South African Statistical Journal
Vol. 54, No. 1, 25–43
https://doi.org/10.37920/sasj.2020.54.1.3

LINEAR INFERENCE UNDER MATRIX-STABLE ERRORS

D. R. Jensen
Virginia Polytechnic Institute
e-mail: *djensen@vt.edu*

Linear inference is the foundation stone for much of theoretical and applied statistics. In practice errors often have excessive tails and are lacking the moments required in conventional usage. For random vector responses such errors often are modeled via spherical α-stable distributions with stability index $\alpha \in (0, 2]$, arising in turn through central limit theory but converging to non-Gaussian limits. Earlier work [Jensen, D.R. (2018). *Biom. Biostat. Int. J.* **7**: 205–210] reexamined conventional linear models under n-dimensional α-stable responses, to the effect that Ordinary Least Square (OLS) solutions and residual vectors under α-stable errors also have α-stable distributions, whereas F ratios remain exact in level and power as for Gaussian errors. The present study generalizes those findings to include multivariate linear models having matrix responses of order $(n \times k)$. Topics in inference focus on both location and scale matrices, the latter in connection with analogs of simple, multiple, and canonical correlations without benefit of second moments, seen nonetheless to gauge degrees of association under α-stable symmetry.

Key words: Central limit theory, Excessive errors, Linear inference, Stable laws.

1. Introduction

In the model $\{\boldsymbol{Y} = \boldsymbol{X}\mathfrak{B} + \mathcal{E}\}$ the rows of $\boldsymbol{Y}(n \times k)$ are observed responses, $\boldsymbol{X}(n \times r)$ the regressors, $\mathfrak{B}(r \times k)$ the regression coefficients, and $\mathcal{E}(n \times k)$ an array of random errors. Classical linear inference is cast in the currency of means, variances, correlations, skewness and kurtosis, requiring moments to fourth order. All are met under the traditional requirement that the rows of $\boldsymbol{Y}$ be mutually independent and Gaussian. To the contrary, errors having excessive tails often lack first or second moments; venues for these include acoustics, image processing, radar tracking, biometrics, portfolio analysis, risk management, and others. For an overview see Samorodnitsky and Taqqu (1994). In addition, normality is often deemed to be nonrobust, whereas dependent but uncorrelated responses are encountered in time series and econometrics. In consequence, the classical foundations must be reworked, to include dependent sequences having excessive tails. Accordingly, this study continues the work of Jensen and Good (1981) on structured distributions, to include α-stable matrix distributions.

Non-Gaussian alternatives in wide usage include $\mathsf{S}_n(\mathbf{0}, \boldsymbol{\Sigma})$ as elliptically contoured on $\mathbb{R}^n$ having location-scale parameters $(\mathbf{0}, \boldsymbol{\Sigma})$, as in Fang and Anderson(1990), Fang and Zhang (1990), Gupta and Varga (1993), and numerous archival sources. Excessive tails often are modeled via symmetric α-stable (SαS) distributions $\mathsf{S}_n^{\alpha}(\mathbf{0}, \boldsymbol{\Sigma})$ in $\mathsf{S}_n(\mathbf{0}, \boldsymbol{\Sigma})$ with index $\alpha \in (0, 2\,]$. As foundations to practice, these are the limit distributions for standardized vector sums, namely, Gaussian limits ($\alpha = 2$), Cauchy limits ($\alpha = 1$), and corresponding stable limits otherwise. Despite the venues cited, α-stable errors

MSC2010 subject classifications. 62E15, 62H15, 62J20.

have seen limited usage for want of closed expressions for their density functions, known in selected cases but topics of continuing research. Developments here rest heavily instead on the characteristic functions of distributions, their representations, and their inversion into α-stable densities. Even here a divide emerges between independent, identically distributed (i.i.d.) α-stable sequences, and dependent SαS sequences, these having disparate limit properties as reported in Jensen (2017).

This study undertakes extensions of Jensen (2018) for α-stable distributions in $\mathbb{R}^n$, to include α-stable matrix distributions in $\mathbb{F}_{n\times k}$ of order $(n \times k)$, and their elliptical matrix versions $\mathsf{S}_{n,k}(\mathbf{0}, \boldsymbol{\Xi})$. A key development entails embedding $\mathbb{F}_{n\times k} \to \mathbb{R}^{nk}$, enabling in turn the transfer to $\mathsf{S}_{n,k}(\mathbf{0}, \boldsymbol{\Xi})$ of essential properties widely known for $\mathsf{S}_n(\mathbf{0}, \boldsymbol{\Sigma})$ on $\mathbb{R}^n$. In addition, many findings of the present study are genuinely nonparametric, in applying for all or portions of SαS distributions in the range $\alpha \in (0, 2]$, and thus remaining distribution-free within that class. An outline follows.

Notation and technical foundations are provided next. Subsequent sections develop essentials for structured matrix distributions, and their special role in regard to $\{\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\mathfrak{B}} + \boldsymbol{\mathcal{E}}\}$ subject to SαS errors. In particular, topics in estimation and hypothesis testing are presented separately for the location-scale parameters $(\boldsymbol{\mathfrak{B}}, \boldsymbol{\Xi})$, where $\boldsymbol{\Xi} = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$, to include models having Cauchy errors. A comprehensive collection of supporting materials is contained for completeness in Appendix A.

2. Preliminaries

2.1 Notation

Spaces of note include $\mathbb{R}^n$ as Euclidean n-space; $\mathbb{S}_n$ and $\mathbb{S}_n^+$ as the real symmetric $(n\times n)$ matrices and their positive definite varieties; and $\mathbb{F}_{n\times k}$ as the real matrices of order $(n\times k)$. Vectors and matrices are set in bold type; the transpose, inverse, trace, and determinant of $\boldsymbol{A}$ are $\boldsymbol{A}'$, $\boldsymbol{A}^{-1}$, $\mathrm{tr}(\boldsymbol{A})$, and $|\boldsymbol{A}|$; the unit vector in $\mathbb{R}^n$ is $\mathbf{1}_n = [1, \ldots, 1]'$; and $\boldsymbol{I}_n$ is the $(n \times n)$ identity. For $\boldsymbol{A} \in \mathbb{S}_n^+$, $\mathtt{Ch}(\boldsymbol{A}) = [\alpha_1 \geq \ldots \geq \alpha_n]$ are its characteristic values, $\boldsymbol{A}^{\frac{1}{2}}$ its spectral square root, and $\mathtt{Cnd}(\boldsymbol{A}) = \alpha_1/\alpha_n$ its condition number as in von Neumann and Goldstine (1947). Moreover, $\mathrm{Diag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k)$ is a block-diagonal array, and for $\boldsymbol{A}(n \times n)$ and $\boldsymbol{B}(k \times k)$, the Kronecker product is $\boldsymbol{A} \otimes \boldsymbol{B} = [a_{ij}\boldsymbol{B}]$ of order $(nk \times nk)$. In addition, essential matrix orderings are as follow.

Definition 1. (i) Matrices $(\boldsymbol{G}, \boldsymbol{H})$ in $\mathbb{S}_k$ are ordered as $\boldsymbol{G} \geq_L \boldsymbol{H}$ if and only if $(\boldsymbol{G} - \boldsymbol{H})$ is positive semidefinite; see Loewner (1934);

(ii) Matrices $(\boldsymbol{A}, \boldsymbol{B})$ in $\mathbb{F}_{n\times k}$ are ordered as $\boldsymbol{A} \geq \boldsymbol{B}$ if and only if $\boldsymbol{A}'\boldsymbol{A} \geq_L \boldsymbol{B}'\boldsymbol{B}$; see Jensen (1984);

(iii) The orderings $\boldsymbol{G} >_L \boldsymbol{H}$ and $\boldsymbol{A} > \boldsymbol{B}$ are strict when $(\boldsymbol{G} - \boldsymbol{H})$ and $(\boldsymbol{A}'\boldsymbol{A} - \boldsymbol{B}'\boldsymbol{B})$ are positive definite.

To continue, let $\boldsymbol{Y} \in \mathbb{F}_{n\times k}$ be random. Its law of distribution is $\mathcal{L}(\boldsymbol{Y})$, expected values $\mathtt{E}(\boldsymbol{Y})$ and dispersion matrix $\mathtt{V}(\boldsymbol{Y})$ when defined. Its characteristic function ($\mathtt{Chf}$) with argument $\boldsymbol{T} \in \mathbb{F}_{n\times k}$ is $\phi_{\boldsymbol{Y}}(\boldsymbol{T}) = \mathtt{E}[\exp(\boldsymbol{i}\,\mathrm{tr}\,\boldsymbol{Y}\boldsymbol{T}')]$ and $\boldsymbol{i} = \sqrt{-1}$. See Lukacs and Laha (1964). Conventions for arranging the elements of $\mathtt{V}(\boldsymbol{Y}) = \boldsymbol{\Xi}$, of order $(nk \times nk)$, are addressed in the mapping $\mathtt{J} : \mathbb{F}_{n\times k} \to \mathbb{R}^{nk}$ as follows and in detail in Appendix A.1.

Definition 2. Let $\mathtt{J} : \mathbb{F}_{n\times k} \to \mathbb{R}^{nk}$ take matrices into vectors in indicial order, i.e. for $\boldsymbol{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n]' \in \mathbb{F}_{n\times k}$, then $\mathtt{J}(\boldsymbol{Z}) = [\boldsymbol{z}_1', \boldsymbol{z}_2', \ldots, \boldsymbol{z}_n']' = \boldsymbol{z}_0 \in \mathbb{R}^{nk}$, serving to juxtapose its rows.

Given $\mathtt{V}(\boldsymbol{z}) = \boldsymbol{\Xi} \in \mathbb{S}_n^+$, the function $\Delta^2_{\boldsymbol{\Xi}}(\boldsymbol{x},\boldsymbol{y}) = [(\boldsymbol{x}-\boldsymbol{y})'\boldsymbol{\Xi}^{-1}(\boldsymbol{x}-\boldsymbol{y})]$ is the Mahalanobis (1936) generalized squared distance between $(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}^n$. Essential to this study are the following

Definition 3. Take $\mathcal{L}(\boldsymbol{Z}) \in \mathsf{S}^{\alpha}_{n,k}(\boldsymbol{0},\boldsymbol{I}_{nk})$ to be spherical α-stable on $\mathbb{F}_{n\times k}$ with location-scale parameters $(\boldsymbol{0},\boldsymbol{I}_{nk})$ and stable index $\alpha \in (0, 2\,]$, having moments of order $\alpha+\epsilon$ for $\epsilon > 0$ and $(0 < \alpha < 2)$. Its $\mathtt{Chf}$ is $\phi_{\boldsymbol{Z}}(\boldsymbol{T}) = \exp[\gamma(\mathrm{tr}\,\boldsymbol{T}'\boldsymbol{T})^{\frac{\alpha}{2}}]$ with $\boldsymbol{T} \in \mathbb{F}_{n\times k}$ and $\gamma < 0$. The $\mathtt{Chf}$s for row i and element Z_{ij} are $\phi_{\boldsymbol{Z}_i}(\boldsymbol{t}_i) = \exp\{\gamma(\boldsymbol{t}_i'\boldsymbol{t}_i)^{\frac{\alpha}{2}}\}$ and $\phi_{Z_{ij}}(t_{ij}) = \exp[\gamma|\,t_{ij}|^{\alpha}]$, respectively. Affine transformations yield $\mathsf{S}^{\alpha}_{n,k}(\boldsymbol{\Theta},\boldsymbol{\Xi})$ as elliptical matrix versions; and for $(0 < \alpha < 2)$ the elements $\boldsymbol{\Xi} = [\xi_{ij}]$ serve to gauge degrees of association among elements of $\boldsymbol{Y}$, *in lieu* of undefined correlations. See Appendix A.3.

The concept of *mode* for matrix distributions is essential. We have the following extension to $\mathbb{F}_{n\times k}$ from Dharmadhikari and Joag Dev (1988) on $\mathbb{R}^n$.

Definition 4. A distribution $\mathbf{P}$ on $\mathbb{F}_{n\times k}$ is said to be *monotone unimodal* about $\mathbf{0} \in \mathbb{F}_{n\times k}$ if for every $\mathtt{Y} \in \mathbb{F}_{n\times k}$ and every set C in the class $\mathsf{C}^0_{n,k}$ convex and symmetric about $\mathbf{0} \in \mathbb{F}_{n\times k}$, i.e. $\mathsf{C} \in \mathsf{C}^0_{n,k}$ implies $-\mathsf{C} \in \mathsf{C}^0_{n,k}$, then $\mathbf{P}[\mathsf{C} + k\mathtt{Y}]$ is nonincreasing in $k \in [\,0,\infty)$.

2.2 Central limit theory

This has to do with i.i.d. matrix sequences $\{[\boldsymbol{Z}_1,\boldsymbol{Z}_2,\boldsymbol{Z}_3,\ldots];\ \boldsymbol{Z}_i \in \mathbb{F}_{n\times k}\}$; their partial sums $\overline{\boldsymbol{Z}}_N = N^{-1}[\boldsymbol{Z}_1 + \ldots + \boldsymbol{Z}_N]$; and their distributions as $N \to \infty$. Under conditions on $[\boldsymbol{Z}_1,\boldsymbol{Z}_2,\boldsymbol{Z}_3,\ldots]$ and $c > 0$, the limit distributions $\lim\limits_{N\to\infty} \mathcal{L}(c\overline{\boldsymbol{Z}}_N) \stackrel{\mathrm{def}}{=} \mathcal{L}_{\infty}(c\overline{\boldsymbol{Z}}_N)$ are α-stable. In short, the collection

$$\{\mathsf{S}^{\alpha}_{n,k}(\boldsymbol{0},\boldsymbol{\Xi});\ \boldsymbol{\Xi} \in \mathbb{S}^+_{nk},\ \alpha \in (0,2]\}$$

exhausts the limit distributions for centered i.i.d. sequences $[\boldsymbol{Z}_1,\boldsymbol{Z}_2,\boldsymbol{Z}_3,\ldots]$ on $\mathbb{F}_{n\times k}$; specifically, Gaussian limits at $\alpha = 2$, Cauchy limits at $\alpha = 1$, and α-stable limits otherwise.

To continue, designate by $\mathfrak{D}_{\alpha}$ the *domain of attraction* of $\boldsymbol{Z}_i \in \mathbb{F}_{n\times k}$ such that the limit $\mathcal{L}_{\infty}(c\overline{\boldsymbol{Z}}_N)$ is in $\mathsf{S}^{\alpha}_{n,k}(\boldsymbol{0},\boldsymbol{\Xi})$. Specifically, the distributions $\mathfrak{D}_2$ attracted to Gaussian limits comprise all distributions on $\mathbb{F}_{n\times k}$ having second moments. Similarly, distributions in $\mathfrak{D}_1$ have matrix Cauchy limits. More generally, domains of attraction to SαS distributions in $\mathbb{R}^n$ have been studied in Rvačeva (1962), Kuelbs and Mandrekar (1974), and De Haan and Resnick (1979), to include Lindeberg conditions in Barbosa and Dorea (2009). Berry–Esseén bounds on rates of convergence to stable limits are studied in Rachev and Rüschendorf (1992) and Paulauskas (2009). These in turn carry forward to encompass SαS distributions on $\mathbb{F}_{n\times k}$ through the Duality theory of Appendix A.1.

2.3 Wishart matrices

Consider the scale model $\boldsymbol{\Xi} = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$ as in Muirhead (1982; pp.89–90), together with the Gaussian distribution $\mathrm{N}_{n,k}(\boldsymbol{0},\boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$. A prominent derivation is that of the Wishart matrix $\boldsymbol{W} = \boldsymbol{Y}'\boldsymbol{Y}$ on $\mathbb{S}^+_k$ with distribution $\mathtt{W}_k(n,\boldsymbol{\Sigma})$ having n degrees of freedom and scale parameters $\boldsymbol{\Sigma}$. As this remains of interest even for SαS distributions, its $\mathtt{Chf}$ with argument $\boldsymbol{T} \in \mathbb{S}^+_k$ is $\phi_{\boldsymbol{W}}(\boldsymbol{T}) = |\boldsymbol{I}_n - 2\boldsymbol{i}\,\boldsymbol{T}\boldsymbol{\Sigma}|^{-\frac{n}{2}}$. See Anderson (1984; p.253).

Cumulative probabilities for $\boldsymbol{W} \in \mathbb{S}^+_k$ by convention are $P(W_{ij} \le c_{ij})$ with arguments $\{c_{ij}\}$. However, it is instructive to consider that probabilities might accumulate as $P(\boldsymbol{W} \preceq_L \boldsymbol{D})$ in the sense of Definition 2, with matrix argument $\boldsymbol{D} \in \mathbb{S}^+_k$. The following shows for $\boldsymbol{\Omega} \succeq_L \boldsymbol{\Sigma}$ that $\mathtt{W}_k(\nu,\boldsymbol{\Sigma})$ is more concentrated about $\mathbf{0} \in \mathbb{S}^+_k$ than $\mathtt{W}_k(\nu,\boldsymbol{\Omega})$ in the following sense.

Theorem 1. *Consider* $\mathtt{W}_k(\nu, \boldsymbol{\Omega})$ *and* $\mathtt{W}_k(\nu, \boldsymbol{\Sigma})$ *such that* $\boldsymbol{\Omega} \geq_L \boldsymbol{\Sigma}$, *together with induced measures* $P_{\boldsymbol{\Omega}}(\cdot)$ *and* $P_{\boldsymbol{\Sigma}}(\cdot)$ *on* $\mathbb{S}_n^+$. *Then for each* $\boldsymbol{D} \in \mathbb{S}_k^+$,

$$P_{\boldsymbol{\Sigma}}(\boldsymbol{W} \leq_L \boldsymbol{D}) \geq P_{\boldsymbol{\Omega}}(\boldsymbol{W} \leq_L \boldsymbol{D}).$$

Proof. Take $(\boldsymbol{Y}, \boldsymbol{Z})$ to be centered, and suppose that $\mathcal{L}(\boldsymbol{Y}) = \mathrm{N}_{n,k}(\boldsymbol{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ and $\mathcal{L}(\boldsymbol{Z}) = \mathrm{N}_{n,k}(\boldsymbol{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Omega})$. For $\boldsymbol{a} \in \mathbb{R}^k$ then $\mathcal{L}(\boldsymbol{Y}\boldsymbol{a}) = \mathrm{N}_n(\boldsymbol{0}, \boldsymbol{I}_n \times \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a})$, and similarly $\mathcal{L}(\boldsymbol{Z}\boldsymbol{a}) = \mathrm{N}_n(\boldsymbol{0}, \boldsymbol{I}_n \times \boldsymbol{a}'\boldsymbol{\Omega}\boldsymbol{a})$. Moreover, $\{\boldsymbol{Y}'\boldsymbol{Y} \leq_L \boldsymbol{D}\}$ if and only if $\{\boldsymbol{a}'\boldsymbol{Y}'\boldsymbol{Y}\boldsymbol{a} \leq \boldsymbol{a}'\boldsymbol{D}\boldsymbol{a}\}$ for every $\boldsymbol{a} \in \mathbb{R}^k$, and similarly for $\{\boldsymbol{Z}'\boldsymbol{Z} \leq_L \boldsymbol{D}\}$. In view for both ratios that $\mathcal{L}(\boldsymbol{a}'\boldsymbol{Y}'\boldsymbol{Y}\boldsymbol{a}/\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a})$ and $\mathcal{L}(\boldsymbol{a}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{a}/\boldsymbol{a}'\boldsymbol{\Omega}\boldsymbol{a})$ are chi-squared having n degrees of freedom, together with the fact that $\boldsymbol{\Omega} \geq_L \boldsymbol{\Sigma}$ implies $\boldsymbol{\Omega} = \boldsymbol{\Sigma} + \boldsymbol{A}$ with $\boldsymbol{A}$ positive semidefinite, it follows that

$$\begin{aligned} &P(\boldsymbol{a}'\boldsymbol{Y}'\boldsymbol{Y}\boldsymbol{a} \leq \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}\, c^2) = P(\boldsymbol{a}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{a} \leq \boldsymbol{a}'\boldsymbol{\Omega}\boldsymbol{a}\, c^2) = \\ &P(\boldsymbol{a}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{a} \leq \boldsymbol{a}'(\boldsymbol{\Sigma} + \boldsymbol{A})\boldsymbol{a}\, c^2) \geq P(\boldsymbol{a}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{a} \leq \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}\, c^2) \\ &\Longleftrightarrow P(\boldsymbol{Y}'\boldsymbol{Y} \leq_L \boldsymbol{D}) \geq P(\boldsymbol{Z}'\boldsymbol{Z} \leq_L \boldsymbol{D}) \end{aligned}$$

on replacing $c^2 = \boldsymbol{a}'\boldsymbol{D}\boldsymbol{a}$ for every $\boldsymbol{a} \in \mathbb{R}^k$, to complete the proof. ∎

3. Essentials for $S\alpha S$ distributions

Fundamentals for these distributions are established next, to be followed by an inventory of their stochastic ordering properties.

3.1 Elementary properties

As noted, SαS densities in $\mathbb{R}^n$ are known in selected cases only, to be complemented here for densities on $\mathbb{F}_{n\times k}$. Essentials are given next for spherical distributions centered at $\boldsymbol{0}$, namely $\mathtt{S}^{\alpha}_{n,k}(\boldsymbol{0}, \boldsymbol{I}_{nk})$; location and scale changes follow subsequently. Here $\phi_{\boldsymbol{Y}}(\boldsymbol{T}) = \exp[-\frac{1}{2}\mathrm{tr}(\boldsymbol{T}'\boldsymbol{T})]$ and $g_{n,k}(\mathtt{Y}; \boldsymbol{0}, \boldsymbol{I}_{nk}) = (2\pi)^{-\frac{nk}{2}} \exp[-\frac{1}{2}\mathrm{tr}(\mathtt{Y}'\mathtt{Y})]$ are the Gaussian `Chf` and density for $\mathcal{L}(\boldsymbol{Y}) \in \mathrm{N}_{n,k}(\boldsymbol{0}, \boldsymbol{I}_{nk})$. To continue, the provisional SαS density is $f^{\alpha}_{n,k}(\mathtt{Z}; \boldsymbol{0}, \boldsymbol{I}_{nk})$ for $\mathcal{L}(\boldsymbol{Z}) \in \mathtt{S}^{\alpha}_{n,k}(\boldsymbol{0}, \boldsymbol{I}_{nk})$, and $\phi_{\boldsymbol{Z}}(\boldsymbol{T}; \alpha)$ its `Chf`. The following properties are essential, where $N = nk$ is the effective dimension.

Theorem 2. *Let* $\mathcal{L}(\boldsymbol{Z}) \in \mathtt{S}^{\alpha}_{n,k}(\boldsymbol{0}, \boldsymbol{I}_{nk})$ *have the* `Chf` $\phi_{\boldsymbol{Z}}(\boldsymbol{T}; \alpha)$ *and density function* $f^{\alpha}_{n,k}(\mathtt{Z}; \boldsymbol{0}, \boldsymbol{I}_{nk})$ *where defined. Then the following properties hold.*

(i) $\mathcal{L}(\boldsymbol{Z}) \in \mathtt{S}^{\alpha}_{n,k}(\boldsymbol{0}, \boldsymbol{\Xi})$ *is absolutely continuous on* $\mathbb{F}_{n\times k}$, *having a density function* $f^{\alpha}_{n,k}(\mathtt{Z}; \boldsymbol{0}, \boldsymbol{\Xi})$, *if and only* $\boldsymbol{\Xi}$ *has full rank;*

(ii) The Gaussian mixture $\phi_{\boldsymbol{Z}}(\boldsymbol{T}; \alpha) = \int_0^{\infty} e^{-s^2\mathrm{tr}(\boldsymbol{T}'\boldsymbol{T})/2}\, d\Psi(s; \alpha)$ *for* `Chf`*s holds with* $\Psi(s; \alpha)$ *as a mixing* `cdf.` *on* $[0, \infty)$;

(iii) The Gaussian mixture $f^{\alpha}_{n,k}(\mathtt{Z}; \boldsymbol{0}, \boldsymbol{I}_{nk}) = \int_0^{\infty} g_{n,k}(\mathtt{Z}; \boldsymbol{0}, s^{-2}\boldsymbol{I}_{nk})\, d\Psi(s; \alpha)$ *for densities holds with* $\Psi(s; \alpha)$ *as the mixing* `cdf.` *as before;*

(iv) $\mathcal{L}(\boldsymbol{Z}) \in \mathtt{S}^{\alpha}_{n,k}(\boldsymbol{0}, \boldsymbol{I}_{nk})$ *is monotone unimodal with mode at* $\boldsymbol{0}$, *for each* $\alpha \in (0, 2]$.

Proof. Conclusion (i) is Theorem 6.5.4 of Press (1972) for $\mathbb{R}^n$, extended here for $\mathbb{F}_{n\times k}$ by Duality as in Appendix A.1. Conclusion (ii) invokes a result of Hartman and Wintner (1940), namely, the process $\{Z_t; t = 1, 2 \ldots\}$ is spherically invariant if and only if, for each n and $\boldsymbol{Z} = [Z_1, \ldots, Z_n]$, the Chf $\phi_{\boldsymbol{Z}}(\boldsymbol{t})$ is a scale mixture of Gaussian Chfs on $\mathbb{R}^n$. It is clear from Appendix A.1 that $\boldsymbol{Z}$ is spherical on $\mathbb{F}_{n\times k}$ if and only if the mapping $\mathtt{J}(\boldsymbol{Z})$ of Definition 2 is spherical on $\mathbb{R}^{nk}$, to which the Hartman and Wintner (1940) result applies directly to give conclusion (ii). To continue, $f(\mathsf{Z}) = (2\pi)^{-nk} \int_{\mathbb{F}_{n\times k}} e^{-i\mathrm{tr}(\boldsymbol{T}'\mathsf{Z})} \phi_{\boldsymbol{Z}}(\boldsymbol{T}) \Lambda(d\boldsymbol{T})$ is the standard inversion formula for Chfs to densities in $\mathbb{F}_{n\times k}$ with $\Lambda(\cdot)$ as Lebesgue measure. From conclusion (ii) we thus recover

$$f^{\alpha}_{n,k}(\mathsf{Z}; \boldsymbol{0}, \boldsymbol{I}_{nk}) = \frac{1}{(2\pi)^{nk}} \int_{\mathbb{F}_{n\times k}} e^{-i\mathrm{tr}(\boldsymbol{T}'\mathsf{Z})} \int_0^{\infty} e^{-s^2 \mathrm{tr}(\boldsymbol{T}'\boldsymbol{T})/2} d\Psi(s;\alpha) \Lambda(d\boldsymbol{T}).$$

Reversing the order of integration inverts the Gaussian Chf to give conclusion (iii). Conclusion (iv) attributes to Wolfe (1975) on $\mathbb{R}^n$ thus $\mathbb{F}_{n\times k}$ by Duality. ■

3.2 Stochastic orderings

Essential stochastic orderings are known for distributions in $\mathsf{S}_n(\boldsymbol{0}, \boldsymbol{\Sigma})$ as $\boldsymbol{\Sigma}$ is allowed to vary; by Duality these extend to include $\mathsf{S}^{\alpha}_{n,k}(\boldsymbol{0}, \boldsymbol{\Xi})$ on $\mathbb{F}_{n\times k}$ as follows.

Definition 5. A probability measure $\mu(\cdot)$ on $\mathbb{F}_{n\times k}$ is said to be *more peaked* about $\boldsymbol{0} \in \mathbb{F}_{n\times k}$ than $\nu(\cdot)$ if, for every set $\mathsf{C} \in \mathsf{C}^0_{n,k}$ as in Definition 4, the inequality $\mu(\mathsf{C}) \geq \nu(\mathsf{C})$ holds, to be designated as $\mu \geq_{\mathrm{P}} \nu$. See Birnbaum (1948) for distributions on $\mathbb{R}^1$, and Sherman (1955) for distributions on $\mathbb{R}^n$.

Essential peakedness orderings for SαS distributions on $\mathbb{F}_{n\times k}$ are as follow.

Theorem 3. *Let* $\mu(\cdot\,;\boldsymbol{\Xi}) \in \mathsf{S}^{\alpha}_{n,k}(\boldsymbol{0}, \boldsymbol{\Xi})$ *and* $\nu(\cdot\,;\boldsymbol{\Gamma}) \in \mathsf{S}^{\alpha}_{n,k}(\boldsymbol{0}, \boldsymbol{\Gamma})$. *Then* $\mu \geq_{\mathrm{P}} \nu$, *that is,* $\mu(\mathsf{C}; \boldsymbol{\Xi}) \geq \nu(\mathsf{C}; \boldsymbol{\Gamma})$ *for every* $\mathsf{C} \in \mathsf{C}^0_{n,k}$, *if and only if* $\boldsymbol{\Gamma} \geq_L \boldsymbol{\Xi}$ *in the sense of Definition 1.*

Proof. That $\boldsymbol{\Gamma} \geq_L \boldsymbol{\Xi}$ implies $\mu \geq_{\mathrm{P}} \nu$, is established in Das Gupta et al. (1971) and Fefferman et al. (1972) for distributions in $\mathsf{S}_n(\boldsymbol{0}, \boldsymbol{\Sigma})$, and thus for $\mathsf{S}^{\alpha}_n(\boldsymbol{0}, \boldsymbol{\Sigma})$ on $\mathbb{R}^n$ by inclusion. The converse was shown in Theorem 1 of Jensen (1984); both are extended here to include $\mathsf{S}^{\alpha}_{n,k}(\boldsymbol{0}, \boldsymbol{\Omega})$ by Duality as in Appendix A.1. ■

4. Linear models under $S\alpha S$ errors

4.1 The structure: $\boldsymbol{\Xi} = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$

The mapping $\mathtt{J} : \mathbb{F}_{n\times k} \to \mathbb{R}^{nk}$ of Appendix A.1 gives $\mathtt{V}(\boldsymbol{Y}) \stackrel{\text{def}}{=} \mathtt{V}(\mathtt{J}(\boldsymbol{Y})) = \boldsymbol{\Xi} \in \mathbb{S}^+_{nk}$ under second moments, where the elements of $\boldsymbol{Y}(n\times k)$ may be reported in nk distinct units. Instead, here the rows of $\boldsymbol{Y}$ are k-dimensional responses consistently across the n rows. On partitioning $\mathtt{V}(\boldsymbol{Y}) = [\boldsymbol{\Xi}_{ij}(k\times k)]$, it is clear for row $\boldsymbol{Y}_i$ that $\mathtt{V}(\boldsymbol{Y}_i) = \boldsymbol{\Xi}_{ii} = \boldsymbol{\Sigma}$ say, as diagonal blocks, whereas the off-diagonal blocks $\mathtt{Cov}(\boldsymbol{Y}_i, \boldsymbol{Y}_j) = \boldsymbol{\Xi}_{ij}$, as cross-covariances, are necessarily in the same units as $\boldsymbol{\Xi}_{ii}$. In the absence of second moments take $\boldsymbol{\Xi}$ instead to be scale parameters. Accordingly, these facts support the basic structure $\boldsymbol{\Xi} = [\omega_{ij}\boldsymbol{\Sigma}] = \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}$ with $\boldsymbol{\Omega}(n\times n)$ and $\boldsymbol{\Sigma}(k\times k)$. Further taking $\boldsymbol{\Xi} = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$, as is done subsequently, not only mimics the classical Gaussian model at $\alpha = 2$ having independent rows, but it goes beyond in that the rows of $\boldsymbol{Y}$ in $\mathsf{S}_{n,k}(\boldsymbol{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ may be dependent but uncorrelated, as in multivariate time series and econometrics. Nonetheless, it remains to ask whether non-Gaussian members of $\mathsf{S}_{n,k}(\boldsymbol{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ might have independent rows. To the contrary, extending a result of Maxwell (1860), these are independent and uncorrelated if and only if Gaussian, as shown in Appendix A.4. Nonetheless, confusion on this persists in the literature; see Remark 5, Appendix A.4.

4.2 Properties of the solutions

Here $\mathcal{L}(\boldsymbol{Y}) \in \mathsf{S}^{\alpha}_{n,k}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ with $(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ as centering and scale parameters, where $\{\boldsymbol{Y} \in \mathbb{F}_{n\times k}, \boldsymbol{X} \in \mathbb{F}_{n\times r}, \mathfrak{B} \in \mathbb{F}_{r\times k}\}$. The classical OLS solutions, as minimally dispersed unbiased linear estimates, are available here only for $\alpha = 2$, whereas alternative moment criteria necessarily are subject to moment constraints. Specifically, for scalars $(\widehat{\theta}, \theta) \in \mathbb{R}^1$ under loss $L(\widehat{\theta}, \theta) = |\widehat{\theta} - \theta|$, the risk $R(\widehat{\theta}) = \mathsf{E}[L(\widehat{\theta}, \theta)]$ is undefined for $\alpha < 1$ as for Cauchy errors at $\alpha = 1$. Moreover, loss functions $(|\widehat{\theta} - \theta|^{\kappa})$ are concave for $\{\kappa < \alpha < 1\}$, and convex for $\{1 < \kappa \leq \alpha \leq 2\}$, the latter at issue in attaining global optima. Versions of these, i.e. minimal risk estimation for matrix models, not only would require knowledge regarding α, but also optimizing algorithms. Instead we seek what might be salvaged from classical linear models under the constraints of SαS errors.

Minimizing $Q(\mathfrak{B}) = \mathrm{tr}(\boldsymbol{Y} - \boldsymbol{X}\mathfrak{B})'(\boldsymbol{Y} - \boldsymbol{X}\mathfrak{B})$ as $\mathfrak{B}$ varies yields the OLS solution $\widehat{\mathfrak{B}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ for $\mathfrak{B}$ and $\boldsymbol{X}\widehat{\mathfrak{B}}$ for approximating $\boldsymbol{Y}$. We adopt much stronger minimizing properties as follow.

Theorem 4. *Suppose* $\boldsymbol{Y} = \boldsymbol{X}\mathfrak{B} + \mathcal{E}$ *together with the ordering* $(\mathbb{F}_{n\times k}, \succeq)$ *of Definition 2. Then*

(i) $\widehat{\mathfrak{B}}$ *is minimizing on* $(\mathbb{F}_{n\times k}, \succeq)$ *in the sense that* $(\boldsymbol{Y} - \boldsymbol{X}\mathfrak{B}) \succeq (\boldsymbol{Y} - \boldsymbol{X}\widehat{\mathfrak{B}})$ *for every* $\mathfrak{B} \in \mathbb{F}_{r\times k}$;

(ii) $\psi(\boldsymbol{Y} - \boldsymbol{X}\widehat{\mathfrak{B}}) \leq \psi(\boldsymbol{Y} - \boldsymbol{X}\mathfrak{B})$ *for every* ψ *in the class* Ψ *of functions monotone increasing on* $(\mathbb{F}_{n\times k}, \succeq)$;

(iii) $\widehat{\mathfrak{B}}$ *is the minimum-norm solution to* $\min_{\mathfrak{B}\in\mathbb{F}_{r\times k}} ||\boldsymbol{Y} - \boldsymbol{X}\mathfrak{B}||_{\phi}$ *for every unitarily invariant norm* $||\cdot||_{\phi}$ *on* $\mathbb{F}_{n\times k}$.

Proof. See Theorem 6 of Jensen (1984). Conclusion (iii) was obtained by Rao (1980) as a consequence of ordering the singular values $\sigma(\boldsymbol{Y} - \boldsymbol{X}\mathfrak{B}) \geq \sigma(\boldsymbol{Y} - \boldsymbol{X}\widehat{\mathfrak{B}})$. ■

On taking $\mathbf{P} = [\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']$, the elements of $\boldsymbol{E} = \mathbf{P}\boldsymbol{Y}$ comprise the observed residuals and $\boldsymbol{S} = \boldsymbol{E}'\boldsymbol{E}/(n - r)$ the matrix of residual mean squares and mean products. We proceed to examine essential properties of $\mathsf{S}^{\alpha}_{n,k}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ as α ranges over $(0, 2]$.

Remark 1. The `Chf` for $\mathcal{L}(\boldsymbol{Y}) \in \mathsf{S}^{\alpha}_{n,k}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ is $\phi_{\boldsymbol{Y}}(\boldsymbol{T}) = \exp\{\boldsymbol{i}\,\mathrm{tr}\,\boldsymbol{T}'\boldsymbol{X}\mathfrak{B} + \gamma[\mathrm{tr}\,\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}']^{\frac{\alpha}{2}}\}$ from Definition 3. Moreover, expressions to follow are simplified on letting $(\boldsymbol{X}'\boldsymbol{X})^{-1} = \mathbf{H}$.

The following properties are fundamental.

Theorem 5. *Given* $\mathcal{L}(\boldsymbol{Y}) = \mathsf{S}^{\alpha}_{n,k}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$, *consider* $[\widehat{\mathfrak{B}}, \boldsymbol{E}]$ *with* $\boldsymbol{E} = \mathbf{P}\boldsymbol{Y}$ *as the matrix of residuals, and* $\boldsymbol{S} = \boldsymbol{E}'\boldsymbol{E}/(n - r)$. *Then*

(i) $\mathcal{L}(\widehat{\mathfrak{B}}, \boldsymbol{E}) \in \mathsf{S}^{\alpha}_{r+n,k}([\mathfrak{B}, \mathbf{0}], \boldsymbol{\Omega})$, *with* $\boldsymbol{\Omega} = \mathrm{Diag}(\mathbf{H}, \mathbf{P}) \otimes \boldsymbol{\Sigma}$, *a distribution on* $\mathbb{F}_{(r+n)\times k}$ *of rank n*;

(ii) The marginals are $\mathcal{L}(\widehat{\mathfrak{B}}) \in \mathsf{S}^{\alpha}_{r,k}(\mathfrak{B}, \mathbf{H} \otimes \boldsymbol{\Sigma})$ *centered at* $\mathfrak{B}$ *with scale parameters* $\mathbf{H} \otimes \boldsymbol{\Sigma}$;

(iii) $\mathcal{L}(\boldsymbol{E}) \in \mathsf{S}^{\alpha}_{n,k}(\mathbf{0}, \mathbf{P} \otimes \boldsymbol{\Sigma})$, *a distribution on* $\mathbb{F}_{n\times k}$ *of rank* $(n - r)k$ *centered at* $\mathbf{0}$ *with scale parameters* $\mathbf{P} \otimes \boldsymbol{\Sigma}$;

(iv) *The distribution of* $\boldsymbol{W} = \boldsymbol{E}'\boldsymbol{E}$ *is* α*-Wishart, namely* $\mathtt{W}_k^{\alpha}(\nu, \boldsymbol{\Sigma})$, *having the* Chf

$$\phi_{\boldsymbol{W}}(\boldsymbol{T}; \nu, \alpha) = \int_0^{\infty} |\boldsymbol{I}_n - 2\boldsymbol{i}\, s^{-2}\boldsymbol{T}\boldsymbol{\Sigma}|^{-\frac{\nu}{2}}\, d\Psi(s; \alpha)$$

with argument $\boldsymbol{T} \in \mathbb{S}_k^+$, $\nu = (n - r)$ *degrees of freedom, and* $\Psi(s; \alpha)$ *as in Theorem 1;*

(v) *Let* $\psi(\boldsymbol{Y})$ *be scale-invariant, i.e.* $\psi(\boldsymbol{Y}) = \psi(c\boldsymbol{Y}), c \neq 0$; *then its distribution is identical to that on sampling* $\boldsymbol{Y}$ *from* $\mathrm{N}_{n,k}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$.

Proof. Let $\boldsymbol{L}' = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and $\mathbf{P} = \left[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right]$ to project onto the error space, so that $\boldsymbol{G} = [\boldsymbol{L}, \mathbf{P}]$ operates on $\boldsymbol{Y}$ to give

$$\boldsymbol{Z} = \boldsymbol{G}'\boldsymbol{Y} = \left[\begin{smallmatrix} \boldsymbol{L}' \\ \mathbf{P}' \end{smallmatrix}\right] \boldsymbol{Y} = \left[\begin{smallmatrix} \widehat{\mathfrak{B}} \\ \boldsymbol{E} \end{smallmatrix}\right] \text{ and } \boldsymbol{G}'\boldsymbol{G} = \left[\begin{smallmatrix} \mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{P} \end{smallmatrix}\right],$$

the latter of order $[(n + r) \times (n + r)]$ and rank n. From Remark 1 we have the Chf

$$\begin{aligned}\phi_{\boldsymbol{Z}}(\boldsymbol{U}) &= \exp[\boldsymbol{i}\,\mathrm{tr}(\boldsymbol{G}'\boldsymbol{Y}\boldsymbol{U}')] = \exp[\boldsymbol{i}\,\mathrm{tr}(\boldsymbol{U}'\boldsymbol{G}'\boldsymbol{Y})]\\ &= \phi_{\boldsymbol{Y}}(\boldsymbol{G}\boldsymbol{U}) = \exp\{\boldsymbol{i}\,\mathrm{tr}\,\boldsymbol{U}_1'\mathfrak{B} + \gamma\,[\mathrm{tr}\,\boldsymbol{\Sigma}(\boldsymbol{U}_1'\mathbf{H}\boldsymbol{U}_1 + \boldsymbol{U}_2'\mathbf{P}\boldsymbol{U}_2)]^{\frac{\alpha}{2}}\}\end{aligned}$$

on partitioning $\boldsymbol{U}' = [\boldsymbol{U}_1', \boldsymbol{U}_2']$. Here $\boldsymbol{U}'\boldsymbol{G}'\boldsymbol{X}\mathfrak{B} = \boldsymbol{U}_1'\mathfrak{B}$ since $\mathbf{P}'\boldsymbol{X} = \mathbf{0}$. The result is conclusion (i). Taking in succession $\boldsymbol{U}_2 = \mathbf{0}$, then $\boldsymbol{U}_1 = \mathbf{0}$, gives the marginal Chfs as in conclusions (ii) and (iii). To see conclusion (iv), the Gaussian mixture

$$f_{n,k}^{\alpha}(\mathbf{Y}; \boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}) = \int_0^{\infty} g_{n,k}(\mathbf{Y}; \boldsymbol{X}\mathfrak{B}, s^{-2}\boldsymbol{I}_n \otimes \boldsymbol{\Sigma})\, d\Psi(s; \alpha) \tag{1}$$

updates the location and scale parameters of Theorem 1(iii). On letting $\boldsymbol{W} = \boldsymbol{E}'\boldsymbol{E} = \boldsymbol{Y}'\mathbf{P}\boldsymbol{Y}$ and observing that $\mathbf{P}\boldsymbol{X}\mathfrak{B} = \mathbf{0}$, we proceed to the Chf

$$\phi_{\boldsymbol{W}}(\boldsymbol{T}) = \int_{\mathbb{F}_{n\times k}} e^{\boldsymbol{i}\,\mathrm{tr}(\mathbf{Y}'\mathbf{P}\mathbf{Y})\boldsymbol{T}')} \int_0^{\infty} g_{n,k}(\mathbf{Y}; \boldsymbol{X}\mathfrak{B}, s^{-2}\boldsymbol{I}_n \otimes \boldsymbol{\Sigma})\, d\Psi(s; \alpha)\, \Lambda(d\mathbf{Y})$$

as in conclusion (iv), again on reversing the order of integration. For conclusion (v), make the change of variables $\psi(\boldsymbol{Y}) \to \psi(c\boldsymbol{Y})$ behind the integral in (1). From scale-invariance, i.e. $\psi(\boldsymbol{Y})$ not depending on s, it suffices to take $\Psi(s; \alpha)$ to be the Dirac delta function, $\{d\Psi(s; \alpha) = 0,\ s \neq 1; d\Psi(1; \alpha) = 1;\}$, equal to zero everywhere except for unity and whose integral over the entire real line is equal to one, to show conclusion (v). ■

Remark 2. That $\boldsymbol{\Omega} = \mathrm{Diag}(\boldsymbol{H}, \mathbf{P}) \otimes \boldsymbol{\Sigma}$ is block-diagonal in conclusion (i), assures under SαS errors that $(\widehat{\mathfrak{B}}, \boldsymbol{E})$ are α-nonassociated as in Appendix A.3 Definition 8, well known to be mutually uncorrelated under second moments.

A notable special case is the matrix Cauchy distribution as follows, often cited for its anomalous characteristics.

Corollary 1. *The elliptical Cauchy density for* $\widehat{\mathfrak{B}}$ *on* $\mathbb{F}_{n\times k}$ *is*

$$f_k^1(\widehat{\mathfrak{B}}; \mathfrak{B}, \boldsymbol{X}'\boldsymbol{X}) = c(nk)\left[1 + \mathrm{tr}\,\boldsymbol{\Sigma}^{-1}(\widehat{\mathfrak{B}} - \mathfrak{B})'\boldsymbol{X}'\boldsymbol{X}(\widehat{\mathfrak{B}} - \mathfrak{B})\right]^{-\frac{nk+1}{2}}$$

where $c(nk) = [\Gamma(\frac{nk+1}{2})|\boldsymbol{X}'\boldsymbol{X}|^{\frac{1}{2}}/\pi^{\frac{nk+1}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}]$ *and from Theorem 1(iii)* $d\Psi(s; 1) = \frac{2}{\sqrt{2\pi}}s^{-2}e^{-1/2s^2}$ *is the mixing distribution, namely, an inverted Gamma density.*

Proof. Zellner (1976) gives the multivariate t-density on $\mathbb{R}^n$ with ν degrees of freedom as $f(\boldsymbol{x}) = c(n;\boldsymbol{\Sigma})[1+\frac{1}{\nu}(\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x})]^{-\frac{\nu+n}{2}}$ with $c(n;\boldsymbol{\Sigma}) = \Gamma(\frac{\nu+n}{2})/[\Gamma(\frac{\nu}{2})(\nu\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}]$, namely, Cauchy at $\nu = 1$. The matrix version follows by Duality. ∎

5. Topics: estimation

For distributions $\mathcal{L}(\boldsymbol{Y}) \in \mathtt{S}^{\alpha}_{n,k}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$, we consider in succession particulars in regard to $\mathfrak{B}$ and $\boldsymbol{\Sigma}$.

5.1 Location parameters

Portions of our findings extend beyond Gauss–Markov theory and *OLS*, typically requiring second moments, to include other criteria and the much larger class of *equivariant* estimators. Details follow.

Definition 6. An estimator $\widetilde{\mathfrak{B}}(\boldsymbol{Y})$ for $\mathfrak{B} \in \mathbb{F}_{r\times k}$ is unbiased if $\mathtt{E}(\widetilde{\mathfrak{B}}) = \mathfrak{B}$; is *linearly median-unbiased* if $\{\mathtt{med}(\boldsymbol{a}'\widetilde{\mathfrak{B}}\boldsymbol{b}) = \boldsymbol{a}'\mathfrak{B}\boldsymbol{b};\ (\boldsymbol{a},\boldsymbol{b}) \in \mathbb{R}^r \times \mathbb{R}^k\}$; is *modal-unbiased* if mode $\mathtt{M}(\widetilde{\mathfrak{B}}) = \mathfrak{B}$; and is *translation-equivariant* if for $\{\boldsymbol{Y} \to \boldsymbol{Y} + \boldsymbol{X}\boldsymbol{B}\}$, then $\widetilde{\mathfrak{B}}(\boldsymbol{Y}+\boldsymbol{X}\boldsymbol{B}) = \widetilde{\mathfrak{B}}(\boldsymbol{Y}) + \boldsymbol{B}$ for every $\boldsymbol{B} \in \mathbb{F}_{r\times k}$.

Essential properties include the following, where the peakedness ordering for measures is as in Definition 5. Again it is expedient to let $\boldsymbol{H} = (\boldsymbol{X}'\boldsymbol{X})^{-1}$.

Theorem 6. *For $\mathcal{L}(\boldsymbol{Y}) = \mathtt{S}^{\alpha}_{n,k}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$, consider properties of the OLS solutions $\widehat{\mathfrak{B}} = \boldsymbol{H}\boldsymbol{X}'\boldsymbol{Y}$ without benefit of second moments, of the equivariant estimators $\widetilde{\mathfrak{B}}(\boldsymbol{Y})$ of Definition 6, and of alternative estimators $\check{\mathfrak{B}}$ centered at $\mathfrak{B}$.*

(i) $\widehat{\mathfrak{B}}$ *is unbiased for* $(1 < \alpha \leq 2)$; *is linearly median unbiased; and is modal unbiased for* $\mathfrak{B}$;

(ii) $\widehat{\mathfrak{B}}$ *is most peaked about* $\mathfrak{B}$ *among all unbiased, median-unbiased, and modal-unbiased linear estimators for* $\mathfrak{B}$;

(iii) $\widehat{\mathfrak{B}}$ *is most peaked about* $\mathfrak{B}$ *among all equivariant estimators* $\widetilde{\mathfrak{B}}(\boldsymbol{Y})$.

(iv) Consider designs $(\boldsymbol{X},\boldsymbol{Z})$ such that $\boldsymbol{X} \geq \boldsymbol{Z}$ as in Definition 1(ii). Then $\widehat{\mathfrak{B}}(\boldsymbol{X})$ is more peaked about $\mathfrak{B}$ than $\widehat{\mathfrak{B}}(\boldsymbol{Z})$.

(v) The Mahalanobis metric, $\Delta^2_{\Xi}(\widehat{\mathfrak{B}},\mathfrak{B}) = [\mathrm{tr}\,(\widehat{\mathfrak{B}}-\mathfrak{B})'\boldsymbol{X}'\boldsymbol{X}(\widehat{\mathfrak{B}}-\mathfrak{B})\boldsymbol{\Sigma}^{-1}]$, *serves to quantify the discrepancy between* $(\widehat{\mathfrak{B}},\mathfrak{B})$.

Proof. Conclusion (i) follows since $\widehat{\mathfrak{B}}$ is centered at $\mathfrak{B}$ from Theorem 5(ii) and that first moments are defined for $(1 < \alpha \leq 2)$. Then $\mathcal{L}(\boldsymbol{a}'\widehat{\mathfrak{B}}\boldsymbol{b})$ is symmetric on $\mathbb{R}^1$ and centered at its median $\boldsymbol{a}'\mathfrak{B}\boldsymbol{b}$; moreover, $\mathcal{L}(\widehat{\mathfrak{B}})$ is unimodal with mode $\mathfrak{B}$ from Theorem 1(iv). Conclusion (ii) follows on showing that alternative linear estimators $\check{\mathfrak{B}}$ centered at $\mathfrak{B}$ have inflated scale parameters, i.e. inflated $\mathtt{V}(\check{\mathfrak{B}})$ under second moments. To continue, begin with $\phi_{\boldsymbol{Y}}(\boldsymbol{T}) = \exp[\boldsymbol{i}\,\mathrm{tr}(\boldsymbol{T}'\boldsymbol{X}\mathfrak{B}) - \frac{1}{2}(\mathrm{tr}\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}')^{\frac{\alpha}{2}}]$, and consider $\check{\mathfrak{B}} = \boldsymbol{L}'\boldsymbol{Y}$ with $\boldsymbol{L}' = [\boldsymbol{H}\boldsymbol{X}', \boldsymbol{G}']$, so that

$$\phi_{\check{\mathfrak{B}}}(\boldsymbol{S}) = \exp[\boldsymbol{i}\,\mathrm{tr}(\boldsymbol{S}'\boldsymbol{L}'\boldsymbol{X}\mathfrak{B}) - \frac{1}{2}(\mathrm{tr}\boldsymbol{L}\boldsymbol{S}\boldsymbol{\Sigma}\boldsymbol{S}'\boldsymbol{L}')^{\frac{\alpha}{2}}].$$

Since $\boldsymbol{S}'\boldsymbol{L}'\boldsymbol{X}\mathfrak{B} = \boldsymbol{S}'[\boldsymbol{H}\boldsymbol{X}', \boldsymbol{G}']\boldsymbol{X}\mathfrak{B}$, in order that $\check{\mathfrak{B}}$ should be centered at $\mathfrak{B}$, it is necessary that $\boldsymbol{G}'\boldsymbol{X} = \boldsymbol{0}$.

Table 1. Estimating $\boldsymbol{\Sigma}$, its characteristic values $\mathtt{Ch}(\boldsymbol{\Sigma})$, and $\mathrm{tr}(\boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}^*$ seeks to shrink $\mathtt{Ch}(\boldsymbol{\Sigma}^*)$ towards their arithmetic mean. Here $\mathcal{L}(\boldsymbol{Y}) \in \mathtt{T}_{n,k}^{\nu}(\mathbf{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ is a matrix T distribution on $\mathbb{F}_{n\times k}$ having ν degrees of freedom, a subset of $\mathtt{S}_{n,k}(\mathbf{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$.

Function	**Loss Function**	**Model**	**Reference**
$\mathtt{Ch}(\boldsymbol{\Sigma})$	Quadratic	$\mathtt{T}_{n,k}(\mathbf{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$	Joarder&Ahmed (1996)
$\boldsymbol{\Sigma}$	Entropy	$\mathtt{T}_{n,k}(\mathbf{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$	Joarder&Ali (1997)
$\mathrm{tr}(\boldsymbol{\Sigma})$	Quadratic	$\mathtt{T}_{n,k}(\mathbf{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$	Joarder&Singh (2001)
$\boldsymbol{\Sigma}^*$	Quadratic	$\mathtt{S}_{n,k}(\mathbf{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$	Leung&Ng (2004)

Accordingly, it follows that $\phi_{\breve{\mathfrak{B}}}(S) = \exp[\boldsymbol{i}\,\mathrm{tr}\boldsymbol{S}'\mathfrak{B} - \frac{1}{2}(\mathrm{tr}\boldsymbol{\Omega}\boldsymbol{S}\boldsymbol{\Sigma}\boldsymbol{S}')^{\frac{\alpha}{2}}]$, with $\boldsymbol{\Omega} = \boldsymbol{L}'\boldsymbol{L} = [\boldsymbol{H} + \boldsymbol{G}'\boldsymbol{G}]$, giving $\mathcal{L}(\breve{\mathfrak{B}}) \in \mathtt{S}_{n,k}^{\alpha}(\mathfrak{B}, \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma})$, to be compared with the conventional $\mathcal{L}(\widehat{\mathfrak{B}}) \in \mathtt{S}_{n,k}^{\alpha}(\mathfrak{B}, \boldsymbol{H} \otimes \boldsymbol{\Sigma})$ from Theorem 5(ii). Clearly the matrix $[\boldsymbol{L}'\boldsymbol{L} - \boldsymbol{H}] = \boldsymbol{G}'\boldsymbol{G}$ is positive semidefinite, so that $\boldsymbol{\Gamma} = \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma} \geq_L \boldsymbol{H} \otimes \boldsymbol{\Sigma} = \boldsymbol{\Xi}$. Conclusion (ii) now follows from Theorem 2. Conclusion (iii) stems from Theorem 2.7 of Burk and Hwang (1989) since SαS distributions are unimodal from Theorem 1(iv). Next observe that the scale parameters for $\widehat{\mathfrak{B}}(\boldsymbol{X})$ and $\widehat{\mathfrak{B}}(\boldsymbol{Z})$ are ordered as $((\boldsymbol{X}'\boldsymbol{X})^{-1} \otimes \boldsymbol{\Omega}) \leq_L ((\boldsymbol{Z}'\boldsymbol{Z})^{-1} \otimes \boldsymbol{\Sigma})$, so that conclusion (iv) follows from Theorem 2.

To continue, take $\{\mathcal{L}(\boldsymbol{Z}) \in \mathtt{S}_{r,k}^{\alpha}(\boldsymbol{M}, \boldsymbol{\Xi}); \boldsymbol{\Xi} = \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma})\}$; by Duality the Mahalanobis (1936) metric on $\mathbb{F}_{r\times k}$ is

$$\Delta_{\boldsymbol{\Xi}}^2(\mathbf{X}, \mathbf{Y}) \overset{\text{def}}{=} ||\mathbf{X}, \mathbf{Y}||_{\boldsymbol{\Xi}}^2 = \mathrm{tr}\,[\boldsymbol{\Omega}^{-1}(\mathbf{X} - \mathbf{Y})\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \mathbf{Y})'].$$

Conclusion (v) follows on identifying $(\mathbf{X}, \mathbf{Y}, \boldsymbol{\Omega}^{-1})$ as $(\widehat{\mathfrak{B}}, \mathfrak{B}, \boldsymbol{X}'\boldsymbol{X})$, respectively. ■

Remark 3. Returning to the Cauchy density of Corollary 1, it is seen that the matrix form within square brackets is the squared Mahalanobis (1936) distance metric on $\mathbb{F}_{r\times k}$.

5.2 Scale parameters

Return to $\boldsymbol{S} = \boldsymbol{E}'\boldsymbol{E}/(n-r)$ and the evidence it conveys regarding scale in $\mathtt{S}_{n,k}^{\alpha}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$. Clearly unbiasedness is denied for $(0 < \alpha < 2)$ for want of second moments. Even under second moments, the Table 1 survey is prompted by the cited tendency of the characteristic values of $\boldsymbol{S}$ to be more scattered than those of $\boldsymbol{\Sigma}$. Accordingly, alternatives to the conventional use of $\boldsymbol{S}$ itself are reported in the literature, including those listed in Table 1 and the references cited, amounting in effect to shrinkage operations. For point of reference, recall that $\mathtt{T}_{n,k}^{\nu}(\mathbf{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ at $\nu = 1$ is a matrix Cauchy distribution in $\mathtt{S}_{n,k}^{\alpha}(\mathbf{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ for $\alpha = 1$.

If scatter among the characteristic values $\mathtt{Ch}(\boldsymbol{S})$ is at issue, it would appear doubly so for the heavy-tailed distributions in $\mathtt{S}_{n,k}^{\alpha}(\mathbf{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$. Unfortunately, none of the estimators of Table 1 applies here for $(0 < \alpha < 2)$, despite their theoretical and empirical advantages, because risks for the designated loss functions are undefined for want of second moments. Instead we adopt and illustrate two matrix-theoretic devices for shrinking $\boldsymbol{S}$, neither depending on provisional loss functions whose risks might be undefined.

Table 2. The matrix $\boldsymbol{\Sigma}$; its $\boldsymbol{\Sigma}$-diminishing transformation $\boldsymbol{\Xi}$; its majorized transformation $\boldsymbol{\Omega}$; and the supporting orthogonal matrix $\boldsymbol{Q}$ and doubly stochastic matrix $\mathbf{B}$.

Item	$\boldsymbol{\Sigma}$	$\boldsymbol{\Xi}$	$\boldsymbol{\Omega}$	$\boldsymbol{Q}$	$\mathbf{B}$
Matrices	$\begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 4 & 0 & 0 \\ 0 & 2.5 & 0.5 \\ 0 & 0.5 & 2.5 \end{bmatrix}$	$\frac{1}{\sqrt{2}}\begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix}$	$\begin{bmatrix} \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & 1 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \end{bmatrix}$

Method 1: $\boldsymbol{\Sigma}$-diminishing transformations

Lemma A.2 asserts that $\boldsymbol{\Sigma} \geq_L \boldsymbol{\Xi}$ if and only if $\boldsymbol{\Xi} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{W}'\boldsymbol{W}\boldsymbol{\Sigma}^{\frac{1}{2}}$ such that the singular values of $\boldsymbol{W}$ are bounded above by unity. This applies for diagonal matrices; take $\boldsymbol{\Sigma} - \boldsymbol{\Xi} = \left[\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 1.5 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right]$, the difference having positive diagonals and determinant and thus positive definite, so that $\boldsymbol{\Sigma} >_L \boldsymbol{\Xi}$. Taking $\boldsymbol{\Sigma} = \boldsymbol{Q}\boldsymbol{D}_\sigma^2\boldsymbol{Q}'$ with $\boldsymbol{D}_\sigma = \mathrm{Diag}(\sigma_1, \ldots, \sigma_k)$ and $\boldsymbol{Q}$ orthogonal, suppose that commuting justifies the factorizations

$$\boldsymbol{\Xi} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{W}'\boldsymbol{W}\boldsymbol{\Sigma}^{\frac{1}{2}} = \boldsymbol{Q}\boldsymbol{D}_\sigma\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{D}_\omega\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{D}_\sigma\boldsymbol{Q}' = \boldsymbol{Q}\boldsymbol{D}_\sigma\boldsymbol{D}_\omega\boldsymbol{D}_\sigma\boldsymbol{Q}'$$

with $\boldsymbol{D}_\omega$ as the characteristic values of $\boldsymbol{W}'\boldsymbol{W}$. This provides a venue for choosing $\boldsymbol{D}_\omega$ to a desired effect. Specifically, on taking $\boldsymbol{D}_\omega = \mathrm{Diag}(\frac{\sigma_k^2}{\sigma_1^2}, \ldots, \frac{\sigma_k^2}{\sigma_k^2})$ with its characteristic values reverse-ordered, it follows that $\boldsymbol{\Xi} = \boldsymbol{Q}(\sigma_k^2\boldsymbol{I}_k)\boldsymbol{Q}' = \sigma_k^2\boldsymbol{I}_k$; its characteristic values are $[\sigma_k^2, \ldots, \sigma_k^2]$; and $\boldsymbol{\Sigma} \geq_L \sigma_k^2\boldsymbol{I}_k$. Moreover, this choice is maximally conditioned with $\mathtt{Cnd}(\boldsymbol{\Xi}) = 1$.

Method 2: majorizing transformations

Taking $\boldsymbol{\Omega} = \boldsymbol{Q}\boldsymbol{D}_\omega\boldsymbol{Q}' \in \mathbb{S}_k$ in its spectral form, the classical definition of the matrix-valued function $\Phi : \mathbb{S}_k \to \mathbb{S}_k$ is $\Phi(\boldsymbol{\Omega}) \stackrel{\text{def}}{=} \boldsymbol{Q}[\mathrm{Diag}(\Phi(\omega_1, \ldots, \omega_k))]\boldsymbol{Q}'$. Here we consider the following.

Definition 7. Given $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}_a\boldsymbol{P}'$ and $\boldsymbol{B} = \boldsymbol{Q}\boldsymbol{D}_b\boldsymbol{Q}'$ in $\mathbb{S}_k$, take the diagonal matrices into vectors as $\boldsymbol{D}_a \to \boldsymbol{a}' = [a_1, \ldots, a_k]$ and $\boldsymbol{D}_b \to \boldsymbol{b}' = [b_1, \ldots, b_k]$. Then $\boldsymbol{A}$ is said to *majorize $\boldsymbol{B}$ spectrally* on $\mathbb{S}_k$ if and only if $[a_1, \ldots, a_k]$ majorizes $[b_1, \ldots, b_k]$ on $\mathbb{R}^k$ as in Marshall and Olkin (1979), to be denoted as $\boldsymbol{A} \geq_M \boldsymbol{B}$. Specifically, $\boldsymbol{A} \geq_M \boldsymbol{B}$ on $\mathbb{S}_k$ if and only if $\boldsymbol{a} \geq_M \boldsymbol{b}$ on $\mathbb{R}^k$.

To continue, by a majorizing mapping is meant $\boldsymbol{\Omega} \to \boldsymbol{P}\mathrm{Diag}(\tau(\boldsymbol{\omega}))\boldsymbol{P}'$ in which $\boldsymbol{\omega}$ majorizes $\tau(\boldsymbol{\omega}) = \boldsymbol{\theta}$ on $\mathbb{R}^k$. Specifically, it is known that $\boldsymbol{\omega} \geq_M \boldsymbol{\theta}$ if and only if there is a doubly stochastic matrix $\mathbf{B}(k \times k)$ such that $\boldsymbol{\theta} = \mathbf{B}\boldsymbol{\omega}$; see Marshall and Olkin (1979). Accordingly, Method 2 takes $\boldsymbol{\Sigma} \to \boldsymbol{Q}[\mathrm{Diag}(\mathbf{B}\boldsymbol{\delta})]\boldsymbol{Q}'$ with $\boldsymbol{\delta} = [\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2]'$ as the characteristic values of $\boldsymbol{\Sigma}$.

Case studies

To illustrate, begin with $\boldsymbol{\Sigma} = \boldsymbol{Q}\boldsymbol{D}_\sigma^2\boldsymbol{Q}'$ with factors as listed in Table 2 such that $\mathtt{Ch}(\boldsymbol{\Sigma}) = [5, 3, 1]$ and $\boldsymbol{D}_\sigma^2 = \mathrm{Diag}(5, 3, 1)$. Now taking $\boldsymbol{D}_\omega = \mathrm{Diag}(\frac{1}{5}, \frac{1}{3}, \frac{1}{1})$ with its characteristic values reverse-ordered, Method 1 yields $\boldsymbol{\Xi} = 1 \cdot \boldsymbol{I}_3$ as in Table 2. For the majorizing transformation of Definition 7, take the doubly stochastic matrix $\mathbf{B}$ to be as in Table 2. Then $\mathbf{B}[5, 3, 1]' = [4, 3, 2]'$, namely, the majorized characteristic values to be used in constructing $\boldsymbol{\Omega} = \boldsymbol{Q}\mathrm{Diag}(4, 3, 2)\boldsymbol{Q}'$, as listed in Table 2.

Returning to the sample matrix $\boldsymbol{S} = \boldsymbol{E}'\boldsymbol{E}/(n - r)$, the foregoing constructions apply on treating $\boldsymbol{\Sigma}$ instead as $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{S}$. Both Methods 1 and 2 reduce scatter among the characteristic values of $\widehat{\boldsymbol{\Sigma}} \to \widehat{\boldsymbol{\Xi}}$

and $\widehat{\mathbf{\Sigma}} \to \widehat{\mathbf{\Omega}}$. Observe further that Methods 1 and 2 are intimately connected. Given the target values $[4, 3, 2]$ as having been stipulated apart from majorization, Method 1 nonetheless gives $\widehat{\mathbf{\Sigma}} \to \widehat{\mathbf{\Omega}}$ as in Table 2 on taking $\boldsymbol{D}_{\boldsymbol{\omega}} = \mathrm{Diag}(\frac{4}{5}, \frac{3}{3}, \frac{2}{1})$.

Recall that $\mathbf{\Sigma}^*$ in Table 1 seeks to shrink the characteristic values towards their arithmetic mean, but with undefined risk for $\mathsf{S}^{\alpha}_{n,k}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \mathbf{\Sigma})$ with $(0 < \alpha < 2)$. Clearly this objective is achieved exactly using Method 1, on identifying 3 as the mean of $[5, 3, 1]$ and taking $\boldsymbol{D}_{\boldsymbol{\omega}} = [\frac{3}{5}, \frac{3}{3}, \frac{3}{1}]$. This does in fact reduce the characteristic values precisely to their arithmetic mean, in which case $\mathbf{\Xi}$ of Table 2 becomes $\mathbf{\Xi} = 3\boldsymbol{I}_3$.

In addition, it is relevant to discern the conditioning of $\boldsymbol{S}$ in view of its scattered characteristic values and the vaunted heavy tails of Cauchy distributions. Despite these, the following holds from the scale-invariance of $\mathtt{Cnd}(\boldsymbol{S}) = \lambda_1/\lambda_k$.

Remark 4. Given $\mathcal{L}(\boldsymbol{Y}) \in \mathsf{S}^{\alpha}_{n,k}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \mathbf{\Sigma})$, the sample condition number $\mathtt{Cnd}(\boldsymbol{S}) = \lambda_1/\lambda_k$, as the ratio of its characteristic values, has properties identical to those in sampling from $\mathrm{N}_{n,k}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \mathbf{\Sigma})$.

6. Topics: testing hypotheses

Normal-theory tests regarding location and scale parameters are myriad and far reaching, covering many pages in contemporary text books. Our objectives are elementary: For selected tests, to demonstrate the remarkable feature that they remain valid for every $\{\mathcal{L}(\boldsymbol{Y}) \in \mathsf{S}^{\alpha}_{n,k}(\mathbf{0}, \boldsymbol{I}_n \otimes \mathbf{\Sigma}); 0 < \alpha < 2\}$, and to identify the key point to be checked by users for the many procedures not covered here for lack of space. Accordingly, it remains to reexamine hypothesis testing under $\mathsf{S}\alpha\mathsf{S}$ errors in $\mathbb{F}_{n\times k}$. We first consider tests regarding $\mathfrak{B}$, then for selected functions of $\mathbf{\Sigma}$.

6.1 Location parameters

Tests regarding $\mathtt{M}$ in $\mathcal{L}(\boldsymbol{Y}) \in \mathrm{N}_{n,k}(\mathtt{M}, \boldsymbol{I}_n \otimes \mathbf{\Sigma})$ entail $(k \times k)$ matrices, $\mathbf{H}$ due to hypothesis and $\mathbf{E}$ due to error, both as quadratic and bilinear matrix forms of type $\boldsymbol{Y}'\boldsymbol{A}\boldsymbol{Y}$. The Invariance Principle stipulates designated transformation groups, and it serves to extract the essentials of a problem through the symmetries of group operations. Among invariant tests, the principal designees are identified in Table 3, where rejection rules pertain to normal-theory critical values. These have been developed and their properties studied in detail over the past century and beyond. There is a plethora of tests, as none is uniformly most powerful.

Guided by invariance, these tests entail roots of the determinantal equation $|\mathbf{H} - \lambda\mathbf{E}| = 0$, equivalently, the values $\mathtt{Ch}(\mathbf{H}\mathbf{E}^{-1}) = [\lambda_1, \ldots, \lambda_s]$ with s as the rank of $\mathbf{H}$. Requiring neither moments nor the need to demonstrate that likelihood estimators are maximal, the Invariance Principle nonetheless applies for distributions in $\mathsf{S}^{\alpha}_{n,k}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \mathbf{\Sigma})$.

Invariant tests for $H_0 : \mathfrak{B} = \mathfrak{B}_0$ vs. $H_1 : \mathfrak{B} \neq \mathfrak{B}_0$ utilize $\mathbf{H} = (\widehat{\mathfrak{B}} - \mathfrak{B}_0)'\boldsymbol{X}'\boldsymbol{X}(\widehat{\mathfrak{B}} - \mathfrak{B}_0)$ due to hypothesis, and $\mathbf{E} = \boldsymbol{E}'\boldsymbol{E}$ due to error, the latter comprising the residual sums of squares and products as in Theorem 5. Observe that $\mathbf{H}$ is the squared Mahalanobis (1936) metric for $(\widehat{\mathfrak{B}} - \mathfrak{B}_0)$ on $\mathbb{F}_{r\times k}$ as in Theorem 6(v). As moments are not invoked, these apply verbatim for $\mathsf{S}^{\alpha}_{n,k}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \mathbf{\Sigma})$. Numerous variations have been studied, to include tests on rows and columns of $\mathfrak{B}$ and, more generally, $H_0 : \boldsymbol{A}'\mathfrak{B}\boldsymbol{B} = \mathbf{\Delta}$ with $\{\boldsymbol{A}, \boldsymbol{B}, \mathbf{\Delta}\}$ specified, all entailing prescribed versions of $\mathbf{H}$ and $\mathbf{E}$. It is essential in practice that the vast compendium of normal-theory linear tests applies verbatim for distributions in $\mathsf{S}^{\alpha}_{n,k}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \mathbf{\Sigma})$ as follows, despite excessive tails and the dearth of available

Table 3. The principal test procedures in multivariate linear inference, having matrix $\mathbf{H}$ of rank s due to hypothesis and $\mathbf{E}$ due to error, with $\mathrm{Ch}(\mathbf{HE}^{-1}) = [\lambda_1, \ldots, \lambda_s]$, together with rejection rules for normal-theory tests at level α.

Item	Description	Rule	Designation
Λ	$\prod_{i=1}^{s} \frac{1}{1+\lambda_i} = \frac{\lvert\mathbf{E}\rvert}{\lvert\mathbf{H}+\mathbf{E}\rvert}$	$\Lambda < \Lambda_\alpha$	Wilks
U	$\sum_{i=1}^{s} \lambda_i = \mathrm{tr}(\mathbf{HE}^{-1})$	$U > U_\alpha$	Lawley–Hotelling
V	$\sum_{i=1}^{s} \frac{\lambda_i}{1+\lambda_i} = \mathrm{tr}\,[\mathbf{H}(\mathbf{H}+\mathbf{E})^{-1}]$	$V > V_\alpha$	Pillai
θ	$\frac{\lambda_1}{1+\lambda_1}$	$\theta > \theta_\alpha$	Roy

moments.

Theorem 7. *Given $\mathcal{L}(\boldsymbol{Y}) \in \mathsf{S}^{\alpha}_{n,k}(\boldsymbol{X}\boldsymbol{\mathfrak{B}}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ together with invariant tests utilizing $\mathbf{H}$ and $\mathbf{E}$ through $\mathrm{Ch}(\mathbf{HE}^{-1})$.*

(i) For all $\mathcal{L}(\boldsymbol{Y}) \in \mathsf{S}^{\alpha}_{n,k}(\boldsymbol{X}\boldsymbol{\mathfrak{B}}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$, all such procedures are identical in level and power to those in sampling from $\mathrm{N}_{n,k}(\boldsymbol{X}\boldsymbol{\mathfrak{B}}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$;

(ii) Accordingly, exact critical values from special aid tables and algorithms, as well as a surfeit of approximations, apply for all $\mathcal{L}(\boldsymbol{Y}) \in \mathsf{S}^{\alpha}_{n,k}(\boldsymbol{X}\boldsymbol{\mathfrak{B}}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$.

Proof. Observe that $\boldsymbol{Y} \to c\boldsymbol{Y}$ for $c \neq 0$ generates $\boldsymbol{Y} \to (c^2\mathbf{H})(c^2\mathbf{E})^{-1} = \mathbf{HE}^{-1}$, so that the conclusions now follow from Theorem 5(v). ∎

6.2 Scale parameters

The second-moment matrix $\boldsymbol{\Sigma}$ supports essential concepts for dependence. For $\mathcal{L}(\boldsymbol{Y}) = \mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, these include simple, multiple, partial and canonical correlation parameters, as well as the independence of subsets $\boldsymbol{Y}' = [\boldsymbol{Y}_1', \boldsymbol{Y}_2']$ through the diagonal structure $\boldsymbol{\Sigma} = \mathrm{Diag}(\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22})$. Those same parameters in $\{\mathcal{L}(\boldsymbol{Y}) \in \mathsf{S}^{\alpha}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}); 0 < \alpha < 2\}$ nonetheless continue to quantify the stochastic affinity between pairs, to be called *α-association*. Specifically, with $\{\rho = \sigma_{12}/(\sigma_{11}\sigma_{22})^{\frac{1}{2}}\}$, the pair (Y_1, Y_2) becomes increasingly indistinguishable, hence more α-associated, in the sense that

$$\{P(|Y_1 - Y_2| \leq c) \uparrow 1 \text{ as } \rho \uparrow 1; \text{ for every } c > 0\}.$$

For further details, Lemma A.3 pertains also to corresponding analogs of multiple and canonical correlations. In addition, that $\boldsymbol{\Sigma} = \mathrm{Diag}(\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22})$ under $\{\mathcal{L}(\boldsymbol{Y}) \in \mathsf{S}^{\alpha}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}); 0 < \alpha < 2\}$, designates the subsets to be *α-unassociated* rather than independent; but with second moments in $\mathsf{S}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, to be dependent but uncorrelated as in time series and econometrics.

Our objectives again are to illustrate the validity of selected tests in view of the many procedures not covered. Specifically, the canonical correlations and conventional tests for these as in Table 4, where a typical row of $\boldsymbol{Y} \in \mathbb{F}_{n\times k}$ is partitioned as $\{[\boldsymbol{Y}_1'(1\times p), \boldsymbol{Y}_2'(1\times q)]; p+q=k\}$, and rejection rules pertain to normal-theory critical values. The following results are fundamental.

Table 4. The principal procedures for testing $H_0 : \boldsymbol{\Sigma} = \mathrm{Diag}(\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22})$ in terms of the squared sample canonical coefficients $[r_1^2, r_2^2, \ldots, r_s^2]$ between $\boldsymbol{Y}_1(p \times 1)$ and $\boldsymbol{Y}_2(q \times 1)$ with $s = \min(p, q)$, together with rejection rules for normal-theory tests at level α.

Item	Description	Rule	Designation
Λ	$\prod_{i=1}^{s}(1 - r_i^2)$	$\Lambda < \Lambda_\alpha$	Wilks
U	$\sum_{i=1}^{s} \frac{r_i^2}{1-r_i^2}$	$U > U_\alpha$	Lawley–Hotelling
V	$\sum_{i=1}^{s} r_i^2$	$V > V_\alpha$	Pillai
θ	r_1^2	$\theta > \theta_\alpha$	Roy

Theorem 8. *Take* $\{\mathcal{L}(\boldsymbol{Y}) \in \mathrm{S}_{n,k}^{\alpha}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}); 0 < \alpha < 2\}$; *let* $\boldsymbol{S} = \boldsymbol{E}'\boldsymbol{E}/(n - r)$; *and consider the simple, multiple, and canonical* α*-association parameters* $\{\rho_{ij},\ R(Y_1, \boldsymbol{Y}_2),\ \rho_c(\boldsymbol{Y}_1, \boldsymbol{Y}_2)\}$ *of Lemma A.3. Then*

(i) Normal-theory tests for $\{\rho_{ij},\ R(Y_1, \boldsymbol{Y}_2),\ \rho_c(\boldsymbol{Y}_1, \boldsymbol{Y}_2)\}$ *and their properties, when based on* $\boldsymbol{S}$ *from* $\mathrm{S}_{n,k}^{\alpha}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$, *are identical to those from* $\mathrm{N}_{n,k}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$;

(ii) In particular, the Table 4 procedures, to include exact and approximate critical values, each has level and power identical to that if sampled from $\mathrm{N}_{n,k}(\boldsymbol{X}\mathfrak{B}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$.

Proof. The quantities at issue are scale-invariant as $\boldsymbol{Y} \to c\boldsymbol{Y}$ for $c \neq 0$. For example, the sample canonical coefficients are the singular values of $\boldsymbol{S}_{11}^{-\frac{1}{2}} \boldsymbol{S}_{12} \boldsymbol{S}_{22}^{-\frac{1}{2}}$ from $\boldsymbol{S} = [\boldsymbol{S}_{ij}]$, clearly invariant as $\boldsymbol{Y} \to c\boldsymbol{Y}$ and $\boldsymbol{S} \to c^2\boldsymbol{S}$. The assertions now follow from Theorem 5(v). ■

Remark 5. For *SaS* distributions with $\alpha \in (0, 2)$, it is remarkable not only that the probabilistic notions of Lemma A.3 may supplant the classical moment-based notions of dependence, but that conventional normal-theory properties should continue to hold.

7. Summary and discussion

As noted, many essentials of the present study are genuinely nonparametric, in applying for all or portions of the SαS distributions and thus distribution-free within that class. Especially in hypothesis testing, this has the ultimate advantage to obviate the need to derive distributions of test statistics from largely unknown stable density functions and, from them, to devise algorithms and special aid tables for finding cutoff rules, apart from the serious limitations of simulated values. In addition, Theorem 6(iv) addresses the matter of design efficiency beyond the confines of normal-theory inference, in showing for the design $\boldsymbol{X}$ dominating $\boldsymbol{Z}$ that $\widehat{\mathfrak{B}}(\boldsymbol{X})$ is more peaked about $\mathfrak{B}$ than $\widehat{\mathfrak{B}}(\boldsymbol{Z})$.

The present study offers further insight into the role of SαS distributions in practice. Normal-theory procedures long have been applied as large-sample approximations in distributions attracted to Gaussian limits. Specifically, Berry–Esséen bounds on rates of convergence to Gaussian limits are studied with applications in Jensen (1973, 1977), with special reference to linear models in Jensen and Mayer (1975) and Jensen et al. (1975).

Findings cited here complement those studies for distributions attracted to heavy tailed SαS laws, as well as rates of convergence to stable limits, as cited in Section 2.2. By showing that many standard data-analytic procedures carry over in essence under significantly weakened assumptions, this study gives further credence to the widely and correctly held view that Gauss-Markov estimation and normal theory inferences extend considerably beyond the confines of the classical theory and its needless reliance on moments to fourth order.

A. Appendix

A.1 Duality: imbedding $\mathbb{F}_{n\times k} \to \mathbb{R}^{nk}$

Note first that shifted distributions $\mathcal{L}(\boldsymbol{Y}) \in \mathsf{S}_{n,k}(\boldsymbol{\Theta}, \boldsymbol{\Xi})$ having elliptical contours derive from spherical distributions through affine transformations. These in turn have a density in $\mathbb{F}_{n\times k}$ if and only if $\boldsymbol{\Xi}$ is of full rank; otherwise $\mathcal{L}(\boldsymbol{Z})$ is concentrated in a subspace of $\mathbb{F}_{n\times k}$ of dimension equal to the rank of $\boldsymbol{\Xi}$. Specifically, given $\mathcal{L}(\boldsymbol{Z}) \in \mathsf{S}_{n,k}(\boldsymbol{0}, \boldsymbol{I}_{nk})$; let $\boldsymbol{Y} = \boldsymbol{A}'\boldsymbol{Z}\boldsymbol{B} + \boldsymbol{\Theta}$ of orders $\boldsymbol{A}'(r \times n)$ and $\boldsymbol{B}(k \times s)$; then $\mathcal{L}(\boldsymbol{Y}) \in \mathsf{S}_{r,s}(\boldsymbol{\Theta}, \boldsymbol{\Gamma} \otimes \boldsymbol{\Sigma})$ on $\mathbb{F}_{r\times s}$, with $\boldsymbol{\Gamma} = \boldsymbol{A}'\boldsymbol{A}$ and $\boldsymbol{\Sigma} = \boldsymbol{B}'\boldsymbol{B}$. See Jensen and Good (1981).

Details in regard to the mapping $\mathtt{J} : \mathbb{F}_{n\times k} \to \mathbb{R}^{nk}$ of Definition 2 are as follow, to include that $\mathbb{R}^{nk}$ and $\mathbb{F}_{n\times k}$ are isomorphic. This in turn supports the transfer to $\mathbb{F}_{n\times k}$ of widely studied symmetric distributions on $\mathbb{R}^{nk}$.

Lemma 1. *Consider $\mathcal{L}(\boldsymbol{x}) \in \mathsf{S}_{nk}(\boldsymbol{0}, \boldsymbol{I}_{nk})$ on $\mathbb{R}^{nk}$ and $\mathcal{L}(\boldsymbol{Y}) \in \mathsf{S}_{n,k}(\boldsymbol{0}, \boldsymbol{I}_{nk})$ on $\mathbb{F}_{n\times k}$, both symmetric about their respective origins with scale parameters $\boldsymbol{I}_{nk}$.*

(i) Distributions on $\mathbb{F}_{n\times k}$ and $\mathbb{R}^{nk}$ correspond through $\boldsymbol{Y} = \mathtt{J}^{-1}(\boldsymbol{C}'\boldsymbol{x} + \gamma)$ and conversely, with $\boldsymbol{C}$ of order $(nk \times nk)$ and $\gamma \in \mathbb{R}^{nk}$;

(ii) Under first moments $\mathtt{E}(\boldsymbol{Y}) = \mathtt{J}^{-1}(\gamma) = \boldsymbol{\Theta} \in \mathbb{F}_{n\times k}$;

(iii) Under second moments $\mathtt{V}(\boldsymbol{Y}) \overset{def}{=} \mathtt{V}(\boldsymbol{Y}_{\mathtt{J}}) = \boldsymbol{C}'\boldsymbol{C} = \boldsymbol{\Xi}$ of order $(nk \times nk)$;

(iv) In short, $\{\mathcal{L}(\boldsymbol{Y}) \in \mathsf{S}_{n,k}(\boldsymbol{\Theta}, \boldsymbol{\Xi})$ on $\mathbb{F}_{n\times k}\}$ if and only if $\{\mathcal{L}(\boldsymbol{x}) \in \mathsf{S}_{nk}(\gamma, \boldsymbol{\Xi})$ on $\mathbb{R}^{nk}\}$ with $\boldsymbol{x} = \mathtt{J}(\boldsymbol{Y})$ and $\boldsymbol{Y} = \mathtt{J}^{-1}(\boldsymbol{x})$;

(v) Sets convex and symmetric about their respective origins, namely, $\mathsf{C}^0_{nk} \subset \mathbb{R}^{nk}$ and $\mathsf{C}^0_{n,k} \subset \mathbb{F}_{n\times k}$ as in Definition 4, correspond one-to-one;

(vi) The measurable subsets $\mathfrak{B}_{n,k}$ of $\mathbb{F}_{n\times k}$ and $\mathfrak{B}_{nk}$ of $\mathbb{R}^{nk}$ correspond one-to-one;

(vii) The c-spheres $\mathsf{S}_{n,k}(c) \subset \mathbb{F}_{n\times k}$ and $\mathsf{S}_{nk}(c) \subset \mathbb{R}^{nk}$, together with related c-balls, correspond one-to-one.

Proof. Conclusions (i)-(iv) attribute to Jensen and Good (1981). Conclusions (v)-(vii) follow on looking at images $\mathtt{J}(\cdot) = I(\cdot)$ in $\mathbb{R}^{nk}$ and preimages $\mathtt{J}^{-1}I(\cdot)$ in $\mathbb{F}_{n\times k}$. In particular, clearly the spheres $\mathsf{S}_{n,k}(c) = \{\boldsymbol{Y} : \mathrm{tr}(\boldsymbol{Y}'\boldsymbol{Y}) = c^2\}$ in $\mathbb{F}_{n\times k}$ and $\mathsf{S}_{nk}(c) = \{\boldsymbol{Y}_{\mathtt{J}} : \boldsymbol{Y}_{\mathtt{J}}'\boldsymbol{Y}_{\mathtt{J}} = c^2\}$ in $\mathbb{R}^{nk}$ are so related, as are the corresponding balls of radius c. ∎

A.2 Example

An alternative view is seen on partitioning $\{\boldsymbol{Y} = \boldsymbol{X}\mathfrak{B} + \mathcal{E}\}$ in $\mathbb{F}_{n\times k}$ as

$$[\mathtt{Y}_1, \dots, \mathtt{Y}_k] = \boldsymbol{X}[\beta_1, \dots, \beta_k] + [\mathcal{E}_1, \dots, \mathcal{E}_k]$$

and imbedding column wise as $\mathtt{K}(\boldsymbol{Y}) = [\mathtt{Y}_1', \mathtt{Y}_2', \dots, \mathtt{Y}_k']' = \mathtt{Y}_0 \in \mathbb{R}^{nk}$, and $\mathtt{K}(\mathfrak{B}) = [\beta_1', \beta_2', \dots, \beta_k']' = \beta_0 \in \mathbb{R}^{rk}$. Next taking $\boldsymbol{X}_0 = \mathrm{Diag}(\boldsymbol{X}, \boldsymbol{X}, \dots \boldsymbol{X})$ having k blocks, it follows that the model on $\mathbb{F}_{n\times k}$ transforms into $\mathtt{K}(\boldsymbol{Y}) = \boldsymbol{X}_0\, \mathtt{K}(\mathfrak{B}) + \mathtt{K}(\mathcal{E})$ on $\mathbb{R}^{nk}$, namely

$$\begin{bmatrix} \mathtt{Y}_1 \\ \mathtt{Y}_2 \\ \dots \\ \mathtt{Y}_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{X} & \mathbf{0} & \dots, & \mathbf{0} \\ \mathbf{0}, & \boldsymbol{X} & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{X} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \\ \dots \\ \mathcal{E}_k \end{bmatrix}.$$

Equivalence holds as follows.

- The OLS solutions $\widehat{\beta_0}$ in $\mathbb{R}^{nk}$ and $\widehat{\mathfrak{B}}$ in $\mathbb{F}_{n\times k}$ correspond one-to-one as $\widehat{\beta_0} = \mathtt{K}(\widehat{\mathfrak{B}})$ and $\widehat{\mathfrak{B}} = \mathtt{K}^{-1}(\widehat{\beta_0})$.
- $\mathcal{L}(\widehat{\mathfrak{B}}) \in \mathsf{S}_{n,k}(\mathfrak{B}, (\boldsymbol{X}'\boldsymbol{X})^{-1} \otimes \boldsymbol{\Sigma})$ in $\mathbb{F}_{n\times k}$ if and only if $\mathcal{L}(\widehat{\beta_0}) \in \mathsf{S}_{nk}(\beta_0, (\boldsymbol{X}'\boldsymbol{X})^{-1} \otimes \boldsymbol{\Sigma})$ in $\mathbb{R}^{nk}$.

A.3 $\boldsymbol{\Sigma}$-diminishing transformations

We seek to diminish $\boldsymbol{\Sigma}$ in the sense of the ordering of Definition 1(i). The following is one such venue.

Lemma 2. *The transformation $T : \boldsymbol{\Sigma} \to \boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}'$ is $\boldsymbol{\Sigma}$-diminishing as in Definition 1(i), if and only if $\boldsymbol{T} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{W}'\boldsymbol{\Sigma}^{-\frac{1}{2}}$, where $\boldsymbol{W} \in F_{k\times k}$ is a matrix whose singular values are bounded above by unity. In particular, the class $\tau(\boldsymbol{\Sigma})$ consisting of these is given by*

$$\tau(\boldsymbol{\Sigma}) = \{\boldsymbol{T} : \boldsymbol{T} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{W}'\boldsymbol{\Sigma}^{-\frac{1}{2}};\ \sigma(\boldsymbol{W}) \le 1\}.$$

Proof. See Theorem 1 of Jensen and Ramirez (1990). ∎

A.4 Degrees of association

The lack of second moments vitiates the classical simple, multiple, and canonical correlations in SαS distributions, the elements of $\boldsymbol{\Sigma}$ serving instead as parameters of scale. As to whether $\{\rho_{ij} = \sigma_{ij}/(\sigma_{ii}\sigma_{jj})^{\frac{1}{2}}\}$ again might serve to quantify affinity between random variables, a definitive answer is supplied in the following.

Lemma 3. *Consider $\mathcal{L}(\boldsymbol{Z}) \in \mathsf{S}_n^{\alpha}(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ together with the simple, multiple, and canonical correlation type quantities based on $\boldsymbol{\Sigma}$. Partition $\boldsymbol{Z} \in \mathbb{R}^n$ variously as $\boldsymbol{Z}' = [Z_1, \boldsymbol{Z}_2']$ and $\boldsymbol{Z}' = [\boldsymbol{Z}_1', \boldsymbol{Z}_2']$ of orders $\{\boldsymbol{Z}_1'(1 \times p), \boldsymbol{Z}_2'(1 \times q);\ p + q = n\}$. Then for each $\mathcal{L}(\boldsymbol{Z}) \in \mathsf{S}_n^{\alpha}(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ with $(0 < \alpha < 2)$, the following properties hold.*

(i) The entities $\{\rho_{ij} = \sigma_{ij}/(\sigma_{ii}\sigma_{jj})^{\frac{1}{2}}\}$ serve to quantify the stochastic affinity *between (Z_i, Z_j), the degrees of affinity increasing with ρ_{ij};*

(ii) Parameters of type $R(Z_1, \boldsymbol{b}'\boldsymbol{Z}_2)$, as analogs of multiple correlations, serve to quantify affinity *between $(Z_1, \boldsymbol{b}'\boldsymbol{Z}_2)$, the degree increasing with $R(Z_1, \boldsymbol{b}'\boldsymbol{Z}_2)$;*

(iii) *Values of* $\rho(\boldsymbol{a}'\boldsymbol{Z}_1, \boldsymbol{b}'\boldsymbol{Z}_2)$, *as analogs to canonical correlations, serve to quantify* affinity *between* $(\boldsymbol{a}'\boldsymbol{Z}_1, \boldsymbol{b}'\boldsymbol{Z}_2)$, *the degree increasing with* $\rho(\boldsymbol{a}'\boldsymbol{Z}_1, \boldsymbol{b}'\boldsymbol{Z}_2)$.

Proof. For conclusion (i) take (Z_1, Z_2) with $\boldsymbol{\delta} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right]$. Let $U = (Z_1 - Z_2)$; then $\mathcal{L}(U)$ clearly is symmetric about 0 with scale parameter $\sigma_U = 2(1-\rho)$. A result of Birnbaum (1946) shows for each $c > 0$ that $P(U \in (-c, c))$ is decreasing in σ_U thus increasing in ρ. Equivalently, $P(|Z_1 - Z_2| \le c) \uparrow 1$ as $\rho \uparrow 1$, identifying the sense in which (Z_1, Z_2) become increasingly indistinguishable with increasing values of ρ. Conclusions (ii) and (iii) follow on identifying (Z_i, Z_j) first with the pair $(Z_1, \boldsymbol{b}'\boldsymbol{Z}_2)$, then with $(\boldsymbol{a}'\boldsymbol{Z}_1, \boldsymbol{b}'\boldsymbol{Z}_2)$ as the canonical variables. ∎

Definition 8. For $\mathcal{L}(\boldsymbol{Z}) \in \mathtt{S}_n^{\alpha}(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ with $(0 < \alpha < 2)$, the analogs of simple, multiple, and canonical correlations are called α-*association* parameters.

A.5 Independence versus non-association

For $\mathcal{L}(\boldsymbol{Y}) \in \mathtt{S}_{n,k}(\mathbf{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ the structure $\mathtt{V}(\boldsymbol{Y}) = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}$ asserts under second moments that the rows of $\boldsymbol{Y}$ are mutually uncorrelated. Suppose that they are also independent. Maxwell (1860) showed for $\mathcal{L}(\boldsymbol{Z}) \in \mathtt{S}_n(\mathbf{0}, \sigma^2\boldsymbol{I}_n)$ that elements of the spherical vector $\boldsymbol{Z} = [Z_1, \ldots, Z_n]'$ are mutually uncorrelated and independent if and only if Gaussian. It is relevant to ask whether this might carry over to include the row vectors of $\mathcal{L}(\boldsymbol{Y}) \in \mathtt{S}_{n,k}(\mathbf{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$. An affirmative answer follows.

Lemma 4. *Given* $\mathcal{L}(\boldsymbol{Y}) \in \mathtt{S}_{n,k}(\mathbf{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$ *having second moments, so that the rows of* $\boldsymbol{Y}$ *are mutually uncorrelated. Then they are mutually independent if and only if Gaussian.*

Proof. Temporarily fix $\boldsymbol{T} = [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n]' \in \mathbb{F}_{n\times k}$, and let $\boldsymbol{U}' = [U_1, \ldots, U_n]$ such that $\{U_i = \boldsymbol{y}_i'\boldsymbol{t}_i; 1 \le i \le n\}$ with $[\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]$ as the columns of $\boldsymbol{Y}'$. Here $\{\mathtt{Var}(U_i|\boldsymbol{t}_i) = \boldsymbol{t}_i'\boldsymbol{\Sigma}\boldsymbol{t}_i = c_i^2; 1 \le i \le n\}$. Consider, as the $\mathtt{Chf}$ for $\boldsymbol{U}$ with arguments $\boldsymbol{s} = [s_1, \ldots, s_n]$ and $\boldsymbol{T}$ fixed, the function

$$\phi_{\boldsymbol{U}}(\boldsymbol{s}|\,\boldsymbol{T}) = \mathtt{E}\left[e^{i(s_1U_1+\ldots+s_nU_n)}\right] = \mathtt{E}\left[e^{i(s_1\boldsymbol{t}_1'\boldsymbol{y}_1+\ldots+s_n\boldsymbol{t}_n'\boldsymbol{y}_n)}\right].$$

Then $\mathcal{L}([U_1/c_1, \ldots, U_n/c_n]|\,\boldsymbol{T})$ is spherical on $\mathbb{R}^n$ and, by Maxwell's (1860) result, are independent if and only if Gaussian, in which case $[U_1, \ldots, U_n]$ are themselves independent Gaussian, and the factorization

$$\phi_{\boldsymbol{U}}(\boldsymbol{s}|\,\boldsymbol{T}) = \mathtt{E}\left[e^{i(s_1\boldsymbol{t}_1'\boldsymbol{y}_1+\ldots+s_n\boldsymbol{t}_n'\boldsymbol{y}_n)}\right] = \Pi_{i=1}^{n}\,\mathtt{E}\left[e^{is_i\boldsymbol{t}_i'\boldsymbol{y}_i}\right]$$

holds. On extending the device of Cramér and Wold (1936), we now fix $[s_1, \ldots, s_n]$ at unity and recognize the resulting expression as the joint $\mathtt{Chf}$ for the n rows of $\boldsymbol{Y}$ with arguments $[\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n]$, now independent by factorization. ∎

Remark 6. There is confusion about this in the literature. With $\boldsymbol{X}' = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$, while stating that $\mathcal{L}(\boldsymbol{X}) \in S_{n,k}(\boldsymbol{M}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$, Leung and Ng (2004) assumed the columns of $\boldsymbol{X}'$ to be a random sample, thus independent, from $\mathcal{L}(\boldsymbol{x}_i) \in \mathtt{S}_k(\boldsymbol{m}_i, \boldsymbol{\Sigma})$. In view of Lemma A.4, this is tantamount to assuming instead that $\mathcal{L}(\boldsymbol{X}) = \mathrm{N}_{n,k}(\boldsymbol{M}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma})$.

Acknowledgement. Thanks are extended to Professor Donald E. Ramirez for helpful exchanges on these topics.

References

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York.

Barbosa, E. G. and Dorea, C. C. (2009). A note on the Lindeberg condition for convergence to stable laws in mallows distance. *Bernoulli*, **15**, 922–924.

Berk, R. and Hwang, J. T. (1989). Optimality of the least squares estimator. *Journal of Multivariate Analysis*, **30**, 245–254.

Birnbaum, Z. (1948). On random variables with comparable peakedness. *The Annals of Mathematical Statistics*, **19**, 76–81.

Cramér, H. and Wold, H. (1936). Some theorems on distribution functions. *Journal of the London Mathematical Society*, **1**, 290–294.

Das Gupta, S., Eaton, M. L., Olkin, I., Perlman, M., Savage, L. J., and Sobel, M. (1972). Inequalities on the probability content of convex regions for elliptically contoured distributions. *In Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, volume 2. 241–265.

De Haan, L. and Resnick, S. I. (1979). Derivatives of regularly varying functions in rd and domains of attraction of stable distributions. *Stochastic Processes and their Applications*, **8**, 349–355.

Dharmadhikari, S. and Joag-Dev, K. (1988). *Unimodality, convexity, and applications*. Elsevier, New York.

Fang, K.-T. and Anderson, T. W. (1990). *Statistical Inference in Elliptically Contoured and Related Distributions*. Allerton Press, New York.

Fang, K.-T. and Zhang, Y.-T. (1990). *Generalized Multivariate Analysis*, volume 19. Science Press and Springer-Verlag, Beijing/Berlin.

Fefferman, C., Jodeit, M., and Perlman, M. D. (1972). A spherical surface measure inequality for convex sets. *In Proceedings of the American Mathematical Society*, volume 33. 114–119.

Gupta, A. K. and Varga, T. (1993). *Elliptically Contoured Models in Statistics*. Kluwer Academic, Dordrecht.

Hartman, P. and Wintner, A. (1940). On the spherical approach to the normal distribution law. *American Journal of Mathematics*, **62**, 759–779.

Jensen, D. R. (1973). Monotone bounds on the chi-squared approximation to the distribution of Pearson's X^2 statistics. *Australian Journal of Statistics*, **15**, 65–70.

Jensen, D. R. (1977). On approximating the distributions of Friedman's χ_r^2 and related statistics. *Metrika*, **24**, 75–85.

Jensen, D. R. (1984). Invariant ordering and order preservation. *In Inequalities in Statistics and Probability*. Institute of Mathematical Statistics, 26–34.

Jensen, D. R. (2017). Limits and inferences for alpha-stable variables. *Biostatistics and Biometrics Open Access Journal*, **4**, 1–4.

Jensen, D. R. (2018). Linear inference under alpha-stable errors. *Biometrics and Biostatistics International Journal*, **7**, 205–210.

Jensen, D. R. and Good, I. J. (1981). Invariant distributions associated with matrix laws under structural symmetry. *Journal of the Royal Statistical Society: Series B*, **43**, 327–332.

Jensen, D. R. and Mayer, L. S. (1975). Normal-theory approximations to tests for linear hypotheses. *The Annals of Statistics*, 429–444.

Jensen, D. R., Mayer, L. S., and Mayers, R. (1975). Optimal designs and large-sample tests for linear hypotheses. *Biometrika*, **62**, 71–78.

Jensen, D. R. and Ramirez, D. E. (1990). Dispersion-diminishing transformations. *Communications in Statistics – Theory and Methods*, **19**, 3259–3266.

Joarder, A. and Singh, S. (2001). Estimation of the trace of the scaled covariance matrix of a multivariate t-model using a known information. *Metrika*, **54**, 53–58.

Joarder, A. H. and Ahmed, S. E. (1996). Estimation of the characteristic roots of the scale matrix. *Metrika*, **44**, 259–267.

Joarder, A. H. and Ali, M. M. (1997). Estimation of the scale matrix of a multivariate t-model under entropy loss. *Metrika*, **46**, 21–32.

Klett, J. (1972). *Applied Multivariate Analysis*. Holt, Rinehart & Winston, New York.

Kuelbs, J. and Mandrekar, V. (1974). Domains of attraction of stable measures on a Hilbert space. *Studia Mathematica*, **50**, 149–162.

Leung, P. L. and Ng, F. Y. (2004). Improved estimation of a covariance matrix in an elliptically contoured matrix distribution. *Journal of Multivariate Analysis*, **88**, 131–137.

Löwner, K. (1934). Über monotone Matrixfunktionen. *Mathematische Zeitschrift*, **38**, 177–216.

Lukacs, E. and Laha, R. G. (1964). *Applications of Characteristic Functions*. Hafner Press, New York.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *In Proceedings of the National Institute of Science of India*, volume 2. 49–55.

Marshall, A. W., Olkin, I., and Arnold, B. C. (1979). *Inequalities: Theory of Majorization and its Applications*. Springer, Academic Press.

Maxwell, J. C. (1860). Illustrations of the dynamical theory of gases. Part 1. On the motions and collisions of perfectly elastic spheres. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **19**, 19–32.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, New York.

Paulauskas, V. (2009). On the rate of convergence to bivariate stable laws. *Lithuanian Mathematical Journal*, **49**, 426.

Rachev, S. and Rüschendorf, L. (1992). A new ideal metric with applications to multivariate stable limit theorems. *Probability Theory and Related Fields*, **94**, 163–187.

Rao, C. R. (1980). Matrix approximations and reduction of dimensionality in multivariate statistical analysis multivariate statistical analysis. *In* Krishnaiah, P. (Editor) *Multivariate Analysis*. North Holland Press, Amsterdam, 3–22.

Rvačeva, E. L. (1961). On domains of attraction of multi-dimensional distributions. *Selected Translations in Mathematical Statistics and Probability*, **2**, 183–205.

Samorodnitsky, G. and Taqqu, M. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall, New York.

SHERMAN, S. (1955). A theorem on convex sets with applications. *The Annals of Mathematical Statistics*, **26**, 763–767.

VON NEUMANN, J. AND GOLDSTINE, H. H. (1947). Numerical inverting of matrices of high order. *Bulletin of the American Mathematical Society*, **53**, 1021–1099.

WOLFE, S. J. (1975). On the unimodality of spherically symmetric stable distribution functions. *Journal of Multivariate Analysis*, **5**, 236–242.

ZELLNER, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-*t* error terms. *Journal of the American Statistical Association*, **71**, 400–405.

Manuscript received 2019-08-27, revised 2019-11-04, accepted 2019-11-06.