(q_{\tau_l}\right)}, \quad \textit{for} \quad r \neq c. \tag{21}$$

Proof. See Appendix A. □

Remark 6. *The expression $a \wedge b$ stands for the minimum of a and b. As regards the calculation of the sparsity function $s(\tau) = f\left(F^{-1}(\tau)\right)$ we refer to Koenker (2005) and Dominicy and Veredas (2013).*

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

Assumption 7. *There exists a unique/unknown true value $\vartheta_0 \subset \Theta$ such that the sample function of projectional quantiles equal the theoretical one. That is $\vartheta = \vartheta_0 \Leftrightarrow \hat{\boldsymbol{\Phi}} = \boldsymbol{\Phi}_{\vartheta_0}$.*

Assumption 8. *ϑ_0 is the unique minimiser of $\left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\vartheta}^{R}\right)' \mathbf{W}_{\vartheta} \left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\vartheta}^{R}\right)$.*

Assumption 9. *Let $\hat{\boldsymbol{\Omega}}$ be the sample variance–covariance matrix of $\hat{\boldsymbol{\Phi}}$ and $\boldsymbol{\Omega}_{\vartheta}$ be the variance–covariance matrix of $\boldsymbol{\Phi}_{\vartheta}$, then $\hat{\boldsymbol{\Omega}}$ converges to $\boldsymbol{\Omega}_{\vartheta}$.*

Assumption 10. *The matrix* $\mathbf{\Omega}_\vartheta$ *is non–singular.*

Assumption 11. *The matrix* $\left(\frac{\partial \mathbf{\Phi}_\vartheta}{\partial \vartheta'}\mathbf{W}_\vartheta\frac{\partial \mathbf{\Phi}_\vartheta}{\partial \vartheta}\right)$ *is non–singular.*

The asymptotic properties of functions of quantiles and projection quantiles are established by the next theorem.

Theorem 12. *Under the hypothesis of Theorem 5 and assumptions 6–9, we have*

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

as $n \to \infty$, *where* $\mathbf{\Omega}_\vartheta = \frac{\partial \mathbf{\Phi}_\vartheta}{\partial \mathbf{q}'}\boldsymbol{\eta}\frac{\partial \mathbf{\Phi}_\vartheta}{\partial \mathbf{q}}$, $\mathbf{q} = (\mathbf{q}_{\boldsymbol{\tau}_1,\mathbf{u}_1}, \mathbf{q}_{\boldsymbol{\tau}_2,\mathbf{u}_2}, \dots, \mathbf{q}_{\boldsymbol{\tau}_K,\mathbf{u}_K})'$, $\boldsymbol{\eta}$ *is the variance–covariance matrix of the projectional quantiles* $\mathbf{q}$ *defined in Theorem 5 and* $\frac{\partial \mathbf{\Phi}_\vartheta}{\partial \mathbf{q}} = diag\left\{\frac{\partial \mathbf{\Phi}_\vartheta}{\partial \mathbf{q}_{\boldsymbol{\tau}_1,\mathbf{u}_1}}, \frac{\partial \mathbf{\Phi}_\vartheta}{\partial \mathbf{q}_{\boldsymbol{\tau}_2,\mathbf{u}_2}}, \dots, \frac{\partial \mathbf{\Phi}_\vartheta}{\partial \mathbf{q}_{\boldsymbol{\tau}_K,\mathbf{u}_K}}\right\}$.

Proof. See Appendix A. ☐

Next theorem shows the asymptotic properties of the MMSQ estimator.

Theorem 13. *Under the hypothesis of Theorem 5 and assumptions 6–10, we have*

$$\sqrt{n}\left(\hat{\vartheta} - \vartheta\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \left(1 + \frac{1}{R}\right)\mathbf{D}_\vartheta\mathbf{W}_\vartheta\mathbf{\Omega}_\vartheta\mathbf{W}'_\vartheta\mathbf{D}'_\vartheta\right), \tag{22}$$

as $n \to \infty$, *where* $\mathbf{D}_\vartheta = \left(\frac{\partial \mathbf{\Phi}_\vartheta}{\partial \vartheta'}\mathbf{W}_\vartheta\frac{\partial \mathbf{\Phi}_\vartheta}{\partial \vartheta}\right)^{-1}\frac{\partial \mathbf{\Phi}_\vartheta}{\partial \vartheta}$, $\mathbf{\Omega}_\vartheta$ *is defined as in Theorem 12 and* $\mathbf{W}_\vartheta$ *is an appropriately chosen positive definite weighting matrix.*

Proof. See Appendix A. ☐

The next corollary shows how to choose the optimal weighting matrix $\mathbf{W}_\vartheta$.

Corollary 14. *Under the hypothesis of Theorem 5 and assumptions 6–10, the optimal weighting matrix is*

$$\mathbf{W}^*_\vartheta = \mathbf{\Omega}^{-1}_\vartheta, \tag{23}$$

where $\mathbf{\Omega}_\vartheta$ *is defined as in Theorem 12. The efficient method of simulated quantiles estimator E–MMSQ has the following asymptotic distribution*

$$\sqrt{n}\left(\hat{\vartheta} - \vartheta\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \left(1 + \frac{1}{R}\right)\left(\frac{\partial \mathbf{\Phi}_\vartheta}{\partial \vartheta'}\mathbf{\Omega}^{-1}_\vartheta\frac{\partial \mathbf{\Phi}_\vartheta}{\partial \vartheta}\right)^{-1}\right), \tag{24}$$

as $n \to \infty$.

5 Handling sparsity

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 variance–covariance 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

$$p_\lambda(|\gamma|) = \begin{cases} \lambda|\gamma| & \text{if } |\gamma| \leq \lambda \\ \frac{1}{a-1}\left(a\lambda|\gamma| - \frac{\gamma^2}{2}\right) - \frac{\lambda^2}{2(a-1)} & \text{if } \lambda < \gamma \leq a\lambda \\ \frac{\lambda^2(a+1)}{2} & \text{if } a\lambda < |\gamma|, \end{cases} \tag{25}$$

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 penalised MMSQ estimator minimises the penalised MMSQ objective function, defined as follows

$$\hat{\vartheta} = \arg\min_{\vartheta} \mathcal{Q}^\star(\vartheta), \tag{26}$$

where $\mathcal{Q}^\star(\vartheta) = \left(\hat{\boldsymbol{\Phi}}_{\mathbf{u}} - \tilde{\boldsymbol{\Phi}}^R_{\mathbf{u},\vartheta}\right)' \mathbf{W}_\vartheta \left(\hat{\boldsymbol{\Phi}}_{\mathbf{u}} - \tilde{\boldsymbol{\Phi}}^R_{\mathbf{u},\vartheta}\right) + n\sum_{i<j} p_\lambda(|\sigma_{ij}|)$ is the penalised distance between $\hat{\boldsymbol{\Phi}}_{\mathbf{u}}$ and $\tilde{\boldsymbol{\Phi}}^R_{\mathbf{u},\vartheta}$ and $\hat{\boldsymbol{\Phi}}_{\mathbf{u}}$, $\tilde{\boldsymbol{\Phi}}^R_{\mathbf{u},\vartheta}$ are those defined in Section 3. 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 $\vartheta_0 = \left(\vartheta_0^1, \vartheta_0^0\right)$ be the true value of the unknown parameter ϑ, where $\vartheta_0^1 \in \mathbb{R}^s$ is the subset of non–zero parameters and $\vartheta_0^0 = 0 \in \mathbb{R}^{k-s}$ and let $\mathcal{A} = \left\{(i,j) : i < j, \sigma_{ij,0} \in \vartheta_0^1\right\}$. The following definition of oracle estimator is given in Zou (2006).

Definition 15. *An oracle estimator* $\hat{\vartheta}_{\text{oracle}}$ *has the following properties:*

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

(ii) asymptotic normality: $\sqrt{n}\left(\hat{\vartheta}^1_{\text{oracle}} - \vartheta_0^1\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, *as* $n \to \infty$, *where* $\boldsymbol{\Sigma}$ *is the variance covariance matrix of* ϑ_0^1.

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

Theorem 16. *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* $\hat{\vartheta}$ *of* $\mathcal{Q}^\star(\vartheta)$ *in*

(26) *with* $\|\hat{\vartheta} - \vartheta_0\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. *Furthermore, we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\hat{\vartheta}^0 = 0\right) = 1. \tag{27}$$

Proof. See Appendix A. ☐

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

Theorem 17. *Given the SCAD penalty function* $p_\lambda\left(|\sigma_{ij}|\right)$, *for a sequence* $\lambda_n \to 0$ *and* $\sqrt{n}\lambda_n \to \infty$ *as* $n \to \infty$, *then* $\hat{\vartheta}^1$ *has the following asymptotic distribution:*

$$\sqrt{n}\left(\hat{\vartheta}^1 - \vartheta_0^1\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \left(1 + \frac{1}{R}\right)\left(\frac{\partial \boldsymbol{\Phi}_\vartheta}{\partial \vartheta^{1\prime}} \boldsymbol{\Omega}_{\vartheta_0^1}^{-1} \frac{\partial \boldsymbol{\Phi}_\vartheta}{\partial \vartheta^1}\right)^{-1}\right), \tag{28}$$

as $n \to \infty$.

Proof. See Appendix A. ☐

6 Simulated experiments

6.1 Elliptical transformation of Stable distributions

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

$$\mathbf{Y} =^d \boldsymbol{\xi} + \mathcal{R}\boldsymbol{\Gamma}\mathbf{U}, \tag{29}$$

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 $\mathbb{S}^{m-1} = \{\mathbf{u} \in \mathbb{R}^m : \mathbf{u}'\mathbf{u} = 1\}$ 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 $Z_2 \sim \mathcal{S}_{\frac{\alpha}{2}}\left(\xi, \omega, \delta\right)$ 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

$$\mathbf{Y} \sim \mathcal{ESD}_m\left(\alpha, \boldsymbol{\xi}, \boldsymbol{\Omega}\right), \tag{30}$$

with characteristic function

$$\begin{aligned} \psi_{\mathbf{Y}}\left(\mathbf{t}\right) &= \mathbb{E}\left(\exp\left\{i\mathbf{t}'\mathbf{Y}\right\}\right) \\ &= \exp\left\{i\mathbf{t}'\boldsymbol{\xi} - \left(\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}\right)^{\frac{\alpha}{2}}\right\}. \end{aligned} \tag{31}$$

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) and McNeil et al. (2015) for further details. Simulating from an ESD is straightforward, indeed let $\bar{\varpi}_\alpha = \left(\cos\frac{\pi\alpha}{4}\right)^{\frac{2}{\alpha}}$, 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

$$\mathbf{Y} = \boldsymbol{\xi} + \zeta^{\frac{1}{2}}\mathbf{X}, \tag{32}$$

where $\zeta \sim \mathcal{S}_{\frac{\alpha}{2}}\left(0, \bar{\varpi}_\alpha, 1\right)$ 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

$$\begin{aligned}
\psi_{\mathbf{Y}}\left(\mathbf{t}\right) &= \mathbb{E}\left(\exp\left\{i\mathbf{t}'\mathbf{Y}\right\}\right)\\
&= \mathbb{E}_\zeta\mathbb{E}\left(\exp\left\{i\mathbf{t}'\boldsymbol{\xi} + i\zeta^{\frac{1}{2}}\mathbf{t}'\mathbf{X}\right\} \mid \zeta\right)\\
&= \mathbb{E}_\zeta\mathbb{E}\left(\exp\left\{i\mathbf{t}'\boldsymbol{\xi} - \frac{\zeta\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}}{2}\right\} \mid \zeta\right)\\
&= \exp\left\{i\mathbf{t}'\boldsymbol{\xi} - \left(\frac{1}{2}\right)^{\frac{\alpha}{2}}\left(\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}\right)^{\frac{\alpha}{2}}\right\}, \qquad \alpha \neq 1,
\end{aligned} \tag{33}$$

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 $\zeta \sim \mathcal{S}_{\frac{\alpha}{2}}\left(0, \bar{\varpi}_\alpha, 1\right)$ with $0 < \alpha \leq 2$ is

$$\begin{aligned}
\psi_\zeta^*\left(A\right) &= \mathbb{E}\left(\exp\left\{-A\zeta\right\}\right)\\
&= \begin{cases} \exp\left\{-\frac{\left(\bar{\varpi}_\alpha\right)^{\frac{\alpha}{2}}}{\cos\frac{\pi\alpha}{4}}A^{\frac{\alpha}{2}}\right\}, & \alpha \neq 1\\ \exp\left\{\frac{2\bar{\varpi}_\alpha}{\pi}A\log\left(A\right)\right\}, & \alpha = 1. \end{cases}
\end{aligned} \tag{34}$$

The Elliptical Stable distribution is a particular case of multivariate Stable distribution so it admits finite moments if $\mathbb{E}\left[\zeta^p\right] < \infty$ 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 $\mathbb{E}\left(\mathbf{Y}\right) = \boldsymbol{\xi}$, 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.

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

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

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, we can 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

$$Z_{ij,\mathbf{u}} = \mathbf{u}'\boldsymbol{\xi}_{ij} + \sqrt{\mathbf{u}'\mathbf{\Omega}_{ij}\mathbf{u}}Z, \tag{35}$$

where $Z \sim \mathcal{ESD}_1(\alpha, 0, 1)$. Following Definition 4, in order to find the optimal directions we need to compute

$$\mathbf{u}_{\max} = \arg\max_{\mathbf{u}\in\mathbb{S}^1} q_{\tau\mathbf{u}}\left(\mathbf{Z}_{ij}\right), \tag{36}$$

where $q_{\tau\mathbf{u}}(\mathbf{Z}_{ij})$ 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 (35), it holds

$$\mathbf{u}_{\max} = \arg\max_{\mathbf{u}\in\mathbb{S}^1} \mathbf{u}'\boldsymbol{\xi}_{ij} + \sqrt{\mathbf{u}'\mathbf{\Omega}_{ij}\mathbf{u}}, \tag{37}$$

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

$$\mathcal{L}(\mathbf{u}, \lambda) = \mathbf{u}'\boldsymbol{\xi}_{ij} + \sqrt{\mathbf{u}'\mathbf{\Omega}_{ij}\mathbf{u}} - \lambda\left(\|\mathbf{u}\| - 1\right). \tag{38}$$

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

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial u_1} &= \frac{\left(\omega_{ii}^2 u_1 + \omega_{ij}u_2\right)}{\sqrt{\omega_{ii}^2 u_1^2 + \omega_{jj}^2 u_2^2 + 2\omega_{ij}u_1u_2}} - 2\lambda u_1 = 0\\ \frac{\partial\mathcal{L}}{\partial u_2} &= \frac{\left(\omega_{jj}^2 u_2 + \omega_{ij}u_1\right)}{\sqrt{\omega_{ii}^2 u_1^2 + \omega_{jj}^2 u_2^2 + 2\omega_{ij}u_1u_2}} - 2\lambda u_2 = 0\\ \frac{\partial\mathcal{L}}{\partial\lambda} &= u_1^2 + u_2^2 - 1 = 0,\end{aligned} \tag{39}$$

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 (39), we solve for u_1

$$\begin{aligned}
&u_1^2 + \frac{u_1^2}{4}\left(-\frac{\omega_{ii}^2 - \omega_{jj}^2}{\omega_{ij}} \pm \sqrt{\left(\frac{\omega_{ii}^2 - \omega_{jj}^2}{\omega_{ij}}\right)^2 + 4}\right)^2 = 1 \\
&u_1^2\left[1 + \frac{1}{4}\left(-\frac{\omega_{ii}^2 - \omega_{jj}^2}{\omega_{ij}} \pm \sqrt{\left(\frac{\omega_{ii}^2 - \omega_{jj}^2}{\omega_{ij}}\right)^2 + 4}\right)^2\right] = 1 \\
&u_1 = \pm\frac{1}{\sqrt{\left[1 + \frac{1}{4}\left(-\frac{\omega_{ii}^2 - \omega_{jj}^2}{\omega_{ij}} \pm \sqrt{\left(\frac{\omega_{ii}^2 - \omega_{jj}^2}{\omega_{ij}}\right)^2 + 4}\right)^2\right]}},
\end{aligned} \tag{40}$$

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 4.

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

$$\boldsymbol{\Sigma}_2^s = \begin{pmatrix} 0.5 & 0.9 \\ 0.9 & 2 \end{pmatrix}, \tag{41}$$

for $m = 2$, and

$$\boldsymbol{\Sigma}_5^s = \begin{pmatrix} 0.25 & 0.25 & 0.4 & 0 & 0 \\ 0.25 & 0.5 & 0.4 & 0 & 0 \\ 0.4 & 0.4 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2.55 \\ 0 & 0 & 0 & 2.55 & 4 \end{pmatrix}, \tag{42}$$

for $m = 5$. We also consider two different sample sizes $n = 500, 2000$ and we fix $R = 200$. We also consider a simulation example of dimension $m = 12$, with $n = 500$ and $R = 50$ where the location parameters are equal to zero, as in previous examples, while the scale matrix is that considered in Wang (2015) and reported below

$$\boldsymbol{\Sigma}_{12}^{s} = \\ \begin{pmatrix} 0.239 & 0.117 & 0 & 0 & 0 & 0 & 0 & 0.031 & 0 & 0 & 0 & 0 \\ 0.117 & 1.554 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.362 & 0.002 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.002 & 0.199 & 0.094 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.094 & 0.349 & 0 & 0 & 0 & 0 & 0 & 0 & -0.036 \\ 0 & 0 & 0 & 0 & 0 & 0.295 & -0.229 & 0.002 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.229 & 0.715 & 0 & 0 & 0 & 0 & 0 \\ 0.031 & 0 & 0 & 0 & 0 & 0.002 & 0 & 0.164 & 0.112 & -0.028 & -0.008 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.112 & 0.518 & -0.193 & -0.09 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.028 & -0.193 & 0.379 & 0.167 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.008 & -0.09 & 0.167 & 0.159 & 0 \\ 0 & 0 & 0 & 0 & -0.036 & 0 & 0 & 0 & 0 & 0 & 0 & 0.207 \end{pmatrix} \tag{43}$$

In Table 1, we report estimation results obtained over $1,000$ replications for $m = 2$, for all the values of α with $n = 500, 2000$ while in Tables 2, 3, 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 $i = 1, 2, \ldots, m$ for which they display lower values than expected, which means that in those cases the asymptotic standard errors are underestimated.

7 Conclusions

In this paper we present the multivariate extension of the method of simulated quantiles proposed in Dominicy and Veredas (2013). 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. Projectional quantiles along optimal directions are then introduced in order to carry the information over the parameters of interest in an efficient way. We establish the consistency and the asymptotic distribution of the proposed MMSQ estimator. We also introduce a penalised version of the MMSQ using the SCAD ℓ_1–penalty of Fan and Li (2001) into the MMSQ objective function in order to achieve sparse estimation of the scaling matrix. We extend the asymptotic theory and we show that the sparse–MMSQ estimator enjoys the oracle properties under mild regularity conditions. The method is illustrated using synthetic datasets from Elliptical Stable distribution. Further simulation studies on the penalized version of the MMSQ and real data applications will be considered in further developments.

Acknowledgements

The authors are grateful to prof. G. Guagnano for her useful suggestions and comments.

A Appendix A: Proofs

Proof. Theorem 5.

1. The proof of this result can be found in Cramér (1946).
2. Without loss of generality we can consider τ_1, τ_2 and Z_1, Z_2. 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:

$$\begin{aligned} \hat{q}_{\tau_1,Z_1} - q_{\tau_1,Z_1} &= \frac{1}{n}\sum_{i=1}^{n} \frac{\tau_1 - \mathbb{1}_{\left[z_{i,1} \leq q_{\tau_1}\right]}}{f_{Z_1}\left(q_{\tau_1}\right)} + R_{n,1} \\ \hat{q}_{\tau_2,Z_2} - q_{\tau_2,Z_2} &= \frac{1}{n}\sum_{i=1}^{n} \frac{\tau_2 - \mathbb{1}_{\left[z_{i,2} \leq q_{\tau_2}\right]}}{f_{Z_2}\left(q_{\tau_2}\right)} + R_{n,2} \end{aligned} \tag{44}$$

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

$$\begin{aligned} &Var\left(\hat{q}_{\tau_1,Z_1} - q_{\tau_1,Z_1}\right) = Var\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\tau_1 - \mathbb{1}_{\left[z_{i,1} \leq q_{\tau_1}\right]}}{f_{Z_1}\left(q_{\tau_1}\right)} + R_{n,1}\right) = \\ &= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\tau_1 - \mathbb{1}_{\left[z_{i,1} \leq q_{\tau_1}\right]}}{f_{Z_1}\left(q_{\tau_1}\right)} + R_{n,1}\right)^2\right] = \\ &= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\tau_1 - \mathbb{1}_{\left[z_{i,1} \leq q_{\tau_1}\right]}}{f_{Z_1}\left(q_{\tau_1}\right)}\right)^2 + 2R_{n,1}\frac{1}{n}\sum_{i=1}^{n} \frac{\tau_1 - \mathbb{1}_{\left[z_{i,1} \leq q_{\tau_1}\right]}}{f_{Z_1}\left(q_{\tau_1}\right)} + R_{n,1}^2\right] = \\ &= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\tau_1 - \mathbb{1}_{\left[z_{i,1} \leq q_{\tau_1}\right]}}{f_{Z_1}\left(q_{\tau_1}\right)}\right)^2\right] + 2R_{n,1}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \frac{\tau_1 - \mathbb{1}_{\left[z_{i,1} \leq q_{\tau_1}\right]}}{f_{Z_1}\left(q_{\tau_1}\right)}\right] + R_{n,1}^2 = \\ &= \frac{1}{n^2 f_{Z_1}\left(q_{\tau_1}\right)^2}\mathbb{E}\left[\left(\sum_{i=1}^{n} \tau_1 - \mathbb{1}_{\left[z_{i,1} \leq q_{\tau_1}\right]}\right)^2\right] + \\ &\quad \frac{2R_{n,1}}{n f_{Z_1}\left(q_{\tau_1}\right)}\mathbb{E}\left[\sum_{i=1}^{n} \tau_1 - \mathbb{1}_{\left[z_{i,1} \leq q_{\tau_1}\right]}\right] + R_{n,1}^2 = \\ &= \frac{1}{n^2 f_{Z_1}\left(q_{\tau_1}\right)^2} Var\left(\sum_{i=1}^{n} \tau_1 - \mathbb{1}_{\left[z_{i,1} \leq q_{\tau_1}\right]}\right) + R_{n,1}^2 = \\ &= \frac{\tau_1\left(1-\tau_1\right)}{f_{Z_1}\left(q_{\tau_1}\right)^2} + R_{n,1}^2 \end{aligned} \tag{45}$$

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}
& 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)= \\
& cov\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_1-\mathbb{1}_{\left[z_{i,1}\leq q_{\tau_1}\right]}}{f\left(q_{\tau_1}\right)}+R_{n,1},\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_2-\mathbb{1}_{\left[z_{i,2}\leq q_{\tau_2}\right]}}{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}_{\left[z_{i,1}\leq q_{\tau_1}\right]}}{f\left(q_{\tau_1}\right)}+R_{n,1}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_2-\mathbb{1}_{\left[z_{i,2}\leq q_{\tau_2}\right]}}{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}_{\left[z_{i,1}\leq q_{\tau_1}\right]}}{f\left(q_{\tau_1}\right)}\sum_{i=1}^{n}\frac{\tau_2-\mathbb{1}_{\left[z_{i,2}\leq q_{\tau_2}\right]}}{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}_{\left[z_{i,2}\leq q_{\tau_2}\right]}}{f\left(q_{\tau_2}\right)}\right]+ \\
& \quad \mathbb{E}\left[R_{n,2}\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_1-\mathbb{1}_{\left[z_{i,1}\leq q_{\tau_1}\right]}}{f\left(q_{\tau_1}\right)}\right]+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}_{\left[z_{i,1}\leq q_{\tau_1}\right]}}{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}_{\left[z_{i,2}\leq q_{\tau_2}\right]}}{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}_{\left[z_{i,2}\leq q_{\tau_2}\right]}\right]+ \\
& \quad \frac{R_{n,2}}{nf\left(q_{\tau_1}\right)}\mathbb{E}\left[\sum_{i=1}^{n}\tau_1-\mathbb{1}_{\left[z_{i,1}\leq q_{\tau_1}\right]}\right]+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}_{\left[z_{i,2}\leq q_{\tau_2}\right]}}{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}_{\left[z_{i,1}\leq q_{\tau_1}\right]}}{f\left(q_{\tau_1}\right)}\right]+ \\
& \quad \frac{1}{n^2}\mathbb{E}\left[\sum_{i=1}^{n}\frac{\mathbb{1}_{\left[z_{i,1}\leq q_{\tau_1}\right]}}{f\left(q_{\tau_1}\right)}\sum_{i=1}^{n}\frac{\mathbb{1}_{\left[z_{i,2}\leq q_{\tau_2}\right]}}{f\left(q_{\tau_2}\right)}\right]+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}_{\left[z_{i,2}\leq q_{\tau_2}\right]}\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}_{\left[z_{i,1}\leq q_{\tau_1}\right]}\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}_{\left[z_{i,1}\leq q_{\tau_1}\right]}\sum_{i=1}^{n}\mathbb{1}_{\left[z_{i,2}\leq q_{\tau_2}\right]}\right]+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}{f\left(q_{\tau_1}\right)f\left(q_{\tau_2}\right)}\mathbb{E}\left[\mathbb{1}_{\left[z_2\leq q_{\tau_2}\right]}\right]- \\
& \quad \frac{\tau_2}{f\left(q_{\tau_1}\right)f\left(q_{\tau_2}\right)}\mathbb{E}\left[\mathbb{1}_{\left[z_1\leq q_{\tau_1}\right]}\right]+\frac{1}{f\left(q_{\tau_1}\right)f\left(q_{\tau_2}\right)}\mathbb{E}\left[\mathbb{1}_{\left[z_1\leq q_{\tau_1}\right]}\mathbb{1}_{\left[z_2\leq q_{\tau_2}\right]}\right]+ \\
& \quad R_{n,1}R_{n,2}=
\end{aligned}
$$

$$= \frac{\tau_1\tau_2}{f(q_{\tau_1})f(q_{\tau_2})} - 2\frac{\tau_1\tau_2}{f(q_{\tau_1})f(q_{\tau_2})} + \frac{F_{Z_1,Z_2}(\mathbf{q}_\tau, \mathbf{\Sigma}_{Z_1,Z_2})}{f(q_{\tau_1})f(q_{\tau_2})} + R_{n,1}R_{n,2} =$$

$$= -\frac{\tau_1\tau_2}{f(q_{\tau_1})f(q_{\tau_2})} + \frac{F_{Z_1,Z_2}(\mathbf{q}_\tau, \mathbf{\Sigma}_{Z_1,Z_2})}{f(q_{\tau_1})f(q_{\tau_2})} + R_{n,1}R_{n,2} \tag{46}$$

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

3. Using the Bahadur representation

$$\begin{aligned}
&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) = \\
&= cov\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_i - \mathbb{1}_{[z_i \le 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}_{[z_i \le q_{\tau_j,\mathbf{u}}]}}{f\left(q_{\tau_j,\mathbf{u}}\right)} + R_{n,2}\right) = \\
&= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_k - \mathbb{1}_{[z_i \le 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}_{[z_i \le q_{\tau_j,\mathbf{u}}]}}{f\left(q_{\tau_j,\mathbf{u}}\right)} + R_{n,2}\right)\right] = \\
&= \frac{1}{n^2}\mathbb{E}\left[\sum_{i=1}^{n}\frac{\tau_k - \mathbb{1}_{[z_i \le q_{\tau_k,\mathbf{u}}]}}{f(q_{\tau_k,\mathbf{u}})}\sum_{i=1}^{n}\frac{\tau_j - \mathbb{1}_{[z_i \le q_{\tau_j,\mathbf{u}}]}}{f\left(q_{\tau_j,\mathbf{u}}\right)}\right] + \\
&\quad R_{n,1}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\tau_j - \mathbb{1}_{[z_i \le q_{\tau_j,\mathbf{u}}]}}{f\left(q_{\tau_j,\mathbf{u}}\right)}\right] + \\
&\quad R_{n,2}\mathbb{E}\left[\sum_{i=1}^{n}\frac{\tau_k - \mathbb{1}_{[z_i \le q_{\tau_k,\mathbf{u}}]}}{f(q_{\tau_k,\mathbf{u}})}\right] + R_{n,1}R_{n,2} = \\
&= \frac{1}{f(q_{\tau_k,\mathbf{u}})f\left(q_{\tau_j,\mathbf{u}}\right)}\mathbb{E}\left[\left(\tau_k - \mathbb{1}_{[z_i \le q_{\tau_k,\mathbf{u}}]}\right)\left(\tau_j - \mathbb{1}_{[z_i \le q_{\tau_j,\mathbf{u}}]}\right)\right] + R_{n,1}R_{n,2} = \\
&= \frac{1}{f(q_{\tau_k,\mathbf{u}})f\left(q_{\tau_j,\mathbf{u}}\right)}\left[\tau_k\tau_j - \tau_k\mathbb{E}\left[\mathbb{1}_{[z_i \le q_{\tau_j,\mathbf{u}}]}\right] - \tau_j\mathbb{E}\left[\mathbb{1}_{[z_i \le q_{\tau_k,\mathbf{u}}]}\right]\right] + \\
&\quad \frac{1}{f(q_{\tau_k,\mathbf{u}})f\left(q_{\tau_j,\mathbf{u}}\right)}\left[\mathbb{E}\left[\mathbb{1}_{[z_i \le q_{\tau_k,\mathbf{u}}]}\mathbb{1}_{[z_i \le q_{\tau_j,\mathbf{u}}]}\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\left(q_{\tau_j,\mathbf{u}}\right)} + R_{n,1}R_{n,2}
\end{aligned} \tag{47}$$

☐

Proof. Theorem 12. The function $\hat{\mathbf{\Phi}}$ is assumed to be continuously differentiable, so Delta method applies

$$\hat{\mathbf{\Phi}} \approx \mathbf{\Phi}_\theta + \frac{\partial \mathbf{\Phi}_\theta}{\partial \mathbf{q}}(\hat{\mathbf{q}} - \mathbf{q}) \tag{48}$$

Then

$$Var\left(\hat{\mathbf{\Phi}}\right) \approx Var\left(\frac{\partial \mathbf{\Phi}_\theta}{\partial \mathbf{q}}\hat{\mathbf{q}}\right) = \frac{\partial \mathbf{\Phi}_\theta}{\partial \mathbf{q}}' Cov(\hat{\mathbf{q}})\frac{\partial \mathbf{\Phi}_\theta}{\partial \mathbf{q}}. \tag{49}$$

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

Proof. Theorem 13. The first order condition of (6) is

$$\frac{1}{R}\sum_{r=1}^{R}\frac{\partial\tilde{\mathbf{\Phi}}_{\theta}^{r}}{\partial\theta}\mathbf{W}_{\bar{\theta}}\left(\hat{\mathbf{\Phi}}_{\theta}-\frac{1}{R}\sum_{r=1}^{R}\tilde{\mathbf{\Phi}}_{\theta}^{r}\right)=0 \tag{50}$$

where $\bar{\theta}$ is a consistent estimate of θ. Let us consider the first order Taylor expansion around the true parameter θ_0

$$\begin{aligned}\frac{1}{R}\sum_{r=1}^{R}\frac{\partial\tilde{\mathbf{\Phi}}_{\theta_0}^{r}}{\partial\theta}\mathbf{W}_{\bar{\theta}}&\left(\hat{\mathbf{\Phi}}_{\theta}-\frac{1}{R}\sum_{r=1}^{R}\tilde{\mathbf{\Phi}}_{\theta_0}^{r}\right)\\&-\frac{1}{R}\sum_{r=1}^{R}\frac{\partial\tilde{\mathbf{\Phi}}_{\theta_0}^{r}}{\partial\theta}\mathbf{W}_{\bar{\theta}}\frac{1}{R}\sum_{r=1}^{R}\frac{\partial\tilde{\mathbf{\Phi}}_{\theta_0}^{r}}{\partial\theta}\left(\hat{\theta}-\theta_0\right)=o_p\left(1\right)\end{aligned} \tag{51}$$

From this equation we get

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right)\approx\left(\frac{\partial\tilde{\mathbf{\Phi}}_{\theta}'}{\partial\theta}\mathbf{W}_{\bar{\theta}}\frac{\partial\tilde{\mathbf{\Phi}}_{\theta}}{\partial\theta}\right)^{-1}\frac{\partial\tilde{\mathbf{\Phi}}_{\theta}}{\partial\theta}\mathbf{W}_{\bar{\theta}}\sqrt{n}\left(\hat{\mathbf{\Phi}}-\frac{1}{R}\sum_{r=1}^{R}\tilde{\mathbf{\Phi}}_{\theta_0}^{r}\right) \tag{52}$$

From Theorem 12

$$\sqrt{n}\left(\hat{\mathbf{\Phi}}-\frac{1}{R}\sum_{r=1}^{R}\tilde{\mathbf{\Phi}}_{\theta_0}^{r}\right)\to^{d}\mathcal{N}\left(\mathbf{0},\left(1+\frac{1}{R}\right)\mathbf{\Omega}_{\theta}\right) \tag{53}$$

and $\tilde{\mathbf{\Phi}}_{\theta_0}^{r}$ converges to $\mathbf{\Phi}_{\theta}$. Moreover since $\bar{\theta}$ is consistent the matrix $\mathbf{W}_{\bar{\theta}}$ converges to $\mathbf{W}_{\theta}$. From these results we get

$$\begin{aligned}&Var\left(\sqrt{n}\left(\hat{\theta}-\theta\right)\right)\to\\&\left(1+\frac{1}{R}\right)\left[\left(\frac{\partial\mathbf{\Phi}_{\theta}'}{\partial\theta}\mathbf{W}_{\theta}\frac{\partial\mathbf{\Phi}_{\theta}}{\partial\theta}\right)^{-1}\frac{\partial\mathbf{\Phi}_{\theta}}{\partial\theta}\right]\mathbf{W}_{\theta}\mathbf{\Omega}_{\theta}\mathbf{W}'\left[\left(\frac{\partial\mathbf{\Phi}_{\theta}'}{\partial\theta}\mathbf{W}_{\theta}\frac{\partial\mathbf{\Phi}_{\theta}}{\partial\theta}\right)^{-1}\frac{\partial\mathbf{\Phi}_{\theta}}{\partial\theta}\right]'\end{aligned} \tag{54}$$

□

Proof. Theorem 16. 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 $\|\vartheta-\vartheta_0\|\leq Mn^{-\frac{1}{2}}$ for some finite constant M. In order to prove the result in equation (59), let us consider the first order condition of equation (26) and its first order taylor expansion

$$\begin{aligned}\frac{\partial\mathcal{Q}\left(\vartheta\right)}{\partial\vartheta}&=-2\frac{\partial\tilde{\mathbf{\Phi}}_{\vartheta}^{R}}{\partial\vartheta}\mathbf{W}_{\vartheta}\left(\hat{\mathbf{\Phi}}-\tilde{\mathbf{\Phi}}_{\vartheta}^{R}\right)+n\mathbf{v}\\&\approx-2\frac{\partial\tilde{\mathbf{\Phi}}_{\vartheta_0}^{R}}{\partial\vartheta}\mathbf{W}_{\vartheta}\left(\hat{\mathbf{\Phi}}-\tilde{\mathbf{\Phi}}_{\vartheta_0}^{R}\right)+2\frac{\partial\tilde{\mathbf{\Phi}}_{\vartheta_0}^{R}}{\partial\vartheta'}\mathbf{W}_{\vartheta}\frac{\partial\tilde{\mathbf{\Phi}}_{\vartheta_0}^{R}}{\partial\vartheta}\left(\vartheta-\vartheta_0\right)+n\mathbf{v},\end{aligned} \tag{55}$$

where $\mathbf{v}=\left(\mathbf{0};p_{\lambda_n}'\left(|\sigma_{ij}|\right)\operatorname{sgn}\left(\sigma_{ij}\right),i<j\right)$. 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 $p'_{\lambda_n}(|\sigma_{ij}|)$ with respect to $|\sigma_{ij}|$ is given by

$$p'_{\lambda_n}(|\sigma_{ij}|) = \begin{cases} \lambda_n & \text{if } |\sigma_{ij}| \leq \lambda_n \\ \frac{(a\lambda_n - |\sigma_{ij}|)}{a-1} & \text{if } \lambda_n < |\sigma_{ij}| \leq a\lambda_n \\ 0 & \text{if } a\lambda_n < |\sigma_{ij}|, \end{cases} \tag{56}$$

and it holds

$$\lim_{|\sigma_{ij}| \to 0} \frac{p'_{\lambda_n}(|\sigma_{ij}|)}{\lambda_n} = 1. \tag{57}$$

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

$$n\lambda_n \operatorname{sgn}(\sigma_{ij}) \frac{p'_{\lambda_n}(|\sigma_{ij}|)}{\lambda_n} = n\lambda_n \operatorname{sgn}(\sigma_{ij}). \tag{58}$$

We rewrite (55) as follows

$$\frac{\partial \mathcal{Q}(\vartheta)}{\partial \vartheta} = n\lambda_n \left\{ \lambda_n^{-1}\mathbf{v} - \mathcal{O}_p\left(n^{-\frac{n}{2}}\lambda_n^{-1}\right), \right\} \tag{59}$$

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 (59). This means that the sign of $\frac{\partial \mathcal{Q}(\vartheta)}{\partial \sigma_{ij}}$ 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

$$\begin{aligned} \mathcal{Q}(\vartheta_0) - \mathcal{Q}(\vartheta) = {} & \left(\hat{\mathbf{\Phi}} - \tilde{\mathbf{\Phi}}^R_{\vartheta_0}\right)' \mathbf{W}_{\vartheta_0} \left(\hat{\mathbf{\Phi}} - \tilde{\mathbf{\Phi}}^R_{\vartheta_0}\right) - \left(\hat{\mathbf{\Phi}} - \tilde{\mathbf{\Phi}}^R_{\vartheta}\right)' \mathbf{W}_{\vartheta} \left(\hat{\mathbf{\Phi}} - \tilde{\mathbf{\Phi}}^R_{\vartheta}\right) \\ & + n \sum_{i<j} \left[p_\lambda\left(|\sigma^0_{ij}|\right) - p_\lambda\left(|\sigma_{ij}|\right) \right] \\ \approx {} & 2 \frac{\partial \tilde{\mathbf{\Phi}}^R_{\vartheta_0}}{\partial \vartheta} \mathbf{W}_{\vartheta_0} \left(\hat{\mathbf{\Phi}} - \tilde{\mathbf{\Phi}}^R_{\vartheta_0}\right) (\vartheta - \vartheta_0) \\ & + (\vartheta - \vartheta_0)' \left[-2 \frac{\partial \tilde{\mathbf{\Phi}}^R_{\vartheta_0}}{\partial \vartheta'} \mathbf{W}_{\vartheta_0} \frac{\partial \tilde{\mathbf{\Phi}}^R_{\vartheta_0}}{\partial \vartheta} \right] (\vartheta - \vartheta_0) \\ & - n \sum_{i<j} \left(p'_{\lambda_n}(|\sigma_{ij}|) \operatorname{sgn}(\sigma_{ij}) \left(\sigma_{ij} - \sigma^0_{ij}\right) + p''_{\lambda_n}(|\sigma_{ij}|) \left(\sigma_{ij} - \sigma^0_{ij}\right)^2 \right), \end{aligned}$$

where $p''_{\lambda_n}(|\sigma_{ij}|)$ stands for the second derivative. For n large enough the summation term in equation (60) is negligible since $\sigma_{ij} \neq 0$ and

$$\begin{aligned} &\lim_{n\to\infty} p'_{\lambda_n}(|\sigma_{ij}|) = 0 \\ &\lim_{n\to\infty} p''_{\lambda_n}(|\sigma_{ij}|) = 0. \end{aligned} \tag{60}$$

The same holds for the fist term. The matrix

$$-2 \frac{\partial \tilde{\mathbf{\Phi}}^R_{\vartheta_0}}{\partial \vartheta'} \mathbf{W}_{\vartheta_0} \frac{\partial \tilde{\mathbf{\Phi}}^R_{\vartheta_0}}{\partial \vartheta}, \tag{61}$$

is negative definite and for n large it dominates the other terms, therefore $\mathcal{Q}(\vartheta_0) - \mathcal{Q}(\vartheta) \leq 0$. This implies that there exist a local minimizer $\hat{\vartheta}$ such that $\|\hat{\vartheta} - \vartheta_0\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. □

Proof. Theorem 17. Let us consider the first order Taylor expansion with respect to ϑ_0^1 of the first order condition computed in equation (60)

$$\begin{aligned}\frac{\partial Q(\vartheta)}{\partial \vartheta^1} &= -2\frac{\partial \tilde{\boldsymbol{\Phi}}_{\vartheta}^{R}}{\partial \vartheta^1}\mathbf{W}_{\vartheta^1}\left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\vartheta}^{R}\right) + n\mathbf{v}\\ &= -2\frac{\partial \tilde{\boldsymbol{\Phi}}_{\vartheta_0}^{R}}{\partial \vartheta^1}\mathbf{W}_{\vartheta_0^1}\left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\vartheta_0}^{R}\right) + 2\left(\frac{\partial \tilde{\boldsymbol{\Phi}}_{\vartheta_0}^{R}}{\partial \vartheta^{1\prime}}\mathbf{W}_{\vartheta_0^1}\frac{\partial \tilde{\boldsymbol{\Phi}}_{\vartheta_0}^{R}}{\partial \vartheta^1}\right)\left(\vartheta^1 - \vartheta_0^1\right)\\ &\quad + n\mathbf{v}_0 + n\mathbf{P}_0\left(\vartheta^1 - \vartheta_0^1\right) = 0, \end{aligned} \tag{62}$$

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

$$\begin{aligned}&2\left(\frac{\partial \tilde{\boldsymbol{\Phi}}_{\vartheta_0}^{R}}{\partial \vartheta^{1\prime}}\mathbf{W}_{\vartheta_0^1}\frac{\partial \tilde{\boldsymbol{\Phi}}_{\vartheta_0}^{R}}{\partial \vartheta^1}\right)\left(\vartheta^1 - \vartheta_0^1\right) + n\mathbf{v} + n\mathbf{P}\left(\vartheta^1 - \vartheta_0^1\right)\\ &\quad = 2\frac{\partial \tilde{\boldsymbol{\Phi}}_{\vartheta_0}^{R}}{\partial \vartheta^1}\mathbf{W}_{\vartheta_0^1}\left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\vartheta_0}^{R}\right)\sqrt{n}\left[2\left(\frac{\partial \tilde{\boldsymbol{\Phi}}_{\vartheta_0}^{R}}{\partial \vartheta^{1\prime}}\mathbf{W}_{\vartheta_0^1}\frac{\partial \tilde{\boldsymbol{\Phi}}_{\vartheta_0}^{R}}{\partial \vartheta^1}\right) + n\mathbf{P}_0\right]\\ &\qquad \times \left\{\vartheta^1 - \vartheta_0^1 + \left[2\left(\frac{\partial \tilde{\boldsymbol{\Phi}}_{\vartheta_0}^{R}}{\partial \vartheta^{1\prime}}\mathbf{W}_{\vartheta_0^1}\frac{\partial \tilde{\boldsymbol{\Phi}}_{\vartheta_0}^{R}}{\partial \vartheta^1}\right) + n\mathbf{P}_0\right]^{-1} n\mathbf{v}_0\right\}\\ &\quad = 2\frac{\partial \tilde{\boldsymbol{\Phi}}_{\vartheta_0}^{R}}{\partial \vartheta^1}\mathbf{W}_{\vartheta_0^1}\sqrt{n}\left(\hat{\boldsymbol{\Phi}} - \tilde{\boldsymbol{\Phi}}_{\vartheta_0}^{R}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \frac{\partial \boldsymbol{\Phi}_{\vartheta_0}}{\partial \vartheta^1}\mathbf{W}_{\vartheta_0^1}\boldsymbol{\Omega}_{\vartheta_0}\mathbf{W}'_{\vartheta^1}\frac{\partial \boldsymbol{\Phi}_{\vartheta_0}}{\partial \vartheta^{1\prime}}\right). \end{aligned} \tag{63}$$

Since $\mathbf{v}_0$ and $\mathbf{P}_0$ vanish asymptotically, we apply the same argument of Theorem 13 to complete the proof. □

Appendix B: tables

		$n = 500$			$n = 2000$		
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 1: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the locations $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_d)$, scale matrix $\boldsymbol{\Omega} = \{\omega_{ij}\}$, with $i, j = 1, 2, \ldots, d$ and $i \leq j$ and tail parameter α of the bivariate Elliptical Stable distribution. The results reported above are obtained using $1,000$ 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
μ_1	0.00	-0.0008	0.0281	0.9409	0.0010	0.0470	0.9376
μ_2	0.00	0.0011	0.0406	0.9479	-0.0001	0.0625	0.9315
μ_3	0.00	-0.0024	0.0533	0.9550	0.0021	0.0984	0.9406
μ_4	0.00	-0.0055	0.0785	0.9409	0.0177	0.5072	0.9527
μ_5	0.00	0.0023	0.1149	0.9389	0.0278	1.0030	0.9436
σ_{11}	0.5000	-0.0047	0.0312	0.7688	-0.0015	0.0160	0.8187
σ_{22}	0.7071	0.0040	0.0393	0.7678	0.0019	0.0214	0.7795
σ_{33}	1.0000	-0.0058	0.0547	0.7247	-0.0033	0.0316	0.7402
σ_{44}	1.4142	0.0022	0.0801	0.7337	0.0040	0.0479	0.7422
σ_{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
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.90**	-0.0387	0.1539	0.9891	-0.0092	0.0571	0.7480
μ_1	0.00	0.0030	0.0920	0.9659	-0.0007	0.0142	0.9540
μ_2	0.00	0.0052	0.0768	0.9628	-0.0028	0.0192	0.9560
μ_3	0.00	0.0123	0.3309	0.9566	-0.0012	0.0275	0.9580
μ_4	0.00	0.0130	0.3594	0.9395	0.0029	0.1348	0.9570
μ_5	0.00	0.0306	0.5551	0.9333	0.0000	0.2171	0.9410
σ_{11}	0.5000	-0.0062	0.0297	0.7628	-0.0024	0.0160	0.8150
σ_{22}	0.7071	-0.0021	0.0375	0.7736	0.0000	0.0180	0.7790
σ_{33}	1.0000	-0.0063	0.0499	0.7271	-0.0034	0.0249	0.7420
σ_{44}	1.4142	-0.0021	0.0779	0.7876	0.0017	0.0367	0.7520
σ_{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 2: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the locations $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_d)$, scale matrix $\boldsymbol{\Omega} = \{\omega_{ij}\}$, with $i, j = 1, 2, \ldots, d$ and $i \leq j$ and tail parameter α of the Elliptical Stable distribution in dimension 5. The results reported above are obtained using $1,000$ replications for three different values of $\alpha = (1.70, 1.90)$.

		$n = 500$			$n = 2000$		
Par.	True	BIAS	SSD	ECP	BIAS	SSD	ECP
α	**1.95**	-0.0662	0.2190	0.9854	-0.0278	0.1222	0.9910
μ_1	0.00	-0.0011	0.0284	0.9610	-0.0010	0.0140	0.9550
μ_2	0.00	0.0012	0.0415	0.9463	-0.0029	0.0193	0.9565
μ_3	0.00	-0.0051	0.0572	0.9382	-0.0019	0.0274	0.9475
μ_4	0.00	-0.0162	0.2342	0.9431	-0.0025	0.0400	0.9520
μ_5	0.00	-0.0221	0.4252	0.9496	-0.0094	0.0596	0.9385
σ_{11}	0.5000	-0.0040	0.0293	0.7724	-0.0031	0.0136	0.8186
σ_{22}	0.7071	0.0004	0.0416	0.7333	-0.0012	0.0184	0.7676
σ_{33}	1.0000	-0.0005	0.0516	0.7561	-0.0033	0.0254	0.7481
σ_{44}	1.4142	-0.0010	0.0728	0.7577	0.0012	0.0409	0.7271
σ_{55}	2.0000	-0.0042	0.1085	0.7236	0.0021	0.0493	0.7661
σ_{12}	0.7071	-0.0276	0.1349	0.9496	-0.0043	0.0602	0.9880
σ_{13}	0.8000	-0.0783	0.1569	0.9154	-0.0231	0.0670	0.9835
σ_{14}	0.00	0.0183	0.1326	0.9463	0.0044	0.0574	0.9550
σ_{15}	0.00	0.0226	0.1373	0.8862	-0.0034	0.0548	0.9100
σ_{23}	0.5657	-0.0339	0.1248	0.9171	-0.0064	0.0614	0.9805
σ_{24}	0.00	0.0146	0.1123	0.9496	0.0019	0.0632	0.9220
σ_{25}	0.00	0.0168	0.1233	0.9268	-0.0036	0.0684	0.9280
σ_{34}	0.00	0.0068	0.1100	0.9659	-0.0006	0.0507	0.9805
σ_{35}	0.00	0.0055	0.1053	0.9707	-0.0078	0.0524	0.9835
σ_{45}	0.9016	-0.1245	0.1787	0.9447	-0.0416	0.0724	0.9910

Table 3: Bias (BIAS), sample standard deviation (SSD), and empirical coverage probability (ECP) at the 95% confidence level for the locations $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_d)$, scale matrix $\boldsymbol{\Omega} = \{\omega_{ij}\}$, with $i, j = 1, 2, \ldots, d$ and $i \leq j$ and tail parameter α of the Elliptical Stable distribution in dimension 5. The results reported above are obtained using $1,000$ replications for $\alpha = 1.95$.

References

Jacob Bien and Robert J. Tibshirani. Sparse estimation of a covariance matrix. *Biometrika*, 98(4):807–820, 2011. ISSN 0006-3444. doi: 10.1093/biomet/asr054. URL `http://dx.doi.org/10.1093/biomet/asr054`.

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

Yves Dominicy and David Veredas. The method of simulated quantiles. *J. Econometrics*, 172(2):235–247, 2013. ISSN 0304-4076. doi: 10.1016/j.jeconom.2012.08.010. URL `http://dx.doi.org/10.1016/j.jeconom.2012.08.010`.

Yves Dominicy, Hiroaki Ogata, and David Veredas. Inference for vast dimensional elliptical distributions. *Comput. Statist.*, 28(4):1853–1880, 2013. ISSN 0943-4062. doi: 10.1007/s00180-012-0384-3. URL `http://dx.doi.org/10.1007/s00180-012-0384-3`.

Paul Embrechts, Rdiger Frey, and Alexander McNeil. Quantitative risk management. *Princeton Series in Finance, Princeton*, 10, 2005.

Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.*, 96(456):1348–1360, 2001. ISSN 0162-1459. doi: 10.1198/016214501753382273. URL `http://dx.doi.org/10.1198/016214501753382273`.

Kai Tai Fang, Samuel Kotz, and Kai Wang Ng. *Symmetric multivariate and related distributions*, volume 36 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, Ltd., London, 1990. ISBN 0-412-31430-4. doi: 10.1007/978-1-4899-2937-2. URL `http://dx.doi.org/10.1007/978-1-4899-2937-2`.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics (Oxford, England)*, 9(3):432–441, 2008. doi: 10.1093/biostatistics/kxm045. URL `http://doi.org/10.1093/biostatistics/kxm045`.

A. Ronald Gallant and George Tauchen. Which moments to match? *Econometric Theory*, 12(4):657–681, 1996. ISSN 0266-4666. doi: 10.1017/S0266466600006976. URL `http://dx.doi.org/10.1017/S0266466600006976`.

Xin Gao and Hélène Massam. Estimation of symmetry-constrained Gaussian graphical models: application to clustered dense networks. *J. Comput. Graph. Statist.*, 24(4): 909–929, 2015. ISSN 1061-8600. doi: 10.1080/10618600.2014.937811. URL `http://dx.doi.org/10.1080/10618600.2014.937811`.

Christian Gouriéroux and Alain Monfort. *Simulation-based econometric methods*. CORE lectures. Oxford University Press, 1996. ISBN 0198774753.

Christian Gouriéroux, Alain Monfort, and E. Renault. Indirect inference. *Journal of Applied Econometrics*, 8(S1):S85–S118, 1993. ISSN 1099-1255. doi: 10.1002/jae.3950080507. URL `http://dx.doi.org/10.1002/jae.3950080507`.

Marc Hallin, Davy Paindaveine, and Miroslav Šiman. Multivariate quantiles and multiple-output regression quantiles: from L_1 optimization to halfspace depth. *Ann. Statist.*, 38(2):635–669, 2010. ISSN 0090-5364. doi: 10.1214/09-AOS723. URL `http://dx.doi.org/10.1214/09-AOS723`.

Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982. ISSN 0012-9682. doi: 10.2307/1912775. URL `http://dx.doi.org/10.2307/1912775`.

Wenxin Jiang and Bruce Turnbull. The indirect method: inference based on intermediate statistics—a synthesis and examples. *Statist. Sci.*, 19(2):239–263, 2004. ISSN 0883-4237. doi: 10.1214/088342304000000152. URL `http://dx.doi.org/10.1214/088342304000000152`.

Tae-Hwan Kim and Halbert White. On more robust estimation of skewness and kurtosis. *Finance Research Letters*, 1(1):56–73, 2004. ISSN 1544-6123. doi: 10.1016/S1544-6123(03)00003-5. URL `http://dx.doi.org/10.1016/S1544-6123(03)00003-5`.

Roger Koenker. *Quantile regression*, volume 38 of *Econometric Society Monographs*. Cambridge University Press, Cambridge, 2005. ISBN 978-0-521-60827-5; 0-521-60827-9. doi: 10.1017/CBO9780511754098. URL `http://dx.doi.org/10.1017/CBO9780511754098`.

Linglong Kong and Ivan Mizera. Quantile tomography: using quantiles with multivariate data. *Statist. Sinica*, 22(4):1589–1610, 2012. ISSN 1017-0405.

Dennis Kristensen and Yongseok Shin. Estimation of dynamic models with nonparametric simulated maximum likelihood. *J. Econometrics*, 167(1):76–94, 2012. ISSN 0304-4076. doi: 10.1016/j.jeconom.2011.09.042. URL `http://dx.doi.org/10.1016/j.jeconom.2011.09.042`.

Marco J. Lombardi and David Veredas. Indirect estimation of elliptical stable distributions. *Comput. Statist. Data Anal.*, 53(6):2309–2324, 2009. ISSN 0167-9473. doi: 10.1016/j.csda.2008.04.035. URL `http://dx.doi.org/10.1016/j.csda.2008.04.035`.

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

Daniel McFadden. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57(5):995–1026, 1989. ISSN 0012-9682. doi: 10.2307/1913621. URL `http://dx.doi.org/10.2307/1913621`.

Alexander J. McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative risk management*. Princeton Series in Finance. Princeton University Press, Princeton, NJ, revised edition, 2015. ISBN 978-0-691-16627-8. Concepts, techniques and tools.

Nicolai Meinshausen and Peter Bühlmann. High-dimensional graphs and variable selection with the lasso. *Ann. Statist.*, 34(3):1436–1462, 2006. ISSN 0090-5364. doi: 10.1214/009053606000000281. URL `http://dx.doi.org/10.1214/009053606000000281`.

John P. Nolan. Multivariate elliptically contoured stable distributions: theory and estimation. *Comput. Statist.*, 28(5):2067–2089, 2013. ISSN 0943-4062. doi: 10.1007/s00180-013-0396-7. URL `http://dx.doi.org/10.1007/s00180-013-0396-7`.

Davy Paindaveine and Miroslav Šiman. On directional multiple-output quantile regression. *J. Multivariate Anal.*, 102(2):193–212, 2011. ISSN 0047-259X. doi: 10.1016/j.jmva.2010.08.004. URL `http://dx.doi.org/10.1016/j.jmva.2010.08.004`.

Gennady Samorodnitsky and Murad S. Taqqu. *Stable non-Gaussian random processes*. Stochastic Modeling. Chapman & Hall, New York, 1994. ISBN 0-412-05171-0. Stochastic models with infinite variance.

Robert Serfling. Quantile functions for multivariate analysis: approaches and applications. *Statist. Neerlandica*, 56(2):214–232, 2002. ISSN 0039-0402. doi: 10.1111/1467-9574.00195. URL `http://dx.doi.org/10.1111/1467-9574.00195`. Special issue: Frontier research in theoretical statistics, 2000 (Eindhoven).

Robert Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58(1):267–288, 1996. ISSN 0035-9246.

Hao Wang. Scaling it up: Stochastic search structure learning in graphical models. *Bayesian Anal.*, 10(2):351–377, 06 2015. doi: 10.1214/14-BA916. URL `http://dx.doi.org/10.1214/14-BA916`.

Cun-Hui Zhang. Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.*, 38(2):894–942, 2010. ISSN 0090-5364. doi: 10.1214/09-AOS729. URL `http://dx.doi.org/10.1214/09-AOS729`.

Hui Zou. The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.*, 101 (476):1418–1429, 2006. ISSN 0162-1459. doi: 10.1198/016214506000000735. URL `http://dx.doi.org/10.1198/016214506000000735`.