

Robust Semiparametric Efficient Estimators in CES Distributions

Stefano Fortunati

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

Aalto-yliopisto

Friday, September 3rd, 2021

The visit is founded by Institut Français de Finlande, the Embassy of France in Finland, the French Ministry of Education, Higher Education and Research, the Finnish Academy for Science and Letters and the Finnish Society of Sciences and Letters.



INSTITUT
FRANÇAIS
FINLANDE

Outline of the talk

Why semiparametric models?

Semiparametric estimation in CES distributions

Le Cam theory on one-step efficient estimators

A complex-valued R -estimator for shape matrix

Numerical results

Parametric models

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

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

- ▶ The (lack of) knowledge about the random experiment 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* may occur. ¹

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.

Non-parametric models

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

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

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

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

- ▶ 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.

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

$$p_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-1} h((\mathbf{x} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})),$$

- ▶ $h \in \mathcal{G}$, $h: \mathbb{R}_0^+ \rightarrow \mathbb{R}^+$ is the *density generator*,
 - ▶ $\boldsymbol{\mu} \in \mathbb{C}^N$ is the mean value,
 - ▶ $\boldsymbol{\Sigma} \in \mathcal{M}_N$ is the (full rank) scatter matrix.
- ▶ The set of all CES pdfs is a semiparametric model of the form:

$$\mathcal{P}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}, h} \triangleq \left\{ p_{\mathbf{X}} | p_{\mathbf{X}}(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, h), \boldsymbol{\mu} \in \mathbb{C}^N, \boldsymbol{\Sigma} \in \mathcal{M}_N, h \in \mathcal{G} \right\}.$$

- ▶ This semiparametric model is a particular instance of the more general set of *semiparametric group models*.

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.

Examples: Missing data

- ▶ Let $\mathbf{z} \triangleq (\mathbf{x}^T, \mathbf{y}^T)^T$ be a *complete* dataset, where:
 - ▶ \mathbf{x} is the *observed* (available) dataset.
 - ▶ \mathbf{y} is the *unobservable* (missing) dataset.
- ▶ **Problem:** Estimate $\boldsymbol{\theta} \in \Theta$ from the observed dataset \mathbf{x} when the pdf p_Y of the missing data \mathbf{y} is unknown.
- ▶ The pdf p_X of the observed dataset can be expressed as:

$$p_X(\mathbf{x}|\boldsymbol{\theta}) = \int_{\mathcal{Y}} p_{X,Y}(\mathbf{x}, \mathbf{y}|\boldsymbol{\theta}) d\mathbf{y} = \int_{\mathcal{Y}} p_{X|Y}(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}) p_Y(\mathbf{y}) d\mathbf{y}.$$

- ▶ The set of all the pdfs of the observed dataset \mathbf{x} is a *semiparametric mixture model* of the form :

$$\mathcal{P}_{\boldsymbol{\theta}, p_Z} \triangleq \{p_X | p_X(\mathbf{x}|\boldsymbol{\theta}, p_Y), \boldsymbol{\theta} \in \Theta, p_Y \in \mathcal{K}\}.$$

Examples: Non-linear regression

- ▶ Let us consider the general non-linear regression model:

$$\mathbf{x} = f(\mathbf{z}, \boldsymbol{\theta}) + \epsilon,$$

- ▶ $\boldsymbol{\theta} \in \Theta$: parameter vector to be estimated,
 - ▶ $f \in \mathcal{F}$: possibly unknown non-linear function,
 - ▶ \mathbf{z} : random vector with possibly unknown pdf $p_Z \in \mathcal{K}$,
 - ▶ ϵ : random noise with possibly unknown pdf $p_\epsilon \in \mathcal{E}$.
- ▶ The set of all pdfs for \mathbf{x} is a semiparametric model of the form:

$$\mathcal{P}_{\boldsymbol{\theta}, f, p_Z, p_\epsilon} \triangleq \{p_X(\mathbf{x}|\boldsymbol{\theta}, f, p_Z, p_\epsilon), \boldsymbol{\theta} \in \Theta, f \in \mathcal{F}, p_Z \in \mathcal{K}, p_\epsilon \in \mathcal{E}\}.$$

- ▶ This model is a general form of a *semiparametric regression model*.

Examples: Autoregressive processes

- ▶ Consider the AR(p) process:

$$x_n = \sum_{i=1}^p \theta_i x_{n-i} + w_n, \quad n \in (-\infty, \infty)$$

- ▶ $\boldsymbol{\theta} \triangleq [\theta_1, \dots, \theta_p]$: parameter vector to be estimated.
 - ▶ w_n : i.i.d. innovations with unknown pdf $p_w \in \mathcal{W}$,
- ▶ Let $\mathbf{x} \in \mathbb{R}^N$ a vector of N observations from an AR(p).
- ▶ The set of all possible pdfs for $\mathbf{x} \in \mathbb{R}^N$ is a semiparametric model:

$$\mathcal{P}_{\boldsymbol{\theta}, p_w} \triangleq \{p_X | p_X(\mathbf{x} | \boldsymbol{\theta}, p_w), \boldsymbol{\theta} \in \Theta, p_w \in \mathcal{W}\}.$$

Outline of the talk

Why semiparametric models?

Semiparametric estimation in CES distributions

Le Cam theory 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: ⁴

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

- ▶ $h \in \mathcal{G}$, $h: \mathbb{R}_0^+ \rightarrow \mathbb{R}^+$ is the *density generator*,
 - ▶ $\boldsymbol{\mu} \in \mathbb{C}^N$ is the location vector,
 - ▶ $\Sigma \in \mathcal{M}_N$ is the (full rank) scatter matrix.
-
- ▶ Note that Σ and h are not jointly identifiable:

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

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

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

⁴E. 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: ⁵

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

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

⁵The 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. $\underline{\text{vec}}(\mathbf{A}) \triangleq [\mathbf{a}_{11}, \underline{\text{vec}}(\mathbf{A})^T]^T$.

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 $\boldsymbol{\theta}_0 \triangleq (\boldsymbol{\mu}_0^T, \boldsymbol{\mu}_0^H, \text{vec}(\mathbf{V}_{1,0})^T)^T$)?
 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: ⁶
 - ▶ *Semiparametric efficient score vector* $\bar{\mathbf{s}}_{\boldsymbol{\theta}_0, h_0}$,
 - ▶ *Semiparametric Fisher Information Matrix (SFIM)* $\bar{\mathbf{I}}(\boldsymbol{\theta}_0 | h_0)$.

⁶S. 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.

- ▶ By using the Wirtinger calculus, the “parametric” score vector for θ_0 is:

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

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

$$\bar{\mathbf{s}}_{\theta_0, h_0} = [\bar{\mathbf{s}}_{\boldsymbol{\mu}_0}^T, \bar{\mathbf{s}}_{\boldsymbol{\mu}_0^*}^T, \bar{\mathbf{s}}_{\text{vec}(\mathbf{v}_{1,0})}^T]^T = \mathbf{s}_{\theta_0, h_0} - \Pi(\mathbf{s}_{\theta_0, h_0} | \mathcal{T}_{h_0}).$$

- ▶ $\Pi(\mathbf{s}_{\theta_0, h_0} | \mathcal{T}_{h_0})$ indicates the orthogonal projection⁷ of \mathbf{s}_{θ_0} on the nuisance tangent space \mathcal{T}_{h_0} of $\mathcal{P}_{\boldsymbol{\theta}, h}$ evaluated at h_0 .
- ▶ $\Pi(\mathbf{s}_{\theta_0, h_0} | \mathcal{T}_{h_0})$ tells us the loss of information on the estimation of $\boldsymbol{\theta}_0$ due to the lack of knowledge of h_0 .

⁷P.J. Bickel, C.A.J. Klaassen, Y. Ritov and J.A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, 1993.

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

- ▶ 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 μ_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 $\boldsymbol{\mu} = \mathbf{0}$ without any loss of generality.
- ▶ In fact, even if $\boldsymbol{\mu} \neq \mathbf{0}$, we can always obtain the “centered data” as:

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

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

- ▶ In the rest of the talk, 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

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

Outline of the talk

Why semiparametric models?

Semiparametric estimation in CES distributions

Le Cam theory on one-step efficient estimators

A complex-valued R -estimator for shape matrix

Numerical results

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

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

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

- ▶ The Maximum Likelihood estimator for θ is:

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

- ▶ Solving the optimization problem may result to be a prohibitive task.
- ▶ In some cases, $\hat{\theta}_{ML}$ may not even exist.

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

- ▶ Recall the definition of score vector:

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

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

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

- ▶ Under Cramér-type regularity conditions ⁸ if $\hat{\boldsymbol{\theta}}_{ML}$ exists, then it satisfies:

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

⁸ 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. $E_0\{\mathbf{s}_{\boldsymbol{\theta}_0, h_0}^T(\mathbf{x})\mathbf{s}_{\boldsymbol{\theta}_0, h_0}(\mathbf{x})\} < \infty$,
3. The FIM $\mathbf{I}(\boldsymbol{\theta}) \triangleq \int \mathbf{s}_{\boldsymbol{\theta}, h_0}(\mathbf{x})\mathbf{s}_{\boldsymbol{\theta}, h_0}^T(\mathbf{x})p_X(\mathbf{x}|\boldsymbol{\theta})d\mathbf{x}$ is non-singular and continuous in $\boldsymbol{\theta}$.

Note that in the original work of Le Cam, these assumptions are strongly relaxed at a cost of a more abstract definition of score vector.

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

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

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

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

- ▶ $\nabla_{\theta}^T \Delta_{\tilde{\theta}, h_0}$ indicates the Jacobian matrix of Δ_{θ, h_0} at $\tilde{\theta}$.

Key point. It can be shown that:

$$\nabla_{\theta}^T \Delta_{\theta, h_0} \equiv -L^{1/2} \mathbf{I}_{h_0}(\theta) + o_P(1),^9 \quad \forall \theta \in \Theta,$$

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

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

⁹ Let x_l be a sequence of random variables. Then $x_l = o_P(1)$ if $\lim_{l \rightarrow \infty} \Pr \{ |x_l| \geq \epsilon \} = 0, \forall \epsilon > 0$ (convergence in probability to 0).

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

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

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

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

Properties:

P1 \sqrt{L} -consistency:

$$\sqrt{L}(\hat{\theta} - \theta_0) = O_P(1),^{10}$$

P2 Asymptotic normality and efficiency:

$$\sqrt{L}(\hat{\theta} - \theta_0) \underset{L \rightarrow \infty}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_{h_0}(\theta_0)^{-1}),$$

where $\mathbf{I}_{h_0}(\theta_0)^{-1} \equiv \text{CRB}(\theta_0)$.

¹⁰ 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*: $\boldsymbol{\theta}_0 = \text{vec}(\mathbf{V}_{1,0})$ has to be estimated in the presence of the unknown density generator h_0 .
- ▶ Let us introduce the *efficient central sequence* as:

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

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: ¹¹

$$\overline{\Delta}_{\theta, \hat{h}^*}(\mathbf{z}_1, \dots, \mathbf{z}_L) \Big|_{\theta = \hat{\theta}_{ML}} = \mathbf{0}.$$

where \hat{h}^* 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.

¹¹A. 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 $\mathbf{V}_{1,0}$.^{12,13}
- ▶ 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 estimator $\widehat{\mathbf{V}}_1$ as:

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

where $\widehat{\mathbf{V}}_1^*$ is any preliminary \sqrt{L} -consistent estimator of $\mathbf{V}_{1,0}$.

2. Robustify $\widehat{\mathbf{V}}_1$ by using a distribution-free, rank based, procedure.

¹² 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.

¹³ 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

A complex-valued R -estimator for shape matrix

Numerical results

Efficient central sequence and SFIM

► It can be shown that:

1. The efficient central sequence:

$$\bar{\Delta}_{\mathbf{V}_1, h_0} = -L^{-1/2} \mathbf{L}_{\mathbf{V}_1} \sum_{l=1}^L Q_l \psi_0(Q_l) \text{vec}(\mathbf{u}_l \mathbf{u}_l^H).$$

2. The Semiparametric FIM

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

► $Q_l \triangleq \mathbf{z}_l^H \mathbf{V}_1^{-1} \mathbf{z}_l \stackrel{d}{=} Q \sim P_{Q, h_0},$

► $\psi_0(q) \triangleq d \ln h_0(q) / dq,$

► $\mathbf{u}_l \sim \mathcal{U}(\mathbb{C}S^N),$

► $\mathbf{P} = [\mathbf{e}_2 | \mathbf{e}_3 | \dots | \mathbf{e}_{N^2}]^T,$

► $\Pi_{\text{vec}(\mathbf{I}_N)}^\perp = \mathbf{I}_{N^2} - N^{-1} \text{vec}(\mathbf{I}_N) \text{vec}(\mathbf{I}_N)^T,$

► $\mathbf{L}_{\mathbf{V}_1} = \mathbf{P} \left(\mathbf{V}_1^{-T/2} \otimes \mathbf{V}_1^{-1/2} \right) \Pi_{\text{vec}(\mathbf{I}_N)}^\perp$

► **Note that** $\psi_0(q)$ and the cdf P_{Q, h_0} of Q_l depends on the true and unknown $h_0!$

Rank-based approximation

- ▶ **Main idea:** Distribution-free approximation of the: ¹⁴
 - ▶ Efficient central sequence:

$$\tilde{\Delta}_{\hat{\mathbf{V}}_1^*} \triangleq -L^{-1/2} \mathbf{L}_{\hat{\mathbf{V}}_1^*} \sum_{l=1}^L K_h \left(\frac{r_l^*}{L+1} \right) \text{vec}(\hat{\mathbf{u}}_l^* (\hat{\mathbf{u}}_l^*)^H),$$

- ▶ Semiparametric FIM:

$$\hat{\Upsilon} \triangleq \hat{\alpha} \mathbf{L}_{\hat{\mathbf{V}}_1^*} \mathbf{L}_{\hat{\mathbf{V}}_1^*}^H,$$

- ▶ $\hat{\alpha}$ is a consistent estimator of $\alpha \triangleq \frac{E\{\mathcal{Q}^2 \psi_0(\mathcal{Q})^2\}}{N(N+1)}$ based on the following property of $\tilde{\Delta}_{\hat{\mathbf{V}}_1^*}$:

$$\tilde{\Delta}_{\hat{\mathbf{V}}_1^* + L^{-1/2} \mathbf{H}^0} - \tilde{\Delta}_{\hat{\mathbf{V}}_1^*} = -\alpha \mathbf{L}_{\hat{\mathbf{V}}_1^*} \mathbf{L}_{\hat{\mathbf{V}}_1^*}^H \text{vec}(\mathbf{H}^0) + o_P(1),$$

where \mathbf{H}^0 is a “small” Hermitian matrix s.t. $[\mathbf{H}^0]_{1,1} = 0$.

¹⁴ M. Hallin and B. J. M. Werker, “Semi-parametric efficiency, distribution-freeness and invariance,” *Bernoulli*, vol. 9, no. 1, pp. 137–165, 2003.

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: ¹⁵

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

- ▶ $\widehat{\mathbf{\Upsilon}}$ is an approximation of $\bar{\mathbf{I}}(\underline{\text{vecs}}(\mathbf{V}_1)|h_0)$.
- ▶ $\widetilde{\Delta}_{\widehat{\mathbf{V}}_1^*}$ is a distributionally-free approximation of the efficient central sequence $\bar{\Delta}_{\mathbf{V}_1}$.
- ▶ This R -estimator has the following desirable properties:
 1. *distributionally-robust* and
 2. *semiparametric efficient*,

¹⁵S. 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)

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

- ▶ $\{r_l^*\}_{l=1}^L$ are the ranks of the r. v. $\hat{Q}_l^* \triangleq \mathbf{z}_l^T [\widehat{\mathbf{V}}_1^*]^{-1} \mathbf{z}_l$,
- ▶ $\hat{\mathbf{u}}_l^* \triangleq \frac{[\widehat{\mathbf{V}}_1^*]^{-1/2} \mathbf{z}_l}{\sqrt{\hat{Q}_l^*}}$,
- ▶ $K_h(\cdot)$ is a “score” function based on $h \in \mathcal{G}$,
- ▶ $\hat{\alpha}$ is a consistent estimator of $\alpha \triangleq \frac{E\{Q^2 \psi_0(Q)^2\}}{N(N+1)}$,
- ▶ $\widehat{\mathbf{V}}_1^*$ 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: ¹⁶

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

where:

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

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

¹⁶S. 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 ¹⁷ for a discussion on how to build score functions.

¹⁷S. 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

A complex-valued R -estimator for shape matrix

Numerical results

Simulation set-up

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

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

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

where:

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

- ▶ The estimators $\hat{\boldsymbol{\mu}}_{Ty}$ and $\hat{\mathbf{V}}_{1, Ty}$ 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:

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

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

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

Simulation set-up

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

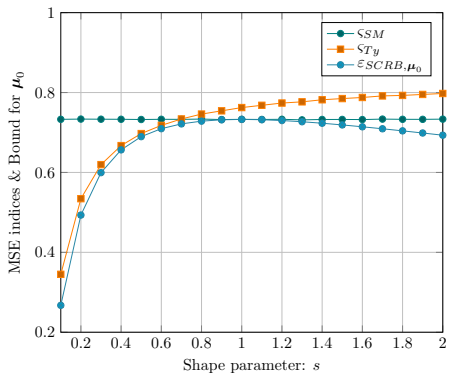
$$\varsigma_{\gamma} \triangleq \|E\{(\hat{\boldsymbol{\mu}}_{\gamma} - \boldsymbol{\mu}_0)(\hat{\boldsymbol{\mu}}_{\gamma}^a - \boldsymbol{\mu}_0^a)^H\}\|_F,$$

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

and γ indicates the relevant estimator at hand.

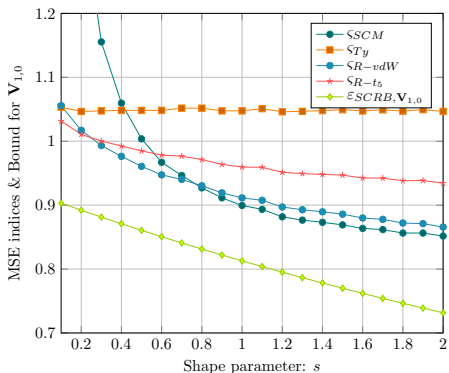
- ▶ As lower bounds, we use $\varepsilon_{CRB,\boldsymbol{\mu}_0} \triangleq \|\text{SCRb}(\boldsymbol{\mu}_0|h_0)\|_F$ and $\varepsilon_{SCRb,\mathbf{V}_{1,0}} \triangleq \|\text{SCRb}(\underline{\text{vec}}(\mathbf{V}_{1,0})|h_0)\|_F$.

MSE on μ_0 for GG data ($L = 5N$)



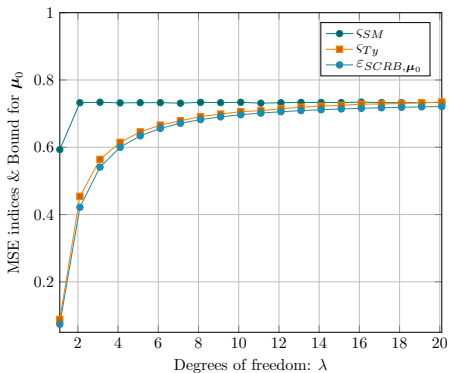
- ▶ $\hat{\mu}_{T_y}$ is almost efficient in heavy-tailed data ($0 < s < 1$),
- ▶ $\hat{\mu}_{T_y}$ outperforms $\hat{\mu}_{SM}$ that it is known to be non robust,
- ▶ $\hat{\mu}_{SM}$ is efficient in the Gaussian case ($s = 1$), and tends to have better performance than $\hat{\mu}_{T_y}$ for $s > 1$.

MSE on $V_{1,0}$ for GG data ($L = 5N$)



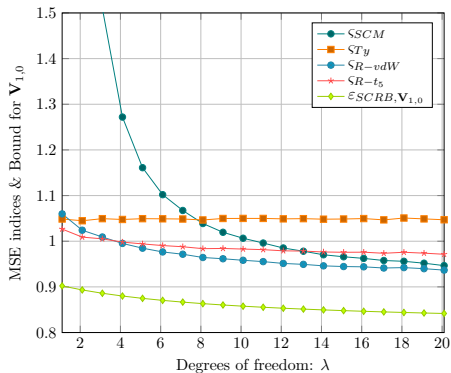
- ▶ $\hat{V}_{1,R-vdW}$ and $\hat{V}_{1,R-t5}$ outperform $\hat{V}_{1,Ty}$ for every values of s , i.e. for both heavy-tailed and light-tailed data.
- ▶ They also outperform the sample covariance matrix $\hat{V}_{1,SCM}$ in the presence of heavy-tailed data ($0 < s < 1$), while their MSE is of the same order for $s > 1$.

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



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

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



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

Finite-sample breakdown point (BP)

- ▶ Let $Z = \{\mathbf{z}_I\}_{I=1}^L \sim CES(\mathbf{0}, \mathbf{V}_1, h_0)$ be the “pure” GG data set.
- ▶ Let $Z_\varepsilon = \{\mathbf{z}_I\}_{I=1}^L \sim f_{Z_\varepsilon}$ be the ε -contaminated data set s.t.:

$$f_{Z_\varepsilon}(\mathbf{z}|\mathbf{V}_1, h_0, \varrho) = (1 - \varepsilon)CES(\mathbf{0}, \mathbf{V}_1, h_0) + \varepsilon q_Z(\varrho),$$

where $\varepsilon \in [0, 1/2]$ is a contamination parameter and the outlier $\tilde{\mathbf{z}} = \tau^{-1}\mathbf{u} \sim q_Z(\varrho)$ where $\tau \sim \text{Gam}(\varrho, 1/\varrho)$.

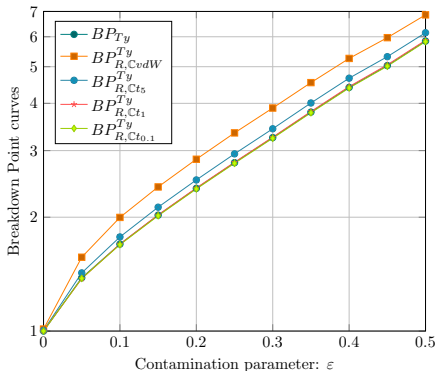
- ▶ Then the finite-sample BP curves can be evaluated as: ¹⁸

$$BP(\varepsilon) \triangleq \max \{ \lambda_1(\varepsilon), 1/\lambda_N(\varepsilon) \},$$

where $\lambda_i(\varepsilon)$ is the i -th ordered eigenvalue of the matrix $[\hat{\mathbf{V}}(Z)]^{-1}\hat{\mathbf{V}}(Z_\varepsilon)$, s.t. $\lambda_1(\varepsilon) \geq \dots \geq \lambda_N(\varepsilon)$.

¹⁸L. D’umberg and D. E. Tyler, “On the breakdown properties of some multivariate M -functionals,” *Scandinavian Journal of Statistics*, vol. 32, no. 2, pp. 247–264, 2005.

Finite-sample breakdown point (BP)



- ▶ All the BP curves, related to the resulting R -estimator, remain close to the Tyler's one for every value of ε .
- ▶ t_ν -scores with a small value of ν lead to more robust estimators.

Empirical Influence Function (EIF)

- ▶ Let $Z = \{\mathbf{z}_l\}_{l=1}^L \sim CES(\mathbf{0}, \mathbf{V}_1, h_0)$ be the “pure” GG data set.
- ▶ Let $\tilde{\mathbf{z}} = \tau^{-1}\mathbf{u}$ be the outlier where $\tau \sim \text{Gam}(\varrho, 1/\varrho)$.¹⁹
- ▶ The empirical influence function (EIF) can be evaluated as:²⁰

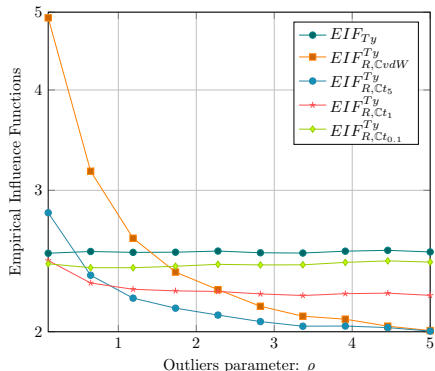
$$EIF \triangleq (L + 1) \|\hat{\mathbf{V}}(Z) - \hat{\mathbf{V}}(Z, \tilde{\mathbf{z}})\|_F.$$

- ▶ The EIF gives us a measure of the impact that a single outlier $\tilde{\mathbf{z}}$ has on the shape matrix estimator $\hat{\mathbf{V}}$ when it is added to the “pure” data set Z .

¹⁹Note that we can obtain “arbitrarily large” outlier by generating arbitrarily small values of $\tau \sim \text{Gam}(\varrho, 1/\varrho)$ by letting $\varrho \rightarrow 0$.

²⁰C. Croux, “Limit behavior of the empirical influence function of the median,” *Statistics & Probability Letters*, vol. 37, no. 4, pp. 331 – 340, 1998.

Empirical Influence Function (EIF)



- ▶ The EIFs of the proposed R -estimator remain bounded and close to the one of the Tyler's estimator for arbitrarily large values of $\|\tilde{\mathbf{z}}\|$ ($\rho \rightarrow 0$).

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

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

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.²¹

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

²¹Note 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 $\{x_l\}_{l=1}^L$ 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 $\{1, 2, \dots, L\}$,
3. For each $l = 1, \dots, L$, $K\left(\frac{r_l}{L+1}\right) = K(u_l) + o_P(1)$, where $K \in \mathcal{K}$ and $u_l \sim \mathcal{U}[0, 1]$ is a random variable uniformly distributed in $(0, 1)$.