A theoretical comparison between moments and L-moments

Nicolás Hernández Pérez, University of Toronto

Santiago Carrillo Menéndez, Universidad Autonoma de Madrid

Luis Seco, University of Toronto

Abstract

Despite its popularity in applied statistics, standard measures of shape have long been recognized to be unsatisfactory, due to their extreme sensitivity to outliers and poor sample efficiency. These difficulties seem to be largely overcome by a new system: the L-moments. During the last decade several authors have established the superior performance of L-moments over classical moments based on heuristic studies, but until present no formal explanation has been provided. We address these issues from a theoretical viewpoint. Our comparative programme is focussed on two aspects, which highlight the statistical performance of a descriptive measure: qualitative robustness and global efficiency. L-moments are treated as members of a general class of descriptive measures that are shown to outperform conventional moments based on these criteria. Consequently, they may be considered as appealing substitutes as L-moments to replace the standard measures. Since the results obtained hold for rather large nonparametric sets of distribution functions, they unify previous heuristic studies.

1 Introduction

It is a standard practice in statistics to describe the shape of the distribution of a population by means of a finite set of quantities summarizing the location, dispersion, skewness, peakedness and tail behavior of the unknown population. Classical measures of distributional shape has been defined by means of algebraic moments of different orders, resulting in the mean to estimate location, the variance to measure the spread, and the standardized measures of skewness and kurtosis.

Despite the popularity of algebraic moments both in data description and more formal statistical procedures, they are known to suffer from several drawbacks. First, sample moments tend to be very sensitive to a few extreme observations. Second, the asymptotic efficiency of sample moments is rather poor specially for distributions with fat tails. The last property is an immediate consequence of the fact that the asymptotic variances of these estimators are mainly determined by higher order moments, which will tend to be rather large or even unbounded, for heavy tail distributions.

Moment-based measures are just a particular, but not exhaustive, means of summarizing qualitative features of the shape of a distribution. The notions of dispersion, skewness and kurtosis are rather abstract and therefore can be described in countless ways. Among these, the approach based on partial orderings has proved to be the most effective one. Partial orderings have been proposed by several authors (see, for example. Van Zwet(1964), Bickel and Lehmann(1975, 1976) and MacGillivray(1986)) for one probability distribution to be more dispersed, more skewed, more kurtotic than another. A real valued functional defined on a given set of distributions, must in principle, preserve the ordering in question, to be reasonably called a measure of "location", "dispersion", "skewness", or "kurtosis". More formal definitions are given in Appendix.

The focuss of this paper is centered around an alternative set of descriptive measures, which seem to largely overcome the sampling drawbacks of classical measures. These are the so called, L-moments. The L-moments were formally introduced by Hosking(1990), as linear combinations of the order statistics of a population. Like classical moments, the L-moments provide intuitive information about the shape of a general distribution, which can be consistently estimated from their sample values.

Motivated by the more satisfactory sampling behavior of L-moments estimators, several authors (Hosking (1990), Hosking et al (1997) have advocated the use of L-moments over classical moments. These conclusions have been drawn on the basis of superior empirical performance over specific sets of data, and have later been supported by formal simulation experiments performed over a finite range of selected distributions (Sankarasubramanian and Srinivasan (1999)). In the present paper we address this issue on a theoretical basis.

The paper is organized as follows: Section 2 introduces the L-moments and establish some of their main properties. Section 3 focusses on the concept of *comparative robustness*, as a useful tool for comparing descriptive measures in terms of stability against outliers. Based on this concept, a formal comparison of L-moments with the conventional system is established. Section 4 reviews the basic framework that allows a theoretical treatment of the efficiency problem. In section 5, positive bounds for the relative efficiency of L-moments to conventional moments are derived over several familiar sets of distributions.

2 L-moments

We review some elementary properties and definitions. Let $F(x)$ be the distribution function of a random variable X. We shall use the symbol M_p to denote the moment functional of order p given by

$$M_p(F) = E(X^p) = \int_{-\infty}^{+\infty} x^p dF(x)$$

We shall denote by $Q(u)$ the associated quantile function of the distribution F, defined by

$$Q(u) \;=\; \inf\,\{x:\; F(x) \;\geq\; u\}$$

In what follows we shall assume that both F and Q are continuously differentiable. Let $q(u) = Q'(u)$, be the density quantile function of F. The random variables $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ shall denote the order statistics associated to the distribution F. Sometimes we shall use the same notation for the ordered values of a single sample of size n. The L-moments of a population, $L_r(F),\ r = 1, 2, \ldots$, were originally defined by Hosking(1990), as linear combinations of the expectations of $X_{i:n}$

$$L_r(F) = \frac{1}{r}\sum_{j=0}^{r-1}(-1)^j \binom{r-1}{j} E(X_{r-j:r})\,, \qquad r = 1, 2, \ldots$$

Descriptive measures based on L moments were introduced by Hosking (1990) together with a clarifying discussion of their intuitive meanings. The first L-moment $L_1(F)$ is the mean of F a measure of location; $L_2(F)$ is a scale measure, being half the value of Gini's mean difference. Other shape features are obtained by normalizing higher order L-moments by $L_2(F)$. The L-moment ratios

$$\tau_r(F) \;=\; \frac{L_r(F)}{L_2(F)}$$

are scale free measures of the shape of a distribution. In particular, $\tau_3(F)$ and $\tau_4(F)$ are measures of skewness and kurtosis respectively in the formal sense described in Appendix. Formal proofs of these facts may be found in Hosking (1996).

From a technical viewpoint, it is often more convenient to adopt an equivalent definition of L-moments as the Fourier coefficients of the quantile function $Q(u)$ in terms of orthogonal polynomials on the interval [0,1] given by Hosking (1990),

$$L_r(F) \;=\; \int_0^1 Q(u)P_{r-1}(u)du, \qquad r = 1, 2, \ldots \tag{1}$$

where $P_r(u)$ are the shifted Legendre orthogonal polynomials on the interval $(0, 1)$. By definition

$$P_{r-1}(u) \;=\; \sum_{j=0}^{r-1} p_{r,j}\, u^j$$

where

$$p_{r,j} = \frac{(-1)^{r-j}(r+j)!}{(j!)^2(r-j)!}$$

From equation 1, it follows that L-moments fall into the general class of L-functionals of the form

$$L(F) = \int_0^1 Q(u)J(u)du\,,$$

where $J(u)$ is a bounded and measurable function on $[0, 1]$. Thus, the estimation of L-moments from their sample analogs $L_r(F_n)$, fits well in the general theory of L-estimates (see, for example, Serfling (1980)). Under certain assumptions over the distribution F and $J(u)$, it is known that

$$n^{1/2}(L_r(F) - L_r(F_n) \to \ N(0, \sigma^2(F))\,,$$

where

$$\sigma^2(F) \ = \ \int_0^1 \int_u^1 J(u)J(v)\, u(1-v)\, q(u)q(v)dvdu \tag{2}$$

3 Robustness

Experience with real data has shown that L-moments are less sensitive to outliers than the classical moments. A satisfactory explanation to this fact can be provided by means of the concept of comparative robustness first introduced by Bickel and Lehman (1975). In this section we extend their results to include comparisons between moments and a general class of L-functionals. .

For any distribution function F and any constant $M > 0$, define two distribution functions $F^+(x, M)$ and $F_-(x, M)$ given by

$$F^+(x, M) \ = \ \begin{cases} 0 & x \ < \ 0 \\ F(M) & 0 \ \le x \ \le M \\ F(x) & M < x \end{cases}$$

$$F^-(x, M) \ = \ \begin{cases} F(x) & x \ \le \ -M \\ F(-M) & -M \ \le x \ < 0 \\ 1 & 0 \le x \end{cases}$$

For a functional τ, a topology is defined on its set of definition as follows.

Definition 1. *We shall define the τ-topology by the following mode of convergence: $F_k \xrightarrow{\tau} F$ if*

1. $F_k \to F$ *in law*
2. $\tau(F_k) \to \tau(F)$
3. $\lim_{M\to+\infty} \limsup_k \{|T(F_k^+(x, M))| \ + \ |T(F_k^-(x, M))|\} \ = \ 0.$

Definition 2. *Given two functionals* τ_1 *and* τ_2*, we shall say that* τ_2 *is more robust than* τ_1 *if*

a) τ_2 *is continuous with respect to the topology induced by* τ_1*.*

b) τ_1 *is not continuous with respect to the topology induced by* τ_2*.*

The previous criteria provides a theoretical framework to establish comparisons between descriptive measures based on its sensitivity to outliers.

In the next result, for convenience we shall assume that $0 < F(0-) \leq F(0) < 1$. The proofs for the case $F(0-) = 0$ and $F(0) = 1$ are similar, and will not be included here.

Theorem 1. *Let* τ *be any functional of the form*

$$\tau(F) \;=\; \int_0^1 J(u)Q(u)du,$$

where $J(t)$ *is a bounded measurable function on the interval* $(0, 1)$*.*

Then, for all $p > 1$*,* τ *is more robust than* M_p*.*

The proof hinges in the following proposition.

Proposition 1. *Let* M_r *and* M_p *be two moment functionals, with* $r > p \geq 1$*. Then,* M_p *is more robust than* M_r*.*

Proof. First note that

$$M_p(F^+(x, M)) = \int_{F(M)}^1 (Q)^p(u)du \quad \text{and} \quad M_p(F^-(x, M)) = \int_0^{F(M)} (Q)^p(u)du$$

Let $F_k \to F$ in law for which $M_r(F_k) \to M_r(F)$. Let $Q(u)$ and $Q_k(u)$ denote the quantile functions of F and F_k respectively. To show that $M_p(F_k) \to M_p(F)$, note that

$$\begin{aligned} |M_p(F) \;-\; M_p(F_k) \;\leq\; & |\int_\delta^{1-\delta} \{ (Q)^p(u) - (Q_k)^p(u)\} \, du \,| \;+\; |\int_0^\delta (Q)^p(u) \, du \,| \\ & + |\int_0^\delta (Q_k)^p(u) \, du \,| + |\int_{1-\delta}^1 (Q)^p(u) \, du \,| + |\int_{1-\delta}^1 (Q_k)^p(u) \, du \,| \end{aligned}$$

for any $0 < \delta < 1$.

Since convergence in law of F_k implies bounded pointwise convergence for Q_k to Q on the intervals $(\delta, 1-\delta)$, it follows from the dominated convergence theorem that the first term on the right hand side converges to 0 for any fixed δ as $n \to \infty$.

Using the fact that $\int_0^1 |(Q)^p(u)|du < \infty$ we can find δ_0 such that for all $0 < \delta \leq \delta_0$,

$|\int_0^\delta (Q)^p(u)du| \leq \epsilon$ and $|\int_{1-\delta}^1 (Q)^p(u)du| \leq \epsilon$.

To handle the term $|\int_0^\delta (Q_k)^p(u)du|$ note that

$$|\int_0^\delta (Q_k)^p(u)\,du| \;\leq\; \delta^{r-p/r} \cdot (\int_0^\delta |Q_k(u)|^r\,du\,)^{p/r}$$

by Holder's Inequality. Therefore, we need only to establish that for all $\delta \leq \delta_0$ there exists k_δ such that: $\int_0^\delta |Q_k(u)|^r du \;\leq\; \epsilon$, for all $k \geq k_\delta$.

By the pointwise convergence of Q_k to F, we have that for δ sufficiently small and sufficiently large, $F_k(u) \leq 0$ for all $u \in\; (0, \delta)$. Hence,

$$\int_0^\delta |Q_k(u)|^r du \;=\; |\int_0^\delta (Q_k(u)\,)^r\,du\,|$$

Now, using the fact: $\lim_{\delta\to 0}\; \limsup_k |\int_0^\delta (Q_k)^r(u)du\,| = 0$, it can be easily argued that this implies the existence of δ_0 such that for each $0 \leq \delta \leq \delta_0$ we have: $|\int_0^\delta (Q_k)^r(u)du| \leq \epsilon$, taking k sufficiently large. This proves the desired assertion.

The term $|\int_{1-\delta}^1 (Q_k)^r(u)du|$ can be handled similarly, to complete the proof of the first part.

To show that M_r is not continuous in the M_p-topology, consider the sequence of distributions F_k given by

$$F_k(x) = (1-\alpha_k)F_0(x) + \alpha_k\Phi_k(x),$$

where $\alpha_k = \frac{1}{\sqrt{k}}$ and $F_0(x)$ and $\Phi_k(x)$ are the distributions supported on $[1, +\infty)$ given by

$$F_0(x) = \exp(1-x)\,dx \quad \text{and} \quad \Psi_k(x) = \frac{1}{r+1/k} \cdot x^{-(r+1+1/k)}\,dx$$

respectively. From the fact: $r - p > 0$ it follows immediately that $F_k(x) \to F_0(x)$ in the M_p-topology. On the other hand,

$$\int_1^{+\infty} x^r dF_k(x) = \frac{\sqrt{k}-1}{\sqrt{k}} \int_1^{+\infty} dF_0(x) + \frac{1}{\sqrt{k}} \cdot \frac{1}{r+1/k} \cdot k \quad \longrightarrow \quad +\infty$$

as $k \to \infty$. This completes the proof. □

Proof of Theorem 1. From the previous proposition, it follows that the mean is more robust than the other moments of higher order. Here, we shall show that τ is, at least, as robust as the mean, from which the theorem will follow .

Let $F_k \to F$ in law for which $M_1(F_k) \to M_1(F)$. To show that $\tau(F_k) \to \tau(F)$ note that

$$\begin{aligned} |\tau(F) - \tau(F_k)| \;\; \leq \;\; & |\int_\delta^{1-\delta} J(u)(Q(u) - Q_k(u))du\,| + |\int_0^\delta J(u)Q(u)du\,| \\ & + |\int_0^\delta J(u)Q_k(u)du\,| + |\int_{1-\delta}^1 J(u)Q(u)du\,| + |\int_{1-\delta}^1 J(u)Q_k(u)du\,| \end{aligned}$$

for any $0 < \delta \leq 1$.

Since convergence in law of F_k implies bounded pointwise convergence of Q_k to Q on the intervals $(\delta, 1-\delta)$, it follows from the dominated convergence theorem that the first term on the right hand side converges to 0 for any fixed δ as $n \to \infty$.

Since $|J(u)|$ is bounded and $\int_0^1 |Q(u)|du < \infty$, we can find δ_0 such that for all $0 < \delta \leq \delta_0$ it follows,

$$|\int_0^\delta Q(u)J(u)du\,| \leq \epsilon \text{ and } |\int_{1-\delta}^1 Q(u)J(u)du\,| \leq \epsilon$$

To handle the term $|\int_0^\delta J(u)Q_k(u)du\,|$ note that

$$|\int_0^\delta J(u)Q_k(u)du\,| \;\le\; M\int_0^\delta |Q_k(u)|du$$

for some constant $M > 0$. By the pointwise convergence of Q_k to F we have that for δ sufficiently small and n sufficiently large, $F_k(u) \le 0$ for all $u \in (0, \delta)$. Hence,

$$\int_0^\delta |Q_k(u)|du \;=\; |\int_0^\delta Q_k(u)du\,|$$

Now, using the fact: $\lim_{\delta\to 0}\ \limsup_k\{|\int_0^\delta Q_k(u)du\,|\} = 0$, it can be easily argued that this implies the existence of δ_0 such that for each $< 0 \le \delta \le \delta_0$ we have: $|\int_0^\delta Q_k(u)du\,| \le \epsilon$, taking n sufficiently large. This proves the desired assertion.

The term $|\int_{1-\delta}^1 Q_k(u)\,du\,|$ can be handled similarly, to complete the proof.

Corollary 1. *If $p > 1$, then the functionals L_q are more robust than M_p, for all q.*

4 Global Efficiency of Descriptive Measures

Suppose we are given two functionals τ_1 and τ_2, defined over a large set of distribution functions Λ, and we are interested in quantifying and comparing the accuracy of their "natural" estimators, not only at a particular distribution F, but rather over large enough subsets of Λ. By "natural" estimators we mean the value of the functional τ at the empirical distribution function F_n.

All the functionals stated below should be assumed to be defined over a suitably large set of distributions Λ, whose precise definition will depend on the context.

A natural way of quantifying the accuracy of the estimator $\tau_n = \tau(F_n)$ is given by suitable scaling the asymptotic variance of this estimator, that is the standardized variance proposed by Bickel and Lehmann (1976).

Definition 3. *Let us assume that* $n^{1/2}(\tau(F_n) - \tau(F)) \sim N(0, \sigma^2(F))$. *The standardized asymptotic variance of* $\tau(F_n)$, *is defined to be the quantity*

$$\sigma^{st}(F) = \frac{\sigma^2(F)}{\tau^2(F)}$$

Based on the previous concept, the accuracy of $\tau_1(F_n)$ to $\tau_2(F_n)$ is judged by means of the ratio of their standardized asymptotic variances,

$$\varepsilon(\tau_1(F),\ \tau_2(F)) = \frac{\sigma_1^{st}(F)}{\sigma_2^{st}(F)}$$

The quantity $\varepsilon(\tau_1,\ \tau_2,\ F)$ is termed: the relative efficiency of τ_1 to τ_2 at the distribution F. The infimun of these ratios over a set Λ can be interpreted as an index of the global efficiency of τ_1 to τ_2 over distributions in the set Λ.

Definition 4. *Consider two functionals* τ_1, τ_2 *defined over a set* Λ *of distribution functions. The relative global efficiency of* τ_2 *to* τ_1 *over the set* Λ, *is defined to be the quantity*

$$\varepsilon(\tau_1, \tau_2,\ \Lambda) = \inf_{F \in \Lambda} \varepsilon(\tau_1,\ \tau_2,\ F)$$

Based on the previous definition we shall say that τ_2 is more globally efficient than τ_1 over the set Λ if

$$\varepsilon(\tau_1, \tau_2,\ \Lambda) > 0$$

and

$$\varepsilon(\tau_2, \tau_1,\ \Lambda) = 0.$$

In this setup, we shall show that L-moments outperform conventional moments over large enough sets of distributions. Our analysis focusses on the *original* measures instead of the standardized versions, i.e., we compare L-moments with moments. The reasons behind this simplification are mainly technical due to the less tractable expressions for the asymptotic variances of the standardized measures. From the results found, one may reasonably expect these properties to continue to hold in formal comparisons between descriptive measures.

5 Moments vs L-moments

Formal efficiency comparisons between descriptive measures were first formally addressed by Bickel and Lehmann (1975) in a series of papers in the 70's. They found that measures of scale in the class of pth absolute power deviations given by

$$\zeta_p \;=\; \left(\int_0^1 |Q|^p\,(t)\,dt\right)^{1/p}$$

for $1 \leq \; p \; < 2$, were more globally efficient, in a sense that will be made precise later, than the standard deviation functional over the set Λ of all symmetric distributions. Our approach is analogous, but we consider alternative classes of descriptive measures, as well as other sets larger than Λ.

To analyze the behavior of the global index $\varepsilon(L_q, M_p,\;\Lambda)$, over large enough sets Λ, some restrictions need to be imposed. The main difficulty arises from the existence of distributions, such as the uniform law, for which L-moments of order higher than 2 are zero, which invalidates the definition of the standardized variance for these functionals. This problem is not completely avoided by restricting our attention to sets where the functionals considered do not vanish, and an additional assumption will be needed as we shall see later.

For the reasons mentioned above, we shall find it convenient to distinguish comparisons involving the second L-moment l_2 from comparisons involving higher order L-moment functionals.

In the discussion that follows we shall assume that the expectation μ of all distributions considered is known. Since L-moment estimators are not affected by the randomness of this parameter as the estimators of M_p are, one may reasonably expect that the results obtained in the next sections should continue to hold to the more general setting when μ is unknown.

5.1 The second L-moment versus moments

In this section we show that L_2 outperforms the moment functionals M_p over the sets

Λ_1 = {All symmetric absolutely continuous distributions F with finite pth moment}

Λ_2 = {All unimodal absolutely continuous distributions F with finite pth moment and $M_p(F) \neq 0$},

where $p > 1$, will be made precise from the context. Clearly for distributions in Λ_1 it only makes sense to consider functionals M_p when p is even, and this shall be implicitly assumed in our discussion.

The above efficiency property of l_2, continue to hold for a large class of scale measures, which is introduced next.

Definition 5. *We shall define $\mathcal{L}_s$ to be the class of scale functionals given by*

$$\tau(F) = \int_0^1 J(u)Q(u)du$$

where $J(u)$ is a bounded measurable function satisfying:

a) $J(u) \geq 0$ for all $u \in (1/2,\, 1]$

b) $J(u) = -J(1-u)$, for all $u \in (1/2,\, 1]$

Formal efficiency comparisons between moments and functionals in the class $\mathcal{L}_s$ can be established without major technical difficulties, by connecting both classes with a well known measure of scale: the absolute standard deviation, $\zeta(F) = E\,|X|$. The main link comes from the formula,

$$\sigma^2(\zeta, F) = E\,(X^2) - E^2|X| = \int_0^1 \int_u^1 J^*(u)J^*(v)\,u(1-v)\,q(u)q(v)dvdu \quad (3)$$

where $J^*(u)$ is defined by

$$J^*(u) = \begin{cases} -1 & u \in [0,\, u_0) \qquad u_0 = F(0) \\ 1 & u \in [u_0,\, 1] \end{cases} \quad (4)$$

Now, for our main result. Although not stated explicitly, we shall assume that the conditions guaranteeing convergence of the sample estimates for all the functionals considered are satisfied.

Theorem 2. *Let $\tau \in \mathcal{L}_s$ with score function $J(u)$. Let us assume that*

$$\lim_{u \to 1^-} J(u) \;\; \text{exists and is strictly positive} \tag{5}$$

Then, if $p > 1$,

(*a*) $\varepsilon(M_p, \tau, \Lambda_i) \;>\; 0$ *for each* $i = 1, 2$.

(*b*) *There exists a sequence of distribution functions $F_k \in \Lambda_i$ such that*

$$\varepsilon(\tau(F_k), M_p(F_k)) \;\to\; 0, \quad \text{as } n \to \infty.$$

Consequently, τ is more globally efficient than M_p over the sets Λ_i.

Before presenting the proof we shall need the following lemmas.

Lemma 1. *Let F be an absolutely continuous distribution and*

$$\zeta(F) = E\,|X| \;<\; \infty$$

.

If $\tau(F) \;=\; \int_0^1 J(u)Q(u)du$ is a functional in the class $\mathcal{L}_s$, then

$$\frac{\tau(F)}{\zeta(F)} \;\geq\; \frac{1}{2} \inf_{t \in (0,\,1)} \frac{|\psi(t)|}{t(1-t)},$$

where,

$$\psi(t) \;=\; \begin{cases} \int_0^t J(u)du & 0 \;\leq\; t \;\leq\; u_0 \\ \int_t^1 J(u)du & u_0 \;\leq\; t \;\leq\; 1 \end{cases} \tag{6}$$

Proof. Without loss of generality we may assume $E\,(X) \;=\; 0$. Let $q(u)$ be the derivative of $Q(u)$. We have

$$\begin{aligned} \tau(F) = \int_0^1 J(u)Q(u)du \;\; &= \;\; \int_0^{u_0} J(u)\,(-\int_u^{u_0} q(t)dt)\,du \;\;+\;\; \int_{u_0}^1 J(u)\,(-\int_{u_0}^u q(t)dt)du \\ &= \;\; -\int_0^{u_0}(\int_0^t J(u)du)q(t)dt \;\;+\;\; \int_{u_0}^1(\int_t^1 J(u)du)q(t)dt \\ &= \;\; \int_0^1 \psi(t)q(t)dt \end{aligned} \tag{7}$$

where $\psi(t)$ is given in (6).

To obtain a similar expression for $\zeta(F)$, let $u_0 = F(0)$. ¿From the formula: $\zeta(F) = \int_0^1 Q(u)J^*(u)du$, where $J^*(u)$ is given in (4), it follows that

$$\zeta(F) = \int_0^1 \psi^*(t)q(t)dt, \tag{8}$$

where

$$\psi^*(t) = \begin{cases} t & 0 \le t \le u_0 \\ 1-t & u_0 \le t \le 1 \end{cases}$$

From equation (7), we have

$$\begin{aligned} \tau(F) &= \int_0^{u_0} \psi(t)q(t)dt + \int_{u_0}^1 \psi(t)q(t)dt \\ &\ge \inf_{t\in(0,u_0)} \frac{|\psi(t)|}{t(1-t)} \int_0^{u_0} t(1-t)q(t)dt + \inf_{t\in(u_0,1)} \frac{|\psi(t)|}{t(1-t)} \int_{u_0}^1 t(1-t)q(t)dt) \\ &\ge \inf_{t\in(0,1)} \frac{|\psi(t)|}{t(1-t)} (1-u_0) \int_0^{u_0} t\,q(t)dt + \inf_{t\in(0,1)} \frac{|\psi(t)|}{t(1-t)} u_0 \int_{u_0}^1 (1-t)q(t)dt \end{aligned}$$

It is easy to see that (8) together with the condition: $E(X) = 0$, imply

$$\int_0^{u_0} t\,q(t)dt = \int_{u_0}^1 (1-t)q(t)dt$$

Consequently,

$$\begin{aligned} \tau(F) &\ge \inf_{t\in(0,1)} \frac{|\psi(t)|}{t(1-t)} \int_0^{u_0} t\,q(t)dt \\ &= \inf_{t\in(0,1)} \frac{|\psi(t)|}{t(1-t)} \frac{\zeta(F)}{2}, \end{aligned}$$

which proves the assertion. □

Lemma 2. ***(Bickel and Lehman (1976))***

Let V be any nonnegative random variable and

$$\mu_\beta = E(V^\beta) \ < \ \infty.$$

Then, if $1 \le \alpha \le \beta$*,*

$$\frac{\mu_{2\alpha}}{\mu_\alpha^2} \ \le \ \frac{\mu_{2\beta}}{\mu_\beta^2},$$

with equality if and only if V *is a positive constant with probability 1.*

Proof of Theorem 2. To prove (a) let $\Lambda \ = \ \Lambda_1 \cup \Lambda_2$. We divide the proof into two main tasks. First we show that the absolute standard deviation ζ , is more globally efficient over Λ than M_p. The second part involves showing that

$$\varepsilon(\zeta, \tau, \ \Lambda) \ > \ 0. \tag{9}$$

For the first task, let $F \in \Lambda$. Applying standard asymptotic theory of sample moments (see for example Serfling (1980)), we obtain

$$\begin{aligned}
\varepsilon(M_p(F), \, \zeta(F)) \ \ &= \ \ \left(\frac{E\,X^{2p}}{E^2\,X^p} \ - \ 1 \right) / \left(\frac{E\,X^2}{E^2\,|X|} \ - \ 1 \right) \\
&\ge \ \ \left(\frac{E\,X^{2p}}{E^2\,|X|^p} \ - \ 1 \right) / \left(\frac{E\,X^2}{E^2\,|X|} \ - \ 1 \right) \\
&\ge \ \ 1
\end{aligned}$$

where the last inequality follows from lemma 2.

Now, for the second task. By definition

$$\varepsilon(\zeta(F), \tau(F)) \ = \ \frac{\tau^2(F)}{\zeta^2(F)} \cdot \frac{\sigma^2(\zeta, F)}{\sigma^2(\tau, F)},$$

where $\sigma^2(\zeta, F)$ and $\sigma^2(\tau, F)$ stand for the asymptotic variances of $\zeta(F)$, and $\tau(F)$ respectively. Lemma 1 together with condition 5 on the score function imply that $\inf_{\{F \in \Lambda\}} \frac{\zeta(F)^2}{\tau(F)^2} \ > \ 0$. Thus, it is sufficient to show

$$\inf_{F \in \Lambda} \frac{\sigma^2(\zeta, F)}{\sigma^2(\tau, F)} \ > \ 0 \tag{10}$$

1- $F \in \Lambda_1$. Using the symmetry of F and the "antisymmetry" property of the score functions, the integrals given in (2) and (3) can be expressed more conveniently as follows

$$\begin{aligned}
\sigma^2(\tau, F) &= \int_0^1 \int_u^1 J(u)J(v)\, u(1-v)\, q(u)q(v) dvdu \\
&= 2 \int_0^{1/2} \int_u^{1-u} J(u)J(v)\, u(1-v)\, q(u)q(v) dvdu \\
&= 2 \left(\int_0^{1/2}\int_u^{1/2} J(u)J(v)\, u(1-v)\, q(u)q(v) dvdu \right. \\
&\quad \left. + \int_0^{1/2}\int_{1/2}^{1-u} J(u)J(v)\, u(1-v)\, q(u)q(v) dvdu \right) \\
&= 2 \left(\int_0^{1/2}\int_u^{1/2} J(u)J(v)\, u(1-v)\, q(u)q(v) dvdu \right. \\
&\quad \left. - \int_0^{1/2}\int_u^{1/2} J(u)J(v)\, u\, v\, q(u)q(v) dvdu \right) \\
&= 2 \int_0^{1/2}\int_u^{1/2} J(u)J(v)\, u(1-2v)\, q(u)q(v) dvdu \qquad (11)
\end{aligned}$$

Similarly for $\sigma^2(\zeta, F)$ we have

$$\sigma^2(\zeta, F) = 2 \int_0^{1/2}\int_u^{1/2} u(1-2v)\, q(u)q(v) dvdu \qquad (12)$$

Thus,

$$\frac{\sigma^2(\zeta, F)}{\sigma^2(\tau, F)} = \frac{\int_0^{1/2}\int_u^{1/2} u(1-2v)\, q(u)q(v) dvdu}{\int_0^{1/2}\int_u^{1/2} J(u)J(v)\, u(1-2v)\, q(u)q(v) dvdu} \geq \frac{1}{M^2}. \qquad (13)$$

2- $F \in \Lambda_2$. Denote by μ the mean of F. We shall find positive lower bounds for the ratios $\frac{\sigma^2(\mu,F)}{\sigma^2(\tau,F)}$ and $\frac{\sigma^2(\zeta,F)}{\sigma^2(\mu,F)}$, from which the conclusion in (10) will follow.

Since $\mu(F_n)$ can be regarded as an L-statistic, Formula (2) applies (see also appendix 1). Thus,

$$\frac{\sigma^2(\mu, F)}{\sigma^2(\tau, F)} = \frac{\int_0^1 \int_u^1 u(1-v)\, q(u)q(v)dvdu}{\int_0^1 \int_u^1 J(u)J(v)\, u(1-v)\, q(u)q(v)dvdu} \geq \frac{1}{M^2}.$$

To bound $\frac{\sigma^2(\zeta,F)}{\sigma^2(\mu,F)}$, we show that for unimodal distributions

$$\frac{E^2(X)}{E\,(X^2)} < c < 1 \tag{14}$$

where c is a constant independent of F.

To see this, denote by m_0 the mode of F and assume, without loss of generality, that $m_0 \leq 0$. It follows that the function $F^*(x)$ given by

$$F^*(x) = \int_0^x dF(t) \quad x \geq 0,$$

is concave in $[0, +\infty)$. Thus, by Khinchin's theorem (see Dharmadhikari et al, chapter 1), it admits the following representation

$$F^*(x) = \int_0^{+\infty} W_a(x) d\lambda(a)\,, \tag{15}$$

where $W_a(x)$ are the distribution functions of random variables W_a following a uniform law on the interval $[0, a]$ and λ is a positive and finite measure. Applying equation (15) and Holder's inequality we have

$$\begin{aligned}\frac{(\int_0^{+\infty} xdF^*(x))^2}{\int_0^{+\infty} x^2 dF^*(x)} &= \frac{\left(\int_0^{+\infty} [\int_0^{+\infty} xdW_a(x)]\, d\lambda(a)\right)^2}{\int_0^{+\infty} [\int_0^{+\infty} x^2 dW_a(x)]\, d\lambda(a)} \\ &= \frac{3}{4} \frac{(\,\int_0^{+\infty} a\, d\lambda(a)\,)^2}{\int_0^{+\infty} a^2\, d\lambda(a)} \\ &\leq \frac{3}{4}\, \lambda[0, +\infty) \end{aligned} \tag{16}$$

Using the fact: $F[0, +\infty) = \lambda[0, +\infty)$, it can be easily argued that (16) implies

$$\int_0^{+\infty} x\,dF(x) \;<\; \frac{3}{4}\int_0^{+\infty} x^2\,dF(x).$$

This together with Holder's Inequality again gives

$$\begin{aligned}
E\,(X^2) \;&=\; \int_{-\infty}^{0} x^2\,dF(x) \;+\; \int_0^{+\infty} x^2\,dF(x) \\
&\geq\; \frac{4}{3\,F(-\infty,0)}(\int_{-\infty}^{0} |x|\,dF(x))^2 \;+\; \frac{1}{F(0,+\infty)}(\int_0^{+\infty} |x|\,dF(x))^2 \\
&=\; \frac{4}{3\,F(-\infty,0)}\left(\frac{1}{2}\int_{-\infty}^{+\infty} |x|\,dF(x)\right)^2 \;+\; \frac{1}{F(0,+\infty)}\left(\frac{1}{2}\int_{-\infty}^{+\infty} |x|\,dF(x)\right)^2 \\
&=\; (\,\frac{1}{3\,F(-\infty,0)} - \frac{1}{4\,F(0,+\infty)}\,)\,E^2\,(|X|)
\end{aligned}$$

Finally, it can be easily checked that for all $0 \,<\alpha\,< 1$

$$\frac{1}{3\,\alpha} - \frac{1}{4\,(1-\alpha)} \;>\; c \;>\; 1.$$

This proves (14), and thus assertion (a).

For the proof of (b) note that from (a) and the relation:

$$\varepsilon(\zeta(F_k), M_p(F_k)) \;=\; \varepsilon(\tau(F_k),\, M_p(F_k))\cdot\varepsilon(\zeta(F_k),\tau(F_k)),$$

statement (b) will follow if

$$\varepsilon(\zeta(F_k), M_p(F_k)) = \left(\frac{E\,X_k^2}{E^2\,|X_k|} \;-\; 1\right)/\left(\frac{E\,X_k^{2p}}{E^2\,X_k^{p}} \;-\; 1\right) \to 0 \qquad (17)$$

for a suitable sequence $F_k \in \Lambda_1 \cap \Lambda_2$. Set $p_k = p + (k+1)/k$ and let $\rho_k(x)$ be the sequence of probability density functions given by

$$\rho_k(x) = \begin{cases} c_k\,|x|^{-p_k} & x < -1 \\ c_k & -1 \leq x \leq 1 \\ c_k\,x^{-p_k} & x > 1\,, \end{cases}$$

where $c_k \;=\; \frac{2p_k}{p_k - 1}$ is a normalizing constant. It is clear that the resulting sequence of distribution functions will lie in the set $\Lambda_1 \cap \Lambda_2$. Now straightforward calculations lead to the desired conclusion. This completes the proof.

Corollary 2. *The second L-moment functional is more globally efficient than any moment functional M_p, over the sets $\Lambda_i,\ i = 1, 2$, for all $p > 1$.*

Remark 1. *Comparisons between the class $\mathcal{L}_s$ and the class of pth absolute deviations, τ_p^2, is straightforward in view of the inequality*

$$\varepsilon(\tau_p^2(F),\, \zeta(F)) \geq \frac{1}{p^2}\,,$$

which follows from lemma 2. Consequently, the results from theorem 2, continue to hold when τ_p^2 plays the role of M_p, for all $p > 1$.

5.2 Higher order L-moments vs moments

As it was mentioned at the beginning of section 5, the existence of non-degenerate distributions for which the index $\varepsilon(M_p(F), L_q(F))$ is undefined, forces us to make some restrictions on the sets Λ that we may consider. These restrictions do not seem to exclude from the analysis, distributions of most interest in risk management and finance, namely, those displaying heavy tails and asymmetries.

In this slightly modified setup, the results from the previous section continue to hold, for a large class of location invariant functionals, which we shall introduce next.

Definition 6. *We shall define $\mathcal{L}^*$ to be the class of functionals τ^* given by*

$$\tau^* = \int_0^1 J^*(u) Q(u)\, du,$$

where $J^*(u)$ is a bounded and measurable function satisfying

$$\int_0^1 J^*(u)\, du \;=\; 0$$

The class $\mathcal{L}^*$ can be regarded as an extension of the class $\mathcal{L}_s$, that allows to account for higher order features of the shape of F other than dispersion. Bounds for $\varepsilon(M_p, \tau^*, \Lambda)$ are possible, by imposing more stringent assumptions on the set Λ.

We shall denote by $s(F)$ a measure of scale in the class $\mathcal{L}_s$, which will be assumed to satisfy the conditions of theorem 2. Through the entire discussion s will be kept fixed.

Definition 7. *For any $\epsilon > 0$ and $\tau^* \in \mathcal{L}^*$, the set $\Lambda_{\tau^*}(\epsilon)$ is defined by*

$$\Lambda_{\tau^*}(\epsilon) \;=\; \{F \,\in\, \Lambda : \; \tfrac{|\tau^*(F)|}{s(F)} \;\geq\; \epsilon\}$$

The sets $\Lambda_{\tau^*}(\epsilon)$ are scale free, a desirable property in most applications. It does impose some restrictions on the shape of the distributions, but they seem to include the majority of cases when data deviates form normality. For example, taking l_2 as the measure of scale and τ^* the fourth L-moment, then according to the definition above we are only excluding those F having small L-kurtosis. The later is a fairly common feature displayed by many data sets in practice.

Theorem 3. *Let Λ be the set of all unimodal distributions having finite pth moment. Let $\tau^* \in \mathcal{L}^*$.*

Then, for any $\epsilon > 0$ and $p > 1$,

(a) $\varepsilon(M_p, \tau^*, \Lambda_{\tau^*}(\epsilon)) \;>\; 0$

(b) There exists a sequence of distribution functions $F_k \in \Lambda$ such that

$$\varepsilon(\tau^*(F_k),\, M_p(F_k)) \;\rightarrow\; 0.$$

Consequently, τ^* is more globally efficient than M_p over the sets $\Lambda_{\tau^*}(\epsilon)$.

Additionally, if the score function J^* is anti-symmetric, i.e., $J^*(u) = -J^*(1-u)$, then the above statements continue to hold when Λ is the set of all symmetric absolutely continuous distributions having finite pth moment.

Proof. The proof is almost the same as that for theorem 2. The only difference arises in showing that

$$\inf_{F\in\Lambda} \frac{(\tau^*(F))^2}{(\zeta(F))^2} \;>\; 0$$

As it was argued in theorem 2, the existence of positive lower bounds for the ratio $\frac{s^2(F)}{\zeta^2(F)}$, is a consequence of lemma 1 and condition (5) on the score function of the functional s. On the other hand, by definition of the set $\Lambda_{\tau^*}(\epsilon)$, we have: $\frac{(\tau^*(F))^2}{(s(F))^2} \geq \epsilon^2$. This gives the result.

The proof for the set of symmetric distributions goes through the same. □

Appendix. Coherent Shape Measures

Location. Let X, Y be random variables with distribution functions F_X, F_Y and quantile functions Q_X, Q_Y respectively. We shall say Y is stochastically larger than X, $(X \prec Y)$ if

$$Q_X(u) \ \leq \ Q_Y(u)$$

for all $0 < u < 1$.

Definition 8. *By a measure of location μ we shall understand a real functional defined on a subset Λ of distribution functions satisfying the following axioms:*

Shape Coherence. If $X \ \prec \ Y$ then, $\mu(X) \ \leq \ \mu(Y)$

Translation. For any real number, we have

$$\mu(X + c) \ = \ \mu(X) + c$$

Reflection Invariance.

$$-\mu(X) \ = \ \mu(-X)$$

Scale. The following ordering was introduced by Bickel and Lehman (1976). F_Y will be called more dispersed than F_X $(F_X \prec F_Y)$ if

$$Q_X(1 - \alpha) - Q_X(\alpha) \leq Q_Y(1 - \alpha) - Q_Y(\alpha)$$

for all $0 < \alpha < 1/2$.

Definition 9. *By a measure of scale σ, we shall understand a positive functional defined on a subset Λ of cdfs that is consistent with the following axioms:*

Shape Coherence. If F_Y is more dispersed than F_X, then $\sigma(F_X) \leq \sigma(F_Y)$.

Location Invariance. For any real number c,

$$\sigma(X + c) = \sigma(X)$$

Scalability. For any real number a,

$$\sigma(a \cdot X) = |a| \cdot \sigma(X).$$

Skewness The reference ordering for the skewness structure of asymmetric distribution was introduced by Van Zweet (1964). F_X will be called less skewed to the right than F_Y ($F_X \prec F_Y$) if and only if $Q_Y(F_X(x))$ is convex

Definition 10. *By a measure of skewness γ we shall understand a real functional defined on a given set Λ of distribution functions that is consistent with the following axioms:*

Shape Coherence. If $F_X \prec F_Y$ then, $\gamma(F_X) \leq \gamma(F_Y)$.

Symmetry. If F_X is a symmetric distribution then, $\gamma(F_X) = 0$

Location-Scale Invariance. For all a, b real numbers,

$$\gamma(a \cdot X + b) = \gamma(X).$$

Kurtosis The reference ordering for kurtosis of symmetric distributions was introduced by Van Zweet (1964). Let F_X, F_Y be symmetric distributions functions. F_X will be called less kurtotic than F_Y ($F_X \prec F_Y$) if and only if $Q_Y(F_X(x))$ is convex for $x > m_{F_X}$ and concave for $x < m_{F_X}$

Definition 11. *By a measure of kurtosis κ we shall understand a real functional defined on a given subset of symmetric quantile functions satisfying the following axioms:*

Shape Coherence. If $F_X \prec F_Y$ then, $\tau(F_X) \leq \tau(F_Y)$.

Location-Scale Invariance. For all a, b real numbers,

$$\kappa(a\,X + b) = \kappa(X).$$

References

[1] Bickel, P. J., and Lehmann, E. L. (1975). Descriptive statistics for non-parametric models. II. Location. *Annals of Statistics*, **3**, 5, 1045-1069.

[2] Bickel, P. J., and Lehmann, E. L. (1976). Descriptive statistics for non-parametric models. III. Dispersion. *Annals of Statistics*, **4**, 1139-1158.

[3] MacGillivray, H. L. (1986). Skewness and Asymmetry: Measures and Orderings. *The Annals of Statistics*, **14**, 994-1011.

[4] Hosking, J.R.M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society*, Series, B, **52** 105-124.

[5] Hosking, J.R.M., and Wallis, J.R. (1997). *Regional Frequency Analysis.*

[6] Sankarasubramanian, A., and Srinivasan, K. (1999). Investigation and comparison of sampling properties of L-moments and conventional moments. *Journal of Hydrology*, **218**, 13-34.

[7] Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics.* John Wiley. Sons, Inc.

[8] Van Zwet, W. R. (1964). Convex transformations of random variables. *Tract 7*, Mathematisch Centrum, Amsterdam.