Robust Semiparametric Efficient Joint Estimators of Location and Shape Matrix CES Distributions

Stefano Fortunati

Université Paris-Saclay, CNRS, CentraleSupélec, L2S & DR2I-IPSA, France

Le Séminaire Palaisien

Tuesday, June 1st, 2021

Outline of the talk

Why semiparametric models?

Semiparametric estimation in CES distributions

Le Cam thory on one-step efficient estimators

A complex-valued R-estimator for shape matrix

Numerical results

Parametric models

- A parametric model $\mathcal{P}_{\boldsymbol{\theta}}$ is defined as a set of pdfs that are parametrized by a finite-dimensional parameter vector $\boldsymbol{\theta}$:

$$\mathcal{P}_{\boldsymbol{\theta}} \triangleq \{p_X(\mathbf{x}_1, \ldots, \mathbf{x}_L|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q\}.$$

- The (lack of) knowledge about the phenomenon of interest is summarized in $\boldsymbol{\theta}$ that needs to be estimated.

- **Pros**: Parametric inference procedures are generally "simple" due to the finite dimensionality of $\boldsymbol{\theta}$.

- **Cons**: A parametric model could be too restrictive and a *misspecification problem* may occur. [1]

[1] S. Fortunati, F. Gini, M. S. Greco and C. D. Richmond, "Performance Bounds for Parameter Estimation under Misspecified Models: Fundamental Findings and Applications", *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 142-157, Nov. 2017.

Non-parametric models

- A non-parametric model $\mathcal{P}_p$ is a collection of pdfs possibly satisfying some functional constraints (i.e. *symmetry*):

$$\mathcal{P}_p \triangleq \{p_X(\mathbf{x}_1, \ldots, \mathbf{x}_L) \in \mathcal{K}\},$$

where $\mathcal{K}$ is some constrained set of pdfs.

- **Pros**: The risk of model misspecification is minimized.
- **Cons**: In non-parametric inference we have to face with infinite-dimensional estimation problem.
- **Cons**: Non-parametric inference may be a prohibitive task due to the large amount of required data.

Semiparametric models

- A semiparametric model $\mathcal{P}_{\boldsymbol{\theta},g}$ is a set of pdfs characterized by a finite-dimensional parameter $\boldsymbol{\theta} \in \Theta$ along with a *function*, i.e. an infinite-dimensional parameter, $g \in \mathcal{G}$: [2]

$$\mathcal{P}_{\boldsymbol{\theta},g} \triangleq \{p_X(\mathbf{x}_1,\ldots,\mathbf{x}_L|\boldsymbol{\theta},g), \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^q, g \in \mathcal{G}\}.$$

- Usually, $\boldsymbol{\theta}$ is the (finite-dimensional) parameter of interest while g can be considered as a nuisance parameter.

- **Pros**: All parametric signal models involving an unknown noise distribution are semiparametric models.

- **Cons**: Tools from functional analysis are needed.

[2] P.J. Bickel, C.A.J Klaassen, Y. Ritov and J.A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, 1993.

Outline of the talk

Why semiparametric models?

Semiparametric estimation in CES distributions

Le Cam thory on one-step efficient estimators

A complex-valued R-estimator for shape matrix

Numerical results

CES distributions

- A CES distributed random vector $\mathbf{z} \in \mathbb{C}^N$ admits a pdf: [3]

$$p_Z(\mathbf{z}) = |\boldsymbol{\Sigma}|^{-1} h((\mathbf{z}-\boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})) \triangleq CES_N(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h).$$

  - $h \in \mathcal{G}$, $h : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ is the *density generator*,
  - $\boldsymbol{\mu} \in \mathbb{C}^N$ is the location vector,
  - $\boldsymbol{\Sigma} \in \mathcal{M}_N$ is the (full rank) scatter matrix.

- Note that $\boldsymbol{\Sigma}$ and h are not jointly identifiable:

$$CES_N(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h(t)) \equiv CES_N(\boldsymbol{\mu}, c\boldsymbol{\Sigma}, h(ct)), \ \forall c > 0.$$

- To avoid this, we introduce the *shape matrix* as:

$$\mathbf{V}_1 \triangleq \boldsymbol{\Sigma}/[\boldsymbol{\Sigma}]_{1,1}.$$

[3] E. Ollila, D. E. Tyler, V. Koivunen and H. V. Poor, "Complex Elliptically Symmetric Distributions: Survey, New Results and Applications", *IEEE Trans. on Signal Processing*, vol. 60, no. 11, pp. 5597-5625, Nov. 2012.

CES distributions as semiparametric model

- The set of all CES pdfs is a semiparametric model of the form:

$$\mathcal{P}_{\boldsymbol{\theta},h} = \Big\{ p_Z | p_Z(\mathbf{z}|\boldsymbol{\theta}, h) = |\mathbf{V}_1|^{-1} \times \\ h((\mathbf{z}-\boldsymbol{\mu})^H \mathbf{V}_1^{-1}(\mathbf{z}-\boldsymbol{\mu})); \boldsymbol{\theta} \in \Theta, h \in \mathcal{G} \Big\},$$

- h plays the role of a *infinite-dimensional* nuisance parameter.

- By means of the Wirtinger calculus, the *finite-dimensional* parameter vector to be estimated can be cast as: [4]

$$\boldsymbol{\theta} \triangleq (\boldsymbol{\mu}^T, \boldsymbol{\mu}^H, \underline{\mathrm{vec}}(\mathbf{V}_1)^T)^T \in \Theta \subseteq \mathbb{C}^q,$$

where $q = N(N+2) - 1$ $(= 2N + N^2 - 1)$.

[4] The operator $\underline{\mathrm{vec}}(\mathbf{A})$ defines the $N^2 - 1$-dimensional vector obtained from $\mathrm{vec}(\mathbf{A})$ by deleting its first element, i.e. $\mathrm{vec}(\mathbf{A}) \triangleq [a_{11}, \underline{\mathrm{vec}}(\mathbf{A})^T]^T$.

Two starting questions

- Let $\{\mathbf{z}_l\}_{l=1}^{L}$ be a set of i.i.d. CES distributed vectors such that $\mathbb{C}^N \ni \mathbf{z}_l \sim p_0 \equiv CES_N(\boldsymbol{\mu}_0, \mathbf{V}_{1,0}, h_0)$, $\forall l$.

- **Goal**: joint estimate of $\boldsymbol{\mu}_0$ and $\mathbf{V}_{1,0}$ in the presence of an unknown density generator h_0.

  1. What is the impact of not knowing h_0 on the joint estimation of $(\boldsymbol{\mu}_0, \mathbf{V}_{1,0})$ (note that $\boldsymbol{\theta}_0 \triangleq (\boldsymbol{\mu}_0^T, \boldsymbol{\mu}_0^H, \underline{\mathrm{vec}}(\mathbf{V}_{1,0})^T)^T$)?

  2. What is the (asymptotic) impact that the lack of knowledge of $\boldsymbol{\mu}_0$ has on the estimation of $\mathbf{V}_{1,0}$ and vice versa?

- We need to introduce: [5]
  - *Semiparametric efficient score vector* $\bar{\mathbf{s}}_{\boldsymbol{\theta}_0, h_0}$,
  - *Semiparamatric Fisher Information Martix* (SFIM) $\bar{\mathbf{I}}(\boldsymbol{\theta}_0 | h_0)$.

[5] S. Fortunati, F. Gini, M. S. Greco, A. M. Zoubir and M. Rangaswamy, "Semiparametric CRB and Slepian-Bangs Formulas for Complex Elliptically Symmetric Distributions,", *IEEE Trans. on Signal Processing*, vol. 67, no. 20, pp. 5352-5364, 2019.

Semiparametric efficient score vector

- By using the Wirtinger calculus, the "parametric" score vector for $\boldsymbol{\theta}_0$ is:

$$[\mathbf{s}_{\boldsymbol{\theta}_0,h_0}]_i \triangleq \partial \ln p_Z(\mathbf{z};\boldsymbol{\theta},h_0)/\partial\theta_i^*|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\,,\ \ i=1,\ldots,q.$$

- The semiparametric efficient score vector is then given by:

$$\bar{\mathbf{s}}_{\boldsymbol{\theta}_0,h_0} = [\bar{\mathbf{s}}^T_{\boldsymbol{\mu}_0}, \bar{\mathbf{s}}^T_{\boldsymbol{\mu}_0^*}, \bar{\mathbf{s}}^T_{\underline{\mathrm{vec}}(\mathbf{V}_{1,0})}]^T = \mathbf{s}_{\boldsymbol{\theta}_0,h_0} - \Pi(\mathbf{s}_{\boldsymbol{\theta}_0,h_0}|\mathcal{T}_{h_0}).$$

- $\Pi(\mathbf{s}_{\boldsymbol{\theta}_0,h_0}|\mathcal{T}_{h_0})$ indicates the orthogonal projection of $\mathbf{s}_{\boldsymbol{\theta}_0}$ on the nuisance tangent space $\mathcal{T}_{h_0}$ of $\mathcal{P}_{\boldsymbol{\theta},h}$ evaluated at h_0.

- $\Pi(\mathbf{s}_{\boldsymbol{\theta}_0,h_0}|\mathcal{T}_{h_0})$ tells us the loss of information on the estimation of $\boldsymbol{\theta}_0$ due to the lack of knowledge of h_0.

Impact of h_0 on the estimation of μ_0 and $\mathbf{V}_{1,0}$

- It can be shown that:
  1. $\Pi(\mathbf{s}_{\boldsymbol{\mu}_0}|\mathcal{T}_{h_0}) = \mathbf{0}$,
  2. On the contrary, $\Pi(\mathbf{s}_{\underline{\mathrm{vec}}(\mathbf{V}_{1,0})}|\mathcal{T}_{h_0}) \neq \mathbf{0}$.

- **Answer to Point 1)**
  1. The lack of knowledge of h_0 does not have any impact on the (asymptotic) estimation of the location parameter $\boldsymbol{\mu}_0$,
  2. It does have an impact of the estimation of $\mathbf{V}_{1,0}$.

- A good estimator of $\mathbf{V}_{1,0}$ should have the following properties:
  1. It is able to handle the missing knowledge of h_0: **distributional robustness**.
  2. Its Mean Squared Error (MSE) achieves the Semiparametric Cramér-Rao Bound (SCRB): **semiparametric efficiency**.

Impact of μ_0 on the estimation of $\mathbf{V}_{1,0}$

- The SFIM for the joint estimation of $\boldsymbol{\mu}_0$ and $\mathbf{V}_{1,0}$ is:

$$\bar{\mathbf{I}}(\boldsymbol{\theta}_0|h_0) \triangleq E_0\{\bar{\mathbf{s}}_{\boldsymbol{\theta}_0,h_0}\bar{\mathbf{s}}^H_{\boldsymbol{\theta}_0,h_0}\} = \begin{pmatrix} \bar{\mathbf{I}}(\boldsymbol{\mu}_0|h_0) & \mathbf{0}_{2N\times(N^2-1)} \\ \mathbf{0}_{(N^2-1)\times 2N} & \bar{\mathbf{I}}(\mathbf{V}_{1,0}|h_0) \end{pmatrix}.$$

- The cross-information terms between the location $\boldsymbol{\mu}_0$ and the shape matrix $\mathbf{V}_{1,0}$ are equal to zero.

- **Answer to Point 2)**:
  In estimating the shape matrix, $\boldsymbol{\mu}_0$ can be substituted by any $\sqrt{L}$-consistent estimators $\widehat{\boldsymbol{\mu}}$ without any impact on the (asymptotic) performance of the estimator of $\mathbf{V}_{1,0}$.

The semiparametric estimation of $\mathbf{V}_{1,0}$

- Answers 1) and 2) allow us to assume $\boldsymbol{\mu} = \mathbf{0}$ without any loss of generality.

- In fact, even if $\boldsymbol{\mu} \neq \mathbf{0}$, we can always obtain the "centered data" as:

$$\{\mathbf{z}_l\}_{l=1}^{L} \longleftarrow \{\mathbf{z}_l - \widehat{\boldsymbol{\mu}}\}_{l=1}^{L},$$

  where $\widehat{\boldsymbol{\mu}}$ is any $\sqrt{L}$-consistent estimator of $\boldsymbol{\mu}_0$.

- In the rest of the seminar, we will consider the "centered" CES semiparametric model:

$$\mathcal{P}_{\boldsymbol{\theta},h} = \left\{ p_Z | p_Z(\mathbf{z}|\boldsymbol{\theta}, h) = |\mathbf{V}_1|^{-1} h(\mathbf{z}^H \mathbf{V}_1^{-1} \mathbf{z}); \boldsymbol{\theta} \in \Theta, h \in \mathcal{G} \right\},$$

  where

$$\boldsymbol{\theta} \triangleq \underline{\mathrm{vec}}(\mathbf{V}_1) \in \Theta \subseteq \mathbb{C}^d, \quad d = N^2 - 1.$$

Outline of the talk

Le Cam's "one-step" estimators (1/4)

- Let us consider a generic *parametric* model $\mathcal{P}_{\boldsymbol{\theta}}$.

- To fix ideas, we may consider the CES parametric model (h_0 is known):

$$\mathcal{P}_{\boldsymbol{\theta}} = \left\{ p_Z | p_Z(\mathbf{z}|\boldsymbol{\theta}, h_0) = |\mathbf{V}_1|^{-1} h_0(\mathbf{z}^H \mathbf{V}_1^{-1} \mathbf{z}); \boldsymbol{\theta} \in \Theta \right\}.$$

- The Maximum Likelihood estimator for $\boldsymbol{\theta}$ is:

$$\hat{\boldsymbol{\theta}}_{ML} \triangleq \underset{\boldsymbol{\theta} \in \Theta}{\mathrm{argmax}} \sum_{l=1}^{L} \ln p_Z(\mathbf{z}_l|\boldsymbol{\theta}, h_0).$$

- Solving the optimization problem may result to be a prohibitive task.

- In some cases, $\hat{\boldsymbol{\theta}}_{ML}$ may not even exist.

Le Cam's "one-step" estimators (2/4)

- Recall the definition of score vector:

$$[\mathbf{s}_{\boldsymbol{\theta}_0,h_0}]_i \triangleq \partial \ln p_Z(\mathbf{z};\boldsymbol{\theta},h_0)/\partial\theta_i^*|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\,,\;\; i=1,\ldots,d.$$

- Let us define the *central sequence* as:

$$\boldsymbol{\Delta}_{\boldsymbol{\theta},h_0}(\mathbf{z}_1,\ldots,\mathbf{z}_L) \equiv \boldsymbol{\Delta}_{\boldsymbol{\theta},h_0} \triangleq L^{-1/2}\sum\nolimits_{l=1}^{L} \mathbf{s}_{\boldsymbol{\theta},h_0}(\mathbf{z}_l).$$

- Under Cramér-type regularity conditions [6] if $\hat{\boldsymbol{\theta}}_{ML}$ exists, then it satisfies:

$$\boldsymbol{\Delta}_{\boldsymbol{\theta},h_0}(\mathbf{z}_1,\ldots,\mathbf{z}_L)|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ML}} = \mathbf{0}.$$

[6] Let's assume a real sample space $\mathcal{X} \subseteq \mathbb{R}^N$. If for all $\boldsymbol{\theta} \in \Theta$:

1. $p_X(\mathbf{x}|\boldsymbol{\theta})$ is continuously differentiable in $\boldsymbol{\theta}$ for almost all $\mathbf{x} \in \mathcal{X}$,
2. $E_0\{\mathbf{s}^T_{\boldsymbol{\theta}_0,h_0}(\mathbf{x})\mathbf{s}_{\boldsymbol{\theta}_0,h_0}(\mathbf{x})\} < \infty$,
3. The FIM $\mathbf{I}(\boldsymbol{\theta}) \triangleq \int \mathbf{s}_{\boldsymbol{\theta},h_0}(\mathbf{x})\mathbf{s}^T_{\boldsymbol{\theta},h_0}(\mathbf{x})p_X(\mathbf{x}|\boldsymbol{\theta})d\mathbf{x}$ is non-singular and continuous in $\boldsymbol{\theta}$.

Le Cam's "one-step" estimators (3/4)

- A new estimator $\hat{\boldsymbol{\theta}}$ can be obtained by a one-step "Newton-Raphson" iteration:

$$\hat{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}} - \left[\nabla_{\boldsymbol{\theta}}^T \boldsymbol{\Delta}_{\tilde{\boldsymbol{\theta}},h_0}\right]^{-1} \boldsymbol{\Delta}_{\tilde{\boldsymbol{\theta}},h_0},$$

  where $\tilde{\boldsymbol{\theta}}$ is a "good" starting point.

- $\nabla_{\boldsymbol{\theta}}^T \boldsymbol{\Delta}_{\tilde{\boldsymbol{\theta}},h_0}$ indicates the Jacobian matrix of $\boldsymbol{\Delta}_{\boldsymbol{\theta},h_0}$ at $\tilde{\boldsymbol{\theta}}$.

Key point. It can be shown that:

$$\nabla_{\boldsymbol{\theta}}^T \boldsymbol{\Delta}_{\boldsymbol{\theta},h_0} \equiv -L^{1/2}\mathbf{I}_{h_0}(\boldsymbol{\theta}) + o_P(1),^7 \quad \forall \boldsymbol{\theta} \in \Theta,$$

where $\mathbf{I}(\boldsymbol{\theta})$ is the Fisher Information Matrix (FIM):

$$\mathbf{I}_{h_0}(\boldsymbol{\theta}) \triangleq E_{\boldsymbol{\theta},h_0}\left\{\mathbf{s}_{\boldsymbol{\theta},h_0}(\mathbf{z})\mathbf{s}_{\boldsymbol{\theta},h_0}^T(\mathbf{z})\right\}.$$

[7] Let x_l be a sequence of random variables. Then $x_l = o_P(1)$ if $\lim_{l\to\infty} \Pr\{|x_l| \geq \epsilon\} = 0, \forall \epsilon > 0$ (*convergence in probability to 0*).

Le Cam's "one-step" estimators (4/4)

Theorem 1. A "one-step" estimator of $\boldsymbol{\theta}_0$ is defined as:

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{\star} + L^{-1/2}\mathbf{I}_{h_0}(\hat{\boldsymbol{\theta}}^{\star})^{-1}\boldsymbol{\Delta}_{\hat{\boldsymbol{\theta}}^{\star},h_0},$$

where $\hat{\boldsymbol{\theta}}^{\star}$ is any preliminary $\sqrt{L}$-consistent estimator of $\boldsymbol{\theta}_0$.

Properties:

P1 $\sqrt{L}$-consistency:

$$\sqrt{L}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = O_P(1),^{8}$$

P2 Asymptotic normality and efficiency:

$$\sqrt{L}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \underset{L\to\infty}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_{h_0}(\boldsymbol{\theta}_0)^{-1}),$$

where $\mathbf{I}_{h_0}(\boldsymbol{\theta}_0)^{-1} \equiv \mathrm{CRB}(\boldsymbol{\theta}_0)$.

[8] Let x_l be a sequence of random variables. Then $x_l = O_P(1)$ if for any $\epsilon > 0$, there exists a finite $M > 0$ and a finite $L > 0$, s.t. $\Pr\{|x_l| > M\} < \epsilon, \forall l > L$ (*stochastic boundedness*).

Extension to semiparametric models (1/3)

- Theorem 1 is valid in parametric models.
- *Semiparametric extension*: $\boldsymbol{\theta}_0 = \underline{\mathrm{vec}}(\mathbf{V}_{1,0})$ has to be estimated in the presence of the unknown density generator h_0.
- Let us introduce the *efficient central sequence* as:

$$\overline{\boldsymbol{\Delta}}_{\boldsymbol{\theta},h_0}(\mathbf{z}_1,\ldots,\mathbf{z}_L) \equiv \overline{\boldsymbol{\Delta}}_{\boldsymbol{\theta},h_0} \triangleq L^{-1/2}\sum_{l=1}^{L}\bar{\mathbf{s}}_{\boldsymbol{\theta},h_0}(\mathbf{z}_l),$$

where $\bar{\mathbf{s}}_{\boldsymbol{\theta},h_0} \triangleq \mathbf{s}_{\boldsymbol{\theta},h_0} - \Pi(\mathbf{s}_{\boldsymbol{\theta},h_0}|\mathcal{T}_{h_0})$ is the efficient score vector.

- Let us also recall the SFIM:

$$\bar{\mathbf{I}}(\boldsymbol{\theta}|h_0) \triangleq E_{\boldsymbol{\theta},h_0}\{\bar{\mathbf{s}}_{\boldsymbol{\theta},h_0}(\mathbf{z})\bar{\mathbf{s}}_{\boldsymbol{\theta},h_0}(\mathbf{z})^T\}.$$

Extension to semiparametric models (2/3)

- The natural "semiparametric" generalization of the (parametric) ML estimating equations would be: [9]

$$\overline{\boldsymbol{\Delta}}_{\boldsymbol{\theta},h}(\mathbf{z}_1,\ldots,\mathbf{z}_L)\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ML},h=\hat{h}^\star} = \mathbf{0}.$$

where , $\hat{h}^\star$ is a preliminary $\sqrt{L}$-consistent, *non-parametric*, estimator of the nuisance function h.

- Unfortunately, it is quite difficult to find an estimator of h_0 that converges at the $O_P(L^{-1/2})$ rate characterizing most of the parametric estimators.

- Roughly speaking, the non-parametric function estimation requires much more data than the ones needed to estimate a finite-dimensional parameter.

[9] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, 1998

Extension to semiparametric models (3/3)

- Hallin, Oja and Paindaveine proposed a different approach to obtain a semiparametric efficient estimator of $\mathbf{V}_{1,0}$.[10,11]

- The basic idea is to split the semiparametric estimation of $\mathbf{V}_{1,0}$ in two parts:

  1. Assume that h_0 is known and apply Theorem 1 to obtain a "clairvoyant" semiparametric estimatior $\hat{\boldsymbol{\theta}}_s$ as:

  $$\underline{\mathrm{vec}}(\widehat{\mathbf{V}}_1) = \underline{\mathrm{vec}}(\widehat{\mathbf{V}}_1^\star) + L^{-1/2}\bar{\mathbf{I}}(\widehat{\mathbf{V}}_1^\star|h_0)^{-1}\overline{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^\star,h_0},$$

  where $\hat{\boldsymbol{\theta}}^\star$ is any preliminary $\sqrt{L}$-consistent estimator of $\boldsymbol{\theta}_0$.

  2. Robustify $\hat{\boldsymbol{\theta}}_s$ by using a distribution-free, rank based, procedure.

[10] M. Hallin, H. Oja, and D. Paindaveine, "Semiparametrically efficient rank-based inference for shape II. optimal *R*-estimation of shape," *The Annals of Statistics*, vol. 34, no. 6, pp. 2757–2789, 2006.

[11] M. Hallin and B. J. M. Werker, "Semi-parametric efficiency, distribution-freeness and invariance," *Bernoulli*, vol. 9, no. 1, pp. 137–165, 2003.

Outline of the talk

A semiparametric efficient R-estimator (1/3)

- Building upon the results of Hallin, Oja and Paindaveine, a complex-valued R-estimator of $\mathbf{V}_{1,0}$ can be obtained as: [12]

$$\underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,R}) = \underline{\mathrm{vec}}(\widehat{\mathbf{V}}_1^\star) + L^{-1/2}\widehat{\boldsymbol{\Upsilon}}^{-1}\widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^\star}.$$

  - $\widehat{\boldsymbol{\Upsilon}}$ is an approximation of $\bar{\mathbf{I}}(\underline{\mathrm{vecs}}(\mathbf{V}_1)|h_0)$.
  - $\widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^\star}$ is a distributionally-free approximation of the efficient central sequence $\overline{\boldsymbol{\Delta}}_{\mathbf{V}_1}$.

- This R-estimator has the following desirable properties:
  1. *distributionally-robust* and
  2. *semiparametric efficient*,

[12] S. Fortunati, A. Renaux, F. Pascal, "Robust semiparametric efficient estimators in complex elliptically symmetric distributions", *IEEE Transactions on Signal Processing*, vol. 68, pp. 5003-5015, 2020.

A semiparametric efficient R-estimator (2/3)

$$\underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,R}) = \underline{\mathrm{vec}}(\widehat{\mathbf{V}}_1^\star) + \frac{1}{L\hat{\alpha}}\left[\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}\mathbf{L}_{\widehat{\mathbf{V}}_1^\star}^H\right]^{-1} \times \mathbf{L}_{\widehat{\mathbf{V}}_1^\star}\sum\nolimits_{l=1}^{L} K_h\left(\frac{r_l^\star}{L+1}\right)\mathrm{vec}(\hat{\mathbf{u}}_l^\star(\hat{\mathbf{u}}_l^\star)^H),$$

- $\{r_l^\star\}_{l=1}^L$ are the ranks of the r. v. $\hat{Q}_l^\star \triangleq \mathbf{z}_l^T[\widehat{\mathbf{V}}_1^\star]^{-1}\mathbf{z}_l$,
- $\hat{\mathbf{u}}_l^\star \triangleq \frac{[\widehat{\mathbf{V}}_1^\star]^{-1/2}\mathbf{z}_l}{\sqrt{\hat{Q}_l^\star}}$,
- $K_h(\cdot)$ is a *score* function based on $h \in \mathcal{G}$,
- $\hat{\alpha}$ is a data-dependent "cross-information" term,
- $\widehat{\mathbf{V}}_1^\star$ is a preliminary $\sqrt{L}$-consistent estimator of $\mathbf{V}_{1,0}$.

A semiparametric efficient R-estimator (3/3)

- The previous "vectorized" version of the R-estimator requires the unnecessary calculation of $N^2 \times N^2$ matrices with a consequent waste of computational resources.

- To overcome this problem, we have recently proposed the following "matrix" version of the same R-estimator: [13]

$$\widehat{\mathbf{V}}_{1,R} = \widehat{\mathbf{V}}_1^\star + \frac{1}{\hat{\alpha}}\left(\mathbf{W} - [\mathbf{W}]_{1,1}\,\widehat{\mathbf{V}}_1^\star\right)$$

where:

$$\mathbf{W} \triangleq L^{-1/2}(\widehat{\mathbf{V}}_1^\star)^{1/2}\mathbf{R}(\widehat{\mathbf{V}}_1^\star)^{1/2}.$$

$$\mathbf{R} \triangleq \frac{1}{\sqrt{L}}\sum_{l=1}^{L} K_h\left(\frac{r_l^\star}{L+1}\right)\hat{\mathbf{u}}_l^\star(\hat{\mathbf{u}}_l^\star)^H.$$

[13] S. Fortunati, A. Renaux, F. Pascal "Joint Estimation of Location and Scatter in Complex Elliptical Distributions: A robust semiparametric and computationally efficient R-estimator of the shape matrix," *MLSP Special Issue of the Journal of Signal Processing Systems*, 2021.

Two possible score functions

- *van der Waerden* (Gaussian-based) score function:

$$K_{\mathbb{C}vdW}(u) = \Phi_G^{-1}(u), \quad u \in (0,1),$$

  where Φ_G indicates the cdf of a Gamma-distributed random variable with parameters $(N, 1)$.

- t_ν-Student-based score function:

$$K_{\mathbb{C}t_\nu}(u) = \frac{N(2N+\nu)F_{2N,\nu}^{-1}(u)}{\nu + 2NF_{2N,\nu}^{-1}(u)}, \quad u \in (0,1),$$

  where $F_{2N,\nu}(u)$ stands for the Fisher cdf with $2N$ and $\nu \in (0, \infty)$ degrees of freedom.

- We refer to [14] for a discussion on how to build score functions.

[14] S. Fortunati, A. Renaux, F. Pascal, "Robust semiparametric efficient estimators in complex elliptically symmetric distributions", *IEEE Transactions on Signal Processing*, vol. 68, pp. 5003-5015, 2020.

Outline of the talk

Simulation set-up

- Tyler's joint M-estimator $(\hat{\boldsymbol{\mu}}_{Ty}, \widehat{\mathbf{V}}_{1,Ty})$ is given by the convergence points $(k \to \infty)$ of:

$$\hat{\boldsymbol{\mu}}_{Ty}^{(k+1)} = \left[\sum_{l=1}^{L} [\hat{Q}_l^{(k)}]^{-1/2} \right]^{-1} \sum_{l=1}^{L} \left(\hat{Q}_l^{(k)} \right)^{-1/2} \mathbf{z}_l,$$

$$\begin{cases} \widehat{\mathbf{V}}_{Ty}^{(k+1)} = \frac{N}{L} \sum_{l=1}^{L} [\hat{Q}_l^{(k)}]^{-1} (\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{Ty}^{(k)})(\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{Ty}^{(k)})^H. \\ \widehat{\mathbf{V}}_{1,Ty}^{(k+1)} \triangleq \widehat{\mathbf{V}}_{Ty}^{(k+1)} / [\widehat{\mathbf{V}}_{Ty}^{(k+1)}]_{1,1}. \end{cases}$$

where:

$$\hat{Q}_l^{(k)} = (\mathbf{z}_l - \hat{\boldsymbol{\mu}}^{(k)})^H [\widehat{\mathbf{V}}_1^{(k)}]^{-1} (\mathbf{z}_l - \hat{\boldsymbol{\mu}}^{(k)}),$$

- The estimators $\hat{\boldsymbol{\mu}}_{Ty}$ and $\widehat{\mathbf{V}}_{1,Ty}$ are $\sqrt{L}$-consistent under any (unknown) density generator $h \in \mathcal{G}$.

- **They are not semiparametric efficient**.

An R-estimator for $\mathbf{V}_{1,0}$

A robust semiparametric efficient estimator of $\mathbf{V}_{1,0}$ is given by:

$$\widehat{\mathbf{V}}_{1,R} = \widehat{\mathbf{V}}_{1,Ty} + \frac{1}{\hat{\alpha}}\left(\mathbf{W} - [\mathbf{W}]_{1,1}\,\widehat{\mathbf{V}}_{1,Ty}\right),$$

where $(\hat{\boldsymbol{\mu}}_{Ty}, \widehat{\mathbf{V}}_{1,Ty})$ plays the role of preliminary estimator and:

- $\{r_l^\star\}_{l=1}^L$: ranks of $\hat{Q}_l^\star \triangleq (\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{Ty})^H[\widehat{\mathbf{V}}_{1,Ty}]^{-1}(\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{Ty})$,
- $\mathbf{W} \triangleq L^{-1/2}(\widehat{\mathbf{V}}_{1,Ty})^{1/2}$,
- $\mathbf{R} \triangleq \frac{1}{\sqrt{L}}\sum_{l=1}^L K_h\left(\frac{r_l^\star}{L+1}\right)\hat{\mathbf{u}}_l^\star(\hat{\mathbf{u}}_l^\star)^H$,
- $\hat{\mathbf{u}}_l^\star \triangleq (\hat{Q}_l^\star)^{-1/2}[\widehat{\mathbf{V}}_{1,Ty}]^{-1/2}(\mathbf{z}_l - \hat{\boldsymbol{\mu}}_{Ty})$,

Simulation set-up

- We assess the efficiency of three joint estimators:
  1. Sample Mean and Sample Covariance: $(\hat{\boldsymbol{\mu}}_{SM}, \widehat{\mathbf{V}}_{1,SCM})$,
  2. Tyler - Tyler joint estimator: $(\hat{\boldsymbol{\mu}}_{Ty}, \widehat{\mathbf{V}}_{1,Ty})$,
  3. Tyler - R joint estimator: $(\hat{\boldsymbol{\mu}}_{Ty}, \widehat{\mathbf{V}}_{1,R})$.

- We generate the set of non-zero mean data $\{\mathbf{z}_l\}_{l=1}^{L}$ according to a Generalized Gaussian (GG) and a $t-$ distributions.

- Mean Squared Error (MSE) indices:

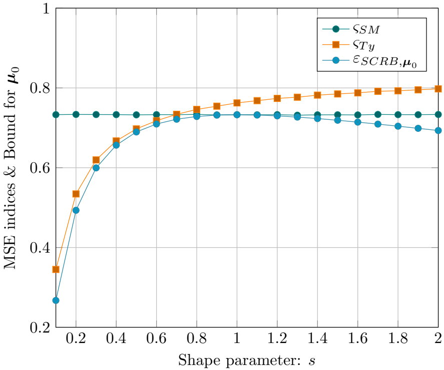
$$\varsigma_\gamma \triangleq ||E\{(\hat{\boldsymbol{\mu}}_\gamma - \boldsymbol{\mu}_0)(\hat{\boldsymbol{\mu}}_\gamma^a - \boldsymbol{\mu}_0^a)^H\}||_F,$$

$$\varsigma_\gamma \triangleq ||E\{\underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,\gamma} - \mathbf{V}_{1,0})\underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,\gamma} - \mathbf{V}_{1,0})^H\}||_F,$$

and γ indicates the relevant estimator at hand.

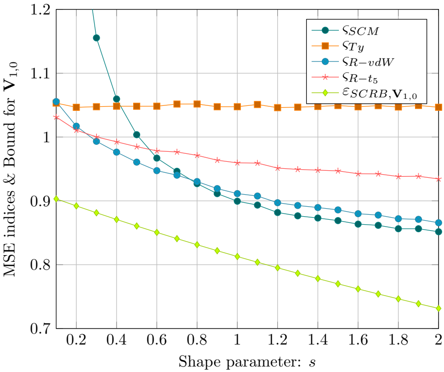
- As lower bounds, we use $\varepsilon_{CRB,\boldsymbol{\mu}_0} \triangleq ||\mathrm{SCRB}(\boldsymbol{\mu}_0|h_0)||_F$ and $\varepsilon_{SCRB,\mathbf{V}_{1,0}} \triangleq ||\mathrm{SCRB}(\underline{\mathrm{vec}}(\mathbf{V}_{1,0})|h_0)||_F$.

MSE on μ_0 for GG data ($L = 5N$)



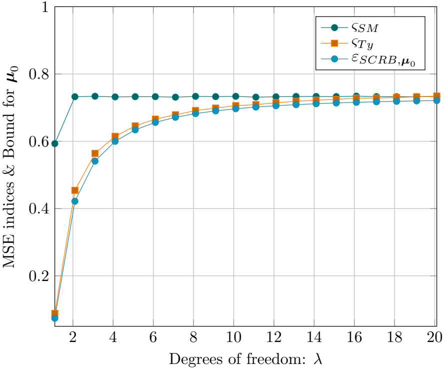
- $\hat{\boldsymbol{\mu}}_{Ty}$ is almost efficient in heavy-tailed data ($0 < s < 1$),
- $\hat{\boldsymbol{\mu}}_{Ty}$ outperforms $\hat{\boldsymbol{\mu}}_{SM}$ that it is known to be non robust,
- $\hat{\boldsymbol{\mu}}_{SM}$ is efficient in the Gaussian case ($s = 1$), and tends to have better performance than $\hat{\boldsymbol{\mu}}_{Ty}$ for $s > 1$.

MSE on $\mathbf{V}_{1,0}$ for GG data ($L = 5N$)



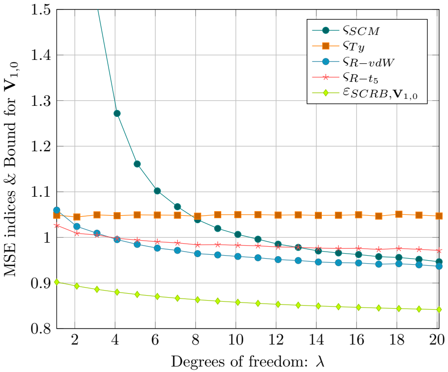
- $\widehat{\mathbf{V}}_{1,R-vdW}$ and $\widehat{\mathbf{V}}_{1,R-t_5}$ outperform $\widehat{\mathbf{V}}_{1,Ty}$ for every values of s, i.e. for both heavy-tailed and light-tailed data.
- They also outperform the sample covariance matrix $\widehat{\mathbf{V}}_{1,SCM}$ in the presence of heavy-tailed data ($0 < s < 1$), while their MSE is of the same order for $s > 1$.

MSE on μ_0 for t-distributed data ($L = 5N$)



- $\hat{\boldsymbol{\mu}}_{Ty}$ is almost efficient in heavy-tailed data (small λ),
- $\hat{\boldsymbol{\mu}}_{Ty}$ outperforms $\hat{\boldsymbol{\mu}}_{SM}$ that it is known to be non robust,
- $\hat{\boldsymbol{\mu}}_{SM}$ tends to be efficient in the Gaussian case ($\lambda \to \infty$).

MSE on $\mathbf{V}_{1,0}$ for t-distributed data ($L = 5N$)



- $\widehat{\mathbf{V}}_{1,R-vdW}$ and $\widehat{\mathbf{V}}_{1,R-t_5}$ outperform $\widehat{\mathbf{V}}_{1,Ty}$ for every values of λ.
- They also outperform the sample covariance matrix $\widehat{\mathbf{V}}_{1,SCM}$ in the presence of heavy-tailed data.

Conclusions

- The wide applicability of the semiparametric framework has been discussed.
- Building upon the Le Cam's "one-step" estimators, a general procedure to obtain semiparametric efficient estimators has been investigated.
- A distributionally robust and nearly semiparametric efficient R-estimator of the shape matrix in CES distributions has been proposed and analyzed.

Current work:
Derivation of an R-estimator of the eigenspace of the scatter matrix in CES-distributed data.

Many thanks for your attention!

Any question?

Backup slides

Ranks (1/2)

- Let $\{x_l\}_{l=1}^{L}$ be a set of L continuous i.i.d. random variables with pdf p_X.

- Define the vector of the *order statistics* as

$$\mathbf{v}_X \triangleq [x_{L(1)}, x_{L(2)}, \ldots, x_{L(L)}]^T,$$

whose entries

$$x_{L(1)} < x_{L(2)} < \cdots < x_{L(L)}$$

are the values of $\{x_l\}_{l=1}^{L}$ ordered in an ascending way.[15]

- The rank $r_l \in \mathbb{N}$ of x_l is the position index of x_l in $\mathbf{v}_X$.

[15] Note that, since $x_l, \forall l$ are continuous random variable the equality occurs with probability 0.

Ranks (2/2)

- Let $\mathbf{r}_X \triangleq [r_1, \ldots, r_L]^T \in \mathbb{N}^L$ be the vector collecting the ranks.
- Let $\mathcal{K}$ be the family of score functions $K : (0,1) \to \mathbb{R}$ that are continuous, square integrable and that can be expressed as the difference of two monotone increasing functions.

Let $\{x_l\}_{l=1}^{L}$ be a set of i.i.d. random variables s.t. $x_l \sim p_X, \forall l$.
Then, we have:

1. The vectors $\mathbf{v}_X$ and $\mathbf{r}_X$ are independent,
2. Regardless the actual pdf p_X, the rank vector $\mathbf{r}_X$ is uniformly distributed on the set of all $L!$ permutations on $\{1, 2, \ldots, L\}$,
3. For each $l = 1, \ldots, L$, $K\left(\frac{r_l}{L+1}\right) = K(u_l) + o_P(1)$, where $K \in \mathcal{K}$ and $u_l \sim \mathcal{U}[0,1]$ is a random variable uniformly distributed in $(0,1)$.