Estimation for Models Defined by Conditions on Their L-Moments

Michel Broniatowski, Alexis Decurninge

Estimation for models defined by conditions on their L-moments

Michel Broniatowski and Alexis Decurninge

Abstract

This paper extends the empirical minimum divergence approach for models which satisfy linear constraints with respect to the probability measure of the underlying variable (moment constraints) to the case where such constraints pertain to its quantile measure (called here semi parametric quantile models). The case when these constraints describe shape conditions as handled by the L-moments is considered and both the description of these models as well as the resulting non classical minimum divergence procedures are presented. These models describe neighborhoods of classical models used mainly for their tail behavior, for example neighborhoods of Pareto or Weibull distributions, with which they may share the same first L-moments. The properties of the resulting estimators are illustrated by simulated examples comparing Maximum Likelihood estimators on Pareto and Weibull models to the minimum Chi-square empirical divergence approach on semi parametric quantile models, and others.

Index Terms

L-moments, divergence, quantile, semiparametric estimation.

I. Motivation and notation

A. First motivations

For univariate distributions, L-moments are expressed as expectations of particular linear combinations of order statistics. Let us consider r independent copies $X_1, ..., X_r$ of a random variable X with $\mathbb{E}(|X|)$ a finite number. The r-th L-moment is defined by

$$\lambda_r = \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k \binom{r-1}{k} \mathbb{E}[X_{r-k:r}] \tag{1}$$

where $X_{1:r} \leq ... \leq X_{r:r}$ denotes the order statistics. The four first L-moment can be considered as a measure of location, dispersion, skewness and kurtosis. Indeed $\lambda_1 = \mathbb{E}(X)$, λ_2 satisfies $\lambda_2 = (1/2)\,\mathbb{E}(|X - Y|)$ with Y an independent copy of X, λ_3 indicates the expected distance between the mean of the extreme terms and the median one in a sample of three i.i.d. replications of X, and λ_4 is an indicator of the expected distance between the extreme terms of a sample of four replicates of X with respect to three times the distance between the two central terms.

L-moments constitute a sound alternative to traditional moments as descriptors of a distribution since only finiteness of $\mathbb{E}(|X|)$ is needed in order to insure their existence. Since their introduction in Hosking's paper in 1990 ([20]), methods based on L-moments have become popular especially in applications dealing with heavy-tailed distributions. As mentioned in [20] and [21]: "The main advantage of L-moments over conventional moments is that L-moments, being linear functions of the data, suffer less from the effect of sampling variability: L-moments are *more robust* than conventional moments to outliers in the data and enable more secure inferences to be made from small samples about an underlying probability distribution". This motivates their success for the inference in models pertaining to the tail behavior of random phenomenons. Let us nevertheless draw the reader's attention on the fact that, since L-moments put positive weights on extreme values, an outlier can still break down L-moments values. However, their degree of nonrobustness is fairly constant contrary to the 2nd, 3rd, 4th central moments which are increasingly sensitive to the outliers.

A further motivation for L-moments is that, unlike conventional moments, if they exist, they always determine the distribution.

In this article, we will consider semi-parametric models conditioned by constraints on a finite number of L-moments. Let us mention two examples of such models describing neighborhoods of the Weibull and the Pareto models, which are classical benchmarks for the description of tail properties.

Example 1.1: We first consider the model which is the family of all the distributions of a random variable (r.v.) X whose second, third and fourth L-moments are given by:

$$\begin{cases} \lambda_2 = \sigma(1-2^{-1/\nu})\Gamma(1+1/\nu) \\ \lambda_3 = \lambda_2[3 - 2\frac{1-3^{-1/\nu}}{1-2^{-1/\nu}})] \\ \lambda_4 = \lambda_2[6 + \frac{5(1-4^{-1/\nu})-10(1-3^{-1/\nu})}{1-2^{-1/\nu}})] \end{cases} \tag{2}$$

LSTA, Université Pierre et Marie Curie, Paris, France

for any $\sigma > 0, \nu > 0$. These distributions share their first L-moments of order 2, 3 and 4 with those of a Weibull distribution with scale and shape parameter σ and ν. When X is substituted by $Y := X + a$ for some real number a then the distribution of Y is Weibull with a shifted support, hence with the same parameters σ and ν as X; the r.v. Y shares the same L-moments λ_r with those of X but for $r = 1$ and the model (2) describes a neighborhood of the continuum of all Weibull distributions on $[a, \infty)$ or on $(-\infty, a]$ when a belongs to $\mathbb{R}$. Hence this model aims at describing a shape constraint on the tail of the distribution of the data, independently of its location.

Example 1.2: Secondly, we consider the model which is the space of the distributions whose second, third and fourth L-moments are given by:

$$\left\{ \begin{array}{lcl} \lambda_2 & = & \frac{\sigma}{(1-\nu)(2-\nu)} \\ \lambda_3 & = & \lambda_2 \frac{1+\nu}{3-\nu} \\ \lambda_4 & = & \lambda_2 \frac{(1+\nu)(2+\nu)}{(3-\nu)(4-\nu)} \end{array} \right. \tag{3}$$

for any $\sigma > 0, \nu \in \mathbb{R}$. These distributions share their first L-moments with those of a generalized Pareto distribution with scale and shape parameter σ and ν. The same remark as in the above example holds; model (3) describes a neighborhood of the whole continuum of Pareto distributions on $[a, \infty)$ or on $(-\infty, a]$ when a belongs to $\mathbb{R}$.

Let us comment on these two examples. They generalize models implicitly considered when applying the method of L-moments by adding a further constraint (see e.g. [13], [32] for studies about the method of L-moments). They constitute neighborhoods of heavy-tailed distributions seen through shape characteristics, namely the 2nd, 3rd, 4th L-moments. They can then be seen as shape-oriented semi-parametric models, the constraints between L-moments expressing a particular tail behavior. Such L-moments neighborhoods show better separability properties then neighborhoods constituted by moment conditions between Weibull distributions and with GEV distributions; see Fig. 6 of [20] where the neighborhoods are constituted by distributions of samples drawn from these families and Fig. 7 of the same paper shows sample distributions of annual maximum hourly rainfall data.

B. Notation

Before any further discussion on the scope of the present paper, a few notation seems useful. For a non decreasing function F with bounded variation on any interval of $\mathbb{R}$ we denote $\mathbf{F}$ the corresponding positive $\sigma-$finite measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. For example when F is the distribution function of a probability measure, then this measure is denoted $\mathbf{F}$ or dF. Denote in this case

$$F^{-1}(u) := \inf \{x \in \mathbb{R} \text{ s.t. } F(x) \geq u\} \text{ for } u \in (0, 1)$$

the generalized inverse of F , a left continuous non decreasing function which is the quantile function of the probability measure $\mathbf{F}$. Denote accordingly $\mathbf{F}^{-1}$ or dF^{-1}, indifferently, the quantile measure with distribution function F^{-1}. If $x_1, .., x_n$ are n realizations of a random variable X with absolutely continuous probability measure $\mathbf{F}$ then the gaps in the empirical distribution function

$$F_n(x) := \frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty, x]}(x_i)$$

are of size $1/n$ and are located on the x_i's; the empirical quantile function satisfies

$$F_n^{-1}(u) = x_{i:n} \text{ when } \frac{i-1}{n} < u \leq \frac{i}{n}$$

and its gaps are given by

$$F_n^{-1}\left((i/n)^+\right) - F_n^{-1}((i/n)) = \mathbf{F}_n^{-1}(i/n) = x_{i+1:n} - x_{i:n}$$

where $x_{1:n} \leq ... \leq x_{n:n}$ denotes the ordered sample; the empirical quantile measure $\mathbf{F}_n^{-1}$ has as its support the uniformly sparsed points $\{1/n, 2/n, .., (n-1)/n\}$ and attributes masses equal to sampled spacings at those points; it follows that the empirical quantile measure is a positive finite measure with finite support. The quantile measure associated with the distribution function F^{-1} is also a positive $\sigma-$finite measure, defined on $(0, 1)$. The above construction defines the quantile measure from the probability measure, but the reciprocal construction will be used, starting from a quantile measure, defining its distribution function, turning to its inverse to define the distribution of a probability measure, and then to the probability measure itself.

II. L-MOMENTS

A. Properties

From the definition given in Section I, all L-moments λ_r but λ_1 are shift invariant, hence independent upon λ_1. If F is continuous, the expectation of the j-th order statistics $X_{j:r}$ is (see David p.33 [11])

$$\mathbb{E}[X_{j:r}] = \frac{r!}{(j-1)!(r-j)!} \int_{\mathbb{R}} x F(x)^{j-1} (1 - F(x))^{r-j} \mathbf{F}(dx). \tag{4}$$

The first four L-moments are

$$\begin{aligned}\lambda_1 &= \mathbb{E}[X]\\ \lambda_2 &= \tfrac{1}{2}\mathbb{E}[X_{2:2}-X_{1:2}]\\ \lambda_3 &= \tfrac{1}{3}\mathbb{E}[X_{3:3}-2X_{2:3}+X_{1:3}]\\ \lambda_4 &= \tfrac{1}{4}\mathbb{E}[X_{4:4}-3X_{3:4}+3X_{2:4}-X_{1:4}].\end{aligned}$$

Remark 2.1: The ratio $\frac{\lambda_2}{\lambda_1}$ is known as the Gini coefficient.

The expectations of the extreme order statistics characterize a distribution: if $\mathbb{E}(|X|)$ is finite, either of the sets $\{\mathbb{E}(X_{1:n}), n=1,..\}$ or $\{\mathbb{E}(X_{n:n}), n=1,..\}$ characterize the distribution of X; see [7] and [22]. Since the moments of order statistics are defined by the family of L-moments, those also characterize the distribution of X.

Remark 2.2: L-moments may be generalized in order to focus on the behavior of a distribution in the tail; Wang [34] considers so called LH moments for statistical analysis of extreme events, defined by

$$\begin{aligned}\lambda_1^k &:= E[X_{k+1:k+1}]\\ \lambda_2^k &:= \frac{1}{2}E[X_{k+2:k+2}-X_{k+1:k+2}]\\ \lambda_3^k &:= \frac{1}{3}E[X_{k+3:k+3}-2X_{k+2:k+3}+X_{k+1:k+1}]\\ \lambda_4^k &:= \frac{1}{4}E[X_{k+4:k+4}-3X_{k+3:k+4}+3X_{k+2:k+4}-X_{k+1:k+4}]\\ &\text{etc.}\end{aligned}$$

When $k=0$, those coincide with L-moments. Moreover, λ_1^k is the location of $X_{k+1:k+1}$; λ_2^k is half the last gap in a sample of size $k+1$ (spreadness); λ_3^k is the asymmetry in the upper tail for large k; λ_4^k describes peakedness on the upper part of the distribution. When k is large, characteristics of the upper part of the distribution are captured by those indices. [34] provides estimators for $\lambda_j^k, 1\leq j\leq 4$, and LH moments for GEV distributions. The methodology which is proposed in this article could be adapted to this setting; also k could be made a function of the size of the observed sample.

Remark 2.3: We can define from the quantile function $F^{-1}:[0;1]\to\mathbb{R}$ an associated measure on $\mathcal{B}([0;1])$

$$\mathbf{F}^{-1}(B)=\int_0^1 \mathbb{1}_{x\in B}dF^{-1}(x)\in\mathbb{R}\cup\{-\infty,+\infty\}.$$

The above integral is a Riemann-Stieltjes integral. It defines a σ-finite measure since F^{-1} has bounded variations on every interval of the form $[a,b]$ with $0<a\leq b<1$. For any $\mathbf{F}^{-1}$-measurable function $a:\mathbb{R}\to\mathbb{R}$, it holds

$$\int_0^1 a(x)dF^{-1}(x)=\int_0^1 a(x)\mathbf{F}^{-1}(dx).$$

Writing the L-moments of a distribution F as an inner product of the corresponding quantile function with a specific complete orthogonal system of polynomials in $L^2(0,1)$ is a cornerstone in the derivation of statistical inference in SPLQ models. The shifted Legendre polynomials define such a system of functions.

Definition 2.1: The shifted Legendre polynomial of order r is

$$L_r(t)=\sum_{k=0}^{r}(-1)^k\binom{r}{k}^2 t^{r-k}(1-t)^k. \tag{5}$$

For $r\geq 1$ define K_r as the integrated shifted Legendre polynomials

$$K_r(t)=\int_0^t L_{r-1}(u)du=-t(1-t)\frac{J_{r-2}^{(1,1)}(2t-1)}{r-1} \tag{6}$$

with $J_{r-2}^{(1,1)}$ the corresponding Jacobi polynomial (see [19])

$$J_{r-2}^{(1,1)}(2t-1)=\frac{\Gamma(r)}{(r-2)!\Gamma(r+1)}\sum_{k=0}^{r-2}\binom{r-2}{k}\frac{\Gamma(r+1+k)}{\Gamma(2+k)}(t-1)^k.$$

The following result holds.

Proposition 2.1: Let F be any cdf and assume that $\int|x|\,dF(x)$ is finite. Then for any $r\geq 1$, it holds

$$\lambda_r=\int_0^1 F^{-1}(t)L_{r-1}(t)dt=\int_0^1 F^{-1}(t)dK_r(t) \tag{7}$$

where the last integral is the Stieltjes integral of F^{-1} with respect to the function $t\mapsto K_r(t)$.

Proof: The proof is based on the following fundamental Lemma (see for example [14]).

Lemma 2.1: Let U be a uniform random variable on [0;1] and X be a random variable with F. Then $F^{-1}(U) =_d X$.

Let $U_1, ..., U_r$ be r independent random variable uniformly distributed on $[0;1]$ and denote by $U_{1:r} \leq ... \leq U_{r:r}$ the ordered statistics. Then

$$(X_{1:r}, ..., X_{r:r}) \stackrel{d}{=} (F^{-1}(U_{1:r}), ..., F^{-1}(U_{r:r})).$$

Hence for $1 \leq j \leq r$

$$\begin{aligned} \mathbb{E}[X_{j:r}] &= \mathbb{E}[F^{-1}(U_{j:r})] \\ &= \frac{r!}{(j-1)!(r-j)!} \int_0^1 F^{-1}(t) t^{j-1} (1-t)^{r-j} dt, \end{aligned}$$

which ends the proof of Proposition 2.1. ■

Before going any further, we present a useful Lemma, the proof of which is also deferred to the Appendix.

Lemma 2.2: Let a be a real-valued function such that $\int_{\mathbb{R}} a(x) dF(x) < \infty$. Then

$$\int_{\mathbb{R}} a(x) d\mathbf{F}(x) = \int_0^1 a(F^{-1}(t)) dt. \tag{8}$$

Similarly if $t \to b(t)$ is a real-valued function such that $\int_0^1 b(t) \mathbf{F}^{-1}(dt) < \infty$, then

$$\int_0^1 b(t) \mathbf{F}^{-1}(dt) = \int_0^1 b(F(x)) dx. \tag{9}$$

Remark 2.4: As a consequence of Lemma 2.2 and equation (7), it holds

$$\lambda_r = \int_0^1 x dK_r(F(x)).$$

This is a particular case of Lemma 8.1.1.A in [29] stipulating that if $J : [0;1] \to \mathbb{R}$ is such that $\int_0^1 J(t) F^{-1}(t) dt$ is finite and if $K(t) := \int_0^t J(u) du$, then

$$\int_0^1 J(t) F^{-1}(t) dt = \int_0^1 x dK(F(x)).$$

Remark 2.5: If we consider a multinomial distribution with support $x_1 \leq x_2 \leq ... \leq x_n$ and associated weights $\pi_1, ..., \pi_n$ ($\sum_{i=1}^n \pi_i = 1$), we get

$$\begin{aligned} \lambda_r &= \sum_{i=1}^n w_i^{(r)} x_i \\ &= \sum_{i=1}^n \left[K_r \left(\sum_{a=1}^i \pi_a \right) - K_r \left(\sum_{a=1}^{i-1} \pi_a \right) \right] x_i \\ &= \int_0^1 L_{r-1}(t) Q_\pi(t) dt \end{aligned}$$

with

$$Q_\pi(t) = \left\{ \begin{array}{ll} x_1 & \text{if} \quad 0 \leq t \leq \pi_1 \\ x_i & \text{if} \quad \sum_{a=1}^{i-1} \pi_a < t \leq \sum_{a=1}^i \pi_a \end{array} \right. .$$

This example illustrates Remark 2.4.

Figure 1 provides the first weight $w_i^{(r)}$ when the x_i's are equally sparsed on $[0,1]$ with equal weights $\pi_1 = ... = \pi_n = 1/n$.

We state the following characterization for the L-moments with order larger or equal to 2.

Proposition 2.2: If $r \geq 2$ and $\int_{\mathbb{R}} |x| \, dF(x) < +\infty$, then

$$\lambda_r = -\int_0^1 K_r(t) \mathbf{F}^{-1}(dt). \tag{10}$$

Proof: This result follows as an application of Fubini-Tonelli Theorem. Indeed

$$\begin{aligned} \lambda_r &= \int_0^1 F^{-1}(t) dK_r(t) \\ &= \int_0^1 \int_0^t \mathbf{F}^{-1}(du) dK_r(t) \\ &= \int_0^1 \int_0^1 \mathbb{1}_{0 \leq u \leq t} \mathbf{F}^{-1}(du) dK_r(t). \end{aligned}$$

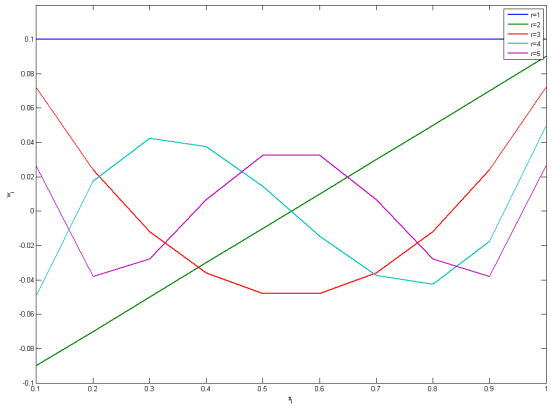


Fig. 1. Weights $w_i^{(r)}$ for the uniform law with a support containing 10 points

This last equality holds since $(u,t) \mapsto \mathbb{1}_{0\leq u\leq t}$ is measurable with respect to the measure $\mathbf{F}^{-1} \times dK_r$ since $\mathbb{E}[X] < \infty$. Applying Fubini-Tonelli Theorem, it holds

$$\begin{aligned}\lambda_r &= \int_0^1 \int_0^1 \mathbb{1}_{0\leq u\leq t} dK_r(t)\mathbf{F}^{-1}(du) \\ &= \int_0^1 \int_0^1 \left[K_r(1) - K_r(u)\right]\mathbf{F}^{-1}(du) \\ &= -\int_0^1 K_r(u)\mathbf{F}^{-1}(du)\end{aligned}$$

since $K_r(1) = 0$ for $r > 1$. ■

Remark 2.6: That (10) does not hold for $r = 1$ follows from the fact that if $G = F(.+a)$ for some $a \in \mathbb{R}$, then $\mathbf{G}^{-1} = \mathbf{F}^{-1}$. Hence, SPLQ models are shift-invariant. This can also be seen setting $r = 1$ in the right-hand side of (10); in this case, the integral is infinite (but if $supp(\mathbf{F})$ is bounded) whereas λ_1 is supposed to be finite.

B. Estimation of L-moments

Let $x_1, ..., x_n$ be iid realizations of a random variable X with distribution F and L-moments λ_r. Define F_n the empirical cdf of the sample and l_r the corresponding plug-in estimator of λ_r,

$$l_r = \int_0^1 F_n^{-1}(t)L_{r-1}(t)dt. \tag{11}$$

This estimator of λ_r is biased as quoted in [20] and [36]. l_r is usually termed as a V-statistics. As noted upon in [20] and [36], the unbiased estimators of L-moments are the following U-statistics

$$l_r^{(u)} = \frac{1}{\binom{n}{r}} \sum_{1\leq i_1<\cdots<i_r\leq n} \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k \binom{r-1}{k} x_{i_{r-k}:n}.$$

Remark 2.7: An alternative definition for l_r as in (11) can be stated as follows. Conditionally on the realizations $x = (x_1, ..., x_n)$, define the uniform distribution on x. Then l_r is the discrete L-moment of order r of this conditional distribution. It can therefore be defined through

$$l_r = \frac{1}{\binom{r+n-1}{n-1}} \sum_{1\leq i_1\leq\cdots\leq i_r\leq n} \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k \binom{r-1}{k} x_{i_{r-k}:n}.$$

Let us now extend Definition 1 of the L-moments as follows. Let $(i_1, ..., i_r)$ be drawn without replacement from $\{1, ..., r\}$. We then define $x_{(i_1)} \leq ... \leq x_{(i_r)}$ the corresponding ordered observations and

$$\lambda_r^{(u)} = \frac{1}{r}\sum_{k=0}^{r-1}(-1)^k \binom{r-1}{k} \mathbb{E}[x_{(i_{r-k})}]$$

where the expectation is taken under the extraction process. Then $\lambda_r^{(u)}$ and $l_r^{(u)}$ coincide.
Although $l_r^{(u)}$ is unbiased, for sake of simplicity only l_r which is asymptotically unbiased, will be used in the sequel.

These two estimators l_r and $l_r^{(u)}$ of the L-moment λ_r have the same asymptotic properties.

Proposition 2.3: Let us suppose that F has finite variance. Then, for any $m \geq 1$

$$\sqrt{n}\left[\begin{pmatrix} l_1 \\ \vdots \\ l_m \end{pmatrix} - \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix}\right] \rightarrow_d \mathcal{N}_m(0, \Lambda)$$

where $\mathcal{N}_m$ denotes the multivariate normal distribution and the elements of Λ are given by

$$\begin{aligned} \Lambda_{rs} &= \iint_{x<y} \Big[L_{r-1}(F(x))L_{s-1}(F(y)) \\ &+ L_{r-1}(F(y))L_{s-1}(F(x))\Big]F(x)(1-F(y))dxdy \end{aligned}$$

Furthermore, the same property holds for $l_1, .., l_r$ substituted by $l_1^{(u)}, ..., l_m^{(u)}$.

Proof: This is a plain consequence of Theorem 6 in [30]. See also [20] for an evaluation of the bias of l_r. ■

III. Models defined by moment and L-moment equations

A. Models defined by moment conditions

We now turn back to our topics.

Models defined as in Examples 1.1 and 1.2 extend the classical parametric ones, and are defined through some constraints on the form of the distributions. They can be paralleled with models defined through moments conditions defined as follows.

Let θ in Θ , an open subset of $\mathbb{R}^d$ and let $g : (x, \theta) \in \mathbb{R} \times \Theta \rightarrow \mathbb{R}^l$ be a l-valued function, each component of which is parametrized by $\theta \in \Theta \subset \mathbb{R}^d$. Define the submodel

$$M_\theta := \left\{\mathbf{F} \text{ s.t. } \int_{\mathbb{R}} g(x, \theta)\mathbf{F}(dx) = 0\right\}$$

and the semi parametric model defined by

$$\mathcal{M} := \bigcup_{\theta\in\Theta} M_\theta. \tag{12}$$

These semiparametric models are defined by l conditions pertaining to l moments of the distributions and are widely used in applied statistics. When the dimension d of the parameter space exceeds l, no plug-in method can achieve any inference on θ; however, various techniques have been proposed in this case; see for example Hansen [18], who defined the Generalized Method of Moments (GMM) and Owen, who defined the so-called empirical likelihood approach [26]. Later, Newey and Smith [25] or Broniatowski and Keziou [5] proposed a refinement of the GMM approach minimizing a divergence criterion over the model. A major feature of models defined by (12) lies in their linearity with respect to the cumulative distribution function (cdf) which brings a dual formulation of the minimization problem. Duality results easily lead to the consistency and the asymptotic normality of the estimators of θ; see [5], [25].

Example 3.1: We can sometimes face distributions with constraints pertaining to the two first moments. For example, Godambe and Thompson [15] considered the distributions verifying $\mathbb{E}[X] = \theta$ and $\mathbb{E}[X^2] = h(\theta)$ with a known function h. Then, with our notations $l = 2$ and $g(x, \theta) = (x - \theta, x^2 - h(\theta))$.

Example 3.2: Consider the distributions F such that for some θ it holds $F(y) = 1 - F(-y) = \theta$ [5]. This corresponds to a moment condition model with $l = 2$ and $g(x, \theta) = (\mathbb{1}_{]-\infty;y]}(x) - \theta, \mathbb{1}_{[y;+\infty[}(x) - \theta)$. The condition on the model is the existence of some θ such that the left and right quantiles of order θ are $-y$ and $+y$ for some given y.

B. Models defined by quantile conditions

Similarly as for models defined by (12), we can introduce semiparametric linear quantile (SPLQ) models through

$$\bigcup_{\theta\in\Theta} L_\theta := \bigcup_{\theta\in\Theta} \left\{\mathbf{F} \text{ s.t. } \int_0^1 F^{-1}(u)k(u, \theta)du = -f(\theta)\right\} \tag{13}$$

where $\Theta \subset \mathbb{R}^d$, $k : (u, \theta) \in [0; 1] \times \Theta \rightarrow \mathbb{R}^l$ and $f : \Theta \rightarrow \mathbb{R}^l$. In the above display, in accordance with the above notation, F^{-1} denotes the generalized inverse function of F, the distribution function of the measure $\mathbf{F}$. We will consider the case when k is a function of u only; this class contains many examples, typically models defined by a finite number of constraints on functions of the moments of the order statistics.

In the following, the r-th shifted Legendre polynomial (see Definition 2.1) is denoted by L_r.

Example 3.3: Turning back to Example 1.1, we define k and f by

$$k(u,\theta) = -\begin{pmatrix} L_2(u) \\ L_3(u) \\ L_4(u) \end{pmatrix}$$

and

$$f(\theta) = \begin{pmatrix} \sigma(1-2^{-1/\nu})\Gamma(1+1/\nu) \\ f_2(\theta)[3-2\frac{1-3^{-1/\nu}}{1-2^{-1/\nu}})] \\ f_2(\theta)[6+\frac{5(1-4^{-1/\nu})-10(1-3^{-1/\nu})}{1-2^{-1/\nu}})] \end{pmatrix}$$

where $\theta = (\sigma, \nu) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$ and $u \in [0;1]$; hence (13) holds.

Example 3.4: Similarly, in case we consider Example 1.2, we define k and f by

$$k(u,\theta) = -\begin{pmatrix} L_2(u) \\ L_3(u) \\ L_4(u) \end{pmatrix}$$

and

$$f(\theta) = \begin{pmatrix} f_2(\theta) \\ f_3(\theta) \\ f_4(\theta) \end{pmatrix} = \begin{pmatrix} \frac{\sigma}{(1+\nu)(2+\nu)} \\ f_2(\theta)\frac{1-\nu}{3+\nu} \\ f_2(\theta)\frac{(1-\nu)(2-\nu)}{(3+\nu)(4+\nu)} \end{pmatrix}$$

where $\theta = (\sigma, \nu) \in \mathbb{R}_+^* \times \mathbb{R}$ and $u \in [0;1]$, in accordance with (13).

Example 3.5: If a distribution is symmetric, it holds that $\lambda_{2r+1} = 0$ for $r \geq 1$.
Defining

$$k(u) = \begin{pmatrix} -L_2(u) \\ -L_3(u) \end{pmatrix},$$

and

$$f(\theta) = \begin{pmatrix} \theta \\ 0 \end{pmatrix},$$

is a way to take symmetry into account while considering the second L-moment.

Example 3.6: As stipulated by Gourieroux [16], financial and insurance risks can be seen as SPLQ constraints. In particular, Wang [35] defined Distortion Risk Measure (DRM) as

$$\text{DRM} = \int g(u)F^{-1}(u)du.$$

where g is the so-called distortion function representing the modeled risk. One example of DRM is the well-known conditional value-at-risk (with parameter $0 < \alpha < 1$) is expressed as $\int F^{-1}(u)\mathbb{1}_{u>\alpha}du$.

$$k(u) = \begin{pmatrix} \mathbb{1}_{u>0.90} \\ \mathbb{1}_{u>0.95} \\ \mathbb{1}_{u>0.99} \end{pmatrix},$$

and

$$f(\theta) = \begin{pmatrix} f_1(\theta) \\ f_2(\theta) \\ f_3(\theta) \end{pmatrix},$$

the resulting model is seen through risk measure and constraints between risks can be taken into account. In [16] for example, the function f comes from a parameterization of the quantile function (which is natural in the modelization of financial risks) and such models provide flexibility with respect to this parameterization, describing a "neighborhood of the risk".

Remark 3.1: The order statistics given by equation (4) can be written as

$$\mathbb{E}[X_{j:r}] = \int_0^1 P_{j:r}(u)F^{-1}(u)du$$

where the polynomials $P_{j:r}$ are given by

$$P_{j:r}(u) = \frac{r!}{(j-1)!(r-j)!}u^{j-1}(1-u)^{r-j}.$$

Any linear combination of moments of order statistics can be written as

$$\sum_{i=1}^{r} a_j \mathbb{E}[X_{j:r}] = \int_0^1 P_a(u) F^{-1}(u) du$$

with coefficients a_j's belonging to $\mathbb{R}$ and

$$P_a(u) = \sum_{i=1}^{r} a_j P_{j:r}(u).$$

These models are SPLQ (see (13)) with

$$\mathcal{L} := \bigcup_\theta L_\theta = \bigcup_\theta \left\{ F \text{ s.t. } \int_0^1 P(u) F^{-1}(u) du = -f(\theta) \right\} \tag{14}$$

where $P : u \in [0;1] \mapsto P(u) \in \mathbb{R}^l$ is an array of l polynomials.

Even if we restrict ourselves in the sequel to models defined by L-moment conditions, we conjecture that the introduced estimators may be used to infer on any SPLQ model and in particular on the model of Example 3.6 and on models defined by L-statistics constraints.

In the present paper we consider models defined by l constraints on their first L-moments, namely satisfying

$$\mathbb{E}\left[\frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} X_{k:r} \right] = f_r(\theta) \quad 1 \le r \le l \tag{15}$$

where Θ is some open set in $\mathbb{R}^d$ and $f_r : \Theta \to \mathbb{R}$ are some given functions defined on Θ , $2 \le r \le l$.
Those models are SPLQ, with $(u, \theta) \mapsto k(u, \theta)$ independent on θ, defined by

$$k(u, \theta) = -L(u) := - \begin{pmatrix} L_1(u) \\ \vdots \\ L_l(u) \end{pmatrix}. \tag{16}$$

It is natural to propose estimation procedures for L-moment condition models based on a minimization of a divergence. Models (13) do not enjoy linearity with respect to the cdf but with respect to the quantile function. Thus, as developed for models defined by (12), we propose to minimize a divergence criterion built on quantiles.
A duality result and the subsequent consistency and asymptotic normality for the corresponding family of estimators are presented in Sections V and VII. Furthermore, Section VI proposes a reformulation of estimators as a minimum of an energy functional.

In the following, the transpose of a vector a will be denoted a^T and if F and G are two cdf's, then $F \ll G$ means that F is absolutely continuous with respect to G. The Lebesgue measure on $\mathbb{R}$ is denoted $d\lambda$ or dx, according to the common use in the context.

IV. Minimum of φ-divergence estimators

Estimation, confidence regions and tests based on moment conditions models have evolved over thirty years. Hansen and Owen respectively proposed the generalized method of moments (GMM) [17] and the empirical likelihood (EL) estimators [26]. Newey and Smith [25] introduced the generalized empirical likelihood (GEL) family of estimators encompassing the previous estimators. They also proposed the dual versions of the GEL estimators, the minimum discrepancy estimators (MD). These estimators are the solution of the minimization of a divergence with constraints corresponding to the model; see also Broniatowski and Keziou [5] for an approach through duality and properties of the inference under misspecification. In the quantiles framework, Gourieroux proposed an adaptation of GMM estimators in [16] for a parametric model seen through its quantile function $F^{-1}(t, \theta)$. In the following, we will consider inference based on divergences in order to present estimators for models defined by L-moments conditions.

A. φ-divergences

Let $\varphi : \mathbb{R} \to [0, +\infty]$ be a strictly convex function with $\varphi(1) = 0$ such that $\text{dom}(\varphi) = \{x \in \mathbb{R} | \varphi(x) < \infty\} := (a_\varphi, b_\varphi)$ with $a_\varphi < 1 < b_\varphi$. If $\mathbf{F}$ and $\mathbf{G}$ are two σ-finite measures of $(\mathbb{R}, B(\mathbb{R}))$ such that $\mathbf{G}$ is absolutely continuous with respect to $\mathbf{F}$, we define the divergence between $\mathbf{F}$ and $\mathbf{G}$ by :

$$D_\varphi(\mathbf{G}, \mathbf{F}) := \int_{\mathbb{R}} \varphi\left(\frac{dG}{dF}(x) \right) \mathbf{F}(dx) \tag{17}$$

where $\frac{dG}{dF}$ is the Radon-Nikodym derivative. When $\mathbf{G}$ is not absolutely continuous with respect to $\mathbf{F}$ then we define $D_\varphi(\mathbf{G},\mathbf{F}) := +\infty$. It is clear that when $\mathbf{F}=\mathbf{G}$, $D_\varphi(\mathbf{G},\mathbf{F})=0$. Furthermore, since φ is supposed to be strictly convex,

$$D_\varphi(\mathbf{G},\mathbf{F})=0 \text{ if and only if } \mathbf{F}=\mathbf{G}.$$

These divergences were independently introduced by Csiszar [9] or Ali and Silvey [1] in the context of probability measures. The definition stated in (17) holds for any σ-finite measures even if our notation refers to probability measures. Indeed in the sequel we will consider divergences between quantile measures which are σ-finite but may be not finite. See Liese [23] who also considered divergences between σ-finite measures.

Example 4.1: The class of power divergences parametrized by $\gamma \geq 0$ is defined through the functions

$$x \mapsto \varphi_\gamma(x) = \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma-1)}.$$

The domain of φ_γ depends on γ. The Kullback-Leibler divergence is associated to $x>0 \mapsto \varphi_1(x) = x\log(x) - x + 1$, the modified Kullback-Leibler (KL_m) divergence to $x>0 \mapsto \varphi_0(x) = -\log(x) + x - 1$, the χ^2-divergence to $x \in \mathbb{R} \mapsto \varphi_2(x) = 1/2(x-1)^2$, etc.

B. M-estimates with L-moments constraints

1) Minimum of φ-divergences for probability measures: A plain approach to inference on θ consists in mimicking the empirical minimum divergence one, substituting the linear constraints with respect to the distribution by the corresponding linear constraints with respect to the quantile measure, and minimizing the divergence between all probability measures satisfying the constraint and the empirical measure $\mathbf{F}_n$ pertaining to the data set. More formally this yields to the following program.

Denote by M the set of all probability measures defined on $\mathbb{R}$. For a given p.m. $\mathbf{F}$ in M we consider the submodel which consists in all p.m's $\mathbf{G}$ in M, absolutely continuous with respect to F, and which satisfy the constraints on their first L-moments for a given $\theta \in \Theta$. Identifying a measure $\mathbf{G}$ with its distribution function G we define

$$L_\theta^{(0)}(\mathbf{F}) = \left\{ \mathbf{G} \in M \mid \mathbf{G} \ll \mathbf{F}, \int_0^1 L(t) G^{-1}(t) dt = -f(\theta) \right\}.$$

Probability measures $\mathbf{G}$ satisfying the constraints and bearing their mass on the sample points belong to $L_\theta^{(0)}(\mathbf{F}_n)$. For any parameter $\theta \in \Theta$, the distance between $\mathbf{F}$ and the submodel $L_\theta^{(0)}(\mathbf{F})$ is defined by

$$D_\varphi(L_\theta^{(0)}(\mathbf{F}),\mathbf{F}) = \inf_{\mathbf{G}\in L_\theta^{(0)}(\mathbf{F})} D_\varphi(\mathbf{G},\mathbf{F}),$$

and its plug-in estimator is

$$D_\varphi(L_\theta^{(0)}(\mathbf{F}_n),\mathbf{F}_n) = \inf_{\mathbf{G}\in L_\theta^{(0)}(\mathbf{F}_n)} D_\varphi(\mathbf{G},\mathbf{F}_n).$$

which measures the distance between the empirical measure $\mathbf{F}_n$ and the class of all the probability measures supported by the sample and which satisfy the L-moment conditions for a given θ.

A natural estimator for θ may be defined by

$$\begin{aligned} \hat{\theta}_n^{(0)} &= \arg\inf_{\theta\in\Theta} D_\varphi(L_\theta^{(0)}(\mathbf{F}_n),\mathbf{F}_n) \\ &= \arg\inf_{\theta\in\Theta} \inf_{\mathbf{G}\in L_\theta^{(0)}(F_n)} \frac{1}{n}\sum_{i=1}^n \varphi(n\mathbf{G}(x_i)). \end{aligned} \tag{18}$$

Unfortunately, existence of this estimator may not hold. Indeed, we cannot assess that $L_\theta^{(0)}(\mathbf{F}_n)$ is not empty : its elements are multinomial distributions $\sum_{i=1}^n w_i \delta_{x_i}$ whose weights $w_1,...,w_n$ are solutions of a family of $l-1$ polynomial algebraic equation of degree l

$$\sum_{i=1}^n K_r\left(\sum_{a=1}^i w_a\right)(x_{i+1:n} - x_{i:n}) = -f_r(\theta);\ 2 < r \leq l.$$

To our knowledge, general conditions of existence for the solutions of such problems do not exist even if we consider signed weights w_i.

Bertail in [2] proposes a linearization of the constraint in (18). We here prefer to switch to a different approach. If we consider the L-moment equation (15), we see that the quantile function plays a similar role as the distribution function in the classical moment equations. We will then change the functional to be minimized in order to be able to use duality for the optimization of the inner step.

2) Minimum of φ-divergences for quantile measures: We have seen that the characterization of the L-moments given by the equation (15) uses the quantile measure $\mathbf{F}^{-1}$, which is defined by the generalized inverse function of F. If $\mathbf{F}^{-1}$ is absolutely continuous, we can define the quantile-density $q(u) = (F^{-1})'(u)$. This density was called "sparsity" function by Tukey [31] as it represents the sparsity of the distribution at the cumulating weight $u \in [0;1]$. This is clear when we look at the empirical version of this measure which is composed by nothing but the increments of the sample. Some other approach, handling properties of the inverse function of $(F^{-1})'$, have been proposed by Parzen [27]. He claims that the inference procedures based on $(F^{-1})'$ possesses inherent robustness properties.

Define

$$K(u) = \begin{pmatrix} K_2(u) \\ \vdots \\ K_l(u) \end{pmatrix}$$

and

$$f(u) := f^{(2:l)}(u) = \begin{pmatrix} f_2(u) \\ \vdots \\ f_l(u) \end{pmatrix}.$$

For any θ in Θ the submodel which consists of all p.m's $\mathbf{G}$ with mass on the sample points is substituted by the set of all quantile measures denoted $\mathbf{G}^{-1}$ which have masses on subsets of $\{1/n, 2/n, .., (n-1)/n\}$ and whose distribution functions coincide with the generalized inverse functions of elements in $L_\theta^{(0)}(\mathbf{F}_n)$.
As in the case of divergence minimization for models constrained by moment conditions, we will relax the positivity for the masses of the quantile measures (see [5]). Let then N be the class of all $\sigma-$finite signed measures on $\mathbb{R}$. Let $L(u) := (L_2(u), .., L_l(u))^T$ for all u in $(0,1)$. Introducing signed measures makes sense when the domain of the function φ is not restricted to $\mathbb{R}^+$, as occurs for the chi-square divergence φ_2 . Making use of (15) define

$$\begin{aligned} L_\theta(\mathbf{F_n}^{-1}) &:= \left\{ \mathbf{G}^{-1} \in N \mid \mathbf{G}^{-1} \ll \mathbf{F}_n^{-1}, \int_0^1 L(u) G^{-1}(u) du = -f(\theta) \right\} \\ &= \left\{ \mathbf{G}^{-1} \in N \mid \mathbf{G}^{-1} \ll \mathbf{F}_n^{-1}, \int_0^1 K(u) \mathbf{G}^{-1}(du) = f(\theta) \right\} \end{aligned}$$

the family of all measures $\mathbf{G}^{-1}$ with support included in $\{1/n, 2/n, .., (n-1)/n\}$ which satisfy the $l-1$ constraints pertaining to the L-moments; see (13). Note that when $\mathbf{F}$ bears an atom then for large enough n, $\mathbf{G}^{-1}$ in $L_\theta(\mathbf{F_n}^{-1})$ has a support strictly included in $\{1/n, 2/n, .., (n-1)/n\}$.
Since the measure $\mathbf{G}^{-1}$ is not necessarily positive, its distribution function G^{-1} is not necessarily a generalized inverse of a function G; we will however inherit of the notation G^{-1} from the case when $\mathbf{G}^{-1}$ is a positive measure to denote its distribution function. If $\mathbf{G}^{-1}$ is positive, the mass of $\mathbf{G}^{-1}$ at point i/n is a spacing $\Delta_i := y_{i+1:n} - y_{i:n}$ where $y_{i:n}$ is the $i-th$ order statistics of the sample $y_1, .., y_n$ generating the empirical distribution function G.

A natural proposal for an estimation procedure in the SPLQ model is then to consider the minimum of a φ-divergence between quantile measures through

$$\hat{\theta}_n = \arg \inf_{\theta \in \Theta} \inf_{\mathbf{G}^{-1} \in L_\theta(\mathbf{F_n}^{-1})} \int_0^1 \varphi \left(\frac{d\mathbf{G}^{-1}}{d\mathbf{F}_n^{-1}}(u) \right) \mathbf{F}_n^{-1}(du) \tag{19}$$

$$= \arg \inf_{\theta \in \Theta} \inf_{\Delta \in A_\theta} D_\varphi(\Delta) \tag{20}$$

with the affine space A_θ defined through

$$A_\theta = \left\{ \Delta := (\Delta_1, .., \Delta_{n-1}) \in \mathbb{R}^{n-1} \text{ such that } \sum_{i=1}^{n-1} K(i/n) \Delta_i = f(\theta) \right\}$$

and

$$D_\varphi(\Delta) = \sum_{i=1}^{n-1} \varphi \left(\frac{\Delta_i}{x_{i+1:n} - x_{i:n}} \right) (x_{i+1:n} - x_{i:n}).$$

Remark 4.1: Since $\mathbf{G}^{-1} \in L_\theta(\mathbf{F_n}^{-1})$, the vector Δ defined the measure $\mathbf{G}^{-1}$ through $\mathbf{G}^{-1}(i/n) = \Delta_i$. When $\mathbf{G}^{-1}$ is a positive σ-finite measure, there exist $y_1 \leq ... \leq y_n$ (defined up to an additive constant) such that $\Delta_i = y_{i+1} - y_i$ with $1 \leq i \leq n-1$. We may consider arbitrary $\sigma-$finite measures denoted $\mathbf{G}^{-1}$, for example when using the χ^2 divergence whose function φ has $\mathbb{R}$ as its domain; in this case the y_i's may be any real numbers, and the mass Δ_i still satisfies $\Delta_i = y_{i+1} - y_i$. Note that the set A_θ is always non void, in contrast with $L_\theta^{(0)}(\mathbf{F}_n)$.

Remark 4.2: The estimation defined by (19) produces estimators $\hat{\theta}_n$ which do not depend on the location of the sample, since a change the sample $(x_i \mapsto x_i + a)_{i=1...n}$ produces, independently on the value of a, the same measure $\mathbf{F}_n^{-1}$ whose mass

on point i/n is the gap $x_{i+1:n} - x_{i:n}$. The minimum discrepancy estimators defined by (20) are invariant with respect to the location of the underlying distribution of the data. Due to this fact, we consider the model defined by L-moments conditions only through equations of the form (15).

Both the constraint and the divergence criterion are expressed in function of $\mathbf{G}^{-1}$ and the constraint is linear with respect to this measure. This allows to use classical duality results in order to efficiently compute the estimator $\hat{\theta}_n$. Before that, we reformulate this criterion as a minimization of an "energy" of transformation of the sample.

V. Dual representations of the divergence under L-moment constraints

The minimization of φ-divergences under linear equality constraint is performed using Fenchel-Legendre duality. It transforms the constrained problems into an unconstrained one in the space of Lagrangian parameters. Let ψ denote the Fenchel-Legendre transform of φ, namely, for any $t \in \mathbb{R}$

$$\psi(t) := \sup_{x \in \mathbb{R}} \{tx - \varphi(x)\}.$$

Let us recall that $dom(\varphi) = (a_\varphi, b_\varphi)$. We can now present a general duality result for the two optimization problems that transform a constrained problem (possibly in an infinite dimensional space) into an unconstrained one in $\mathbb{R}^l$.

Let $C : \Omega \to \mathbb{R}^l$ and $a \in \mathbb{R}^l$. Denote

$$L_{C,a} = \left\{ g : \Omega \to \mathbb{R} \text{ s.t. } \int_\Omega g(t) C(t) \mu(dt) = a \right\}.$$

Proposition 5.1: Let μ be a σ-finite measure on $\Omega \subset \mathbb{R}$. Let $C : \Omega \to \mathbb{R}^l$ be an array of functions such that

$$\int_\Omega \|C(t)\| \mu(dt) < \infty.$$

If there exists some g in $L_{C,a}$ such that $a_\varphi < g < b_\varphi$ μ-a.s. then the duality gap is zero i.e.

$$\inf_{g \in L_{C,a}} \int_\Omega \varphi(g)\, d\mu = \sup_{\xi \in \mathbb{R}^l} \langle \xi, a \rangle - \int_\Omega \psi(\langle \xi, C(x) \rangle) \mu(dx). \tag{21}$$

Moreover, if ψ is differentiable, if μ is positive and if there exists a solution ξ^* of the dual problem which is an interior point of

$$\left\{ \xi \in \mathbb{R}^l \text{ s.t. } \int_\Omega \psi(\langle \xi, C(x) \rangle) \mu(dx) < \infty \right\},$$

then ξ^* is the unique maximum in (21) and

$$\int \psi'(\langle \xi^*, C(x) \rangle) C(x) \mu(dx) = a.$$

Furthermore the mapping $a \mapsto \xi^*(a)$ is continuous.

Proof: The proof is delayed to the Appendix. ∎

Remark 5.1: When $\mathbf{G}^{-1} \ll \mathbf{F}^{-1}$, denoting $g^* = d\mathbf{G}^{-1}/d\mathbf{F}^{-1}$ and assuming $g^* \in L_{K,f(\theta)}$, and when $\mu = \mathbf{F}^{-1}$ it holds

$$\int \varphi(g^*) d\mu = D_\varphi \left(\mathbf{G}^{-1}, \mathbf{F}^{-1} \right).$$

Remark 5.2: Here, the classical assumption of finiteness of μ is replaced by

$$\int_\Omega \|C(x)\| \mu(dx) < \infty$$

which is needed for the application of the dominated convergence Theorem; also we refer to the illuminating paper by Csiszár and Matúš [24] for the description of the geometric tools used in the proof of Proposition 5.1.

We now apply the above Proposition 5.1 to the case when the array of functions C is equal to K, the measure μ is the quantile measure $\mathbf{F}^{-1}$ pertaining to the distribution function F of a probability measure and when the class of functions $L_{C,a}$ is substituted by the class of functions $\mathbf{dG}^{-1}/\mathbf{dF}^{-1}$ when defined. Let $\theta \in \Theta$ and F be fixed. Let us recall that for any reference cdf F

$$L_\theta(\mathbf{F}^{-1}) := \left\{ \mathbf{G}^{-1} \ll \mathbf{F}^{-1} \text{ s.t. } \int_{\mathbb{R}} K(u) \mathbf{G}^{-1}(du) = f(\theta) \right\}. \tag{22}$$

Corollary 5.1: If there exists some $\mathbf{G}^{-1}$ in $L_\theta(\mathbf{F}^{-1})$ such that $a_\varphi < d\mathbf{G}^{-1}/d\mathbf{F}^{-1} < b_\varphi$ $\mathbf{F}^{-1}$-a.s. then

$$\inf_{\mathbf{G}^{-1} \in L_\theta(\mathbf{F}^{-1})} \int_0^1 \varphi\left(\frac{d\mathbf{G}^{-1}}{d\mathbf{F}^{-1}} \right) d\mathbf{F}^{-1} = \sup_{\xi \in \mathbb{R}^l} \langle \xi, f(\theta) \rangle - \int_0^1 \psi(\langle \xi, K(u) \rangle) \mathbf{F}^{-1}(du).$$

Moreover, if ψ is differentiable and if there exists a solution ξ^* of the dual problem which is an interior point of

$$\left\{\xi \in \mathbb{R}^l \text{ s.t. } \int_{\mathbb{R}} \psi(\langle \xi, K(u)\rangle)\mathbf{F}^{-1}(du) < \infty\right\},$$

then ξ^* is the unique maximum in (23) and

$$\int \psi'^*(\langle \xi, K(u)\rangle)K(u)\mathbf{F}^{-1}(du) = f(\theta).$$

Remark 5.3: The above Corollary 5.1 is the cornerstone for the plug-in estimator of $D_\varphi(\mathbf{G}, \mathbf{F})$.

Let us present an other application of the above Proposition 5.1 leading to the same dual problem. Denote by λ the Lebesgue measure on $\mathbb{R}$ and $L'_\theta(F)$ be the set of all functions g defined by

$$L'_\theta(F) = \left\{g : \mathbb{R} \to \mathbb{R} \text{ s.t. } \int_{\mathbb{R}} K(F(x))g(x)\lambda(dx) = f(\theta)\right\},$$

whenever non void.

Corollary 5.2: If there exists some g in $L'_\theta(F)$ such that $a_\varphi < g < b_\varphi$ λ-a.s. then

$$\inf_{g \in L'_\theta(F)} \int_{\mathbb{R}} \varphi(g)\, d\lambda = \sup_{\xi \in \mathbb{R}^l} \langle \xi, f(\theta)\rangle - \int_{\mathbb{R}} \psi(\langle \xi, K(F(x))\rangle)dx. \tag{23}$$

Moreover, if ψ is differentiable and if there exists a solution ξ^* of the dual problem which is an interior point of

$$\left\{\xi \in \mathbb{R}^l \text{ s.t. } \int_{\mathbb{R}} \psi(\langle \xi, K(F(x))\rangle)dx < \infty\right\},$$

then ξ^* is the unique maximizer in (23). It satisfies

$$\int \psi'(\langle \xi^*, K(F(x))\rangle)dx = f(\theta). \tag{24}$$

Proof: We will detail the proof of Corollary 5.2. Corollary 5.1 is proved similarly.
We apply the above Proposition 5.1 for $\Omega = \mathbb{R}$, $\mu = \lambda$, the array of functions C substituted by the array of functions $x \mapsto K(F(x))$ and $a = f(\theta)$.
Consequently, the class of functions g depends upon F, and $L_{C,a} = L'_\theta(F)$.We need then to show that

$$\int_{\mathbb{R}} \|K(F(x))\|dx < \infty.$$

Denote $K := (K_{i_1}, ..., K_{i_l})$ with $i_j \geq 2$ for all j. Recall that from equation (6)

$$K_{i_j}(t) = -t(1-t)\frac{J^{(1,1)}_{i_j-2}(2t-1)}{i_j - 1}.$$

It is clear that there exists $C > 0$ such that $\left|\frac{J^{(1,1)}_{i_j-2}(2t-1)}{i_j-1}\right| < C$. Hence

$$\int_{\mathbb{R}} \|K(F(x))\|dx < lC \int_{\mathbb{R}} F(x)(1-F(x))dx < +\infty$$

since F is the cdf of a random variable with finite expectation. By applying Proposition 5.1, it then holds

$$\inf_{g \in L''_\theta(F)} \int_{\mathbb{R}} \varphi(g)\, d\lambda = \sup_{\xi \in \mathbb{R}^l} \langle \xi, f(\theta)\rangle - \int_{\mathbb{R}} \psi(\langle \xi, K(F(x))\rangle)dx.$$

■

VI. Reformulation as minimum of an energy of deformation

A. The case of models defined by moments constraints

Let us suppose for a while that $\mathbf{F}$ and $\mathbf{G}$ are both absolutely continuous with respect to the Lebesgue measure defined on $\mathbb{R}$. Define the function $T = G \circ F^{-1}$. Then T is differentiable a.e. and $T' = \frac{dT}{d\lambda}$. It holds

$$D_\varphi(\mathbf{G}, \mathbf{F}) = \int_{\mathbb{R}} \varphi\left(\frac{d\mathbf{G}}{d\mathbf{F}}(x)\right)\mathbf{F}(dx) = \int_0^1 \varphi(T'(u))\, du$$

even if $\mathbf{G}$ is not a positive measure, as far as the integrand in the central term of the above display is defined.
The function T can be viewed as a measure of the deformation of $\mathbf{F}$ into $\mathbf{G}$ and

$$E_1(T) = \int \varphi\left(\frac{dT}{d\lambda}\right) d\lambda$$

as an energy of this deformation.

It can be seen that the absolute continuity assumption of both $\mathbf{F}$ and $\mathbf{G}$ with respect to the Lebesgue measure can be relaxed.

Proposition 6.1: Let F and G be two arbitrary cdf's and λ be the Lebesgue measure. Let us define

$$M_\theta(\mathbf{F}) = \left\{ \mathbf{G} \ll \mathbf{F} \text{ s.t. } \int_{\mathbb{R}} g(x,\theta)\mathbf{G}(dx) = 0 \right\}$$

and let $M'_\theta(\mathbf{F})$ denote the class of all functions T defined through

$$M'_\theta(\mathbf{F}) := \left\{ T : [0;1] \to \mathbb{R} \text{ differentiable a.e. on } [0;1] \text{ s.t. } \int_0^1 g(F^{-1}(u),\theta)\frac{dT}{d\lambda}(u)\lambda(du) = 0 \right\}.$$

If there exists $T \in M'_\theta(\mathbf{F})$ such that $a_\varphi < \frac{dT}{d\lambda} < b_\varphi$ and $\mathbf{G} \in M_\theta(\mathbf{F})$ such that $a_\varphi < \frac{d\mathbf{G}}{d\mathbf{F}} < b_\varphi$, then

$$\inf_{G \in M_\theta(\mathbf{F})} \int_{\mathbb{R}} \varphi\left(\frac{d\mathbf{G}}{d\mathbf{F}}(x)\right) \mathbf{F}(dx) = \inf_{T \in M'_\theta(\mathbf{F})} E_1(T).$$

Proof: This results from Proposition 5.1 applieFd twice.
First, if $C = g(.,\theta)$, $a = 0$, $\mu = \mathbf{F}$ and $g = d\mathbf{G}/d\mathbf{F}$, it holds

$$\inf_{G \in M_\theta(\mathbf{F})} \int_{\mathbb{R}} \varphi\left(\frac{d\mathbf{G}}{d\mathbf{F}}(x)\right) \mathbf{F}(dx) = \sup_{\xi \in \mathbb{R}^l} - \int_{\mathbb{R}} \psi\left(\langle \xi, g(x,\theta)\rangle\right) \mathbf{F}(dx).$$

Secondly, if $C = g(F^{-1}(.),\theta)$, $a = 0$, $\mu = \lambda$ and $g = dT/d\lambda$, it holds

$$\inf_{T \in M'_\theta(\mathbf{F})} \int_0^1 \varphi\left(\frac{dT}{d\lambda}\right) d\lambda = \sup_{\xi \in \mathbb{R}^l} - \int_0^1 \psi\left(\langle \xi, g(F^{-1}(u),\theta)\rangle\right) \lambda(du).$$

Lemma 2.2 concludes the proof. ∎

The minimum divergence estimators introduced in [25] and [5] can be expressed in terms of T , introducing the empirical distribution of the sample in place of the true unknown distribution $\mathbf{F}_{\theta_0}$. For each θ in Θ it holds

$$\inf_{\mathbf{G} \in M_\theta(\mathbf{F}_n)} \int_{\mathbb{R}} \varphi\left(\frac{d\mathbf{G}}{d\mathbf{F}_n}\right) \mathbf{F}_n(dx) = \inf_{T \in M'_\theta(\mathbf{F}_n)} E_1(T)$$

and

$$\theta_n := \arg \inf_{\theta \in \Theta} \inf_{T \in M'_\theta(\mathbf{F}_n)} E_1(T).$$

Remark 6.1: Note that if $T \in M'_\theta(\mathbf{F}_n)$, $T : [0;1] \to [0;1]$ is λ-a.e. differentiable and verifies

$$\sum_{i=1}^{n-1} g(x_{i:n},\theta)\left(T\left(\frac{i+1}{n}\right) - T\left(\frac{i}{n}\right)\right) = 0.$$

The plug-in estimator that realizes the minimum of the divergence between a given distribution and the submodel $\mathcal{M}_\theta$ results from the minimum of an energy of a deformation of the uniform grid on $[0,1]$ under constraints involving the observed sample. Therefore the classical minimum divergence approach under moment conditions turns out to be a transformation of the uniform measure on the sample points, represented by the uniform grid on $[0,1]$ onto a projected measure on the same sample points, and the projected measure $\mathbf{G}_n$ which solves the primal problem has support $x_1,..,x_n$ and has a distribution function $G_n = T(F_n)$ where T solves

$$\inf_{T \in M'_\theta(\mathbf{F}_n)} E_1(T).$$

Turning now to the case of models defined by L-moments, we will now see that the approach of Section IV-B2 consists in minimizing a deformation of the support of the distribution of interest instead of its weights.

B. The case of models defined by L-moment constraints

Similarly as for the case of models defined by moment constraints we now see that the solution of the minimum divergence problem (primal problem) holds without assuming $\mathbf{F}^{-1}$ absolutely continuous with respect to the Lebesgue measure.

Proposition 6.2: Let F and G be two arbitrary cdf's. Let $L''_\theta(F)$ denote the class of all functions T which are a.e differentiable on $\mathbb{R}$ defined through

$$L''_\theta(F) := \left\{ T : \mathbb{R} \to \mathbb{R} \text{ differentiable } \lambda\text{-a.e. s.t. } \int_{\mathbb{R}} K(F(x)) \frac{dT}{d\lambda}(x) \lambda(dx) = f(\theta) \right\}.$$

With $L_\theta(\mathbf{F}^{-1})$ defined in (22), if there exists $T \in L''_\theta(F)$ such that $a_\varphi < \frac{dT}{d\lambda} < b_\varphi$ and $\mathbf{G}^{-1} \in L_\theta(\mathbf{F}^{-1})$ such that $a_\varphi < \frac{d\mathbf{G}^{-1}}{d\mathbf{F}^{-1}} < b_\varphi$, it holds

$$\inf_{\mathbf{G}^{-1} \in L_\theta(\mathbf{F}^{-1})} D_\varphi(\mathbf{F}^{-1}, \mathbf{G}^{-1}) = \inf_{T \in L''_\theta(F)} E_1(T).$$

Proof: This results from a combination of Corollaries 5.1 and 5.2. ∎

In the following, we consider the estimator of θ

$$\hat{\theta}_n = \arg \inf_{\theta \in \Theta} \inf_{T \in L''_\theta(\mathbf{F}_n^{-1})} \int_{\mathbb{R}} \varphi \left(\frac{dT}{d\lambda} \right) d\lambda. \tag{25}$$

The estimator $\hat{\theta}_n$ defined in (25) coincides with (19) thanks to the above Proposition 6.2.

Remark 6.2: $\cup_\theta L_\theta(\mathbf{F}^{-1})$ and $\cup_\theta L''_\theta(F)$ both represent the same model with L-moments constraints, seen through a reference measure $\mathbf{F}^{-1}$. This model is either expressed as the space of quantile measures $\mathbf{G}^{-1}$ absolutely continuous with respect to $\mathbf{F}^{-1}$ satisfying the L-moment constraints or as the space of all deformations $\mathbf{F}^{-1} \to T \circ F^{-1}$ of the reference measure $\mathbf{F}^{-1}$ such that the deformed measure satisfies the L-moment constraints. In the second point of view T is differentiable λ-a.e. even if the reference measure is $\mathbf{F}_n^{-1}$.

Remark 6.3: For the set of deformations $L''_\theta(F_n)$ (whenever non void), the duality for finite distributions is expressed through the following equality :

$$\inf_{T \in L''_\theta(F_n)} \int \varphi \left(\frac{dT}{d\lambda} \right) d\lambda = \sup_{\xi \in \mathbb{R}^l} \xi^T f(\theta) - \sum_{i=1}^{n-1} \psi \left(\xi^T K \left(\frac{i}{n} \right) \right) (x_{i+1:n} - x_{i:n}).$$

Notice that we incorporate the requirement that for any T in the model $L''_\theta(F_n)$, $a_\varphi < \frac{dT_1}{d\lambda} < b_\varphi$ λ-a.s. holds.

Example 6.1: If we consider the χ^2-divergence $\varphi(x) = \frac{(x-1)^2}{2}$, then $\psi(t) = \frac{1}{2}t^2 + t$ and the solution ξ_1^* of the equation (24) is

$$\xi_1^* = \Omega^{-1} \left(f(\theta) - \int K(F(x)) d\lambda \right)$$

with

$$\Omega = \int K(F(x)) K(F(x))^T d\lambda.$$

If we set

$$d_n = f(\theta) - \int K(F_n(x)) d\lambda$$

and

$$\Omega_n = \int K(F_n(x)) K(F_n(x)) d\lambda,$$

the estimator shares similarities with the GMM estimator. Indeed

$$\hat{\theta}_n = \arg \inf_{\theta \in \Theta} d_n \Omega_n^{-1} d_n.$$

This divergence should thus be favored for its fast implementation.

Remark 6.4: We did not consider the constraints of positivity classically assumed in moment estimating equations for sake of simplicity of the dual representations. We could suppose that the transformation T is an increasing mapping. It would be the case if the divergence is the Kullback-Leibler one, for example. Indeed, in this case, problem (25) is well defined since $\varphi(x) = +\infty$ for all $x \leq 0$.

VII. Asymptotic properties of the L-moment estimators

In this section, we study the convergence of the estimator equivalently given by (25) and (19). The proof of the two asymptotic theorems are postponed to the Appendix.

Theorem 7.1: Let $x_1, ..., x_n$ be an observed iid sample drawn from a distribution F_0 with finite expectation. Assume that

- there exists θ_0 such that $F_0 \in L_{\theta_0}$, and θ_0 is the unique solution of the equation $f(\theta) = f(\theta_0)$.
- f is continuous and $\Theta \subset \mathbb{R}^d$ is compact.
- the matrix $\Omega_0 = \int K(F_0(x))K(F_0(x))^T dx$ is non singular.
- $\|K'\|$ is bounded in $[0; 1]$ and $K(0) = K(1) = 0$.

Then

$$\hat{\theta}_n \to \theta_0 \text{ in probability as } n \to \infty.$$

We may now turn to the limit distribution of the estimator. Let

- $J_0 = J_f(\theta_0)$ be the Jacobian of f with respect to θ in θ_0
- $M = (J_0^T \Omega^{-1} J_0)^{-1}$
- $H = M J_0^T \Omega^{-1}$
- $P = \Omega^{-1} - \Omega^{-1} J_0 M J_0^T \Omega^{-1}$
- $\Sigma = \iint \left[F(\min(x, y)) - F(x)F(y)\right] K'(F(x))K'(F(y))^T dxdy$.

Theorem 7.2: Let $x_1, ..., x_n$ be an observed iid sample drawn from a distribution F_0 with finite variance. We assume that the hypotheses of Theorem 7.1 holds. Moreover, we assume that

- $\theta_0 \in int(\Theta)$
- J_0 has full rank
- f is continuously differentiable in a neighborhood of θ_0

Then,

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \hat{\xi}_n \end{pmatrix} \to_d \mathcal{N}_{d+l} \left(0, \begin{pmatrix} H\Sigma H^T & 0 \\ 0 & P\Sigma P^T \end{pmatrix}\right)$$

Remark 7.1: The variance assumption in Theorem 7.2 restrict its application with respect to heavy-tailed distributions. In case of distributions with finite expectation and infinite variance, only convergence in probability of the estimators is guaranteed.

The estimator of the minimum of the divergence from $\mathbf{F}$ onto the model, namely $2n \left[\hat{\xi}_n^T f(\hat{\theta}_n) - \int \psi(\hat{\xi}_n^T K(F_n(x))dx\right]$, does not converge to a χ^2-distribution as in the case of moment condition models [25]. However, we can state an alternative result.

Corollary 7.1: Let us assume that the hypotheses of Theorem 7.2 hold.
Let $S_n := n\hat{\xi}_n^T (P_n \Sigma_n P_n^T)^{-1} \hat{\xi}_n$ with P_n and Σ_n the respective empirical versions of P and Σ.
If $P\Sigma P$ is non singular then

$$S_n \to_d \chi^2(l)$$

where $\chi^2(l)$ denotes a chi-square distribution with l degrees of freedom.

Proof: From Theorem 7.2, we have that

$$n^{1/2} \hat{\xi}_n \to_d X = \mathcal{N}_l(0, P\Sigma P)$$

where X denotes such a multivariate Gaussian random vector.
Furthermore

$$P_n \Sigma_n P_n \to_p P\Sigma P.$$

Hence, for n large enough, $P_n \Sigma_n P_n$ is invertible and by Slutsky Theorem

$$n\hat{\xi}_n^T (P_n \Sigma_n P_n)^{-1} \hat{\xi}_n \to_p X^T X =_d \chi^2(l).$$

■

Since the weak convergence of S_n to a chi-square distribution is independent of the value of θ_0, this result may be used in order to build confidence regions related to the semi-parametric model.

VIII. Numerical applications : Inference for Generalized Pareto family

A. Presentation

The Generalized Pareto Distributions (GPD) are known to be heavy-tailed distributions. They are classically parametrized by a location parameter m, which we assume to be 0, a scale parameter σ and a shape parameter ν. They can be defined through their density :

$$f_{\sigma,\nu}(x) = \begin{cases} \frac{1}{\sigma}\left(1+\nu\frac{x}{\sigma}\right)^{-1-1/\nu}\mathbb{1}_{x>0} & \text{if } \nu>0 \\ \frac{1}{\sigma}\exp\left(\frac{x}{\sigma}\right)\mathbb{1}_{x>0} & \text{if } \nu=0 \\ \frac{1}{\sigma}\left(1+\nu\frac{x}{\sigma}\right)^{-1-1/\nu}\mathbb{1}_{-\sigma/\nu>x>0} & \text{if } \nu<0 \end{cases}$$

Let us notice that if $\nu \geq 1$, the GPD does not have a finite expectation. We perform different estimations of the scale and the shape parameter of a GPD from samples with size $n = 100$.
We will estimate the parameters in the model composed by the distributions of all r.v's X whose second, third and fourth L-moments are given by

$$\begin{cases} \lambda_2 &= \frac{\sigma}{(1-\nu)(2-\nu)} \\ \frac{\lambda_3}{\lambda_2} &= \frac{1+\nu}{3-\nu} \\ \frac{\lambda_4}{\lambda_2} &= \frac{(1+\nu)(2+\nu)}{(3-\nu)(4-\nu)} \end{cases} \tag{26}$$

for any $\sigma > 0, \nu \in \mathbb{R}$. These distributions share their first L-moments with those of a GPD with scale and shape parameter σ and ν (see [20]). This estimation will be compared with classical parametric estimators detailed hereafter.

B. Moments and L-moments calculus

The variance and the skewness of the GPD are given by

$$\begin{cases} var &= \mathbb{E}[(X-\mathbb{E}[X])^2] = \frac{\sigma^2}{(1-\nu)^2(1-2\nu)} \\ t_3 &= \mathbb{E}[\left(\frac{X-\mathbb{E}[X]}{\mathbb{E}[(X-\mathbb{E}[X])^2]}\right)^3] = \frac{2(1+\nu)\sqrt{1-2\nu}}{1-3\nu} \end{cases}$$

Let us remark that var and t_3 respectively exist since $\nu < 1/2$ and $\nu < 1/3$.
On the other hand, the first L-moments are given by equation 26. Assuming $\nu < 1$ entails existence of the L-moments.

C. Simulations

We perform $N = 500$ runs of the following estimators

- the estimator proposed in this article (equation (25)) for the χ^2-divergence and the modified Kullback (KL_m) divergence with the constraints pertaining to the L-moments of order $2, 3, 4$
- the estimator defined through the L-moment method, based on the empirical second L-moment $\hat{\lambda}_2$ and the fourth L-moment ratio $\hat{\tau_4} = \frac{\lambda_4}{\lambda_2}$

$$\hat{\nu} = \frac{7\hat{\tau_4} + 3 - \sqrt{(\hat{\tau_4}^2 + 98\hat{\tau_4} + 1}}{2(\hat{\tau_4} - 1)}$$
$$\hat{\sigma} = \hat{\lambda}_2(1-\hat{\nu})(2-\hat{\nu})$$

- the estimator defined through the moment method estimated from the empirical variance $\hat{var}$ and skewness $\hat{t}_3$

$$\hat{\nu} = \frac{2(1+\hat{t}_3)\sqrt{1-2\hat{t}_3}}{1-3\hat{t}_3}$$
$$\hat{\sigma} = \sqrt{\hat{var}(1-\hat{t}_3)^2(1-2\hat{t}_3)}$$

- the MLE defined in the GPD family

We present the following different features for any of the above estimators

- the mean of the N estimates based on the N runs
- the median of the N estimates based on the N runs
- the standard deviation of the N estimates
- the Quantile-Quantile plots (Q-Q plots) of the true quantile against the estimated Generalized Pareto quantile

Finally, we present four different scenarios which illustrate robustness properties of any of the above estimators, as well as their behavior under misspecification:

- a first scenario without outliers : samples of size 30 or 100 are drawn from a GPD

TABLE I
ESTIMATES OF GPD SCALE (TOP) AND SHAPE (BOTTOM) PARAMETERS FOR $\nu = 0.7$ AND $\sigma = 3$ (THE MOMENT METHOD HAS LITTLE SENSE SINCE $\nu > 0.5$) FOR THE FIRST SCENARIO WITHOUT OUTLIERS

	$n = 30$			$n = 100$		
Estimation method	Mean	Median	StD	Mean	Median	StD
χ^2-divergence	4.68	4.41	2.52	3.80	3.75	0.90
KL_m-divergence	6.44	4.77	8.02	4.08	3.95	4.00
L-moment method	5.67	4.98	3.44	3.96	3.80	1.09
Moment method	17.17	10.45	62.95	17.15	11.64	19.52
MLE	3.33	3.17	1.14	3.08	3.07	0.57

	$n = 30$			$n = 100$		
Estimation method	Mean	Median	StD	Mean	Median	StD
χ^2-divergence	0.38	0.39	0.24	0.55	0.55	0.16
KL_m-divergence	0.37	0.38	0.24	0.38	0.37	0.16
L-moment method	0.33	0.38	0.31	0.54	0.56	0.18
Moment method	0.08	0.12	0.12	0.21	0.22	0.06
MLE	0.61	0.63	0.33	0.68	0.69	0.17

TABLE II
ESTIMATES OF GPD SCALE (TOP) AND SHAPE (BOTTOM) PARAMETERS FOR $\nu = 0.7$ AND $\sigma = 3$ FOR A SAMPLE WITH 10% OUTLIERS OF VALUE 300 (THE MOMENT METHOD HAS LITTLE MEANING SINCE $\nu > 0.5$)

	$n = 30$			$n = 100$		
Estimation method	Mean	Median	StD	Mean	Median	StD
χ^2-divergence	12.43	12.24	2.83	12.29	12.21	1.62
KL_m-divergence	24.01	19.36	49.38	27.30	20.99	48.75
L-moment method	22.27	20.83	5.69	21.68	21.03	3.09
Moment method	80.97	76.27	20.89	80.93	76.84	31.09
MLE	3.06	2.88	1.08	2.88	2.86	0.55

	$n = 30$			$n = 100$		
Estimation method	Mean	Median	StD	Mean	Median	StD
χ^2-divergence	0.55	0.55	0.05	0.54	0.54	0.04
KL_m-divergence	0.50	0.52	0.24	0.54	0.49	0.27
L-moment method	0.54	0.54	0.06	0.54	0.53	0.04
Moment method	0.07	0.08	0.02	0.08	0.07	0.03
MLE	1.48	1.44	0.22	1.50	1.49	0.11

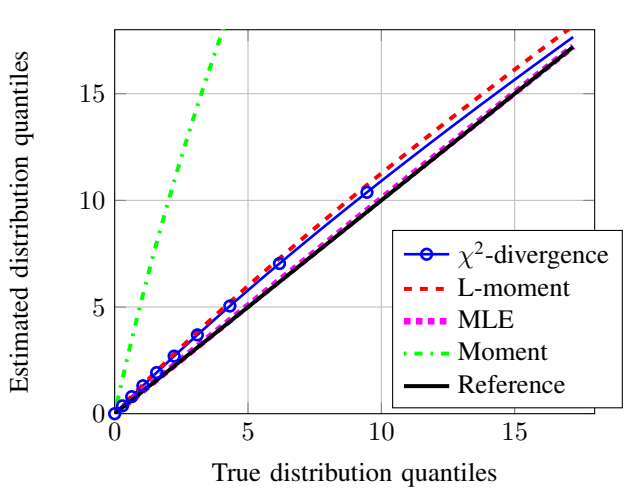
- two more scenarios with 10% outliers : samples of size 27 or 90 are drawn from a GPD. The remaining points are drawn from a Dirac the value of which depends on the shape parameter
- a fourth scenario without outliers but with misspecification : samples of size 30 or 100 are drawn from a Weibull distribution.

Unsurprisingly, the MLE performs well under the model and the L-moment method has an overall better behavior than the classical moment method for the considered heavy-tailed distributions (see Table I). Furthermore, we observe that the χ^2-divergence is more robust than the modified Kullback as indeed expected.
The interesting result lies in their behavior with outliers and misspecification. Indeed, we see that L-moment-based estimators

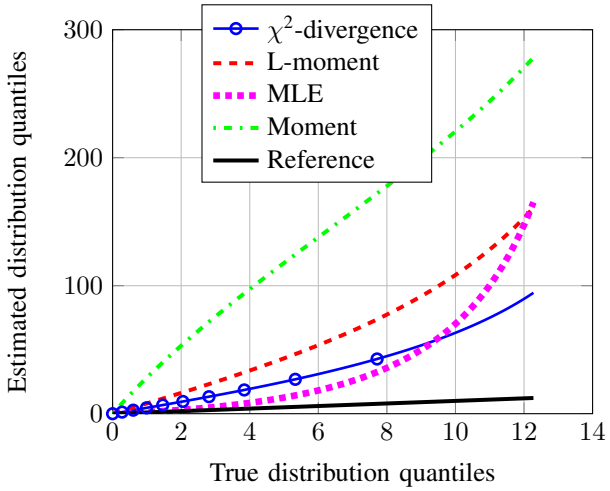
TABLE III
ESTIMATES OF GPD SCALE (TOP) AND SHAPE (BOTTOM) PARAMETERS FOR $\nu = 0.1$ AND $\sigma = 3$ FOR A SAMPLE WITH 10% OUTLIERS OF VALUE 30

	$n = 30$			$n = 100$		
Estimation method	Mean	Median	StD	Mean	Median	StD
χ^2-divergence	4.32	4.23	0.91	4.45	4.42	0.51
KL_m-divergence	5.04	4.90	1.15	5.07	5.08	0.67
L-moment method	5.18	5.04	1.44	5.11	5.04	0.75
Moment method	8.64	8.44	0.92	8.54	8.48	0.50
MLE	3.12	3.08	0.87	3.08	3.05	0.49

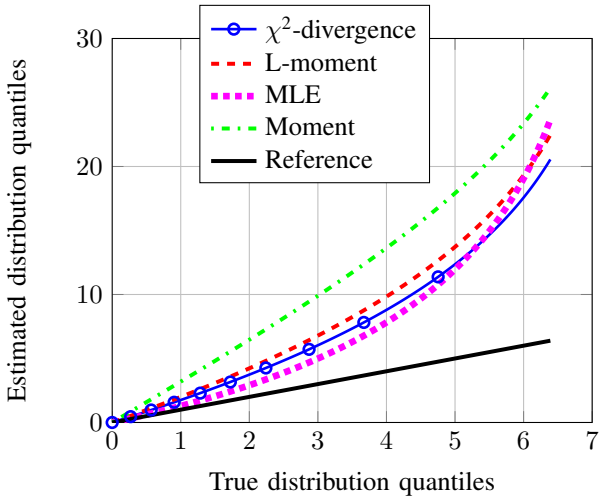
	$n = 30$			$n = 100$		
Estimation method	Mean	Median	StD	Mean	Median	StD
χ^2-divergence	0.27	0.28	0.08	0.27	0.27	0.05
KL_m-divergence	0.25	0.25	0.09	0.24	0.24	0.05
L-moment method	0.24	0.24	0.10	0.24	0.24	0.06
Moment method	0.01	0.02	0.04	0.01	0.02	0.02
MLE	0.56	0.54	0.17	0.55	0.55	0.09



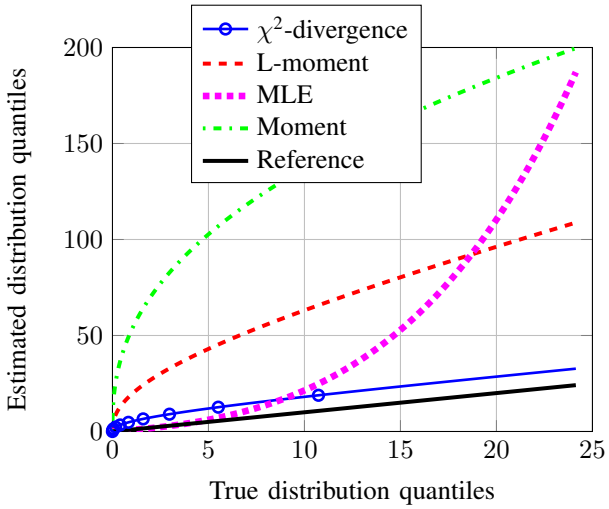
(a) Simulated GPD with $\nu = 0.7$ and $\sigma = 3$



(b) Simulated GPD with $\nu = 0.7$, $\sigma = 3$ and 10% outliers of value 300



(c) Simulated GPD with $\nu = 0.1$, $\sigma = 3$ and 10% outliers of value 30



(d) Simulated Weibull distribution with $\nu = 0.4$ and $\sigma = 3$

Fig. 2. Q-Q plots between true quantiles and estimated quantiles (with $n = 100$) for simulated scenarios

perform well on the shape parameter whereas the MLE provides a good estimation of the scale parameter but overestimates the shape parameter.

Moreover, Figure 2 presents the Q-Q plots between the simulated distribution and the estimated one in order to illustrate the quality of estimation (especially in the upper part of the distribution) for the proposed scenarios (see e.g. [8] for the use of Q-Q plots in contamination scenarios).
The closer to the first bisector, the more accurate the estimation. From a general perspective, MLE tends to better estimate the central part of the distribution (where the majority of the mass is), which corresponds here to low quantiles while L-moments methods concentrate on the upper tail of the distribution.
Figure 2a confirms again the good behavior of the MLE under the model with respect to semi-parametric methods.
Figure 2b and 2c illustrate the robustness of estimators with respect to a contamination by outliers. All methods overestimate the tail of the distribution (the Q-Q plots are asymptotically convex-shaped) under the influence of the outliers. None of the presented methods can then be said *robust* with respect to that misspecification. Thresholding methods would better fit in that situation. We see however that the overestimation of the tail by the MLE is pore important than for L-moment methods.
Figure 2d illustrate the robustness of methods with respect to a model misspecification. Moment and L-moments methods underestimate the distribution tail while MLE overestimate it. These different behaviors could be explained through the importance that the different methods give to extreme values. Indeed, L-moments methods give positive weights to the extreme values themselves while moments methods consider power of these extreme values thus underestimating in a more important

manner the tail behavior of the distribution. However, MLE give logarithmic weights to the tail of the distribution, so that the tail is not enough taken into account under misspecification.
L-moments methods seem then better adapted in that case in order to estimate tail probabilities even in case of misspecification of heavy-tailed distributions.

ACKNOWLEGMENTS

The authors would like to thank the referees for their useful comments and the DGA/MRIS and Thales for their support.

APPENDIX A
PROOF OF LEMMA 2.2

Let us recall that the support of a measure μ defined on $X \subset \mathbb{R}$ is the smallest closed set $C \subset X$ such that

$$U \in \mathcal{B}(X) \text{ and } U \cap C \neq \emptyset \Rightarrow \mu(U \cap C) > 0$$

where $\mathcal{B}(X)$ denotes the Borel sets in X. Let S be the support of $\mathbf{F}^{-1}$. Then $[0;1]\backslash S$ is an open set in $[0;1]$ i.e. a countable union of intervals $\cup_{i\geq 1}]t_{2i}, t_{2i+1}[$ and

$$\begin{aligned}
&\int_0^1 a(F^{-1}(t))dt \\
&= \int_S a(F^{-1}(t))dt + \sum_{i\geq 1} \int_{]t_{2i};t_{2i+1}[} a(F^{-1}(t))dt \\
&= \int_{F^{-1}(S)} a(x)dF(x) + \sum_{i\geq 1} a(F^{-1}(t_{2i}))(t_{2i+1} - t_{2i}) \\
&= \int_{F^{-1}(S)} a(x)dF(x) + \sum_{i\geq 1} \int_{\{F^{-1}(t_{2i})\}} a(x)dF(x) \\
&= \int_{F^{-1}(S)\cup\left(\cup_{i\geq 1}\{F^{-1}(t_{2i})\}\right)} a(x)dF(x).
\end{aligned}$$

The second equality stems from the definition of the quantile as left-continuous function and from the fact that F^{-1} is strictly monotone on S.
Since F^{-1} is constant on the open interval $[t_{2i}; t_{2i+1}[$, $\{F^{-1}(t_{2i})\} = F^{-1}([t_{2i}; t_{2i+1}[)$. Hence

$$\begin{aligned}
F^{-1}(S) \cup \left(\cup_{i\geq 1}\{F^{-1}(t_{2i})\}\right) =& F^{-1}([0;1]) \\
=& \{x \in \mathbb{R} \text{ s.t. there exists } t \text{ with } F^{-1}(t) = x\} = supp(F).
\end{aligned}$$

We conclude the first part of the proof since

$$\int_{supp(F)} a(x)dF(x) = \int_{\mathbb{R}} a(x)dF(x).$$

The second part of the proof can be proved similarly since the above arguments are not particular to a specific measure.

APPENDIX B
PROOF OF PROPOSITION 5.1

The proof is directly adapted from the proof of Theorem II.2 of Csiszár et al. [10].
Let us begin with the fundamental lemma inspired from Theorem 2.9 of Borwein and Lewis [3].

Lemma B.1: Let $C : \Omega \to \mathbb{R}^l$ be an array of bounded functions such that

$$\int_\Omega \|C(x)\| d\mu(x) < \infty.$$

We denote

$$L_{C,a} = \left\{ g \text{ s.t. } \int_\Omega g(t)C(t)d\mu(t) = a \right\}.$$

If there exists some g in $L_{C,a}$ such that $a_\varphi < g < b_\varphi$ μ-a.s and $\int_\Omega \|g(t)C(t)\|d\mu(t) < \infty$, then there exists $a'_\varphi > a_\varphi$, $b'_\varphi < b_\varphi$ and $g_b \in L_{C,a}$ such that $a'_\varphi \leq g_b(x) \leq b'_\varphi$ for all $x \in \Omega$.

Proof: Let L denotes the subspace of $\mathbb{R}^l$ composed by the vectors representable as $\int_\Omega gCd\mu$ for some $g:\Omega\to\mathbb{R}^l$. Let us denote by a_n a decreasing sequence $a_n\to a_\varphi$, by b_n a increasing one $b_n\to b_\varphi$ and let T_n be the set

$$T_n=\{x\in\Omega \text{ s.t. } a_n\le g(x)\le b_n\}.$$

We first claim that, for n large enough

$$L=L_n=\left\{\int_\Omega hCd\mu \text{ with } h(x)=0 \text{ if } x\notin T_n \text{ and } h \text{ bounded}\right\}.$$

Indeed, if not, we can build a sequence of vectors v_n such that $\|v_n\|=1$, $v_n\in L^\perp$ and $v_n\to v\in L$. Furthermore, $v_n\in L^\perp$ means

$$\langle v_n,\int_\Omega hCd\mu\rangle=\int_\Omega h\langle v_n,C\rangle d\mu=0$$

then $\langle v_n,C\rangle=0$ for all $x\in T_n$ μ-a.s. Hence $\langle v,C\rangle=0$ μ-a.s. and $v\in L^\perp$ which contradicts $v\in L$ with $\|v\|=1$.
Let us then fix some n_0 such that $L_{n_0}=L$. We denote by

$$L_n(\delta)=\left\{\int_\Omega hCd\mu \text{ with } h(x)=0 \text{ if } x\notin T_n \text{ and } |h(x)|<\delta \text{ for } x\in\Omega\right\}.$$

Then, the affine hull of $L_n(\delta)$ is the vector space L and $0\in L_n(\delta)$. We can consider the function g_n

$$g_n(x)=\begin{cases} a_n & \text{if} \quad g(x)<a_n\\ g(x) & \text{if} \quad b_n\le g(x)\le a_n\\ b_n & \text{if} \quad g(x)>b_n\end{cases}$$

Then $\|\int_\Omega(g_n-g)Cd\mu\|\to_{n\to\infty}0$. Indeed we can apply the dominated convergence theorem since, for any $x\in\Omega$, $g_n(x)\to g$ and

$$\begin{aligned}\|(g_n(x)-g(x))C(x)\|=&\|\mathbb{1}_{g(x)<a_n}(a_n-g(x))C(x)+\mathbb{1}_{g(x)>b_n}(g(x)-b_n)C(x)\|\\ \le&(\|a_0-g(x)\|+\|b_0-g(x)\|)\|C(x)\|\\ \le&(\|a_0\|+\|b_0\|)\|C(x)\|+2\|g(x)\|\|C(x)\|\end{aligned}$$

which is μ-measurable by hypothesis.
We conclude that $\int_\Omega(g_n-g)Cd\mu\in L_{n_0}(\delta)$ for n large enough because $0\in L_{n_0}(\delta)$. Hence there exists h such that $\int_\Omega(g_n-g)Cd\mu=\int_\Omega hCd\mu$, $|h(x)|=0$ for $x\notin T_{n_0}$ and $|h(x)|<\delta$ for x in T_{n_0}.
Therefore for $x\in\Omega$, $\min(a_n,a_{n_0}-\delta)\le g_n(x)+h(x)\le\min(b_n,b_{n_0}+\delta)$ and $\int_\Omega(g_n+h)Cd\mu=\int_\Omega gCd\mu$. Since δ is arbitrarily small, h is the null function. ∎

We can now prove the duality equality. Let note for $c\in\mathbb{R}^l$ $I(c)=\inf_{\int gCd\mu=c}\int_\Omega\varphi(g)d\mu$ and

$$J(c)=\begin{cases}0 & \text{if } c=a\\ +\infty & \text{otherwise}\end{cases}.$$

Then

$$\inf_{g\in L_{C,a}}\int\varphi(g)d\mu=\inf_{c\in\mathbb{R}^l}I(c)+J(c).$$

Recall that the Fenchel duality theorem ([28] p327) states that if $ri(dom(I))\cap ri(dom(J))\neq\emptyset$ then

$$\inf_{c\in\mathbb{R}^l}I(c)+J(c)=\max_{\xi\in\mathbb{R}^l}-I^*(\xi)-J^*(-\xi).$$

We prove that $ri(dom(I))\cap ri(dom(J))\neq\emptyset$. Note that $ri(dom(J))=\{a\}$. It suffices then to prove that a belongs to $int(dom(I))$ for the topology induced by L. By the above Lemma B.1 there exists g_b such that $a_\varphi<a'_\varphi\le g_b(x)\le b'_\varphi<b_\varphi$ for all $x\in\Omega$. Since $a+L_n(\delta)$ is a neighborhood of a included in $dom(I)$ for δ sufficiently small, it holds that $a\in int(L_n(\delta))\subset int(dom(I))$.

It remains now to compute the conjugates of I and J .

$$\begin{aligned} I^*(\xi) &= \sup_{c\in\mathbb{R}^l} \langle \xi, c\rangle - \inf_{g,\int gCd\mu=c} \varphi(g)d\mu \\ &= \sup_{c\in\mathbb{R}^l} \sup_{g,\int gCd\mu=c} \langle \xi, c\rangle - \varphi(g)d\mu \\ &= \sup_g \langle \xi, \int gCd\mu\rangle - \varphi(g)d\mu \\ &= \sup_g \int \langle \xi, C\rangle g - \varphi(g)d\mu \\ &= \int \psi(\langle \xi, C\rangle)d\mu \end{aligned}$$

This equality is referred to as the integral representation of I^*. The last equality can be rigorously justified (see for example [24]).
Furthermore, $J^*(-\xi) = -\langle \xi, a\rangle$ which closes the first part of the proof, namely

$$\inf_{g\in L_{C,a}} \int_\Omega \varphi(g)\, d\mu = \sup_{\xi\in\mathbb{R}^l} \langle \xi, a\rangle - \int_\Omega \psi(\langle \xi, C(x)\rangle)d\mu.$$

Since we assume ψ differentiable, then $\xi \mapsto \langle \xi, a\rangle - \int_\Omega \psi(\langle \xi, C(x)\rangle)d\mu$ is differentiable as well. It follows that any critical point is the solution of

$$\int_\Omega \psi'(\langle \xi, C(x)\rangle)C(x)d\mu = a.$$

Furthermore, since φ is strictly convex, ψ is strictly concave and for $\xi, \xi' \in \mathbb{R}^l$ and $t \in [0;1]$ it holds

$$\begin{aligned} &\langle (1-t)\xi + t\xi', a\rangle - \int_\Omega \psi(\langle (1-t)\xi + t\xi', C(x)\rangle)d\mu \\ &= \langle (1-t)\xi + t\xi', a\rangle \\ &\qquad - \int_\Omega \psi((1-t)\langle \xi, C(x)\rangle + t\langle \xi', C(x)\rangle)d\mu \\ &< (1-t)\left[\langle \xi, a\rangle - \int_\Omega \psi(\xi, C(x)\rangle)d\mu\right] \\ &\qquad + t\left[\langle \xi', a\rangle - \int_\Omega \psi(\xi', C(x)\rangle)d\mu\right] \end{aligned}$$

i.e. the functional $\xi \to \langle \xi, a\rangle - \int_\Omega \psi(\xi, C(x)\rangle)d\mu$ is strictly convex which proves the uniqueness of ξ^*.
The continuity of $a \mapsto \xi^*(a)$ comes from the implicit function theorem. If we note $D(\xi) = \int \psi'(\langle \xi, C(x)\rangle)C(x)d\mu$ then D is continuously differentiable with a Jacobian given by

$$J_D(\xi) = \int \psi''(\langle \xi, C(x)\rangle)C(x)C(x)^T d\mu$$

which is positive definite thanks to the strict convexity of ψ.

APPENDIX C
PROOF OF THEOREM 7.1

The arguments of this proof and of the following one are similar to the ones given by Newey and Smith in [25] for their Theorem 3.1; the main argument is a Taylor expansion of the functionals in equation (23).
Let us begin with two lemmas,the first one adapted from Theorem 6 due to Stigler [30] and the second due to Van Zwet [33] :

Lemma C.1: Let $x_1, ..., x_n$ be an observed iid sample drawn from a distribution F with finite variance. We denote F_n the empirical distribution of the sample.
Let $A : [0;1] \to \mathbb{R}^l$ be a continuously differentiable function such that A' is bounded $\mathbf{F}^{-1}$-a.e. Then

$$n^{1/2}\left(\int xdA(F_n(x)) - \int xdA(F(x))\right) \to_d N(0, \Sigma_A)$$

where

$$\Sigma_A = \iint \left[F(\min(x,y)) - F(x)F(y)\right] A'(F(x))A'(F(y))^T dxdy.$$

Lemma C.2: Let $1 \le p \le +\infty$, $\frac{1}{p} + \frac{1}{q} = 1$ and suppose that $J \in L_p$ and $F^{-1} \in L_q$. If either

- $1 < p \leq +\infty$ and $\|J\|_p < \infty$
- or $p = 1$ and J is uniformly integrable,

then

$$\int_0^1 J(u)F^{-1}(u)du - \int_0^1 J(u)F_n^{-1}(u)du \to_P 0.$$

If we consider the case where $J = L_r$ and $p = \infty$ and $q = 1$ in Lemma C.2, an immediate consequence is the following Lemma

Lemma C.3: Suppose that $\int_{\mathbb{R}} |x|dF(x) < \infty$ and let $r \geq 2$

$$\int_0^1 K_r(u)d\mathbf{F}_n^{-1}(u) - \int_0^1 K_r(u)d\mathbf{F}^{-1}(u) \to_P 0.$$

In the following, we will note $\frac{dT}{d\lambda}(x) = T'(x)$ for all $x \in \mathbb{R}$.

First step : maximization step

Clearly, it holds

$$\inf_{T \in \cup_\theta L''_\theta(F_n)} \int_{\mathbb{R}} \varphi(T'(x))dx \leq \inf_{T \in L''_{\theta_0}(F_n)} \int_{\mathbb{R}} \varphi(T'(x))dx. \tag{27}$$

By Taylor-Lagrange expansion, there exists some $\delta > 0$ and $D > 0$ such that for any t in $[1-\delta; 1+\delta]$, it holds

$$\varphi(t) \leq \frac{D}{2}(t-1)^2.$$

We may then find an upper bound for the RHS in (27) through the solution of the quadratic case. Let

$$T'_{0,n}(x) := 1 + (f(\theta_0) - m_n)^T \Omega_n^{-1} K(F_n(x))$$

where $m_n := \int K(F_n(x))dx$ and $\Omega_n := \int_{\mathbb{R}} K(F_n(x))K(F_n(x))^T dx$. Since $T'_{0,n} \in L''_\theta(F_n)$, it holds

$$\inf_{T \in L''_{\theta_0}(F_n)} \int_{\mathbb{R}} \varphi(T'(x))dx \leq \int_{\mathbb{R}} \varphi(T'_{0,n}(x))dx.$$

From Lemma C.3, we deduce that $\Omega_n \to \Omega$ in probability. Since Ω is non singular, for n large enough, Ω_n is non singular and $T'_{0,n}$ is well defined.
Since $\|f(\theta_0) - m_n\| = o_P(1)$ from Lemma C.3 and $\|\Omega_n^{-1}\| = O_P(1)$, for almost all $x \in \mathbb{R}$,

$$T'_{0,n}(x) = 1 + o_P(1)$$

and we can apply a Taylor-Lagrange maximization for n large enough

$$\varphi(T'_{0,n}(x)) \leq \frac{D}{2}(f(\theta_0) - m_n)^T \Omega_n^{-1} K(F_n(x))K(F_n(x))^T \Omega_n^{-1}(f(\theta_0) - m_n).$$

By integration in the above display

$$\begin{aligned}\int_{\mathbb{R}} \varphi(T'_{0,n}(x))dx &\leq \frac{D}{2}(f(\theta_0) - m_n)\Omega_n^{-1}(f(\theta_0) - m_n) \\ &\leq \|f(\theta_0) - m_n\|^2 \|\Omega_n^{-1}\|.\end{aligned}$$

Second step : minimization step

Since Θ is compact, and φ is strictly convex, and $\theta \mapsto \inf_{T \in L''_\theta(F_n)} \int_{\mathbb{R}} \varphi(T'(x))dx$ is continuous (see Proposition 5.1), it follows that $\hat{\theta}$ is well defined and the duality equality states

$$\begin{aligned}\inf_{T \in L''_{\hat{\theta}_n}(F_n)} \int_{\mathbb{R}} \varphi(T'(x))dx &= \sup_{\xi \in \mathbb{R}^l} \xi^T f(\hat{\theta}_n) - \int \psi(\xi^T K(F_n(x)))dx \\ &\geq \xi_n^T f(\hat{\theta}_n) - \int \psi(\xi_n^T K(F_n(x)))dx\end{aligned}$$

with

$$\xi_n = \|f(\theta_0) - m_n\| \frac{f(\hat{\theta}_n) - m_n}{\|f(\hat{\theta}_n) - m_n\|}.$$

.
Therefore

$$\xi_n^T K(F_n(x)) = o_P(1) \text{ for a.e } x \in \mathbb{R}.$$

By Taylor-Lagrange expansion, there exists a constant $C > 0$ such that $|\psi(x) - x| < Cx^2$ in a neighborhood of 0. Thus, for n large enough

$$\int \psi(\xi_n^T K(F_n(x)))dx - \xi_n^T m_n < C \int \xi_n^T K(F_n(x))K(F_n(x))^T \xi_n dx = C\xi_n^T \Omega_n \xi_n$$

and

$$\inf_{T \in L''_{\hat{\theta}_n}(F_n)} \int_{\mathbb{R}} \varphi(T'(x))dx > \xi_n^T (f(\hat{\theta}_n) - m_n) - C\xi_n^T \Omega_n \xi_n.$$

Conclusion
Combining the two inequalities, we have

$$\|f(\theta_0) - m_n\| \|f(\hat{\theta}_n) - m_n\| < C\|\Omega_n\| \|f(\theta_0) - m_n\|^2 + \|f(\theta_0) - m_n\|^2 \|\Omega_n^{-1}\| = O_P(\|f(\theta_0) - m_n\|^2)$$

i.e. $\|f(\hat{\theta}_n) - m_n\| = O_P(\|f(\theta_0) - m_n\|)$.
By Lemma C.3, $\|m_n - f(\theta_0)\| = o_P(1)$. Hence, $\|f(\hat{\theta}_n) - f(\theta_0)\| = o_P(1)$.
Since $f(\theta) = f(\theta_0)$ has a unique solution at θ_0, $\|f(\theta) - f(\theta_0)\|$ is bounded away from zero outside some neighborhood of θ_0. Therefore $\hat{\theta}_n$ is inside any neighborhood of θ_0 with probability approaching 1 i.e $\hat{\theta}_n \to \theta_0$ in probability.

APPENDIX D
PROOF OF THEOREM 7.2

First we prove that

$$\hat{\xi}_n = \arg\max_{\xi} \xi^T f(\hat{\theta}_n) - \int \psi(\xi^T K(F_n(x))dx = O_P(n^{-1/2}).$$

Consider

$$\xi_n = \arg \max_{\xi \in \mathbb{R}^l \text{ s.t. } \|\xi\| < n^{-1/4}} \xi^T f(\hat{\theta}_n) - \int \psi(\xi^T K(F_n(x))dx,$$

where the maximum is taken on a ball of radius $n^{-1/4}$. The maximum is attained because of the concavity of the functional

$$U : \xi \mapsto \xi^T f(\hat{\theta}_n) - \int \psi(\xi^T K(F_n(x))dx.$$

For all x in a neighborhood of 0, the inequality $y - \psi(y) < -Cy^2$ for some $C > 0$ holds. For n large enough, as $\|\xi_n\| < n^{-1/4}$ we can claim (as $\psi(0) = 0$)

$$\begin{aligned} 0 &\leq \xi_n^T f(\hat{\theta}_n) - \int \psi(\xi_n^T K(F_n(x))dx \\ &\leq \xi_n^T (f(\hat{\theta}_n) - m_n) - C\xi_n^T \Omega_n \xi_n \\ &\leq \|\xi_n\| . \|f(\hat{\theta}_n) - m_n\| - C\xi_n \Omega_n \xi_n, \end{aligned}$$

with $m_n := \int K(F_n(x))dx$.
Furthermore, there exists $D > 0$ such that $\|\Omega_n\| \geq D > 0$ for n large enough and

$$CD \leq C \frac{\xi_n^T}{\|\xi_n\|} \Omega_n \frac{\xi_n}{\|\xi_n\|} \leq \frac{\|f(\hat{\theta}_n) - m_n\|}{\|\xi_n\|}.$$

It follows that $\xi_n = O_P(n^{-1/2})$ and that ξ_n is an interior point of $\{\xi \in \mathbb{R}^l \text{ s.t. } \|\xi\| < n^{-1/4}\}$; by concavity of the functional U, ξ_n is the unique maximizer, hence $\xi_n = \hat{\xi}_n$.
We write the first order conditions of optimality of $(\hat{\theta}_n - \theta_0, \hat{\xi}_n)$:

$$\begin{cases} (f(\hat{\theta}_n) - f(\theta_0)) + (f(\theta_0) - m_n) - \int \left[\psi'(\hat{\xi}_n^T K(F_n(x)) - 1\right] K(F_n(x))dx = 0 \\ J_f(\hat{\theta}_n)\hat{\xi}_n = 0 \end{cases}$$

A mean value expansion (since $\theta_0 \in int(\Theta)$) gives the existence of $\bar{\xi}$ and $\bar{\theta}$ such that $\|\bar{\xi}\| < \|\hat{\xi}_n\|$ and $\|\bar{\theta} - \theta_0\| < \|\hat{\theta}_n - \theta_0\|$ such that

$$\begin{cases} J_f(\bar{\theta})(\theta - \theta_0) + (f(\theta_0) - m_n) - \left[\int \psi''(\bar{\xi}^T K(F_n(x))K(F_n(x))K(F_n(x))^T dx\right] \hat{\xi}_n = 0 \\ J_f(\hat{\theta}_n)\hat{\xi}_n = 0 \end{cases}.$$

It holds

$$A_n := \left(\begin{array}{cc} J_f(\bar{\theta}) & -\int \psi''(\bar{\xi}^T K(F_n(x))K(F_n(x))K(F_n(x))^T dx \\ 0 & J_f(\hat{\theta}_n) \end{array} \right) \to_p A := \left(\begin{array}{cc} J_0 & -\Omega \\ 0 & J_0 \end{array} \right).$$

By the very definition of A_n,

$$A_n \left(\begin{array}{c} \hat{\theta}_n - \theta_0 \\ \hat{\xi}_n \end{array} \right) = \left(\begin{array}{c} m_n - f(\theta_0) \\ 0 \end{array} \right).$$

As Ω is non singular and J_0 has full rank, A is non singular and its inverse is given by

$$A^{-1} = \left(\begin{array}{cc} H & M \\ P & H - H^T \end{array} \right).$$

Hence by Lemma C.1

$$\sqrt{n} \left(\begin{array}{c} \hat{\theta}_n - \theta_0 \\ \hat{\xi}_n \end{array} \right) = A_n^{-1} \left(\begin{array}{c} \sqrt{n}(m_n - f(\theta_0)) \\ 0 \end{array} \right) \to_d A^{-1} \left(\begin{array}{c} \mathcal{N}_l(0, \Sigma) \\ 0 \end{array} \right),$$

which ends the proof.

REFERENCES

[1] S. M Ali, S.D. Silvey, *A general class of coefficients of divergence of one distribution from another*, Journal of the Royal Statistical Society, Series B, 28 (1), pp 131-142, 1966.

[2] P. Bertail, *Empirical likelihood in some semiparametric models*, Bernouilli, Volume 12, No 2, pp 299-331, 2006

[3] J.M. Borwein, A.S. Lewis, *Duality relationships for entropy-like minimization problems*, SIAM Journal of Control and Optimization, vol. 29, pp 325-338, 1991.

[4] J.M. Borwein, A.S. Lewis, *Partially finite convex programming, Part I: Quasi relative interiors and duality theory*, Mathematical Programming, 57, pp 11-48, 1992.

[5] M. Broniatowski, A. Keziou, *Divergences and duality for estimation and test under moment condition models*, Journal of Statistical Planning and Inference, vol. 142, 9, pp. 2554-2573, 2012.

[6] M. Broniatowski, A. Decurninge, *Estimation for models defined by conditions on their L-moments*, : arXiv:1409.5928, 2014.

[7] L.K.Chan, *On a characterization of distributions by expected values of extreme order statistics*. Amer.Math.Monthly, vol.74, pp 950–951, 1967.

[8] L.K.Choi, P. Hall, B. Presnell, *Rendering parametric procedures more robust by empirically tilting the model*, Biometrika, 87, 2, pp. 453-465, 2000.

[9] I, Csiszár, *Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizitat von Markoffschen Ketten*, Magyar. Tud. Akad. Mat. Kutato Int. Kozl, 8, pp 85-108, 1963.

[10] I. Csiszár, F. Gamboa, E. Gassiat, *MEM pixel correlated solutions for generalized moment and interpolation problems*, IEEE Trans. Inform. Theory, 45(7), pp 2253-2270, 1999.

[11] H. A. David, *Order Statistics*, 2nd edition, New York, Wiley, 1981.

[12] A. Decurninge, *Multivariate quantiles and multivariate L-moments*, : arXiv:1409.6013, 2014.

[13] P. Delicado, M.N. Goria, *A small sample comparison of maximum likelihood, moments and L-moments methods for the asymmetric exponential power distribution*, Computational Statistics & Data Analysis, vol. 52, no. 3, pp. 1661-1673, 2008

[14] P. Embrechts and M. Hofert. *A note on generalized inverses*, Mathematical Methods of Operations Research, 77(3), pp 423-432, 2013.

[15] V.P. Godambe, M.E. Thompson, *An extension of quasi-likelihood estimation*, Journal of Statistical Planning and Inference, 22(2), pp.137-172, 1989.

[16] C. Gourieroux, J. Jasiak, *Dynamic Quantile models*, Journal of econometrics, Vol. 147, 1, pp. 198-205, 2008.

[17] L.P. Hansen, *Large sample properties of generalized method of moments estimators, Econometrica*, Vol. 50, No 4, pp 1029-1054, 1982.

[18] L.P. Hansen, *Finite-Sample Properties of Some Alternative GMM Estimators*, Journal of Business and Economic Statistics, vol. 14, No. 3, pp. 262-280, 1996.

[19] J.R. Hosking, *Some theoretical results concerning L-moments*, Research report RC14492, IBM Research Division, Yorktown Heights, 1989.

[20] J.R. Hosking, *L-moments: analysis and estimation of distributions using linear combinations of order statistics*, Journal of the Royal Statistical Society, vol. 52, No. 1, pp. 105-124, 1990.

[21] J.R. Hosking, *Moments or L Moments? An Example Comparing Two Measures of Distributional Shapes*, The Amer. Stat., vol. 46, No. 3, pp. 186-189, 1992.

[22] A.G. Konheim, *A note on order statistics*, Amer.Math.Mon, vol. 78, p 524, 1971.

[23] F. Liese, *Estimates of Hellinger integrals of infinitely divisible distributions*, Kybernetika, vol. 23, No 3, pp. 227-238, 1987.

[24] I. Csiszár, F. Matúš, *Generalized minimizers of convex functionals, Bregman distance, Pythagorean identities*, Kybernetika, Vol. 48, No 4, pp 637-689, 2012.

[25] W. Newey, R. Smith, *Higher order properties of GMM and generalized empiracal likelihood Estimators*, Econometrica, Vol. 72, No 1, pp. 219-255, 2004.

[26] A. Owen, *Empirical likelihood ratio confidence regions*, Annals of Statistics, vol. 18, No. 1, pp. 90-120, 1990.

[27] E. Parzen, *Nonparametric statistical modelling*, Journal of the American Statistical Association, vol. 74, No. 365, pp 105-121, 1979.

[28] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

[29] R. Serfling, *Approximation Theorems of Mathematical Statistics*, John Wiley and Sons, 1980.

[30] S.M. Stigler, *Linear functions of order statistics with smooth weight functions*, Annals of Statistics, vol 2, pp 676-693, 1974; corrections, vol 7, p. 466, 1979.

[31] J.W. Tukey, *Which Part of the Sample Contains the Information?*, Proceedings of the National Academy of Sciences, 53, pp. 127- 134, 1965.

[32] T.J. Ulrych, D.R. Velis, A.D. Woodbury, M.D. Sacchi, *L-moments and C-moments*, Stochastic Environmental Research and Risk Assessment, no. 1, vol. 14, pp. 50-68, 2000.

[33] W. R. van Zwet, *A strong law for linear functions of order statistics*

[34] Q.J. Wang, *LH moments for statistical analysis of extreme events*, Water Resources Research, 33, 12, pp 2841–2848 , 1997.

[35] S. Wang, *A class of distortion operators for pricing financial and insurance risks*, Journal of Risk and Insurance, 67, pp. 15-36, 2000.

[36] R. Serfling, P. Xiao, *A contribution to multivariate L-moments: L-comoment matrices*, Journal of Multivariate Analysis, 98, pp. 1765-1781, 2007.