

Robust Semiparametric Efficient Estimators in Complex Elliptically Symmetric (CES) Distributions

Stefano Fortunati

Enseignant-chercheur at IPSA
(Candidate member of L2S)

S³ seminar
Friday, October 2sd, 2020



My professional background



Montegiorgio

- ▶ Born in 1983.

Pisa

- ▶ Bachelor (Dec. 2005),
- ▶ Master (June 2008),
- ▶ PhD (June 2012),
- ▶ Post-doc (~ 7 years).

La Spezia (6 months)

- ▶ Visiting researcher.

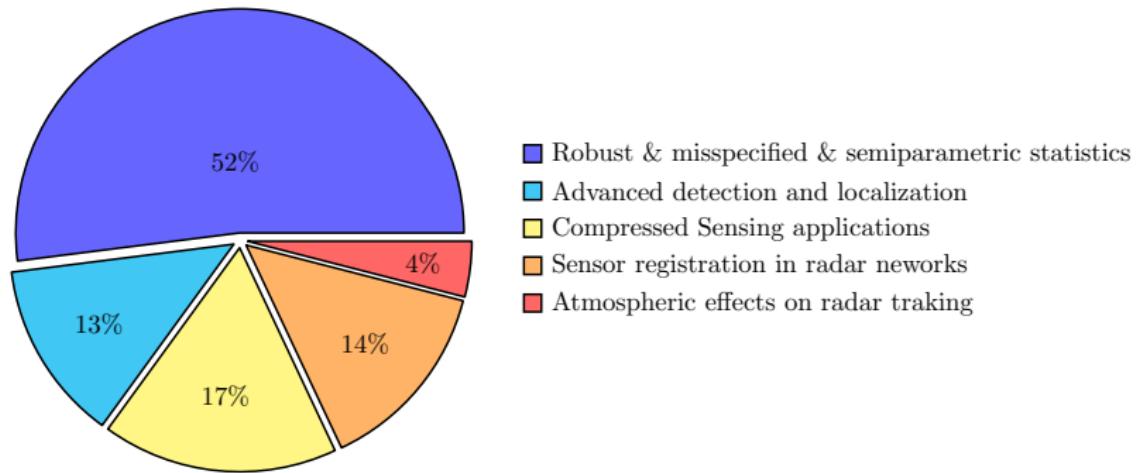
Darmstadt (1 year)

- ▶ Visiting researcher.

Paris

- ▶ Post-doc (~ 1 year),
- ▶ Enseignant-chercheur.

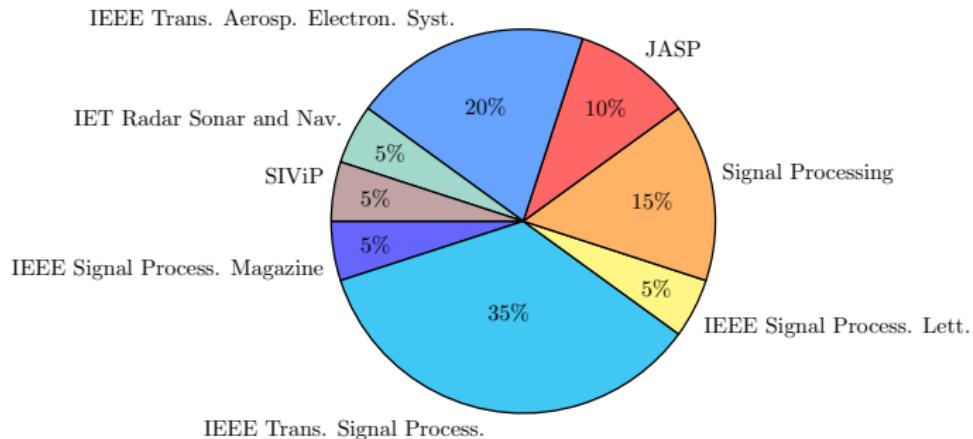
Scientific activities: topics



- ▶ **PhD and first part of my post-doc:**
 - ▶ Radar signal processing,
 - ▶ Compressed sensing applications to sonar and oceanography.
- ▶ **Second part of my post-doc and current work:**
 - ▶ Robust, misspecified and semiparametric statistics,
 - ▶ Covariance matrix estimation in non-Gaussian data.

Scientific activities: Publications

Research: 1 book chapter, 18 journal publications, 30 conference publications.



Conferences:

ICASSP, EUSIPCO, SSP, ISI World Statistics Congress, MLSP, RadarConference...

Scientific activities: Collaborations

- ▶ F. Pascal and A. Renaux, *Université Paris-Saclay, CNRS, CentraleSupélec, L2S, France*,
- ▶ M. N. El Korso, *University Paris Nanterre, France*,
- ▶ F. Gini, S. Greco and L. Sanguinetti *University of Pisa, Italy*,
- ▶ A. M. Zoubir, *Technische Universität Darmstadt, Germany*,
- ▶ Aya Mostafa Ahmed and Aydin Sezgin, *Ruhr Universität Bochum, Germany*,
- ▶ C. D. Richmond, *Arizona State University, USA*,
- ▶ M. Rangaswamy and B. Himed *U.S. AFRL, Sensors Directorate, USA*,
- ▶ R. Grasso, K. LePage and P. Braca, *CMRE, NATO*.

Today's seminar: related papers

Journal

- ▶ 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.

Conferences

- ▶ S. Fortunati, A. Renaux, F. Pascal, "Properties of a new R -estimator of shape matrices", *EUSIPCO 2020*, Amsterdam, the Netherlands, August 24-28, 2020.
- ▶ S. Fortunati, A. Renaux, F. Pascal, "Robust Semiparametric DOA Estimation in non-Gaussian Environment", *2020 IEEE Radar Conference*, Florence, Italy, September 21-25, 2020.
- ▶ S. Fortunati, A. Renaux, F. Pascal, "Robust Semiparametric Joint Estimators of Location and Scatter in Elliptical Distributions", *IEEE MLSP*, Aalto University, Espoo, Finland, September 21-24, 2020.

Outline of the talk

Why semiparametric models?

Semiparametric estimation in CES distributions

Le Cam theory on one-step efficient estimators

The proposed complex-valued R -estimator for shape matrix

Numerical results

- ▶ A parametric model \mathcal{P}_θ is defined as a set of pdfs that are parametrized by a finite-dimensional parameter vector θ :

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

- ▶ The (lack of) knowledge about the phenomenon of interest is summarized in θ that needs to be estimated.
- ▶ **Pros:** Parametric inference procedures are generally “simple” due to the finite dimensionality of θ .
- ▶ **Cons:** A parametric model could be too restrictive and a *misspecification problem*¹ may occur.

¹S. 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.

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

$$\mathcal{P}_p \triangleq \{p_{\mathbf{x}}(\mathbf{x}_1, \dots, \mathbf{x}_M) \in \mathcal{K}\},$$

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.

- ▶ A semiparametric model² $\mathcal{P}_{\theta,g}$ is a set of pdfs characterized by a finite-dimensional parameter $\theta \in \Theta$ along with a *function*, i.e. an infinite-dimensional parameter, $g \in \mathcal{G}$:

$$\mathcal{P}_{\theta,g} \triangleq \{p_{\mathbf{x}}(\mathbf{x}_1, \dots, \mathbf{x}_M | \theta, g), \theta \in \Theta \subseteq \mathbb{R}^q, g \in \mathcal{G}\}.$$

- ▶ Usually, θ 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.

²P.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 theory on one-step efficient estimators

The proposed complex-valued R -estimator for shape matrix

Numerical results

The Complex Elliptically Symmetric (CES) distributions

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

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

- ▶ $h \in \mathcal{G}$, $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is the *density generator*,
 - ▶ $\mu \in \mathbb{C}^N$ is the location vector,
 - ▶ $\Sigma \in \mathcal{M}_N$ is the (full rank) scatter matrix.
-
- ▶ Note that Σ and h are not jointly identifiable:

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

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

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

³E. 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.

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

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

- ▶ 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

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

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

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

Two starting questions

- ▶ Let $\{\mathbf{z}_l\}_{l=1}^L$ be a set of 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 $\boldsymbol{\theta}_0 \triangleq (\boldsymbol{\mu}_0^T, \underline{\text{vec}}(\mathbf{V}_{1,0})^T)^T$)?
 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* $\bar{\mathbf{s}}_{\boldsymbol{\theta}_0}$,
 - ▶ *Semiparametric Fisher Information Matrix (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 θ_0 is:

$$[\mathbf{s}_{\theta_0}]_i \triangleq \partial \ln p_Z(\mathbf{z}; \boldsymbol{\theta}, h_0) / \partial \theta_i^*|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \quad i = 1, \dots, q.$$

- ▶ The semiparametric efficient score vector is then given by:

$$\bar{\mathbf{s}}_{\theta_0, h_0} = [\bar{\mathbf{s}}_{\mu_0}^T, \bar{\mathbf{s}}_{\mu_0^*}^T, \bar{\mathbf{s}}_{\underline{\text{vec}}(\mathbf{v}_{1,0})}^T]^T = \mathbf{s}_{\theta_0} - \Pi(\mathbf{s}_{\theta_0} | \mathcal{T}_{h_0}).$$

- ▶ $\Pi(\mathbf{s}_{\theta_0} | \mathcal{T}_{h_0})$ indicates the orthogonal projection of \mathbf{s}_{θ_0} on the nuisance tangent space \mathcal{T}_{h_0} of $\mathcal{P}_{\theta, h}$ evaluated at h_0 .
- ▶ $\Pi(\mathbf{s}_{\theta_0} | \mathcal{T}_{h_0})$ tells us the loss of information on the estimation of θ_0 due to the lack of knowledge of h_0 .

► It can be shown that:

1. $\Pi(\mathbf{s}_{\mu_0} | \mathcal{T}_{h_0}) = \mathbf{0},$
2. On the contrary, $\Pi(\mathbf{s}_{\text{vec}(\mathbf{v}_{1,0})} | \mathcal{T}_{h_0}) \neq \mathbf{0}.$

► 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 $\mu_0,$
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 μ_0 and $\mathbf{V}_{1,0}$ is:

$$\bar{\mathbf{I}}(\theta_0 | h_0) \triangleq E_0\{\bar{\mathbf{s}}_{\theta_0, h_0} \bar{\mathbf{s}}_{\theta_0, h_0}^H\} = \begin{pmatrix} \bar{\mathbf{I}}(\mu_0 | h_0) & \mathbf{0}_{2N \times (N^2-1)} \\ \mathbf{0}_{(N^2-1) \times 2N} & \bar{\mathbf{I}}(\mathbf{V}_{1,0} | h_0) \end{pmatrix}.$$

- ▶ The cross-information terms between the location μ_0 and the shape matrix $\mathbf{V}_{1,0}$ are equal to zero.
- ▶ **Answer to Point 2):**
In estimating the shape matrix, μ_0 can be substituted by any \sqrt{L} -consistent estimators $\hat{\mu}$ 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 $\mu = \mathbf{0}$ without any loss of generality.
- ▶ In fact, even if $\mu \neq \mathbf{0}$, we can always obtain the “centered data” as:

$$\{\mathbf{z}_l\}_{l=1}^L \leftarrow \{\mathbf{z}_l - \hat{\mu}\}_{l=1}^L,$$

where $\hat{\mu}$ is any \sqrt{L} -consistent estimator of μ_0 .

- ▶ In the rest of the seminar, we will consider the “centered” CES semiparametric model:

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

where

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

Outline of the talk

Why semiparametric models?

Semiparametric estimation in CES distributions

Le Cam theory on one-step efficient estimators

The proposed complex-valued R -estimator for shape matrix

Numerical results

Parametric Le Cam's “one-step” estimators

- ▶ Let us consider a generic *parametric* model \mathcal{P}_θ .
- ▶ To fix ideas, we may consider the CES parametric model (h_0 is known):

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

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

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

- ▶ 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:

$$[\mathbf{s}_{\theta_0}]_i \triangleq \partial \ln p_Z(\mathbf{z}; \boldsymbol{\theta}, h_0) / \partial \theta_i^*|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \quad i = 1, \dots, d.$$

- ▶ Let us define the *central sequence* as:

$$\Delta_{\boldsymbol{\theta}}(\mathbf{z}_1, \dots, \mathbf{z}_L) \equiv \Delta_{\boldsymbol{\theta}} \triangleq L^{-1/2} \sum_{l=1}^L \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{z}_l).$$

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

$$\Delta_{\boldsymbol{\theta}}(\mathbf{z}_1, \dots, \mathbf{z}_L)|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ML}} = \mathbf{0},$$

Le Cam's “one-step” estimators (3/4)

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

$$\hat{\theta} = \tilde{\theta} - [\nabla_{\theta}^T \Delta_{\tilde{\theta}}]^{-1} \Delta_{\tilde{\theta}},$$

where $\tilde{\theta}$ is a “good” starting point.

- ▶ $\nabla_{\theta}^T \Delta_{\tilde{\theta}}$ indicates the Jacobian matrix of Δ_{θ} evaluated at $\tilde{\theta}$.

Key point. It can be shown that:

$$\nabla_{\theta}^T \Delta_{\theta} \equiv -L^{1/2} \mathbf{I}(\theta) + o_P(1),^6 \quad \forall \theta \in \Theta,$$

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

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

⁶ Let x_I be a sequence of random variables. Then $x_I = o_P(1)$ if $\lim_{I \rightarrow \infty} \Pr \{ |x_I| \geq \epsilon \} = 0, \forall \epsilon > 0$ (convergence in probability to 0).

Le Cam's “one-step” estimators (4/4)

Theorem 1. A “one-step” estimator of θ_0 is defined as:

$$\hat{\theta} = \hat{\theta}^* + L^{-1/2} \mathbf{I}(\hat{\theta}^*)^{-1} \Delta_{\hat{\theta}^*},$$

where $\hat{\theta}^*$ is any preliminary \sqrt{L} -consistent estimator of θ_0 .

Properties:

P1 \sqrt{L} -consistency:

$$\sqrt{L} (\hat{\theta} - \theta_0) = O_P(1),^7$$

P2 Asymptotic normality and efficiency:

$$\sqrt{L} (\hat{\theta} - \theta_0) \underset{L \rightarrow \infty}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}(\theta_0)^{-1}),$$

where $\mathbf{I}(\theta_0)^{-1} \equiv \text{CCRB}(\theta_0)$.

⁷ Let x_I be a sequence of random variables. Then $x_I = O_P(1)$ if for any $\epsilon > 0$, there exists a finite $M > 0$ and a finite $L > 0$, s.t. $\Pr \{|x_I| > M\} < \epsilon, \forall I > L$ (*stochastic boundedness*).

Extension to semiparametric models (1/5)

- ▶ Theorem 1 is valid in parametric models.
- ▶ *Semiparametric extension:* $\theta_0 = \underline{\text{vec}}(\mathbf{V}_{1,0})$ has to be estimated in the presence of the unknown density generator h_0 .
- ▶ Let us introduce the *efficient central sequence* as:

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

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

- ▶ Let us also recall the SFIM:

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

- ▶ The natural “semiparametric” generalization of the (parametric) ML estimating equations would be:⁸

$$\Delta_{\theta,h}(z_1, \dots, z_L)|_{\theta=\hat{\theta}_{ML}, h=\hat{h}^*} = \mathbf{0}.$$

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

- ▶ Unfortunately, it is generally impossible 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 estimation of a function requires much more data than the ones needed to estimate a finite-dimensional parameter.

- ▶ Hallin, Oja and Paindaveine proposed a different approach to obtain a semiparametric efficient estimator of \mathbf{V}_1 .⁹
- ▶ The basic idea is to split the semiparametric estimation of \mathbf{V}_1 in two parts:
 1. Assume that h_0 is known and apply Theorem 1 to obtain a "clairvoyant" semiparametric estimator $\hat{\theta}_s$ as:

$$\hat{\theta}_s = \hat{\theta}^* + L^{-1/2} \bar{\mathbf{I}}(\hat{\theta}^* | h_0)^{-1} \bar{\Delta}_{\hat{\theta}^*, h_0},$$

where $\hat{\theta}^*$ is any preliminary \sqrt{L} -consistent estimator of θ_0 .

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

⁹

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.

Extension to semiparametric models (4/5)

- ▶ It can be shown that:

1. The efficient central sequence:

$$\overline{\Delta}_{\mathbf{v}_1, h_0} = -L^{-1/2} \mathbf{L}_{\mathbf{v}_1} \sum_{m=1}^M Q_I \psi_0(Q_I) \text{vec}(\mathbf{u}_I \mathbf{u}_I^H).$$

2. The efficient Semiparametric FIM

$$\bar{\mathbf{I}}(\text{vecs}(\mathbf{V}_1) | h_0) = \frac{E\{\mathcal{Q}^2 \psi_0(\mathcal{Q})^2\}}{N(N+1)} \mathbf{L}_{\mathbf{v}_1} \mathbf{L}_{\mathbf{v}_1}^H.$$

- ▶ $Q_I \triangleq \mathbf{z}_I^H \mathbf{V}_1^{-1} \mathbf{z}_I \stackrel{d}{=} \mathcal{Q} \sim P_{\mathcal{Q}, h_0}$,
 - ▶ $\psi_0(q) \triangleq d \ln h_0(q) / dq$,
 - ▶ $\mathbf{u}_I \sim \mathcal{U}(\mathbb{C}S^N)$,
 - ▶ $\mathbf{P} = [\mathbf{e}_2 | \mathbf{e}_3 | \cdots | \mathbf{e}_{N^2}]$,
 - ▶ $\Pi_{\text{vec}(\mathbf{I}_N)}^\perp = \mathbf{I}_{N^2} - N^{-1} \text{vec}(\mathbf{I}_N) \text{vec}(\mathbf{I}_N)^T$,
 - ▶ $\mathbf{L}_{\mathbf{v}_1} = \mathbf{P} \left(\mathbf{V}_1^{-T/2} \otimes \mathbf{V}_1^{-1/2} \right) \Pi_{\text{vec}(\mathbf{I}_N)}^\perp$
-
- ▶ **Note that** $\psi_0(q)$ and the cdf $P_{\mathcal{Q}, h_0}$ of Q_I depends on the true and unknown h_0 !

- ▶ Is there any way out? Rank-based statistics! ¹⁰
- ▶ In their seminal paper,¹¹ Hallin and Werker proposed an *invariance-based* approach to solve semiparametric estimation problems.
- ▶ **Main idea:** Find a distribution-free approximation of the efficient central sequence $\bar{\Delta}_{\mathbf{v}_1, h_0}$ and of the efficient SFIM $\bar{\mathbf{I}}(\text{vec}_{\mathbf{s}}(\mathbf{V}_1) | h_0)$!

¹⁰ The definition of rank is given in the backup slides.

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

Why semiparametric models?

Semiparametric estimation in CES distributions

Le Cam theory on one-step efficient estimators

The proposed complex-valued R -estimator for shape matrix

Numerical results

A semiparametric efficient R -estimator (1/2)

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

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

- ▶ $\widehat{\boldsymbol{\Upsilon}}$ is an approximation of $\bar{\mathbf{I}}(\underline{\text{vecs}}(\mathbf{V}_1) | h_0)$.
- ▶ $\widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}_1^*}$ 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 (2/2)

$$\begin{aligned}\underline{\text{vec}}(\widehat{\mathbf{V}}_{1,R}) &= \underline{\text{vecs}}(\widehat{\mathbf{V}}_1^*) + \frac{1}{L\hat{\alpha}} \left[\mathbf{L}_{\widehat{\mathbf{V}}_1^*} \mathbf{L}_{\widehat{\mathbf{V}}_1^*}^H \right]^{-1} \\ &\quad \times \mathbf{L}_{\widehat{\mathbf{V}}_1^*} \sum_{l=1}^L K_h \left(\frac{r_l^*}{L+1} \right) \text{vec}(\widehat{\mathbf{u}}_l^* (\widehat{\mathbf{u}}_l^*)^H),\end{aligned}$$

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

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 ¹³ for a discussion on how to build score functions.

¹³ 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

Why semiparametric models?

Semiparametric estimation in CES distributions

Le Cam theory on one-step efficient estimators

The proposed complex-valued R -estimator for shape matrix

Numerical results

Simulation set-up

- ▶ A competing shape matrix estimation: Tyler's one ($k \rightarrow \infty$):

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

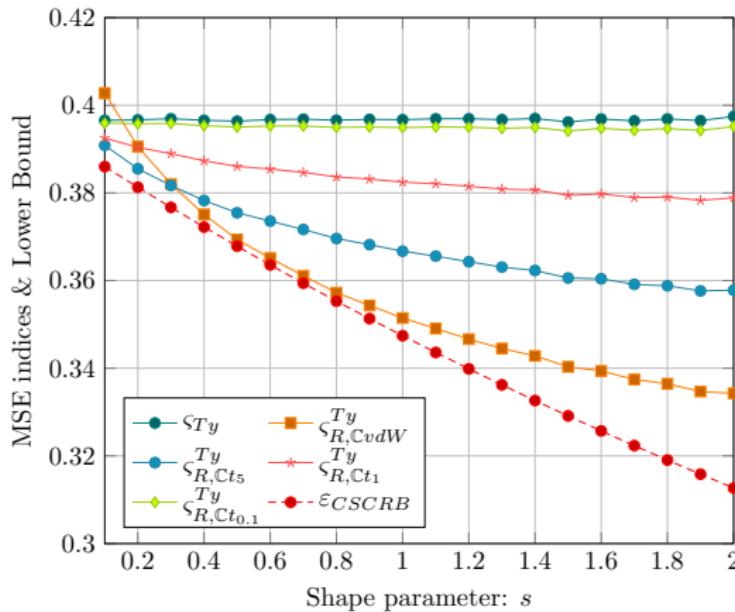
- ▶ Robustness: Yes,
- ▶ Semiparametric efficiency: No.
- ▶ We generate the set of non-zero mean data $\{\mathbf{z}_l\}_{l=1}^L$ according to a Generalized Gaussian (GG).
- ▶ Mean Squared Error (MSE) index and Semiparametric CRB:

$$\varsigma_\gamma^\varphi = \|E\{\text{vec}(\widehat{\mathbf{V}}_{1,\gamma}^\varphi - \mathbf{V}_{1,0})\text{vec}(\widehat{\mathbf{V}}_{1,\gamma}^\varphi - \mathbf{V}_{1,0})^H\}\|_F,$$

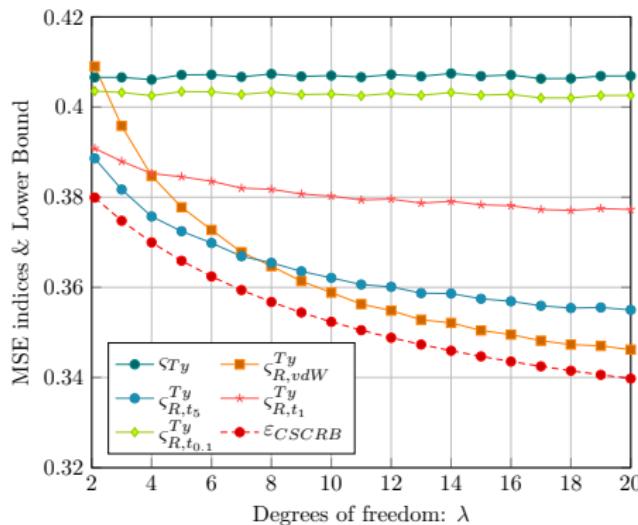
where γ and φ indicate the relevant estimator at hand and

$$\varepsilon_{CSCRB} = \|[CSCRB(\Sigma_0, h_0)]\|_F. \quad (1)$$

GG distribution



- ▶ “Finite-sample” regime: $L = 5N$, $N = 8$.
- ▶ The GG distribution presents heavier tails ($0 < s < 1$) and lighter tails ($s > 1$) compared to the Gaussian one ($s = 1$).

(Real) t -distribution

- ▶ “Finite-sample” regime: $L = 5N$, $N = 8$.
- ▶ When $\lambda \rightarrow \infty$, the t -distribution tends the Gaussian one.

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 discussed.
- ▶ A distributionally robust and nearly semiparametric efficient R -estimator of the shape matrix in Real and Complex ES distributions has been proposed and analyzed.
- ▶ Finally, the finite-sample performance of the R -estimator has been investigated in different scenarios in terms of MSE and robustness to outliers.

With F. Pascal and A. Renaux (L2S):

- ▶ We are working on the derivation of an **efficient estimator** of the “cross-information” term $\hat{\alpha}$.
- ▶ What about the **asymptotic distribution** of the derived the R -estimator?
- ▶ Which is the behavior of the R -estimator as the **data dimension** N goes to infinity?

Future works and collaborations

- ▶ With E. Ollila, Aalto University, Finland:
Is it possible to derive a semiparametric estimator of the eigenspace of the shape matrix?
- ▶ With all those interested in a possible collaboration:
Application of the semiparametric statistic in
 - ▶ *Radar/Sonar processing,*
 - ▶ *Image processing,*
 - ▶ *Distance learning and clustering,*
 - ▶ ...

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

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

whose entries

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

are the values of $\{x_l\}_{l=1}^L$ ordered in an ascending way.¹⁴

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

¹⁴ Note that, since $x_l, \forall l$ are continuous random variable the equality occurs with probability 0.

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

Let $\{x_I\}_{I=1}^L$ be a set of i.i.d. random variables s.t. $x_I \sim p_X, \forall I$. 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 $\{1, 2, \dots, L\}$,
3. For each $I = 1, \dots, L$, $K\left(\frac{r_I}{L+1}\right) = K(u_I) + o_P(1)$, where $K \in \mathcal{K}$ and $u_I \sim \mathcal{U}[0, 1]$ is a random variable uniformly distributed in $(0, 1)$.