A theoretical comparison between moments and L-moments

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

Santiago Carrillo

Marcos Escobar

Nicolás Hernández

Pablo Olivares

Luis Seco

Statistics and Decisions, Vol 19, page 331-366 (2008)

A theoretical comparison between moments and L-moments

Santiago Carrillo, Marcos Escobar, Nicolás Hernández
Pablo Olivares, Luis Seco

Abstract

Despite their 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. 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. Since the results obtained hold for a reasonable large set of distribution functions, they unify previous heuristic studies.

1 Introduction

It is a standard practice in statistics to describe the shape of a probability distribution by mean of a finite set of quantities summarizing the location, dispersion, skewness, peakedness and tail behavior of the unknown population. Classical measures of shape have 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.

Since the establishment of the Method of Moments and their relation to Pearsonian curves by Karl Pearson, measures of shape have become a tool of wide-ranging applicability being used extensively in both, applied and theoretical work where a sensible reduction of a complex phenomenon into a statistical model fully described by a finite set of parameters becomes mandatory.

In the field of empirical finance, measures of distributional shape, usually based on the coefficient of kurtosis and/or skewness, are constantly reported in the financial literature for testing normality of the distribution of financial data. In particular, the sample standardized fourth moment κ has become a widely used evidence of heavy tails in financial series such as daily exchange rates or stock indices. In a range of risk management problems, moment-based methods have been favored to obtain computationally feasible consistent estimates for large families of distributions: mixtures of normal distributions (Hull and White (1998), Duffie and Pan (1997), Parker (2000)), generalized variance gamma distributions (Levin and Tchernitzer (2001)), hyperbolic distributions (Eberlein 1995), and generalized Weibull distributions (Frachot et al (2001)).

Despite the long lasting popularity of classical descriptive measures, they are well known to suffer from several drawbacks. First, sample moments tend to be very sensitive to a few extreme observations. Second, the asymptotic efficiency of these quantities becomes rather poor specially for heavy tail distributions. This 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 descriptive measures are just particular, but not exhaustive, means of summarizing qualitative features of the shape of a distribution. The notion of dispersion, skewness and kurtosis are rather abstract and therefore can be described in countless ways. A sensible approach that attempts to capture the intuition behind these concepts in a more unified manner is based on the idea of partial orderings among distributions. 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". It is precisely this degree of generalization in understanding the shape of a distribution, what enables us to consider not just moments, but also a whole range of alternatives measures describing the same distributional feature.

The focus of this paper is centered around an alternative set of descriptive measures, which seem to largely overcome the sampling drawbacks of the classical moments. 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 convey information about the shape of a general distribution, which can be consistently estimated from their sample values. L-moments offer a number of advantages over conventional moments. First, they can be defined for a wider class of distributions, such as distributions that decay like power laws for which moments of higher order do not exist. Second, they completely characterize the probability law of the random variable X unlike conventional moments. Finally, and most important of all, when compared with their moments analogs, empirical L-moments exhibit a more robust behavior against outliers together with a smaller finite sample variability observed over a wide range of distributions, being more remarkable the differences for heavy tailed distributions.

Hosking's work has been followed by an ever-increasing interest in these quantities, leading to a number of successful applications in a variety of applied fields including: hydrology, where L-moment diagrams have become a favored tool in both the selection and estimation of distributional models of the frequency of floods (Wang (1990), Pearson (1991, 1993), Guttman (1993)); meteorology, in the context of predicting the frequency of maximum wind speeds across related regions (Hosking 1997) among others.

In the above applications, L-moments have been shown to outperform their moment analogs based on robustness and efficiency criteria. In this context, it is worth mentioning the exhaustive simulation analysis carried out by Sankarasubramanian and Srinivasan (1999). They presented a comparison of sampling properties between classical measures of scale and skewness with their L-moment analogs both in terms of relative bias and relative variance. In the vast majority of the situations considered, L-moments statistics outperformed the standard measures. Their results corroborated previous studies by Vogel and Fenessy (1993). This superior finite sample efficiency as well as other desirable properties offered by L-moments have been discussed in more detail by Hosking (1990, 1997).

Despite this considerable amount of empirical evidence, no rigorous explanation has been given so far to these findings. Our work attempts to fill this gap. We carry out a comparative programme between algebraic moments and L-moments, which highlights the statistical performance of an arbitrary descriptive measure: qualitative robustness and global efficiency. The results obtained apply to a larger class of descriptive measures, which can be regarded as appealing substitutes as L-moments for replacing the standard measures.

The paper is organized as follows: Section 2 introduces the L-moments and establishes some of their main properties. Section 3 focuses 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. In section 4 reviews the basic framework that allows a theoretical treatment of the efficiency problem. In subsections 4.1 and 4.2 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

$$\tau_r(F) \ = \ \frac{L_r(F)}{L_2(F)} \quad \text{for} \ \ r = 3, 4, \ldots$$

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. Formal proofs of these facts can be found in Hosking (1996).

From a technical viewpoint, it is often more convenient to adopt an equivalent definition of L-moments as the 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

$$P_r(u) \ = \ \sum_{j=0}^{r} p_{r,j}\, u^j$$

where

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

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

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

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

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

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

For a functional τ, a topology is defined on its domain 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. $\lim_{M\to+\infty} \limsup_k \{|\tau(F_k^+(x,M))| \;+\; |\tau(F_k^-(x,M))|\} \;=\; 0.$

Definition 2. *Given two functionals τ_1 and τ_2 defined over the set of distribution functions, 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.

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

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

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 $k \to \infty$.

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

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

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 Q, we have that for δ sufficiently small and k 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

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

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. $\square$

Proof of Theorem ??. From the previous proposition, it follows that the mean is more robust than 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 $0 < \delta \leq 1$.

By a similar argument than in the proof of proposition **??** it follows that the first term converges to zero. Moreover 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 $|\int_0^\delta J(u)Q_k(u)du\,|$ assume $0 < F(0) < 1$ and note that

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

for some constant $M > 0$. By the pointwise convergence of F_k to F we have that for $\delta > 0$ sufficiently small and k sufficiently large, $Q_k(u) \leq 0$ 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 that $\lim_{\delta\to 0} \limsup_k\{|\int_0^\delta Q_k(u)du\,|\} = 0$, it can be easily argued that this implies the existence of δ_0 such that for each $0 \leq \delta \leq \delta_0$ we have: $|\int_0^\delta Q_k(u)du\,| \leq \epsilon$, taking k sufficiently large. When F has support restricted to the positive or negative set of real numbers the result follows easily. 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

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 mean of the ratio of their standardized asymptotic variances,

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

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

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

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 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.

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 as affected by this parameter as the statistics M_p are, replacing μ by its estimator does not introduce a significative change.
In this setup, we shall show that L-moments outperform conventional moments over certain set of distributions.

4.1 Generalized scale measures versus moments

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

Λ ={All symmetric absolutely continuous distributions F with finite 2p-th moment}

where $p > 1$. Clearly for distributions in Λ it only makes sense to consider functionals M_p when p is even, and this shall be implicitly assumed in our discussion.

In is worth noting that the class Λ contains, in addition to the normal, t-student and doble Laplace distributions, some important elliptical probability laws, see for example **??**. 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* $u \in (1/2,\, 1]$

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*

$$\lim_{u \to 1^-} J(u) \;\; \textit{exists and is strictly positive} \tag{3}$$

Then, if $p > 1$,

(*a*) $\varepsilon(M_p,\, \tau,\, \Lambda) \;>\; 0.$

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

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

Consequently, τ is more globally efficient than M_p over the set Λ.

Before presenting the proof we shall need the following lemmas.

Lemma 1. *Let F be an absolutely continuous distribution in Λ such that $E\,(X) \;=\; 0$. 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,

$$\psi(t) \ = \ \begin{cases} -\int_0^t J(u)du & 0 \ \le \ t \ \le \ 1/2 \\ \int_t^1 J(u)du & 1/2 \ \le \ t \ \le \ 1 \end{cases} \tag{4}$$

Proof. Let $q(u)$ be the derivative of $Q(u)$. We have

$$\begin{aligned} \tau(F) = \int_0^1 J(u)Q(u)du \ &= \ \int_0^{1/2} J(u)\,(-\int_u^{1/2} q(t)dt)\,du \ + \ \int_{1/2}^1 J(u)\,(\int_{1/2}^u q(t)dt)du \\ &= \ - \int_0^{1/2} (\int_0^t J(u)du)q(t)dt \ + \ \int_{1/2}^1 (\int_t^1 J(u)du)q(t)dt \\ &= \ \int_0^1 \psi(t)q(t)dt \end{aligned} \tag{5}$$

where $\psi(t)$ is given in (**??**).

To obtain a similar expression for $\zeta(F)$, from the formula: $\zeta(F) \ = \ \int_0^1 Q(u)J^*(u)du$, where $J^*(u)$ is defined by

$$J^*(u) = \begin{cases} -1 & u \in [0,\, 1/2) \qquad 1/2 = F(0) \\ \ \ 1 & u \in [1/2,\, 1] \end{cases} \tag{6}$$

.

It follows that

$$\zeta(F) = \int_0^1 \psi^*(t)q(t)dt, \tag{7}$$

where

$$\psi^*(t) \ = \ \begin{cases} t & 0 \ \le \ t \ \le \ 1/2 \\ 1-t & 1/2 \ \le \ t \ \le \ 1 \end{cases}$$

From equation (**??**), we have

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

It is easy to see that(**??**) together with the condition $E\,(X) = 0$, imply

$$\int_0^{1/2} t\,q(t)dt = \int_{1/2}^{1} (1-t)q(t)dt$$

.

Consequently,

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

which proves the assertion. □

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

Let X be any nonnegative random variable and

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

Then, if $1 \leq \alpha \leq \beta$,

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

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

Proof of Theorem ??. We divide the proof into two main parts. In the first part we show that the absolute standard deviation ζ, is more globally efficient over Λ than M_p:

1.a. $\varepsilon(M_p(F),\, \zeta(F)) \ > \ 0.$

1.b. There exists a sequence of distribution functions $F_k \in \Lambda$ such that

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

The second part involves showing that:

2.a. $\varepsilon(\zeta, \tau, \ \Lambda) \ > \ 0.$

2.b. There exists a sequence of distribution functions $F_k \in \Lambda$ such that

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

For the first part, (1.a), let $F \in \Lambda$ and $X \sim F$. Applying standard asymptotic theory of sample moments (see for example Serfling (1980)), we obtain

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

where the last inequality follows from lemma **??**.
For part (1.b.), it will be shown that

$$\varepsilon(\zeta(F_k), M_p(F_k)) = \left(\frac{E\,X_k^2}{E^2\,|X_k|} \;-\; 1\right) / \left(\frac{E\,X_k^{2p}}{E^2\,X_k^p} \;-\; 1\right) \to 0 \qquad (8)$$

for a suitable sequence $F_k \in \Lambda$ and $X_k \sim F_k$. Set $p_k = 2p + (k+1)/k$ 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 is the normalizing constant. It is clear that the resulting sequence of distribution functions will lie in the set Λ. Now straightforward calculations lead to the desired conclusion.

Now, for the second task, (2.a.), 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 **??** together with condition **??** on the score function imply that $\inf_{\{F\in\Lambda\}} \frac{\zeta(F)^2}{\tau(F)^2} > 0$. Thus, it is sufficient to show

$$\inf_{F\in\Lambda} \frac{\sigma^2(\tau,F)}{\sigma^2(\zeta,F)} \; > \; 0 \tag{9}$$

Using the symmetry of F and the "antisymmetry" property of the score functions, the integrals given in (**??**) and (**??**) can be expressed more conveniently as follows

$$\begin{aligned}
\sigma^2(\tau,F) &= \int_0^1 \int_u^1 J(u)J(v)\,u(1-v)\,q(u)q(v)dvdu \\
&= 2\int_0^{1/2}\int_u^{1-u} J(u)J(v)\,u(1-v)\,q(u)q(v)dvdu \\
&= 2\left(\int_0^{1/2}\!\!\int_u^{1/2} J(u)J(v)\,u(1-v)\,q(u)q(v)dvdu \right. \\
&\qquad \left. + \int_0^{1/2}\!\!\int_{1/2}^{1-u} J(u)J(v)\,u(1-v)\,q(u)q(v)dvdu\right) \\
&= 2\left(\int_0^{1/2}\!\!\int_u^{1/2} J(u)J(v)\,u(1-v)\,q(u)q(v)dvdu \right. \\
&\qquad \left. - \int_0^{1/2}\!\!\int_u^{1/2} J(u)J(v)\,u\,v\,q(u)q(v)dvdu\right) \\
&= 2\int_0^{1/2}\!\!\int_u^{1/2} J(u)J(v)\,u(1-2v)\,q(u)q(v)dvdu
\end{aligned} \tag{10}$$

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

$$\sigma^2(\zeta,F) = 2\int_0^{1/2}\!\!\int_u^{1/2} u(1-2v)\,q(u)q(v)dvdu \tag{11}$$

Thus,

$$\frac{\sigma^2(\zeta,F)}{\sigma^2(\tau,F)} = \frac{\int_0^{1/2}\int_u^{1/2} u(1-2v)\,q(u)q(v)dvdu}{\int_0^{1/2}\int_u^{1/2} J(u)J(v)\,u(1-2v)\,q(u)q(v)dvdu} \geq \frac{1}{M^2}, \tag{12}$$

where $M = \max_u J^2(u)$.

For the proof of $(2.b)$ note that from $(2.a)$, equation (8) $(\varepsilon(\zeta(F_k), \tau(F_k)) > 0)$, equation (9) 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 using the same sequence F_k as in equation (9). This completes the proof.

Corollary 2. *L-moments are more globally efficient than any moment functional M_p, over the class Λ, for all p even and greater than 1.*

Remark 1. *Comparisons between the class $\mathcal{L}_s$ and the class of p-th 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 **??**. *Consequently, the results from theorem* **??**, *continue to hold when τ_p^2 plays the role of M_p, for all $p > 1$.*

4.2 Higher order measures vs moments

As it was mentioned at the beginning of section 4.1, 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.

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 **??**. Through the entire discussion s will be kept fixed.

Theorem 3. *Let $\tau^* \in \mathcal{L}^*$ and $F_0 \in \Lambda$ such that $\frac{|\tau^*(F)|}{s(F)} \geq \epsilon$, for some $\epsilon > 0$*

Then, for $p > 1$,

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

(b) There exists a sequence of distribution functions $F_k \in \Lambda_{\tau^*}(\epsilon)$ such that

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

Where:

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

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 **??**. 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 **??**, the existence of positive lower bounds for the ratio $\frac{s^2(F)}{\zeta^2(F)}$, is a consequence of lemma **??** and condition (**??**) 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.

□

Remark 2. *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 include the majority of cases. 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 L-kurtosis in the interval $(-\epsilon, \epsilon)$. The later seems to be a fairly small set of probability distributions.*

References

[1] Bickel, P. J., and Lehmann, E. L. (1975). Descriptive statistics for non-parametric models. I. Introduction. II. Location. *Annals of Statistics*, **3**, 1038-1069.

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

[3] Crow, E.L., and Siddiqui, M. M. (1967). Robust estimation of location. *Journal of the American Statistical Association*, **62**, 353-389.

[4] Groeneveld, R.A., and Meeden, G. (1984). Measuring Skewness and Kurtosis. *The Statistician*, **33**, 391-399.

[5] Hull, J., and White, A. (1998). Value at Risk when daily changes in market variables are not normally distributed. *The Journal of Derivatives*, Spring, 9-19.

[6] 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.

[7] Hogg, R. V. (1974). Adaptive Robust Procedures. *Journal of the American Statistical Association.* **69**, 909-921.

[8] Jarque, C. M. and Bera, A. K. (1987). A test for Normality of Observations and Regression Residuals. *International Statistical Review*, **55**, 163-172.

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

[10] McCulloch, J.H. (1986). Simple consistent estimators of stable distribution parameters. *Communications in Statistics-Simulation and Computation*, **15**, 1109-1136.

[11] Pagan, A. (1996). The econometrics of financial markets. *Journal of Empirical Finance*, **3**, 15-102.

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

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