A Contribution to Multivariate L-Moments: L-Comoment Matrices

Robert Serfling* and Peng Xiao†
University of Texas at Dallas

Final Unabridged Version
December 2006

Abstract

Multivariate statistical analysis relies heavily on moment assumptions of second order and higher. With increasing interest in modeling with heavy tailed distributions, however, it is desirable to describe dispersion, skewness, and kurtosis of multivariate distributions under merely first order moment assumptions. Here we present a new method contributing toward this goal in both parametric and nonparametric settings. We extend the univariate L-moments of Hosking (1990), which are analogues of central moments defined for all orders under merely a first moment assumption, by introducing a notion of "L-comoments" similarly analogous to classical central moment notions of covariance, coskewness, and cokurtosis. For certain types of model, this yields correlational analysis not only coherent with classical correlation but also valid and meaningful under just first moment assumptions.

We develop basic properties and estimators for L-comoments, illustrate L-comoment matrices for several multivariate models, examine the behavior of multivariate L-moments as nonparametric descriptive measures in a sampling experiment with a heavy-tailed distribution, and consider certain extensions such as trimmed versions. Also, applications to financial risk analysis and to regional frequency analysis in environmental science are discussed.

AMS 2000 Subject Classification: Primary 62G35 Secondary 62H05.

* *Corresponding author.* Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA. Email: serfling@utdallas.edu. Website: www.utdallas.edu/∼serfling.

† Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA. Email: pengx@utdallas.edu.

Key words and phrases: L-moments; nonparametric; multivariate; coskewness; cokurtosis; financial risk analysis; regional frequency analysis.

1 Introduction

A present limitation of multivariate statistical analysis is heavy reliance on moment assumptions of second order and higher. With increasing attention, however, to the problem of modeling with heavy tailed data, for example in environmental science and financial risk analysis, it is important to rectify this shortcoming. We would like to be able to characterize typical descriptive features of the joint distribution of several covariates, for example not only dispersion but also skewness and kurtosis, without assuming moments of order higher than the first or possibly the second. Here we introduce a new multivariate analysis methodology that contributes toward this goal in both parametric or nonparametric settings.

Our approach is to develop suitable extensions of the univariate "L-moments" (Hosking [22]), which are analogues of the univariate mean and central moments and have similar interpretations but remain well-defined for all orders under merely a first moment assumption and possess other appealing properties as well. To obtain a multivariate extension, we define a notion of "L-comoments" which retains the features of the L-moments and includes analogues of classical central moment notions of covariance, coskewness, and cokurtosis. Accordingly, our multivariate extensions of L-moments for all orders higher than two consist of matrices – L-covariance, L-coskewness, L-cokurtosis, etc. – having roles analogous to the classical covariance matrix.

One important finding in our treatment is that if random variables X and Y are jointly distributed with affinely equivalent marginal distributions and with Y having linear regression on X, then under second moment assumptions the sample L-correlation estimates the same parameter as the classical sample Pearson product-moment correlation but also remains valid and meaningful under only first moment assumptions (see Proposition 3). That is, under these assumptions, the L-correlation provides a coherent extension of the classical correlation. Clearly, this result may be applied not only in parametric, but also in semiparametric and nonparametric, modeling settings.

Section 1.1 provides general background and perspective and describes the nature of our solution, while Section 1.2 presents definitions of the univariate L-moments and their leading features desired as well by L-comoments. These two sections may be read independently. Section 2 provides a needed foundation of basic results, some new, for univariate L-moments. These results, which include treatment of estimators, are instrumental in the formulation and treatment of L-comoments and multivariate

L-moment matrices in Section 3. In Section 4 we provide illustrations with some tractable multivariate models, the normal, Pareto, and Farlie-Gumbel-Morgenstern, and briefly describe some applications, to financial risk management and to regional frequency analysis in environmental science. Section 5 discusses further studies, including an extension of univariate trimmed L-moments to a notion of trimmed L-comoments.

1.1 Background and Perspective

For measuring descriptive features of a univariate distribution, the central moments are very popular, but their use is confined to sufficiently light-tailed distributions. An appealing alternative is provided by the series of L-moments, which have the form of expectations of strategically selected linear functions of order statistics. While the first L-moment is just the ordinary mean, the higher order cases not only measure spread, skewness, kurtosis, etc., just as do the central moments, but also possess attractive properties not shared by the latter. For example, the L-moment of any order k exists under merely a finite first moment assumption, making the entire series of L-moments available for typical heavy-tailed distributions. Further, the L-moments completely determine the parent distribution.

With antecedents in Sillitto [50, 51] and Downton [11], a formal and comprehensive treatment of L-moments was first developed by Hosking [22], who established foundational results supporting a new methodology in data analysis and statistical inference based on L-moments. Parametric fitting of distributions by a "method of L-moments", or exploratory and nonparametric analysis via the L-moments as descriptive measures, may be carried out.

As interest in statistical modeling using heavy-tailed distributions is increasing, so is the importance of the potential offered by the L-moment approach. In some contexts, modeling the frequency of extreme events is of particular concern. In this connection, an extensive L-moment methodology has been developed in support of regional frequency analysis in environmental science, which treats the quantiles of distributions of variables such as annual maximum precipitation, streamflow, or windspeed observed at each site in a given network. Hosking and Wallis [26] provide an excellent exposition. The L-moment approach also has special utility in applications where descriptive estimates more stable than the usual central moments are critically needed. Such concerns arise, for example, in volatility estimation in financial risk management involving market variables such as stock indices, interest rates, etc. (Hosking, Bonti, and Siegel [25]). We return to these applications in Section 6.

Data of current interest typically is multivariate in nature and calls for statistical analysis taking into account relevant underlying dependence structure and geometry. Up to now, however, L-moments have been defined only in the univariate case, and using these marginally cannot adequately capture the features of jointly distributed component variables. What is needed is an extension of the notion of L-moments to the multivariate case. Except for the obvious extension of the univariate mean to the multivariate vector mean, this has remained an open problem, in part for lack of an immediate extension of the notion of linear function of order statistics to higher dimensional space. Indeed, Hosking [22, p. 122] writes: "No extension of L-moments to multivariate distributions is immediately apparent." On the other hand, he also mentions, although without elaboration, that the "seemingly most promising approach" would be to use the notion of concomitants of order statistics to measure association between two random variables. In the present paper, we develop Hosking's insight into an effective solution.

For perspective, we note that classical multivariate analysis revolves around two key parametric entities, the mean vector and the covariance matrix, which extend the univariate mean and variance, respectively. Also, the third and fourth univariate central moments, measuring skewness and kurtosis, respectively, have been extended by Mardia [36] to certain scalar analogues for multivariate data. Further, extending the notion of covariance, very natural notions of (central) "coskewness", "cokurtosis", and even higher-order "comoments" of two jointly distributed random variables have been formulated in financial risk analysis for purposes of characterizing aspects of the response of an asset to market portfolio variations and of providing additional parameterization in the widely used Capital Asset Pricing Model (see, for example, Rubenstein [44], Fang and Lai [14], Christie-David and Chaudhry [7], Dittmar [10], and Jurczenko and Maillet [29, 30].

Our multivariate extension of the univariate L-moments proceeds by introducing, for any ordered pair of random variables (V, W) jointly distributed with finite mean, a general notion of *L-comoment of order* k, $k \geq 2$. In the case $k = 2$, this is the "Gini covariance" already studied by Schechtman and Yitzhaki [46], Yitzhaki and Olkin [55], and Olkin and Yitzhaki [41] in the contexts of income distribution, risk assessment in portfolio theory, and linear regression. The cases $k = 3$ and 4, however, offer novel analogues of the above-mentioned central coskewness and cokurtosis and may be used similarly to them, as well as in other ways, in applications.

Given a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)'$ in $\mathbb{R}^d$ with finite mean, for each $k \geq 2$ the corresponding *k-th multivariate L-moment* is then defined as the $d \times d$ matrix of L-comoments of order k for the ordered pairs (X_i, X_j), $1 \leq i, j, \leq d$. Our multivariate L-moments are thus *matrix-valued* for all orders $k \geq 2$, not merely for $k = 2$ as in the classical case. The 2nd multivariate L-moment may be thought of as a "Gini covariance matrix", an alternative to the classical covariance matrix. For $k = 3$ and 4, the multivariate L-moments represent novel "L-coskewness" and "L-cokurtosis", matrices, respectively. Because the L-comoments are similar in structure and behavior to the univariate L-moments and capture their attractive properties, the matrix-valued multivariate L-moments are effective new descriptive tools having practical utility similar to the widely used classical covariance matrix or, more precisely, to analogous but less known higher-order "central comoment" matrices.

1.2 Univariate L-Moments: Definitions and Features

Essential to our development of L-comoments is an understanding of univariate L-moments. Deferring technical results to Section 2, here we provide definitions and qualitative features. We will see that a host of favorable aspects stem from characterizations as expectations of (a) *L-statistics* (linear functions of order statistics) and (b) kernels defining *U-statistics*. With standard notation $X_{1:k} \leq X_{2:k} \leq \ldots \leq X_{k:k}$ for the ordered observations of a sample of size k from a univariate probability distribution, the kth L-moment is defined as

$$\lambda_k = k^{-1} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} E(X_{k-j:k}). \tag{1}$$

It is immediately evident that the L-moments are *scale equivariant*. The coefficients in the summation in (1) are just those of the binomial expansion of $(1 + (-1))^{k-1} = 0^{k-1}$ and hence sum to 1 or 0 according as $k = 1$ or $k \geq 2$. Thus the first L-moment, the mean $\lambda_1 = E(X_{1:1})$, is *translation equivariant*, while for $k \geq 2$ the L-moments may be characterized as linear contrasts among the expected order statistics and hence, like the central moments, are *translation invariant*: in an obvious notation,

$$\lambda_k(\theta + \eta X) = \eta\, \lambda_k(X), \tag{2}$$

for $\eta > 0$ and arbitrary θ, i.e., for $k \geq 2$ and $\eta > 0$, the k-th L-moment of the distribution $F(\eta^{-1}(x - \theta))$ is η times that of $F(x)$. Also, $\lambda_k(-X) = (-1)^k \lambda_k(X)$.

The coefficients in (1) are also those of the forward difference operator of order $k-1$, $k \geq 2$, in which case λ_k becomes the $(k-1)$th iteration of the forward difference operator applied to the sequence $\{E(X_{j:k}) : j = 1, \ldots, k\}$. Consequently, for $k \geq 2$, λ_k is the $(k-2)$th iteration of the first order differences $\{E(X_{j+1:k}) - E(X_{j:k}) : j = 1, \ldots, k-1\}$ and hence may be expressed as a linear contrast among the *expected spacings* $\chi_{j:k} = E(X_{j+1:k} - X_{j:k})$, $1 \leq j \leq k-1$, from a sample of size k:

$$\lambda_k = k^{-1} \sum_{j=0}^{k-2} (-1)^j \binom{k-2}{j} \chi_{k-1-j:k}. \tag{3}$$

This representation greatly facilitates the interpretation and understanding of L-moments. Although carried somewhat further here, the idea of connecting with expected spacings is due to Sillitto [50].

The 2nd L-moment clearly measures *spread*: $\lambda_2 = \frac{1}{2}E(X_{2:2} - X_{1:2}) = \chi_{1:2}/2$. In fact, it is a well-known such measure: one-half the classical *Gini mean difference* (Gini [17]). Besides its intrinsic interest, λ_2 is used to obtain *scale-free* higher-order descriptive measures,

$$\tau_k = \lambda_k/\lambda_2,\ \ k \geq 3,$$

called *L-moment ratios* (Hosking and Wallis, 1997). Very conveniently for practical use and interpretation, these L-moment coefficients satisfy (Hosking, 1989)

$$-1 \leq \tau_k \leq 1,\ \ k \geq 3. \tag{4}$$

In comparison, the classical central moment analogues (further discussed in Section 3.1.1) do not satisfy any such inequality.

The 3rd L-moment is simply the difference in expectations of the two spacings from a sample of size 3, $\lambda_3 = \frac{1}{3}E(X_{3:3} - 2X_{2:3} + X_{1:3}) = (\chi_{2:3} - \chi_{1:3})/3$, and hence measures *skewness* (unscaled). That it measures skewness may also be seen from its representation as a difference of two location measures, the expected values of the sample mean and the sample median for a sample of size 3: $\lambda_3 = E((X_{1:3} + X_{2:3} + X_{3:3})/3) - E(X_{2:3})$. As pointed out by Hosking [22], by the result of Robbins [43] that the expected range for sample size 3 is three-halves the expected range for sample size 2, we may write $\tau_3 = E(Q_3 - 2Q_2 + Q_1)/E(Q_3 - Q_1)$ in terms of the sample *quartiles* for sample size 3, $Q_1 = (X_{1:3} + X_{2:3})/2$, $Q_2 = X_{2:3}$, and $Q_3 = (X_{2:3} + X_{3:3})/2$, so that τ_3 is a direct analogue of *Bowley's skewness measure* (Bowley [5]).

The 4th L-moment $\lambda_4 = \frac{1}{4}E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4})$ measures *kurtosis*, as argued very nicely by Hosking [22]. This also may be seen very directly via (3) by writing λ_4 as a difference of measures of *spread in the tails* and of *spread in the center*:

$$\lambda_4 = (\chi_{3:4} - 2\chi_{2:4} + \chi_{1:4})/4 = [(\chi_{3:4} + \chi_{1:4})/2 - \chi_{2:4}]/2,$$

namely, the expected average of the outer two spacings minus the expected middle spacing, times one-half.

To interpret the 5th L-moment, we first note that all L-moments of odd order higher than the first are zero in the case of a symmetric distribution and hence may be regarded as "generalized skewness measures" (Hosking [22]). In particular, then, a departure of $\lambda_5 = (\chi_{4:5} - 3\chi_{3:5} + 3\chi_{2:5} - \chi_{1:5})/5$ from zero would seem to indicate skewness combined with departure from unimodality.

For the first four L-moments for a variety of univariate distributions, see Hosking and Wallis [26, Appendix]. In particular, for the Normal(μ, σ^2) distribution, we have $\lambda_1 = \mu$, $\lambda_2 = \pi^{-1/2}\sigma$, $\lambda_3 = 0$, and $\lambda_4 = (30\pi^{-1} \arctan \sqrt{2} - 9)\pi^{-1/2}\sigma$. For the uniform distribution on (a, b), we have $\lambda_1 = (a+b)/2$, $\lambda_2 = (b-a)/6$, and $\lambda_k = 0$, $k \geq 2$. This intuitively appealing property that L-measures for skewness, kurtosis, bimodality, etc., are all zero for uniform distributions is not shared by the central moments.

Let us summarize key features of univariate L-moments:

1. *Existence (finite) for all orders, if first moment finite.*
2. *Distribution determined by its L-moments, if first moment finite.*

   Distinct distributions generate distinct series of L-moments. For the ordinary moments, this holds only under more restrictive conditions (e.g., [49, §1.13]).
3. *L-functional representations, with mutually orthogonal weight functions.*

   L-moments of different orders capture sharply different population features.
4. *Representation as the expected value of an L-statistic (linear function of order statistics), for each choice of order k and sample size $n \geq k$.*

   For $n = k$, this is the defining expression for the kth L-moment and provides the relevant kernel for a U-statistic sample version (item 5 below). For $n \geq k$, this suggests an L-statistic sample version (item 6 below).
5. *U-statistic structure of sample versions.*

   Thus standard U-statistic theory may be applied, for variance computations, martingale representations, almost sure behavior, asymptotic normality, and asymptotic variance estimation.
6. *L-statistic structure of sample versions.*

   This enables computation of the sample kth L-moment based on sample size n in $O(n \log n)$ time, instead of $O(n^k)$ time via the U-statistic representation, and provides another standard approach to asymptotic normality.
7. *Sample versions unbiased as estimators of population L-moments.*

   This follows from each of the U-statistic and L-statistic sample versions, which are equivalent. In comparison, sample central moments of order greater than 1 are biased.
8. *Sample L-moments more stable than central versions, increasingly with higher order.*

   Large deviations from the mean influence the sample central moments with increasing impact $(x-\overline{x})^k$ as the order k increases, while the sample L-moments are only linearly influenced for any order.

2 Univariate L-Moments: Technical Basics

We provide here certain basic results some new, for univariate L-moments, covering representations, estimation, and asymptotic behavior. Besides having independent interest, these results and and methods of proof are instrumental to our development of multivariate extensions in Section 3. Throughout, we consider a univariate distribution function F having quantile function F^{-1} and L-moment sequence $\{\lambda_k\}$.

2.1 Representations

Several different types of representation for L-moments prove useful: as an L-functional, as a covariance, and as a linear function of expected values of order statistics.

2.1.1 An Expression for λ_k as an L-Functional

Substitution into (1) of a standard expression for the expected value of an order statistic (e.g., David and Nagaraja, 2003),

$$\begin{aligned} E(X_{r:n}) &= r\binom{n}{r}\int_0^1 F^{-1}(u)\,u^{r-1}\,(1-u)^{n-r}\,du \qquad (5)\\ &= n\binom{n-1}{r-1}\int_{-\infty}^{\infty} x\,[F(x)]^{r-1}\,[1-F(x)]^{n-r}\,dF(x),\ \ r \le n, \qquad (6)\end{aligned}$$

yields a classical L-functional representation,

$$\lambda_k = \int_0^1 F^{-1}(u)\,P^*_{k-1}(u)\,du, \qquad (7)$$

where

$$P^*_k(u) = \sum_{j=0}^{k} p^*_{k,j}\,u^j,$$

with $p^*_{k,j} = (-1)^{k-j}\binom{k}{j}\binom{k+j}{j}$. For general treatment of L-functionals, see [49, Chap. 8] and [23]. For discussion of (7) in particular, see Hosking [22] and Hosking and Wallis [26, §§2.4–2.5]. The functions $P^*_r(u)$, $r = 0, 1, 2, \ldots$, comprise the *shifted Legendre* system of orthogonal polynomials, i.e., the standard Legendre polynomials defined over the interval $-1 \le u' \le 1$ shifted to $0 \le u \le 1$ via $u' = 2u - 1$. By the orthogonality, the λ_k capture differing types of information about the underlying F. For example, $\lambda_1 = \int_0^1 F^{-1}(u)\,du$ (the mean of F) and $\lambda_2 = \int_0^1 F^{-1}(u)\,(2u-1)\,du$. The latter expression is easily transformed to another well-known representation (Stuart [52]) for the Gini mean difference,

$$\lambda_2 = 2\,\mathrm{Cov}(X, F(X)) = \mathrm{Cov}(X, 2F(X) - 1). \qquad (8)$$

In terms of the usual centered rank function $2F(x) - 1$, λ_2 may be interpreted as the *covariance* of X and its *centered rank*. By the Cauchy-Schwarz inequality, (8) yields a useful inequality comparing the second L-moment with the usual standard deviation:

$$\lambda_2 \leq \sigma/\sqrt{3}. \tag{9}$$

We shall see an application in Section 4.3. In different contexts, (9) was given by Plackett [42] and an equivalent result derived by Schucany, Parr, and Boyer [45]. Next we develop novel and productive covariance expressions for λ_k in general.

2.1.2 An Expression for λ_k as a Covariance

Straightforward transformation in (7) yields

$$\lambda_k = \sum_{j=0}^{k-1} p^*_{k-1,j}\, \beta_j, \tag{10}$$

with $\beta_j = \int_0^1 F^{-1}(u) u^j du = E(XF(X)^j)$, as given in Hosking and Wallis [26], formula (2.36). Using $P^*_0(u) \equiv 1$ and orthogonality of the functions P^*_k, we readily obtain

$$\begin{aligned} \lambda_k &= \mathrm{Cov}(X, P^*_{k-1}(F(X))) + \mathbf{1}\{k = 1\} E(X) \\ &= \begin{cases} E(X), & k = 1; \\ \mathrm{Cov}(X, P^*_{k-1}(F(X))), & k \geq 2. \end{cases} \end{aligned} \tag{11}$$

For $k \geq 2$, equation (11) facilitates an illuminating characterization: the kth L-moment is the covariance of X and a particular function of its *rank* $F(X)$. The case $k = 2$ was noted in (8). For $k = 3$, we have

$$\lambda_3 = -6\, \mathrm{Cov}(X, F(X)(1 - F(X))), \tag{12}$$

expressing λ_3 as the covariance of X and a function symmetric about the median of F. It follows that, as noted earlier, λ_3 is a skewness measure which is zero if the distribution of X is symmetric.

Remark. Via (10), the sequences $\{\lambda_k\}$ and $\{\beta_k\}$ are equivalent, and in turn these are equivalent to $\{E(X_{k:k})\}$, since via (6) we have $\beta_j = E(XF^j(X)) = (j+1)^{-1} E(X_{j+1:j+1})$. By Chan [6], in the case of finite mean, the sequence $\{E(X_{k:k})\}$ determines F and hence so does the sequence of L-moments. See Hosking [22] for further discussion. Also, by (10) we thus have an expression for λ_k in terms of *expected extreme values*: $\lambda_k = \sum_{j=0}^{k-1} p^*_{k-1,j}\, (j+1)^{-1}\, E(X_{j+1:j+1})$. □

2.1.3 An Expression for λ_k in Terms of $\{E(X_{r:n}),\, r = 1, \ldots, n\}$

The use of (6) with the definition of β_j yields

$$E(X_{r:n}) = n\binom{n-1}{r-1}\sum_{j=0}^{n-r}\binom{n-r}{j}(-1)^{n-r-j}\beta_{n-1-j}, \tag{13}$$

which can be inverted via some manipulations to obtain

$$\beta_k = n^{-1}\binom{n-1}{k}^{-1}\sum_{j=k+1}^{n}\binom{j-1}{k}E(X_{j:n}). \tag{14}$$

The use of (14) in (10) then yields the desired representation:

$$\begin{aligned}\lambda_k &= \sum_{j=0}^{k-1} p^*_{k-1,j}\, n^{-1}\binom{n-1}{j}^{-1}\sum_{r=j+1}^{n}\binom{r-1}{j}E(X_{r:n}) \\ &= n^{-1}\sum_{r=1}^{n} w^{(k)}_{r:n}\, E(X_{r:n}), \end{aligned} \tag{15}$$

where

$$w^{(k)}_{r:n} = \sum_{j=0}^{\min\{r-1,k-1\}} (-1)^{k-1-j}\binom{k-1}{j}\binom{k-1+j}{j}\binom{n-1}{j}^{-1}\binom{r-1}{j}.$$

The sample version of (15) and illustrative visual display of $w^{(k)}_{r:n}$ for $n = 19$ and $k \leq 4$ are given by Hosking and Wallis [26, formula (2.59) and Fig. 2.6]. In particular, $w^{(1)}_{r:n} \equiv 1$ and $w^{(2)}_{r:n} = = (2r - n - 1)/(n - 1)$, $r = 1, \ldots, n$. (The symmetry and skew-symmetry of the coefficients $w^{(1)}_{r:n}$ and $w^{(2)}_{r:n}$, respectively, in the index r holds in general, for k even and odd, respectively: $w^{(k)}_{r:n} = (-1)^{k-1}\, w^{(k)}_{n-r+1:n}$, $r = 1, \ldots, n$.)

2.2 Estimation

We now consider estimation of λ_k based on a sample of size n. Estimators of the scaled versions τ_k follow automatically.

2.2.1 An Unbiased L-Statistic Estimator for λ_k

An immediate application of (15) is to suggest the estimator

$$\widehat{\lambda}_k = n^{-1}\sum_{r=1}^{n} w^{(k)}_{r:n}\, X_{r:n}, \tag{16}$$

which is that given by Hosking and Wallis [26], formula (2.59) – an *L-statistic* in form, and *unbiased* for λ_k.

2.2.2 Representation of $\widehat{\lambda}_k$ as a U-Statistic

In particular, for $k = 1$ and 2, formula (16) yields, respectively, $\widehat{\lambda}_1 = \overline{X}$, the sample mean, and (see Serfling [49, p. 263] and Hosking and Wallis [26, p. 30]) $\widehat{\lambda}_2 = \frac{1}{2}G$, where G is the U-statistic known as Gini's mean difference,

$$G = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} |X_i - X_j|.$$

Not only are $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$ thus U-statistics, but also *each* $\widehat{\lambda}_k$ is a U-statistic. To see this, first note from (1) that $\lambda_k = E(h(X_1, \ldots, X_k))$, where

$$h(x_1, \ldots, x_k) = k^{-1} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} x_{k-j:k}. \tag{17}$$

Now, in general, for any kernel $h(x_1, \ldots, x_k)$ which is a linear combination of the order statistics of its arguments, it is not difficult to show that the corresponding U-statistic based on a sample of size n may be expressed also as a linear combination of the order statistics of the full sample. This follows by a straightforward, but tedious, derivation, or by using a technique of Blom [4] to obtain the relevant coefficients in this linear combination from a particular generating polynomial associated with the given kernel. Consequently, the U-statistic based on the kernel (17) is found to agree with the L-statistic representation given by (16).

2.2.3 A Sample Analogue Estimator for λ_k

Let $\widehat{F}_n$ denote the usual sample cdf based on $X_1, \ldots, X_n$ with corresponding quantile function $\widehat{F}_n^{-1}$. Substitution of $\widehat{F}_n^{-1}$ into (7) yields another L-statistic estimator,

$$\widehat{\lambda}_k^* = \int_0^1 \widehat{F}_n^{-1}(u) P_{k-1}^*(u) du = \sum_{i=1}^n c_{n,i} X_{i:n}, \tag{18}$$

where $c_{n,i} = \int_{(i-1)/n}^{i/n} P_{k-1}^*(u) du$. For $k = 1$ the estimators agree, $\widehat{\lambda}_1^* = \hat{\lambda}_1 = \overline{X}$, and for $k = 2$ they are close: $\widehat{\lambda}_2^* = n^{-1}(n-1)\,\widehat{\lambda}_2$. By a straightforward derivation it follows that in general $\widehat{\lambda}_k^* - \widehat{\lambda}_k = o_p(n^{-1})$, $n \to \infty$, yielding equivalence of these estimators with respect to asymptotic distribution theory under typical assumptions (see below). In the sequel we confine attention to the unbiased version $\widehat{\lambda}_k$.

2.2.4 Asymptotic Distribution Theory for $\widehat{\lambda}_k$ and $\widehat{\tau}_k$

From the foregoing results, under second moment conditions on F, standard theory for U-statistics and L-statistics [49, Chaps. 5 and 8] yields that *the vector of the first k L-moments is asymptotically k-variate normal*, with a similar result for the vector of scaled L-moments. These and related results are given by Hosking [22].

3 L-Comoments and Multivariate L-Moments

As discussed in Section 1.1, our multivariate extensions of the univariate L-moments for order $k \geq 2$ are *matrix-valued*, with elements the *L-comoments*. We now introduce these and examine major properties, key inequalities, representations in terms of concomitants of order statistics, suitable estimators, and asymptotic convergence. Concluding this section, we discuss the matrices formed by the L-comoments.

3.1 Definition and Properties of L-Comoments

3.1.1 Preliminary on Central Comoments

Helpful perspective on the L-comoments will be provided by comparison with the central comoments, which we briefly review. Consider a bivariate random variable $(X^{(1)}, X^{(2)})$ having cdf F with marginal distributions F_1 and F_2, means μ_1 and μ_2, and finite central moments $\mu_k^{(1)}$ and $\mu_k^{(2)}$, for $2 \leq k \leq K$. The classical scaled central moments are given by $\psi_k^{(i)} = \mu_k^{(i)}/(\mu_2^{(i)})^{k/2}$, $k \geq 3$, for $i = 1, 2$, the cases $k = 3$ and 4 denoting the classical *skewness* and *kurtosis* coefficients, respectively. The central moment coefficients $\psi_k^{(i)}$, $k \geq 3$, do not satisfy any universal bounds and can have arbitrarily large magnitudes. To interpret sample values, therefore, it is conventional to compare with values from specific reference distributions.

We now introduce related *central comoments*, which are (asymmetric) higher order analogues of covariance that have been developed in financial risk modeling (discussed in detail in Section 4.4). For each $k \geq 2$, the *kth central comoment of $X^{(1)}$ with respect to $X^{(2)}$* is defined as

$$\xi_{k\,[12]} = \mathrm{Cov}(X^{(1)}, (X^{(2)} - \mu_1^{(2)})^{k-1}).$$

(The analogous asymmetric counterpart is denoted $\xi_{k\,[21]}$.) Of course, for the 2nd order case, we have simply $\xi_{2\,[12]} = \xi_{2\,[21]} = \sigma_{12}$, the usual *covariance*. The symmetry in this instance is merely an artifact of the definition of comoments that shows up just for $k = 2$, rather than being a feature necessarily desired for comoments in general. Indeed, for higher order cases one could produce symmetric versions, if desired, by taking signed versions of $\sqrt{\xi_{k\,[12]}\,\xi_{k\,[21]}}$, for example. The ordered pairs $(\xi_{k\,[12]}, \xi_{k\,[21]})$, $k \geq 3$, however, carry greater information while still being simple and therefore are preferred.

Corresponding to the comoments, scale-free versions are given by

$$\psi_{k\,[12]} = \xi_{k\,[12]}/(\mu_2^{(1)})^{1/2}(\mu_2^{(2)})^{(k-1)/2},$$

the case $k = 2$ being the usual correlation coefficient denoted ρ_{12}. In particular, we call $\xi_{3\,[12]}$ and $\xi_{4\,[12]}$ the *coskewness* and *cokurtosis*, respectively, and $\psi_{3\,[12]}$ and $\psi_{4\,[12]}$ the coskewness and cokurtosis *coefficients*, respectively, of $X^{(1)}$ with respect to $X^{(2)}$.

Drawing upon familiarity with covariance, it is straightforward to interpret the central comoments. For example, the *coskewness* $\xi_{3\,[12]}$ of $X^{(1)}$ with respect to $X^{(2)}$ increases or decreases with relatively higher or lower weight, respectively, on points $(x^{(1)}, x^{(2)})$ with positive deviations $x^{(1)} - \mu_1^{(1)}$, for given size of squared deviation $(x^{(2)} - \mu_1^{(2)})^2$. In similar vein, the *cokurtosis* with cubing of deviations $x^{(2)} - \mu_1^{(2)}$ produces still greater sensitivity to tailweight of $X^{(2)}$ along with a signed effect.

A variant notion of kth central comoment is based on k jointly distributed variables $X^{(1)}, \ldots, X^{(k)}$ and given by the tensor of terms

$$E\{(X^{(1)} - \mu_1^{(1)})(X^{(2)} - \mu_1^{(2)}) \cdots (X^{(k)} - \mu_1^{(k)})\}.$$

See Athayde and Flôres [3] and Jurczenko, Maillet, and Merlin [31] for discussion in the context of financial risk modeling.

3.1.2 L-Comoments

Consider a bivariate random variable $(X^{(1)}, X^{(2)})$ having cdf F with finite mean, marginal distributions F_1 and F_2, and L-moment sequences $\{\lambda_k^{(1)}\}$ and $\{\lambda_k^{(2)}\}$. By analogy with the covariance representation (11) for L-moments, and also by analogy with the central comoments, we define associated *L-comoments of order* $k \geq 2$ by

$$\lambda_{k\,[12]} = \mathrm{Cov}(X^{(1)}, P_{k-1}^*(F_2(X^{(2)}))) \tag{19}$$

and

$$\lambda_{k\,[21]} = \mathrm{Cov}(X^{(2)}, P_{k-1}^*(F_1(X^{(1)}))). \tag{20}$$

Here $\lambda_{k\,[12]}$ and $\lambda_{k\,[21]}$ need not be equal. We term these, respectively, the *kth L-comoment of $X^{(1)}$ with respect to $X^{(2)}$* and the *kth L-comoment of $X^{(2)}$ with respect to $X^{(1)}$*. We emphasize the first case, results for the other case being similar.

It is readily checked that the kth L-comoment of $X^{(1)}$ with respect to $X^{(2)}$ is *translation invariant* and *scale equivariant* with respect to transformations of $X^{(1)}$ and *translation and scale invariant* with respect to transformations of $X^{(2)}$. That is,

$$\lambda_{k\,[12]}(\theta + \eta X^{(1)}, \zeta + \beta X^{(2)}) = \eta\, \lambda_{k\,[12]}(X^{(1)}, X^{(2)}), \tag{21}$$

for positive η and β and arbitrary θ and ζ. Appropriate scaled versions are thus given by

$$\tau_{k\,[12]} = \lambda_{k\,[12]}/\lambda_2^{(1)},$$

the analogues of the τ_k. We call these *L-comoment coefficients*. The case $\tau_{2\,[12]}$ we also term the *L-correlation* of $X^{(1)}$ with respect to $X^{(2)}$ and denote by $\rho_{[12]}$.

As discussed with the central comoments, we could construct symmetric versions (see [55] for some examples and comparisons), but the more fundamental notion of an ordered pair of asymmetric comoments is preferred. Fortuitously, in the case of L-comoments we have this option even in the 2nd order case. Indeed, the (asymmetric) 2nd order comoments and related correlations arise instrumentally for a natural decomposition of the 2nd L-moment of an arbitrary sum of random variables: for univariate $Y_1, \ldots, Y_n$ and $S = Y_1 + \cdots + Y_n$,

$$\begin{aligned}
\lambda_2(S) &= 2\,\mathrm{Cov}(S, F_S(S)) \\
&= 2\sum_{i=1}^{n} \mathrm{Cov}(Y_i, F_S(S)) \\
&= \sum_{i=1}^{n} \lambda_{2[12]}(Y_i, S) \\
&= \sum_{i=1}^{n} \rho_{[Y_i,S]}\,\lambda_2(Y_i).
\end{aligned}$$

For $X^{(1)} = X^{(2)}$, the L-comoments reduce to the L-moments: $\lambda_{k\,[12]} = \lambda_{k\,[21]} = \lambda_k^{(1)} = \lambda_k^{(2)}$, which we also denote by $\lambda_{k\,[11]}$ and $\lambda_{k\,[22]}$. On the other hand, for $X^{(1)}$ and $X^{(2)}$ *independent* the L-comoments of all orders $k \geq 2$ take the value 0. (Similar remarks hold for the central comoments in these two cases.)

A convenient tool is that the comoment of $X^{(1)}$ with respect to $X^{(2)}$ can be expressed as that of $E(X^{(1)} \,|\, X^{(2)})$ with respect to $X^{(2)}$. The following proposition and corollary are used to advantage in Section 4.

Proposition 1 *Let $X^{(1)}$ have finite mean. Then, for $k \geq 2$,*

$$\lambda_{k\,[12]} = \mathrm{Cov}(E(X^{(1)} \,|\, X^{(2)}), P_{k-1}^* \circ F_2(X^{(2)})) \tag{22}$$

and, under finiteness of the kth moment of $X^{(2)}$,

$$\xi_{k\,[12]} = \mathrm{Cov}(E(X^{(1)} \,|\, X^{(2)}), (X^{(2)} - \mu_1^{(2)})^{k-1}). \tag{23}$$

Proof. By straightforward steps with conditional expectation, for any measurable function $Q(\cdot)$ we have

$$\begin{aligned}
\mathrm{Cov}(X^{(1)}, Q(X^{(2)})) &= E(X^{(1)}Q(X^{(2)})) - E(X^{(1)})E(Q(X^{(2)})) \\
&= E(E(X^{(1)} \,|\, X^{(2)})Q(X^{(2)})) - E(E(X^{(1)} \,|\, X^{(2)}))E(Q(X^{(2)})) \\
&= \mathrm{Cov}(E(X^{(1)} \,|\, X^{(2)}), Q(X^{(2)})).
\end{aligned}$$

Taking in turn $Q(x) = P^*_{k-1} \circ F_2(x)$ and $Q(x) = (x - \mu_1^{(2)})^{k-1}$ with $k \geq 2$, we obtain (22) and (23). $\square$

If $E(X^{(1)} \,|\, X^{(2)})$ is a *strictly increasing* function $g(X^{(2)})$, say, then $F_{g(X^{(2)})}(g(X^{(2)})) = F_2(X^{(2)})$, and by Proposition 1 $\lambda_{k\,[12]}$ is simply the kth L-moment of $g(X^{(2)})$. In particular, for g *linearly* increasing we obtain

Corollary 2 *Let $X^{(1)}$ have finite mean and linear regression on $X^{(2)}$: $E(X^{(1)} \,|\, X^{(2)}) = a + b\,X^{(2)}$. Then, for $k \geq 2$,*

$$\lambda_{k\,[12]} = b\,\lambda_k^{(2)} \tag{24}$$

and, under finiteness of the kth moment of $X^{(2)}$,

$$\xi_{k\,[12]} = b\,\mu_k^{(2)}. \tag{25}$$

When $X^{(1)}$ has linear regression on $X^{(2)}$ and F_1 and F_2 are affinely equivalent, there hold simple expressions for $\tau_{k\,[12]}$ in terms of $\tau_k^{(1)}$ and $\psi_{k\,[12]}$ in terms of $\psi_k^{(1)}$. In the case $k = 2$ these yield the important result that, under the assumed conditions, the L-correlation $\rho_{[12]}$ not only agrees with the classical Pearson product-moment correlation ρ_{12} but also assumes the same formula in terms of model parameters while remaining well-defined under lesser moment assumptions.

Proposition 3 *Assume (i) $(X^{(1)}, X^{(2)})$ has joint distribution with linear regression of $X^{(1)}$ on $X^{(2)}$: for some constants a and b, $E(X^{(1)} \,|\, X^{(2)}) = a + bX^{(2)}$. Also, assume (ii) the respective marginals F_1 and F_2 are affinely equivalent: for some constants θ and η, $F_2(x) = F_1(\eta^{-1}(x - \theta))$, i.e., $X^{(2)} \stackrel{d}{=} \theta + \eta X^{(1)}$. Then*

$$\rho_{[12]} = b\eta = \rho_{12} \tag{26}$$

holds under second moment assumptions, with the first equality valid as well under only first moment assumptions. Also, for $k \geq 2$,

$$\lambda_{k\,[12]} = b\,\eta\,\lambda_k^{(1)} = \rho_{[12]}\,\lambda_k^{(1)} \tag{27}$$

and thus

$$\tau_{k\,[12]} = \rho_{[12]}\,\tau_k^{(1)}, \tag{28}$$

and, under finiteness of the kth moment of $X^{(2)}$*,*

$$\xi_{k\,[12]} = b\,\eta^k\,\mu_k^{(1)} \tag{29}$$

and thus

$$\psi_{k\,[12]} = b\,\eta\,\psi_k^{(1)} = \rho_{12}\,\psi_k^{(1)}. \tag{30}$$

Proof. Under 1st moment assumptions,

$$\rho_{[12]} = \lambda_{2\,[12]}/\lambda_2^{(1)} = b\lambda_2^{(2)}/\lambda_2^{(1)} = b\eta,$$

the 1st equality being the definition, the 2nd following by Corollary 2 using (i) and (24), and the 3rd following by (ii) and (2). Also, in standard notation, under 2nd moment assumptions we have

$$\rho_{12} = \sigma_{12}/\sigma_1\sigma_2 = b\sigma_2^2/\sigma_1\sigma_2 = b\sigma_2/\sigma_1 = b\eta,$$

where the 2nd equality follows by (i) and the last by (ii). This yields (26), and similar arguments using Corollary 2 lead to (27) and (29). □

Thus, in the setting of Proposition 3, the L-correlation $\rho_{[12]}$ coherently extends the Pearson correlation ρ_{12} to cases when 2nd moments are not finite, and, further, the comoment coefficients are simply described by the correlation coefficient and marginal coefficients. We illustrate in Section 4 with some specific multivariate models. Besides a role in parametric modeling, Proposition 3 also has application, as does the L-moment approach in general, to moment-based nonparametric descriptive analysis.

3.1.3 Key Inequalities for L-Comoments

An important result about L-correlation is that like the Pearson version its values lie between ± 1. Unlike ρ_{12}, however, which attains these extremes only for linear relationships between the variables, $\rho_{[12]}$ more broadly attains these under any strictly monotone relationship. In the same sense that ρ_{12} is said to measure linearity, we may consider $\rho_{[12]}$ to measure monotonicity. This is established rigorously in the following proposition and proof (for useful previous treatments see [46], [47]).

Proposition 4 *In general,*

$$|\lambda_{2\,[12]}| = 2\,|\mathrm{Cov}(X^{(1)}, F_2(X^{(2)})| \le 2\,\mathrm{Cov}(X^{(1)}, F_1(X^{(1)}) = \lambda_2^{(1)} \tag{31}$$

and thus

$$-1 \le \rho_{[12]} \le 1. \tag{32}$$

The upper (lower) bound in (32) *is attained when* $X^{(1)}$ *and* $X^{(2)}$ *are related a.s. through a strictly increasing (decreasing) function, and in the case of continuous distributions this condition is necessary as well.*

Proof. Let F_{12} denote the joint distribution of $X^{(1)}$ and $X^{(2)}$. By a well-known lemma of Hoeffding [20] quoted in [34], for any random variables V and W with finite $E|V|$, $E|W|$, and $E|VW|$, we have

$$\mathrm{Cov}(V,W) = \int\int [F_{V,W}(v,w) - F_V(v)F_W(w)]\,dv\,dw.$$

Transforming by $v = x^{(1)}$ and $w = F_2(x^{(2)})$ and checking that $F_W(w) = F_2(x^{(2)})$ and $F_{V,W}(v,w) = F_{12}(x^{(1)}, x^{(2)})$, we obtain

$$\mathrm{Cov}(X^{(1)}, F_2(X^{(2)})) = \int\int [F_{12}(x^{(1)}, x^{(2)}) - F_1(x^{(1)})F_2(x^{(2)})]\,dx^{(1)}\,dF_2(x^{(2)}). \quad (33)$$

Now, for any jointly distributed X and Y with specified marginals F_X and F_Y, the joint distribution $F_{X,Y}(x,y)$ must satisfy (see [16], [33]) the well-known and quickly derived Fréchet bounds

$$\max\{F_X(x) + F_Y(y) - 1, 0\} \le F_{X,Y}(x,y) \le \min\{F_X(x), F_Y(y)\}.$$

The upper (or lower) bound is attained under the condition that $Y = g(X)$ a.s. for some strictly increasing (or decreasing) function g, since then $F_{g^{-1}(Y)}(g^{-1}(Y)) = F_Y(Y)$ (or $1 - F_Y(Y)$). In the case that F_X and F_Y are continuous distributions, this condition is necessary [48, Theorem 2]. Applying the upper Fréchet bound with (33), we obtain

$$\begin{aligned}
&\mathrm{Cov}(X^{(1)}, F_2(X^{(2)})) \\
&\le \int\int [\min\{F_1(x^{(1)}), F_2(x^{(2)})\} - F_1(x^{(1)})F_2(x^{(2)})]\,dx^{(1)}\,dF_2(x^{(2)}) \\
&= \int\int [\min\{F_1(x), u\} - F_1(x)\,u]\,dx\,du. \qquad (34)
\end{aligned}$$

Also, hypothetically for the moment taking $X^{(2)} = X^{(1)}$, in which case $F_2(X^{(2)}) = F_1(X^{(1)})$ and the joint distribution of $X^{(1)}$ and $F_1(X^{(1)})$ attains the upper bound, the same steps yield

$$\mathrm{Cov}(X^{(1)}, F_1(X^{(1)})) = \int\int [\min\{F_1(x), u\} - F_1(x)\,u]\,dx\,du. \quad (35)$$

Combining (34) and (35), we have

$$\mathrm{Cov}(X^{(1)}, F_2(X^{(2)})) \le \mathrm{Cov}(X^{(1)}, F_1(X^{(1)})).$$

Now using $\max\{a+b-1, 0\} - ab = -[\min\{1-a, b\} - (1-a)b]$, along with the lower Fréchet bound, a similar derivation leads to

$$\mathrm{Cov}(X^{(1)}, F_2(X^{(2)})) \ge -\mathrm{Cov}(X^{(1)}, F_1(X^{(1)})),$$

completing the proof. □

Remark. The quantity in (35) may also be evaluated as

$$
\begin{aligned}
&\int\int [\min\{F_1(x), u\} - F_1(x)\,u]\,dx\,du \\
&= \int \left[\left(\int_0^{F_1(x)} u\,du \right) [1 - F_1(x)] + \left(\int_{F_1(x)}^1 (1-u)\,du \right) F_1(x) \right] dx \\
&= \frac{1}{2} \int F_1(x)[1 - F_1(x)]\,dx,
\end{aligned}
$$

yielding for the 2nd L-moment of a univariate distribution F the representation

$$\lambda_2 = \int F(x)[1 - F(x)]\,dx,$$

also given by Hosking [21] and equivalent to a result for the Gini mean difference obtained in [35]. □

Generalization of Proposition 4 to higher order comoment coefficients is somewhat problematic. Under the assumptions of Proposition 3, however, we can assert that both L-comoment and central comoment coefficients are bounded by corresponding univariate analogues. Using (32) with (28) and the standard inequality for Pearson correlation with (30), we obtain the following result.

Corollary 5 *Under the conditions of Proposition 3, we have for $k \geq 2$*

$$|\tau_{k\,[12]}| \leq |\tau_k^{(1)}| \quad (\leq 1,\ \text{by (4)}) \tag{36}$$

and

$$|\psi_{k\,[12]}| \leq |\psi_k^{(1)}| \quad (\leq \infty). \tag{37}$$

3.1.4 L-Correlation, L-Coskewness, and L-Cokurtosis

In applications the L-comoments of primary interest are those for $k = 2$, 3, and 4. We briefly discuss their special features.

L-correlation. The 2nd L-comoments and the L-correlations $\rho_{[12]}$ and $\rho_{[21]}$ have already been studied by Schechtman and Yitzhaki [46], [47], Yitzhaki and Olkin [55], and Olkin and Yitzhaki [41] as *Gini covariances* and *Gini correlations*, with emphasis on applications in economics. Comparisons are made with the Pearson product-moment and Spearman rank correlation coefficients based on $\mathrm{Cov}(X^{(1)}, X^{(2)})$ and $\mathrm{Cov}(F_1(X^{(1)}), F_2(X^{(2)}))$, respectively, and a "Gini regression analysis" based on Gini covariance is developed.

While $\rho_{[12]}$ and $\rho_{[21]}$ need not agree, each being a covariance between one variable and the rank of another, with differing choices of which variable is the one ranked, they do agree when $(X^{(1)}, X^{(2)})$ are centrally symmetric, i.e., $(X^{(1)}, X^{(2)}) \stackrel{d}{=} (X^{(2)}, X^{(1)})$, or more generally are exchangeable. See Schechtman and Yitzhaki [46], [47] for detailed discussion and examples.

One natural way to interpret the second L-comoment relative to the standard covariance is through the following analogue of (8):

$$\lambda_{2\,[12]} = 2\,\mathrm{Cov}(X^{(1)}, F_2(X^{(2)})) = 2\,\mathrm{Cov}(X^{(1)}, F_2(X^{(2)}) - 1/2).$$

Thus $\lambda_{2\,[12]}$ differs from σ_{12} simply in replacing the deviation $X^{(2)} - \mu_1^{(2)}$ of $X^{(2)}$ from its mean by the deviation $F_2(X^{(2)}) - 1/2$, a scale-free measure of the deviation of $X^{(2)}$ from its *median*. This approach applies equally well to higher-order cases, as seen next.

L-skewness and L-kurtosis. For $\lambda_{3\,[12]}$, the *L-coskewness of* $X^{(1)}$ *with respect to* $X^{(2)}$, we have in analogy with (12)

$$\begin{aligned}
\lambda_{3\,[12]} &= \mathrm{Cov}(X^{(1)}, P_2^*(F_2(X^{(2)}))) \\
&= -6\,\mathrm{Cov}(X^{(1)}, F_2(X^{(2)})(1 - F_2(X^{(2)})) \\
&= 6\,\mathrm{Cov}(X^{(1)}, (F_2(X^{(2)}) - 1/2)^2),
\end{aligned}$$

Thus $\lambda_{3\,[12]}$ differs from its central comoment analogue $\xi_{3\,[12]}$ simply by replacing the deviation $X^{(2)} - \mu_1^{(2)}$ by the deviation $F_2(X^{(2)}) - 1/2$.

For $\lambda_{4\,[12]}$, the *L-cokurtosis of* $X^{(1)}$ *with respect to* $X^{(2)}$, we have

$$\begin{aligned}
\lambda_{4\,[12]} &= \mathrm{Cov}(X^{(1)}, P_3^*(F_2(X^{(2)}))) \\
&= \mathrm{Cov}(X^{(1)}, 20F_2^3(X^{(2)}) - 30F_2^2(X^{(2)}) + 12F_2(X^{(2)}) - 1) \\
&= \mathrm{Cov}(X^{(1)}, 20(F_2(X^{(2)}) - 1/2)^3 - 3(F_2(X^{(2)}) - 1/2)).
\end{aligned}$$

Again the L-comoment differs from its central counterpart by replacing $X^{(2)} - \mu_1^{(2)}$ by $F_2(X^{(2)}) - 1/2$, except that here in addition the particular function of the deviation also changes (slightly, to one less appealing in form, perhaps, but still an odd function and indeed orthogonal to those for other choices of k).

We see that L-comoments provide a hierarchy of intuitively appealing analogues of classical covariance and central comoments. The definitions in terms of the classical covariance operator facilitate useful interpretations and comparisons.

Further understanding of these quantities is provided in the following subsection through useful representations in terms of *concomitants*. In fact, such representations serve as an equivalent way to define the L-comoments.

3.2 Representations for $\lambda_{k\,[12]}$ in Terms of Concomitants

Consider now a sample $\{(X_i^{(1)}, X_i^{(2)}), 1 \le i \le n\}$ from a bivariate distribution $F(x^{(1)}, x^{(2)})$ with marginal distributions F_1 and F_2. Corresponding to the ordered $X^{(2)}$-values $X_{1:n}^{(2)} \le X_{2:n}^{(2)} \le \cdots \le X_{n:n}^{(2)}$, we call the element of $\{X_1^{(1)}, \ldots, X_n^{(1)}\}$ that is paired with $X_{r:n}^{(2)}$ the *concomitant* of $X_{r:n}^{(2)}$ and denote it by $X_{[r:n]}^{(12)}$ (see Yang [53] and David and Nagaraja [8] for general treatments).

It is quickly seen that $E(X_{[r:n]}^{(12)}) = n\, E(X_1^{(1)} \mid X_1^{(2)} = X_{r:n}^{(2)})$, leading immediately to the following useful analogue of (6):

$$E(X_{[r:n]}^{(12)}) = \qquad (38)$$
$$n\binom{n-1}{r-1} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^{(1)}\,[F_2(x^{(2)})]^{r-1}\,[1-F_2(x^{(2)})]^{n-r}\,dF(x^{(1)}, x^{(2)}),\ r \le n.$$

That is, $E(X_{[r:n]}^{(12)}) = n\binom{n-1}{r-1} E(X^{(1)}\,[F_2(X^{(2)})]^{r-1}\,[1-F_2(X^{(2)})]^{n-r})$. We use (38) to establish the following representation, which states that the L-comoments may be defined in terms of expected values of concomitants in exactly the same way that the L-moments are defined in terms of expected values of order statistics. (This does not quite mean, however, that the L-comoments can be called the L-moments of the concomitants.)

Proposition 6 *The kth L-comoment of $X^{(1)}$ with respect to $X^{(2)}$ may be represented as*

$$\lambda_{k\,[12]} = k^{-1} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} E(X_{[k-j:k]}^{(12)}). \qquad (39)$$

Proof. Denote the right-hand side of (39) by L_k. With (38) inserted into (39), similar steps as with the insertion of (6) into (1) in Section 2 yield the following concomitant analogue of (11):

$$L_k = \mathrm{Cov}(X^{(1)}, P_{k-1}^{*}(F_2(X^{(2)}))),\ k \ge 2. \qquad (40)$$

Thus the L_k are in fact the L-comoments defined by (19). □

Proposition 6 immediately yields another proof, communicated by Jon Hosking, of the inequality (31) given in Proposition 4 (although not, however, the statement of necessary and sufficient conditions). We merely apply to (39) the well-known result of Hardy, Littlewood, and Pólya [18, Theorem 368] that, given an ordered sequence $a_1 \le a_2 \le \cdots \le a_I$ and any other sequence $b_1, \ldots, b_I$, *the sum of products* $\sum_{i=1}^{I} a_i b_{\sigma(i)}$ *for a permutation* $(\sigma(1), \ldots, \sigma(I))$ *of* $(1, \ldots, I)$ *attains its maximum (minimum) possible value when the sequence* $b_{\sigma(i)}, \ldots, b_{\sigma(I)}$ *is increasing (decreasing).*

The main role of Proposition 6, however, is to make it straightforward to obtain key results for L-comoments as analogues of those for L-moments, with concomitants in place of order statistics. Thus, for example, defining $\chi^{(12)}_{[j:k]} = E(X^{(12)}_{[j+1:k]}) - E(X^{(12)}_{[j:k]})$, we have for $k \geq 2$ the following analogue of (3):

$$\lambda_{k\,[12]} = k^{-1} \sum_{j=0}^{k-2} (-1)^j \binom{k-2}{j} \chi^{(12)}_{[k-1-j:k]}, \tag{41}$$

providing yet another approach toward interpretation of the L-comoments. Further, using again the derivations of Section 2, we obtain for a sample of size n a direct analogue of (15) and thus the basis for unbiased estimation of comoments:

Proposition 7 *For $k \geq 2$, and with $w^{(k)}_{r:n}$ the same as in (15),*

$$\lambda_{k\,[12]} = n^{-1} \sum_{r=1}^{n} w^{(k)}_{r:n}\, E(X^{(12)}_{[r:n]}). \tag{42}$$

3.3 Estimation of L-Comoments

Proposition 7 yields for the kth L-comoment, $k \geq 2$, the unbiased estimator

$$\widehat{\lambda}_{k\,[12]} = n^{-1} \sum_{r=1}^{n} w^{(k)}_{r:n}\, X^{(12)}_{[r:n]}, \tag{43}$$

which is an *L-statistic in the concomitants*. Further, each $\widehat{\lambda}_{k\,[12]}$ is a *U-statistic*. To see the latter, note from Proposition 6 that

$$\lambda_{k\,[12]} = E(h^{(k)}((X^{(1)}_1, X^{(2)}_1), \ldots, (X^{(1)}_k, X^{(2)}_k))),$$

where

$$h^{(k)}((x^{(1)}_1, x^{(2)}_1), \ldots, (x^{(1)}_k, x^{(2)}_k)) = k^{-1} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} x^{(12)}_{[k-j:k]}. \tag{44}$$

It is routine to show that the discussion of Section 2.2.1 carries over: for a kernel h with bivariate arguments which is a linear combination of the concomitants of one of the components, the corresponding U-statistic based on a sample of size n may be expressed as a linear combination of the concomitants of that component for the full sample. Thus the U-statistic based on the kernel (44) agrees with the L-statistic given by (43).

Illustration: Estimation of 2nd L-Comoment. For $k = 2$ in (43) and (44), we obtain the L- and U-statistic representations

$$\widehat{\lambda}_{2\,[12]} = n^{-1} \sum_{r=1}^{n} \frac{2r - n - 1}{n - 1} X^{(12)}_{[r:n]} = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} (X^{(12)}_{[j:n]} - X^{(12)}_{[i:n]})/2,$$

analogous to expressions for the 2nd L-moment, as expected. (The present U-statistic representation, however, cannot be reexpressed as one-half the Gini mean difference of the concomitants, because the relevant kernel in the concomitants, $(x^{(12)}_{[2:2]} - x^{(12)}_{[1:2]})/2$, is not the same as the kernel $|x^{(12)}_{[2:2]} - x^{(12)}_{[1:2]}|/2$ for the Gini mean difference.) $\square$

The asymptotic distribution of a vector of L-comoment estimators follows from standard theory for U-statistics (Serfling [49]). Noting that the kernel $h^{(i)}$ defined by (44) with $k = i$ is symmetric in its i arguments, and defining

$$g^{(i)}(x^{(1)}, x^{(2)}) = i\, E(h^{(i)}((x^{(1)}, x^{(2)}), (X^{(1)}_2, X^{(2)}_2), \ldots, (X^{(1)}_i, X^{(2)}_i)))$$

and $\zeta_{ij} = \mathrm{Cov}(g^{(i)}(X^{(1)}, X^{(2)}), g^{(j)}(X^{(1)}, X^{(2)}))$, $2 \le i, j \le k$, we have

Proposition 8 *Under second moment assumptions on $X^{(1)}$, for $k \ge 2$ the vector of sample L-comoments $(\widehat{\lambda}_{2\,[12]}, \ldots, \widehat{\lambda}_{k\,[12]})'$ is asymptotically $(k-1)$-variate normal with mean $(\lambda_{2\,[12]}, \ldots, \lambda_{k\,[12]})'$ and covariance matrix $[\zeta_{ij}]/n$.*

(Alternatively, this result follows using (43) with results of Yang [54].) Asymptotic normality of the corresponding vector of scaled versions $\widehat{\tau}_{i\,[12]}$, $2 \le i \le k$, follows by standard results on transformations of asymptotically normal vectors.

3.4 Multivariate L-Moments

For a random d-vector $\boldsymbol{X} = (X^{(1)}, \ldots, X^{(d)})'$, we now define "multivariate L-moments" for all orders $k \ge 1$. The first order multivariate L-moment is simply the *vector mean*

$$\boldsymbol{\lambda}_1 = E(\boldsymbol{X}),$$

assumed finite. For $k \ge 2$, the *kth multivariate L-moment* is the matrix of kth L-comoments for all pairs $(X^{(i)}, X^{(j)})$, $1 \le i, j \le d$:

$$\boldsymbol{\Lambda}_k = (\lambda_{k\,[ij]})_{d \times d},$$

with $\boldsymbol{\Lambda}_2$, $\boldsymbol{\Lambda}_3$, and $\boldsymbol{\Lambda}_4$ the *L-covariance*, *L-coskewness*, and *L-cokurtosis* matrices, respectively. Corresponding versions with scaled elements are given by $\boldsymbol{\Lambda}^*_k = (\tau_{k\,[ij]})$, the *L-comoment coefficient matrices.* The diagonals of $\boldsymbol{\Lambda}_k$ and $\boldsymbol{\Lambda}^*_k$ are the componentwise univariate L-moments and L-moment coefficients, respectively.

In Section 4 we illustrate these matrices in various settings. We also compare with the corresponding *central* versions, denoted by $\boldsymbol{\Xi}_k = (\xi_{k[ij]})$ and $\boldsymbol{\Xi}^*_k = (\psi_{k[ij]})$, respectively, $k \ge 2$ ($\boldsymbol{\Xi}_2$ and $\boldsymbol{\Xi}^*_2$ being the usual covariance and correlation matrices).

4 Illustrations and Applications

In the multivariate case, tractable distributions are fewer and parametric approaches more limited than in the univariate setting. Although univariate L-moments provide a useful alternative to the classical method of moments in parametric model-fitting, and such an approach indeed extends to the multivariate case, the widest and most significant role of multivariate L-moments lies in providing attractive *nonparametric multivariate estimators and descriptive measures*. Using the estimators and theory of Sections 2.2 and 3.3, one may readily compute for a data set sample versions of $\mathbf{\Lambda}_2^*$, $\mathbf{\Lambda}_3^*$, and $\mathbf{\Lambda}_4^*$ and (under second moment assumptions) characterize asymptotic distributions.

As with classical correlation, by Proposition 4 the strengths of the L-correlations between the ith and jth component variables are assessed through comparison with the value 1. No such simple guideline exists in the case of higher orders, however, neither for central comoment nor L-comoment coefficients, nor for the univariate central counterparts. One compensating approach is to rely upon suitable reference multivariate distributions selected as meaningful benchmarks. Another is to introduce productive special assumptions, which may be verified for a particular model or assumed in a nonparametric formulation. In particular, we have

Proposition 9 *Assume that the components of $\boldsymbol{X} = (X^{(1)}, \ldots, X^{(d)})'$ have affinely equivalent marginal distributions and pairwise linear regressions, in the sense of the conditions of Proposition 3. Then marginal L-moment coefficients agree and likewise for marginal central moment coefficients:*

$$\tau_k^{(1)} = \ldots = \tau_k^{(d)} = \tau_k, \ \text{say}, \tag{45}$$

$$\psi_k^{(1)} = \ldots = \psi_k^{(d)} = \psi_k, \ \text{say}, \tag{46}$$

for $k \geq 3$. Further,

$$\rho_{[ij]} = \rho_{ij} = \rho_{[ji]}, \ 1 \leq i, j \leq d,$$

yielding, with $\mathbf{C} = (\rho_{ij}) = (\rho_{[ij]})$,

$$\Lambda_k^* = \tau_k \, \mathbf{C}, \tag{47}$$

$$\Xi_k^* = \psi_k \, \mathbf{C}. \tag{48}$$

This result follows readily from Proposition 3. In each of (47) and (48), the comoment coefficient matrix is simply the product of the univariate moment coefficient of the same order and the correlation matrix $\mathbf{C}$. The central comoment and L-comoment coefficient matrices are both, in this instance, equivalent in structure to the usual correlation matrix, which thus contains all of the multivariate shape information (in the scale-free sense).

In Sections 4.1, 4.2, and 4.3, respectively, we illustrate for three multivariate distributions – normal, Pareto, and Farlie-Gumbel-Morgenstern. The first two are governed by Proposition 9, the third is not. In Sections 4.4 and 4.5, respectively, we indicate the role of multivariate L-moments in portfolio risk analysis in finance and regional frequency analysis in environmental science.

4.1 The Multivariate Normal Distribution

For a d-variate normal model with variances σ_i^2 and covariances σ_{ij}, the assumptions of Propositions 1, 3, and 9 are fulfilled with $b = \sigma_{ij}/\sigma_j^2$, $\eta = \sigma_j/\sigma_i$, and thus $b\eta = \rho_{ij}$. The comoments are given by $\lambda_{k\,[ij]} = (\sigma_{ij}/\sigma_j^2)\lambda_k^{(j)}$ and $\xi_{k\,[ij]} = (\sigma_{ij}/\sigma_j^2)\mu_k^{(j)}$, and the comoment coefficients by $\tau_{k\,[ij]} = \rho_{ij}\tau_k$ and $\psi_{k\,[ij]} = \rho_{ij}\psi_k$, $k \geq 2$. For odd $k \geq 3$, these quantities are all 0. For even k, the central moment coefficients are invariant over parameters and readily found to be $\psi_k = (k-1)(k-3)\cdots 3\cdot 1$. The quantities τ_k are more elusive, explicit expressions for the expected values of order statistics for normal samples in terms of elementary functions being known only for sample sizes ≤ 5. For a range of larger sample sizes, however, these expected values have been computed numerically and tabulated, and approximations are available for indefinitely large sample sizes. See Johnson, Kotz, and Balakrishnan [28, pp. 94–96] for discussion. In particular, the 2nd, 3rd, and 4th normal L-moments were mentioned in Section 1.2 and yield $\tau_3 = 0$, and $\tau_4 = (30\pi^{-1}\arctan\sqrt{2} - 9)$.

4.2 A Multivariate Pareto Distribution

Various forms of multivariate Pareto distribution are treated by Arnold [2]. We consider here his Type II version, given by the d-variate joint cdf

$$F(x^{(1)}, \ldots, x^{(d)}) = 1 - \left[1 + \sum_{i=1}^{d}\left(\frac{x^{(i)} - \theta_i}{\sigma_i}\right)\right]^{-\alpha}, \tag{49}$$

for $x^{(i)} > \theta_i$ and $\sigma_i > 0$, $1 \leq i \leq d$, and $\alpha > 0$. The kth moment is finite if $k < \alpha$. Many typical applications involve heavy-tailed modeling, with α in the range 1 to 2 for quite diverse data sets (see, for example, [2, Appendix A], [28, p. 575], and [37]).

With $\theta_i = \sigma_i$, $1 \leq i \leq d$, (49) reduces to the *Type I Pareto model*, of long-standing use in actuarial science and economics. With $\theta_i \equiv 0$, (49) becomes the *multivariate Lomax distribution* (Nayak [40]) arising in reliability theory for the joint distribution of lifetimes in a system of d components having conditionally independent exponential failure rates of form Δ/σ_i, where Δ is a random "environmental" effect following a gamma distribution with shape parameter α. For general discussion of model (49), see [2] and [33, pp. 380–382 and 603–605].

For *parametric* inference using this model, one may estimate the parameters via the maximum likelihood method. Also, fortuitously, tractable formulas for the L-moments, L-comoments, central moments, central comoments, and related coefficients are available. These, provided below, support method-of-moments approaches to parameter estimation.

We also may apply model (49) to explore, comparatively with central versions, the behavior of the sample L-moments, L-comoments, and related coefficients as *nonparametric descriptive measures* based on data from an unknown and possibly heavy-tailed distribution. Some sampling and simulation results are provided below.

4.2.1 Formulas

For $\boldsymbol{X} = (X^{(1)}, \ldots, X^{(d)})'$ having distribution (49), $X^{(i)}$ has marginal distribution $F_i(x^{(i)}) = 1 - [1 + (\sigma_i^{-1}(x^{(i)} - \theta_i))]^{-\alpha}$ and linear regression on $X^{(j)}$, fulfilling the assumptions and conclusions of Propositions 1, 3, and 9 with $\eta = \sigma_j/\sigma_i$, $b = \sigma_i/\sigma_j\alpha$, and

$$\boldsymbol{C} = \begin{pmatrix} 1 & \alpha^{-1} & \cdots & \alpha^{-1} \\ \alpha^{-1} & 1 & \cdots & \alpha^{-1} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha^{-1} & \cdots & \alpha^{-1} & 1 \end{pmatrix}_{d\times d}.$$

For the distribution F_i we readily obtain from (5) that

$$E(X_{r:n}^{(i)}) = \theta_i + \sigma_i \left(\frac{n!}{(n-r)!(n-1/\alpha)(n-1-1/\alpha)\cdots(n-r+1-1/\alpha)} - 1 \right).$$

This yields $\lambda_1^{(i)} = \theta_i + \sigma_i/(\alpha - 1) = \mu_1^{(i)}$,

$$\lambda_k^{(i)} = \sigma_i\, \alpha\, \frac{\prod_{j=0}^{k-2}(j\alpha + 1)}{\prod_{j=1}^{k}(j\alpha - 1)},\ k \geq 2,$$

and

$$\tau_k = \frac{\prod_{j=0}^{k-2}(j\alpha + 1)}{\prod_{j=3}^{k}(j\alpha - 1)},\ k \geq 3.$$

For computation of the kth central moment, we assume without loss of generality that $\theta_i = 0$ and use [2, (3.3.8)] $E(X^{(i)}) = \sigma_i^k k!/(\alpha-1)\cdots(\alpha-k)$, which leads to

$$\begin{aligned} \mu_k^{(i)} &= \sigma_i^k\, k!\, [(\alpha-1)^k(\alpha-2)\cdots(\alpha-k)]^{-1} \times \\ & \sum_{j=0}^{k} (-1)^{k-j}[(k-j)!]^{-1}(\alpha-1)^{j-1}(\alpha-j-1)\cdots(\alpha-k), \end{aligned}$$

$2 \le k < \alpha$, which in turn determines the central moment coefficient ψ_k. In particular, $\psi_3 = 2(\alpha+1)((\alpha-2)/\alpha)^{1/2}/(\alpha-3)$ and $\psi_4 = 3(3\alpha^2+\alpha+2)(\alpha-2)/\alpha(\alpha-3)(\alpha-4)$.

With $\mathbf{C}$, ψ_k, and τ_k as above, the comoment coefficient matrices for this model are now given by (47) and (48). Here the factors τ_k and ψ_k depend not only upon k but also upon the shape parameter α. The use of (47) requires $\alpha > 1$, while (48) requires $\alpha > k$. We thus obtain for this model an extended correlation analysis, since the formula α^{-1} for all the Pearson correlations under $\alpha > 2$ holds also for all the L-correlations under $\alpha > 1$. The maximal value 1/2 for the correlation under $\alpha > 2$ increases to 1 and becomes approached, as $\alpha \downarrow 1$.

4.2.2 Some Empirical Results

To examine the performance of sample L-moments and L-comoments, with special reference to the case of heavy-tailed data, and to compare with corresponding central versions, we provide a small simulation study using the above Pareto II model. For each of $\alpha = 1.5$, 2.5, 3.5, and 4.5, and for each of sample sizes $n = 50$ and 500, we generated 20,000 samples from the cdf (49) with $d = 3$, $\theta_i \equiv 0$, and $\sigma_i \equiv 1$. Each trivariate observation $\boldsymbol{X} = (X^{(1)}, X^{(2)}, X^{(3)})'$ was obtained via the representation [2, p. 252] $X^{(i)} = W_i/Z$, $1 \le i \le 3$, with independent standard exponential random variables W_1, W_2, and W_3 and gamma$(\alpha, 1)$ random variable Z. For each sample, the L-moments, L-comoments, central moments, central comoments, and corresponding coefficients were computed for orders $k \le 4$. With these simulated data, we can compare, on the basis of 20,000 observations each, the L-versions and central versions of *multivariate nonparametric descriptive measures for spread, skewness, and kurtosis* (taking into account, of course, that each such quantity is measured in a different way by the two versions). Here we present and discuss selected representative results.

Results for L-moments and L-comoments of orders 2–4 as well as for L-correlation are provided for $\alpha = 1.5$ and 4.5 in Tables 1 and 2, respectively. Table 2 also includes corresponding results for central versions (which are defined for $\alpha = 4.5$). For each target parameter and each choice of sample size, we list the population value and, based on the 20,000 sample estimates, the mean estimate (mean), the median estimate (med), the coefficient of variation (CV) of the estimates, and the relative interquartile range (RIQR, defined as IQR/med) of the estimates. The results in the tables support a number of conclusions:

- The CV and RIQR variability measures both decrease as the sample size n increases. However, in the case of $\alpha = 1.5$, the decrease in the CV is only slight, reflecting the higher sensitivity of this measure to extreme observations.
- The CV and RIQR variability measures both increase as the order k increases, with the increase in CV for the central versions very dramatic.

- For $\alpha = 4.5$, comparison of each sample L-moment or sample L-comoment with its central counterpart on the basis of CV and RIQR values, and also in terms of the observed bias (discrepancy between mean estimate and target parameter, and between median estimate and target parameter), indicates that the L-versions are much more stable and efficient than the central versions as estimators of their respective parameters. For order ≥ 3, the central versions are especially erratic. Even though the 4th order central moment is finite, this case of α is still quite heavy-tailed.

- For estimation of correlation, in the case of $\alpha = 4.5$ the sample L-correlation and the sample Pearson correlation are both fairly strong, with pronounced improvement in the case of $n = 500$ versus $n = 50$. Nevertheless, by both the CV and RIQR measures and in terms of the observed bias, the sample L-correlation is distinctly more stable and efficient. In the very heavy-tailed case of $\alpha = 1.5$, however, the sample L-correlation is noticeably less efficient, while the sample Pearson correlation is completely off the mark (figures not included).

- In comparison with the second order sample L-moment and sample L-comoment, the sample L-correlation exhibits considerably greater observed bias (Tables 1 and 2). This effect holds true also for the central versions (Table 2). Even more pronounced for higher orders (figures not included here), this is merely a consequence of the fact that in general the sample central comoment coefficients and the sample L-comoment coefficients are, unfortunately, all *biased* estimators.

- The sample L-comoments for $\lambda_{2[12]}$ and $\lambda_{2[21]}$ (which are equal in the present model) behave very consistently for each case of α.

- *Summary Comment.* For nonparametric moment-based description with data from a possibly heavy-tailed distribution, L-versions offer clear advantages over central versions. The gain increases with increasing order of moments and with increasing heaviness of tails.

Table 1: L-Moment and L-Comoment Sampling Results, $\alpha = 1.5$.

Target Parameter	True Value	Sample Values $n = 50$				Sample Values $n = 500$			
		mean	med	CV	RIQR	mean	med	CV	RIQR
λ_2	1.50	1.47	1.09	1.78	0.67	1.52	1.32	1.70	0.32
$\lambda_{2[12]}$	1.00	0.97	0.56	2.69	1.09	1.01	0.82	2.54	0.49
$\lambda_{2[21]}$	1.00	0.96	0.60	2.93	1.08	1.00	0.82	2.04	0.49
$\rho_{[12]}$	0.67	0.56	0.58	0.37	0.47	0.63	0.63	0.15	0.19
λ_3	1.07	1.05	0.66	2.49	0.95	1.09	0.89	2.37	0.44
$\lambda_{3[12]}$	0.71	0.69	0.32	3.75	1.65	0.73	0.53	3.52	0.69
λ_4	0.86	0.83	0.45	3.11	1.21	0.87	0.67	2.95	0.56
$\lambda_{4[12]}$	0.57	0.54	0.20	4.70	2.27	0.59	0.39	4.37	0.88

4.3 Multivariate Farlie-Gumbel-Morgenstern Distributions

An appealing structure for construction of joint distributions having given marginals was introduced by Morgenstern [38] and extended by Farlie [15]. Considerable further development has led to several variant so-called Farlie-Gumbel-Morgenstern (FGM) classes of distributions. General discussion is provided by Kotz, Balakrishnan, and Johnson [33], whose equation (44.73) gives the particular version we consider here:

$$\begin{aligned} F(x^{(1)}, \ldots, x^{(d)}) = & \\ \prod_{i=1}^{d} F_i(x^{(i)}) \times \Bigg[1 + & \left(\sum_{1 \le i_1 < i_2 \le d} \alpha_{i_1 i_2} \, (1 - F_{i_1}(x^{(i_1)})) \, (1 - F_{i_2}(x^{(i_2)})) \right) + \\ \cdots + & \left(\alpha_{1\,2 \cdots d} \prod_{i=1}^{d} (1 - F_i(x^{(i)})) \right) \Bigg], \end{aligned} \tag{50}$$

with the coefficients $a_{i_1 \cdots i_\ell}$ real-valued and assumed to satisfy the constraints

$$1 + \left(\sum_{1 \le i_1 < i_2 \le d} \alpha_{i_1 i_2} \, \varepsilon_{i_1} \, \varepsilon_{i_2} \right) + \cdots + \alpha_{1\,2 \cdots d} \, \varepsilon_1 \cdots \varepsilon_d \ge 0,$$

for all cases of $\varepsilon_i = \pm 1$, a sufficient condition for $F(x^{(1)}, \ldots, x^{(d)})$ in (50) to be a non-decreasing function of its arguments. The case of mutually independent components is given by $\alpha_{i_1 \cdots i_\ell} \equiv 0$. For $\boldsymbol{X} = (X^{(1)}, \ldots, X^{(d)})'$ having cdf (50), $X^{(i)}$ has marginal distribution $F_i(\cdot)$, thus determining the marginal L-moments and central moments

Table 2: L-Moment, L-Comoment, Moment, and Comoment Sampling Results, $\alpha = 4.5$.

Target Parameter	True Value	Sample Values $n = 50$ mean	med	CV	RIQR	$n = 500$ mean	med	CV	RIQR
λ_2	0.161	0.161	0.155	0.25	0.31	0.161	0.160	0.08	0.10
$\lambda_{2[12]}$	0.036	0.036	0.032	1.04	1.36	0.036	0.035	0.32	0.43
$\lambda_{2[21]}$	0.036	0.036	0.032	1.04	1.37	0.036	0.035	0.32	0.43
$\rho_{[12]}$	0.222	0.209	0.216	0.90	1.21	0.221	0.221	0.29	0.39
λ_3	0.071	0.071	0.065	0.45	0.54	0.071	0.070	0.14	0.18
$\lambda_{3[12]}$	0.016	0.016	0.013	1.93	2.81	0.016	0.015	0.59	0.80
λ_4	0.042	0.042	0.036	0.41	0.77	0.042	0.041	0.19	0.25
$\lambda_{4[12]}$	0.009	0.010	0.007	2.93	4.53	0.009	0.009	0.86	1.15
μ_2	0.147	0.150	0.108	1.39	0.86	0.147	0.135	0.45	0.36
$\xi_{2[12]}$	0.033	0.034	0.019	2.68	1.86	0.032	0.028	0.70	0.63
ρ_{12}	0.222	0.204	0.181	0.98	1.49	0.217	0.267	0.39	0.50
μ_3	0.308	0.413	0.138	7.56	1.55	0.388	0.247	3.72	0.77
$\xi_{3[12]}$	0.068	0.101	0.019	12.5	2.96	0.083	0.046	3.33	1.11
μ_4	3.227	2.517	0.185	25.4	2.50	2.485	0.551	19.7	1.40
$\xi_{4[12]}$	0.717	0.738	0.019	43.3	4.66	0.452	0.091	16.0	1.86

under relevant moment conditions. For derivation of the comoments and comoment coefficients, we use the bivariate distributions

$$F_{ij}(x^{(i)}, x^{(j)}) = F_i(x^{(i)})F_j(x^{(j)})[1 + \alpha_{ij}(1 - F_i(x^{(i)}))(1 - F_j(x^{(j)}))],$$

with $|\alpha_{ij}| \leq 1$, from which it follows [33, p. 56] that $X^{(i)}$ has linear regression on $F_j(X^{(j)})$ with regression coefficient $b = 4\,\alpha_{ij}\,\mathrm{Cov}(X^{(i)}, F_i(X^{(i)})) = 2\,\alpha_{ij}\,\lambda_2^{(i)}$. Applying Proposition 1 with $g = F_j$, we obtain under appropriate moment conditions

$$\lambda_{k\,[ij]} = 2\,\alpha_{ij}\,\lambda_2^{(i)}\,\mathrm{Cov}(F_j(X^{(j)}), P^*_{k-1}(F_j(X^{(j)})))$$

and

$$\xi_{k\,[ij]} = 2\,\alpha_{ij}\,\lambda_2^{(i)}\,\mathrm{Cov}(F_j(X^{(j)}), (X^{(j)} - \mu_1^{(j)})^{k-1}),$$

for $k \geq 2$.

Let us now take all F_j to be *continuous*. Then the covariance in $\lambda_{k\,[ij]}$ is by (11) just the kth L-moment of the uniform$(0,1)$ distribution, which equals $1/6$ for $k = 2$ and 0 for $k \geq 3$, due to the orthogonality of the Legendre polynomials P^*_ℓ. The central comoments $\xi_{k\,[12]}$, however, are nonzero for $k \geq 3$. Treating in detail the case $k = 2$, we have $\lambda_{2\,[ij]} = \alpha_{ij}\,\lambda_2^{(i)}/3$ and $\sigma_{ij} = \xi_{2\,[ij]} = \alpha_{ij}\,\lambda_2^{(i)}\,\lambda_2^{(j)}$, with corresponding correlations $\rho_{[ij]} = \alpha_{ij}/3$ under first moment assumptions and $\rho_{ij} = \alpha_{ij}\lambda_2^{(i)}\lambda_2^{(j)}/\sigma_i\sigma_j$ under second moment assumptions. By (9) we thus obtain the inequality

$$|\rho_{ij}| \leq |\alpha_{ij}|/3 = |\rho_{[ij]}|.$$

The correlations ρ_{ij} and $\rho_{[ij]}$ are different multiples of α_{ij}. The multiplicative factor for ρ_{ij} involves the marginal distributions F_i and F_j through not only their standard deviations but also, interestingly, their second L-moments. For F_i and F_j normal $\rho_{ij} = \alpha_{ij}/\pi$, for F_i and F_j uniform $\rho_{ij} = \alpha_{ij}/3$. On the other hand, for the L-correlation $\rho_{[ij]}$, the multiplicative factor 1/3 does not depend upon the marginal distributions, so that the L-correlation can serve nonparametric or semi-parametric modeling using the FGM structure.

Since $|\alpha_{ij}| \leq 1$, both correlations are no greater than 1/3 in magnitude, limiting practical application of the FGM model to situations involving only weak pairwise linear dependence among the component variables. The weak dependence within this model is further manifest, in a new way, by the L-coskewness, L-cokurtosis, and higher-order L-comoments all being 0, similar to the case of independent variables. Despite the limited possible magnitude of the L-correlation, its availability under merely first moment assumptions usefully extends the range of meaningful FGM modeling with heavy-tailed marginal distributions.

4.4 Modeling for Portfolio Risk Analysis in Finance

Among approaches to portfolio optimization in finance, a central role has long been played by the Capital Asset Pricing Model (CAPM), initially involving just first and second moments but recently including consideration of higher moments. Skewness measures are of interest for evaluation of the downside risk and asymmetric volatility of a portfolio, aspects increasingly important to investors' preferences. Along with variance, measures of kurtosis are instrumental and appealing in accounting for high volatility and uncertainty in returns. Just as the covariance of a particular security and the " market return" measures the contribution of that security to the dispersion of a well-diversified portfolio, the corresponding coskewness and cokurtosis measure the contributions of that asset to the overall skewness and kurtosis of the portfolio. For detailed discussion of the increasing interest in these and higher order comoments in connection with the CAPM in financial risk analysis, see Fang and Lai [14], Harvey and Siddique [19], Christie-David and Chaudhry [7], Jurczenko and Maillet [29, 30], Dittmar [10], Jurczenko, Maillet, and Negrea [32], de Prado, Mailoc, and Peijan [9], Jurczenko, Maillet, and Merlin [31], and Adcock, Jurczenko, and Maillet [1]. In particular, for example, for a d-vector $\boldsymbol{X} = (X^{(1)}, \ldots, X^{(d)})'$ of returns for a portfolio of d assets, Christie-David and Chaudhry (2001) study the corresponding kth central comoment matrices $\Xi_k = (\xi_{k\,[ij]})_{d \times d}$ and also the d-vector $(\xi_{k\,[1*]}, \ldots, \xi_{k\,[d*]})$ of the kth central comoments of portfolio returns taken over each asset with a "market return" $X^{(*)}$, for $k = 2, 3, 4$.

Also increasing is interest in heavy-tailed distributions in modeling stock returns, raising serious concern regarding higher moment assumptions and issues of stability and robustness associated with higher order central moments and comoments. In fact, for the *marginal* distributions of jointly distributed heavy-tailed variables in risk analysis, univariate L-moments have already been applied (see, for example, Hosking, Bonti, and Siegel [25]). Such treatments can be extended using L-comoments. In a nonparametric approach, L-moments and L-comoments may be used descriptively in evaluating risk characteristics of a portfolio. Alternatively, a parametric approach could be explored using, for example, a multivariate Pareto model.

4.5 Modeling for Regional Frequency Analysis in Environmental Science

Many environmental applications involve, for each variable of interest, for example streamflow, separate series of observations taken at different measurement sites within a network. This yields for a given variable multiples samples of similar data, with possible dependence within as well as between samples. One key goal is to estimate the marginal quantile function, especially for the purpose of determining the upper quantile corresponding to occurrence of a specified "extreme" event. In many applications the sample size for a site is too small to enable efficient estimation of upper quantiles, and data for all sites within a suitable region are combined through a "regional frequency analysis" under effective simplifying assumptions. Distributions involving tails heavier than those of normal distributions have been found relevant, and fitting by maximum likelihood or classical moment methods has been problematic. In this context, L-moment methods have proved very effective in providing stable and reliable estimates less sensitive to distributional assumptions and to influence of extreme observations. See Hosking and Wallis [26] for complete exposition of regional frequency analysis via L-moment methodology as it has developed in fields such as hydrology, climatology, and environmental science.

In regional frequency analysis, a major step is to partition a network of sites into *approximately homogeneous regions* of sites with very similar frequency distributions for the variable of interest. For each site, the vector of the first four sample L-moments or L-moment coefficients is obtained and "unusual" sites are identified via a suitable discordancy measure. In many situations, however, there are several variables of interest, for example, streamflow, temperature, precipitation, windspeed, etc., all measured at each site in the network. Instead of regional frequency analysis carried out separately for each variable, each generating a different partition into homogeneous regions, with the multivariate L-moments approach one can now develop an extended regional frequency analysis leading to a single partition into homogeneous regions based on all the variables considered jointly.

5 Further Studies

Besides the lines of development suggested in Sections 4.4 and 4.5, we briefly discuss here several further studies of special interest. All are beyond the scope of the present paper and will be pursued elsewhere.

5.1 Versions for Increased Robustness and Reduced Moment Assumptions

Here we consider two variants of L-moments and L-comoments which are more robust and impose lower moment assumptions.

5.1.1 Trimming

Trimmed L-Moments

As a modification of the L-moments to obtain more robustness, Elamir and Seheult [12] introduce *trimmed L-moments*, given by increasing the conceptual sample size for defining the kth L-moment from k to $k + t_1 + t_2$ and using the k order statistics remaining after trimming the t_1 smallest and t_2 largest observations in the conceptual sample. Thus λ_k given by (1) becomes replaced by

$$\lambda_k^{(t_1,t_2)} = k^{-1} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} E(X_{k+t_1-j:k+t_1+t_2}), \tag{51}$$

$k \geq 1$. Except for $(t_1, t_2) = (0, 0)$, which gives the usual L-moments, the TL-moments exist under weaker moment assumptions (satisfied by the Cauchy distribution, for example) and eliminate the influence of the most extreme observations. The sample TL-moments do not, however, improve upon the asymptotic finite sample breakdown point, 0, of the sample L-moments (or in fact of the sample central moments).

In particular, for $(t_1, t_2) = (1, 1)$, the 1st TL-moment is $\lambda_1^{(1,1)} = E(X_{2:3})$, the expected value of the median from a sample of size 3. See Elamir and Seheult [12], [13] and Hosking [24] for detailed development.

Our definitions of L-comoments and L-comoment coefficients carry over easily to provide TL-comoments and TL-comoment coefficients as direct analogues of (51). We note that asymptotic results are not provided in [12]. However, for (t_1, t_2) fixed as $n \to \infty$, the asymptotic results we have stated for sample L-moments and L-comoments have similar formulations and derivations for these trimmed versions.

L-Moments on Trimmed Samples

Elamir and Seheult [12] also mention without development the alternative approach of defining trimmed L-moments simply as ordinary L-moments defined on a trimmed sample. This yields different versions of trimmed estimators – for example, in this case the usual trimmed mean, which weights each observation equally after trimming the sample. For $(t_1, t_2) = (\beta n, \beta n)$ with $\beta > 0$, the breakdown point improves from 0 to β. Asymptotic normality of these sample versions follows, using the U-statistic representations noted in the present paper, from results of Janssen, Serfling and Veraverbeke [27] for U-statistics defined on trimmed samples.

5.1.2 Quantiles instead of Expectations

A modification of L-moments that eliminates moment restrictions entirely, given by Mudholkar and Hutson [39], consists of replacing each expectation in (1) by a suitable linear combination of quantiles:

$$\lambda_k^{(Q)} = k^{-1} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} \theta_{p,\alpha}(X_{k-j:k}), \tag{52}$$

where $0 \leq \alpha \leq 1/2$, $0 \leq p \leq 1/2$, and

$$\theta_{p,\alpha}(X_{k-j:k}) = pF^{-1}_{X_{k-j:k}}(\alpha) + (1-2p)F^{-1}_{X_{k-j:k}}(1/2) = pF^{-1}_{X_{k-j:k}}(1-\alpha).$$

As with the above trimming approach, here too one may extend our development to define LQ-comoments and related quantities. Starting with the representation (39) of L-comoments in terms of concomitants, as given in Proposition 6, we replace expectations by quantiles to define the *kth LQ-comoment of $X^{(1)}$ with respect to $X^{(2)}$* by

$$\lambda_{k\,[12]}^{(Q)} = k^{-1} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} \theta_{p,\alpha}(X^{(12)}_{[k-j:k]}). \tag{53}$$

5.2 Variances and Covariances of Sample Versions

Exact formulae for the variances and covariances of sample L-moments and TL-moments are developed by Elamir and Seheult [12], [13] following Downton [11].

These have the form of a weighted sum of expected values of order statistics from a conceptual sample. Here we note that these can also be derived the U-statistic representations we have noted above, via standard expressions for the variances and covariances of U-statistics. In the same fashion, exact expressions for the variances and covariances of sample TL-comoments may be obtained.

Elamir and Seheult [12], [13] also provide distribution-free unbiased estimators of these variances and covariances. We note here that, using the relevant weighted sum of expected values from a conceptual sample to define a corresponding kernel for a U-statistic, distribution-free unbiased estimators immediately are given by the corresponding U-statistics.

5.3 Characterization Issues

In the univariate case, as noted in Sections 1.2 and 2.1.2, under the finite mean assumption the series of L-moments determines the distribution. It is of interest to explore this question for the multivariate case. For example, to what extent do the L-moments and L-comoments together determine the bivariate distributions of jointly distributed variates? Such characterization results would delineate the theoretical limitations of the L-comoments as an extension of the univariate case and perhaps suggest further extensions and developments necessary for a "complete" multivariate extension.

Acknowledgments

The authors very appreciatively thank William Asquith, Nick Cox, Jon Hosking, G. L. Thompson, and three anonymous referees for extremely helpful queries and suggestions that have led to many improvements in the exposition. Also, support by National Science Foundation Grants DMS-0103698 and CCF-0430366 is gratefully acknowledged.

References

[1] Adcock, C., Jurczenko, E., and Maillet, B., eds. (2005). *Multi-Moment Capital Asset Pricing Models and Related Topics*. Springer-Verlag, Berlin.

[2] Arnold, B. C. (1983). *Pareto Distributions*. International Cooperative Publishing House, Fairland, Maryland.

[3] Athayde, G. M. and Flôres, Jr., R. G. (2005). On certain geometric aspects of portfolio optimisation with higher moments. In *Multi-Moment Capital Asset*

Pricing Models and Related Topics, C. Adcock, E. Jurczenko, and B. Maillet (eds.), forthcoming. Springer-Verlag, Berlin.

[4] Blom, G. (1980). Extrapolation of linear estimates to larger sample sizes. *Journal of the American Statistical Association* **75** 912–917.

[5] Bowley, A. L. (1902). *Elements of Statistics*, 2nd Edition. P. S. King, London.

[6] Chan, L. K. (1967). On a characterization of distributions by expected values of extreme order statistics. *American Mathematical Monthly* **74** 950–951.

[7] Christie-David, R. and Chaudhry, M. (2001). Coskewness and cokurtosis in futures markets. *Journal of Empirical Finance* **8** 55–81.

[8] David, H. A. and Nagaraja, H. N. (2003). *Order Statistics*, 3rd Edition. Wiley, New York.

[9] de Prado, L., Mailoc, M., and Peijan, A. (2004). Measuring loss potential of hedge fund strategies. *Journal of Alternative Investments* **Summer 2004** 7-31.

[10] Dittmar, R. (2002). Nonlinear pricing kernels, kurtosis preference, and evidence from the cross-section of equity returns. *Journal of Finance* **57** 369–403.

[11] Downton, F. (1966). Linear estimates with polynomial coefficients. *Biometrika* **53** 129–141.

[12] Elamir, E. A. and Seheult, A. H. (2003). Trimmed L-moments. *Computational Statistics and Data Analysis* **43** 299–314.

[13] Elamir, E. A. and Seheult, A. H. (2004). Exact variance structure of sample L-moments. *Journal of Statistical Planning and Inference* **124** 337–359.

[14] Fang, H. and Lai, T.-Y. (1997). Co-kurtosis and capital asset pricing. *The Financial Review* **32** 293–307.

[15] Farlie, D. J. G. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika* **47** 307–323.

[16] Fréchet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon, Section A, Series 3* **14** 53–57.

[17] Gini, C. (1912). Variabilità e Mutabilità. Contributio allo studio delle distribuzioni e delle relazione statistiche. *Studi Economico-Giuridici della Reale Università di Cagliari* **3** 3–159.

[18] Hardy, G. H., Littlewood, J. E. and Pólya, G. (1952). *Inequalities*, 2nd edition. University Press, Cambridge.

[19] Harvey, C. R. and Siddique, A. (2000). Conditional skewness in asset pricing tests. *Journal of Finance* **60** 1263–1295.

[20] Hoeffding, W. (1940). Masstabinvariante Korrelations-theorie. *Schriften Math. Inst. Univ. Berlin* **5** 181–233.

[21] Hosking, J. R. M. (1989). Some Theoretical Results Concerning L-Moments. *Research Report RC14492*. IBM Research Division, Yorktown Heights.

[22] 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.

[23] Hosking, J. R. M. (1998). L-Estimation. In *Handbook of Statistics, Vol. 17*, N. Balakrishnan and C. R. Rao (eds.), pp. 215–235. Elsevier.

[24] Hosking, J. R. M. (2005). Some Theory and Practical Uses of Trimmed L-Moments. *Research Report RC23783*. IBM Research Division, Yorktown Heights.

[25] Hosking, J. R. M., Bonti, G., and Siegel, D. (2000). Beyond the lognormal. *Risk* **13** 59–62.

[26] Hosking, J. R. M. and Wallis, J. R. (1997). *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press.

[27] Janssen, P., Serfling, R. and Veraverbeke, N. (1987). Asymptotic normality of U-statistics based on trimmed samples. *Journal of Statistical Planning and Inference* **16** 63–74.

[28] Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Volume 1, 2nd edition. Wiley, New York.

[29] Jurczenko, E. and Maillet, B. (2001). The three-moment CAPM: theoretical foundations and an asset pricing models comparison in an unified framework. In *Developments in Forecast Combination and Portfolio Choice*, C. L. Dunis, A. Timmerman, and J. Moody (eds.), pp. 239–273. Wiley.

[30] Jurczenko, E. and Maillet, B. (2005). The four-moment capital asset pricing model: some basic results. In *Multi-Moment Capital Asset Pricing Models and Related Topics*, C. Adcock, E. Jurczenko, and B. Maillet (eds.), forthcoming (72 pages). Springer-Verlag, Berlin.

[31] Jurczenko, E., Maillet, B., and Merlin, P. (2005). Hedge funds portfolio selection with higher-order moments: a non-parametric mean-variance-skewness-*kurtosis* efficient frontier. working paper.

[32] Jurczenko, E., Maillet, B., and Negrea, B. (2002). Multimoment approximate option pricing models: a general comparison (Part 1).

[33] Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). *Continuous Multivariate Distributions, Volume 1: Models and Applications*, 2nd edition. Wiley, New York.

[34] Lehmann, E. L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics* **37** 1137–1153.

[35] Lerman, R. I. and Yitzhaki, S. (1984). A note on the calculation and interpretation of the Gini index. *Economics Letters* **15** 163–168.

[36] Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57** 519–530.

[37] Meerschaert, M. M. and Scheffler, H.-P. (2003). Nonparametric methods for heavy tailed vector data: A survey with applications from finance and hydrology. In *Recent advances and trends in nonparametric statistics*, M. G. Akritas and D. N. Politis (eds.), pp. 265-279. Elsevier.

[38] Morgenstern, D. (1956). Einfache Beispiele zweidimensionaler Verteilungen. *Miteilingsblatt für Mathematische Statistik* **8** 234–235.

[39] Mudholkar, G. S. and Hutson, A. D. (1998). LQ-moments: Analogues of L-moments. *Journal of Statistical Planning and Inference* **71** 191–208.

[40] Nayak, T. K. (1987). Multivariate Lomax distribution: properties and usefulness in reliability theory. *Journal of Applied Probability* **24** 170–177.

[41] Olkin, I. and Yitzhaki, S. (1992). Gini regression analysis. *International Statistical Review* **60** 185–196.

[42] Plackett, R. L. (1947). Limits of the ratio of mean range to standard deviation. *Biometrika* **34** 120–122.

[43] Robbins, H. (1944). On the expected values of two statistics. *Annals of Mathematical Statistics* **15** 321–323.

[44] Rubinstein, M. E. (1973). The fundamental theorem of parameter-preference security valuation. *Journal of Financial and Quantitative Analysis* **8** 61–69.

[45] Schucany, W. R., Parr, W. C., and Boyer, J. E. (1978). Correlation structure in Farlie-Gumbel-Morgenstern distributions. *Biometrika* **65** 650–653.

[46] Schechtman, E. and Yitzhaki, S. (1987). A measure of association based on Gini's mean difference. *Communications in Statistics – Theory and Methods* **16** 207–231.

[47] Schechtman, E. and Yitzhaki, S. (1999). On the proper bounds of the Gini correlation. *Economics Letters* **63** 133–138.

[48] Schweizer, B. and Wolff, E. F. (1981). On nonparametric measures of dependence for random variables. *Annals of Statistics* **9** 879–885.

[49] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, New York.

[50] Sillitto, G. P. (1951). Interrelations between certain linear systematic statistics of samples from any continuous population. *Biometrika* **38** 377–382.

[51] Sillitto, G. P. (1969). Derivation of approximants to the inverse distribution function of a continuous univariate population from the order statistics of a sample. *Biometrika* **56** 641–650.

[52] Stuart, A. (1954). The correlation between variate values and ranks in samples from a continuous distribution. *British J. Statist. Psych.* **7** 37–44.

[53] Yang, S. S. (1977). General distribution theory of the concomitants of order statistics. *Annals of Statistics* **5** 996–1002.

[54] Yang, S. S. (1981). Linear functions of concomitants of order statistics with application to nonparametric estimation of a regression function. *Journal of the American Statistical Association* **76** 658–662.

[55] Yitzhaki, S. and Olkin, I. (1991). Concentration indices and concentration curves. In *Stochastic Orders and Decision under Risk* (K. Mosler and M. Scarsini, eds.), pp. 380–392. IMS Lecture Notes – Monograph Series, Volume 19, Hayward, California.