



Some measures of kurtosis and their inference on large datasets

Claudio Giovanni Borroni¹ · Lucio De Capitani¹

Received: 18 June 2021 / Accepted: 18 March 2022 / Published online: 14 April 2022
© The Author(s) 2022, corrected publication 2022

Abstract

This paper deals with the estimation of kurtosis on large datasets. It aims at overcoming two frequent limitations in applications: first, Pearson’s standardized fourth moment is computed as a unique measure of kurtosis; second, the fact that data might be just samples is neglected, so that the opportunity of using suitable inferential tools, like standard errors and confidence intervals, is discarded. In the paper, some recent indexes of kurtosis are reviewed as alternatives to Pearson’s standardized fourth moment. The asymptotic distribution of their natural estimators is derived, and it is used as a tool to evaluate efficiency and to build confidence intervals. A simulation study is also conducted to provide practical indications about the choice of a suitable index. As a conclusion, researchers are warned against the use of classical Pearson’s index when the sample size is too low and/or the distribution is skewed and/or heavy-tailed. Specifically, the occurrence of heavy tails can deprive Pearson’s index of any meaning or produce unreliable confidence intervals. However, such limitations can be overcome by reverting to the reviewed alternative indexes, relying just on low-order moments.

Keywords Pearson’s β_2 · Asymptotic distributions · Coverage · Heavy-tailedness · Skewness

1 Introduction

Undoubtedly, Karl Pearson can be considered as the “father” of kurtosis: indeed, he was the first statistician who firmly stated that several phenomena could be represented by means of frequency distributions which significantly differ from the

✉ Lucio De Capitani
lucio.decapitani1@unimib.it

Claudio Giovanni Borroni
claudio.borroni@unimib.it

¹ Department of Statistics and Quantitative Methods, University of Milano-Bicocca, via Bicocca degli Arcimboldi, 8, 20126 Milano, Italy

Gaussian [see Pearson (1894)]. To represent such distributions, he introduced the well-known “Pearson’s system,” whose elements depart from normality according to the values taken by their moments. Specifically, the standardized third moment β_1 reflects departure from symmetry while the standardized fourth moment β_2 reflects the degree of what Pearson called “flat-toppedness” of the distribution. It was only in Pearson (1905), however, that the term “kurtosis” was coined and that β_2 , as compared with the value 3 taken for the normal distribution, was firstly identified as a measure of kurtosis (the interested reader can refer to Fiori and Zenga (2009) for an exhaustive review of the genesis of β_2 in Pearson’s thought).

Over time, the use of β_2 as a measure of kurtosis was extended, beyond Pearson’s system, to the entire universe of frequency distributions. However, in this passage, some problems arose. First, out of Pearson’s system, β_2 cannot be simply linked to the value taken by the density function at the mean [see Kaplansky (1945)]. Thus, to provide a general interpretation, Darlington [see Darlington (1970)] linked β_2 to the tendency to bimodality, while Finucan [see Finucan (1964)] underlined its concurrent sensitivity to the “tailedness” of the distribution. Second, and more dangerously, β_2 is undefined for distributions with infinite fourth moment.

Despite the lines of criticism above, β_2 is nowadays often considered as the main measure of kurtosis and sometimes it is even identified with such a phenomenon, which reflects something vaguely connected to tails of the distribution. However, alternative measures have been proposed in the literature. There are, actually, large differences among them, motivated by the fact that the definition of kurtosis rests on location, scale and skewness, which are vague concepts as well. For instance, Balanda and McGillivray [see Balanda and MacGillivray (1988)] write that “a shape characteristic that we call kurtosis can be vaguely defined as the location and scale-free movement of probability mass from the shoulders of a distribution into its center and tails,” but, to apply such a logic, one needs both to provide a sharp definition of location, scale and skewness and, with more difficulties, to establish where the shoulders and the tails of a distribution are positioned. The entanglement of kurtosis with skewness is, in effect, the most cumbersome problem in this system of cross-referenced summary measures of a distribution: while it is easy to find scale measures which are independent of location, for instance, kurtosis indexes may behave differently according to the underlying degree of skewness [see Balanda and MacGillivray (1990), Blest (2003), and Jones et al. (2011)].

Today, researchers can thus rely on many alternatives of β_2 to measure kurtosis and, as a result, choosing the one to be computed on a given dataset is hard, both because they are all reasonable and because their efficacy may strongly depend on other unexplored characteristics of the same data. Operationally, there are, in this sense, some open questions about kurtosis: does the performance of the mostly known indexes significantly differ when they are computed on data? Specifically, if data are samples, does the need for a extremely large size prevents to use some particular indexes? How such performances and such indications about the sample size are affected by other possible shape deviations of the underlying population, like skewness?

In this paper, we will try to answer some of the questions addressed above. We will focus on the recent approach to kurtosis due to Zenga [see Zenga (1996) and

Zenga (2006)], but we will use Pearson's β_2 and Geary's α_G [see Geary (1936)], which is also a widespread classical index, as benchmarks. Actually, Zenga (1996) introduced a kurtosis curve and two kurtosis indexes, but the following literature explored their properties just from a descriptive point of view [see Zenga (2006) and the references therein]. After briefly reviewing that literature, we will thus provide the sample asymptotic distribution of the two indexes and introduce two asymptotic confidence intervals for the related population parameters. In our opinion, that can be considered as a first achievement of the paper. In addition, we will report some results of a simulation study aimed at comparing the sample performances of the considered indexes, along with their sensitivity to the sample size and the shape characteristics of the population.

The paper is organized as follows. In Sect. 2, Zenga's approach to kurtosis is reviewed, along with a variant due to De Capitani and Poliscchio [see De Capitani and Poliscchio (2016)], and the kurtosis indexes $K_2(\mu)$, $K_2(\gamma)$, $K_1(\mu)$, $K_1(\gamma)$ are recalled. In Sect. 3, the asymptotic distributions of the estimators of such indexes are derived, along with the ones for Pearson's β_2 and Geary's α_G . In Sect. 4, a simulation study is performed in order to evaluate the accuracy of the approximation provided by the asymptotic theory and the practical applications of the considered indexes on large datasets. Sect. 5 concludes.

2 Recalling Zenga's approach to kurtosis

Zenga (1996) based the study of kurtosis on suitable transformations of a random variable X which produce new distributions possibly characterized by different levels of kurtosis, while maintaining the same median and the same mean absolute deviation around the median. Equivalently, after denoting by γ the median of X , a change in the concentration of the random variables

$$D = (X - \gamma \mid X > \gamma) \quad \text{and} \quad S = (\gamma - X \mid X \leq \gamma) \quad (1)$$

is produced by the transformations above, which, in these sense, resembles those considered in the Pigou-Dalton "principle of transfers". Thus, the appeal of Zenga's approach rests in the opportunity to use all the well-established tools for concentration of positive variables in the context of kurtosis. A good review of this approach is given in Zenga (2006), where the international reader can find details about the original references on the topic. Further useful references are Fiori (2008), Domma (2004) and Fiori (2007), where a careful comparison between the logic of the two indexes K_2 , K_1 (details below) and of Pearson's β_2 is given. More recently, De Capitani and Poliscchio [see De Capitani and Poliscchio (2016)] reviewed Zenga's approach by replacing the median γ with an arbitrary "cutting point" k in the definition of the two variables in (1). Such a substitution, especially if k is chosen as the mean μ , turns out to be a powerful alternative when kurtosis is studied in asymmetric models. In the following, we will then start with a generic k and set $k = \gamma$ or $k = \mu$ when needed.

Let X be a random variable with continuous distribution function F . Assume that X has finite expectation $\mu = E[X]$ and median $\gamma = F^{-1}(0.5)$. Moreover, suppose to measure the dispersion of X by its mean absolute deviation around k :

$$\delta_k = \int |t - k| dF(t) \tag{2}$$

(note that, when unspecified, integrals range from $-\infty$ to $+\infty$). Now, re-define the variables in (1) as

$$D_k = (X - k | X > k) \quad \text{and} \quad S_k = (k - X | X \leq k) \tag{3}$$

and let F_{D_k} and F_{S_k} denote their distribution functions, which can be easily determined from F . Finally, denote by δ_{D_k} and δ_{S_k} the expectations of D_k and S_k and notice that $\delta_k = q_k \delta_{S_k} + (1 - q_k) \delta_{D_k}$, where $q_k = F(k)$. Obviously, D_k and S_k are, by construction, non-negative random variables. Their Lorenz curves are thus:

$$L_{D_k}(p) = \frac{1}{\delta_{D_k}} \int_0^p F_{D_k}^{-1}(t) dt \quad \text{and} \quad L_{S_k}(p) = \frac{1}{\delta_{S_k}} \int_0^p F_{S_k}^{-1}(t) dt .$$

Zenga’s kurtosis diagram of the random variable X is defined as the graph of the following function $G_k : [-1, 1] \rightarrow [0, 1]$,

$$G_k(p) = \begin{cases} L_{S_k}(-p) & -1 \leq p \leq 0 \\ L_{D_k}(p) & 0 \leq p \leq 1 . \end{cases}$$

Consistently, a synthetic index of kurtosis (denoted here as $K_2(k)$ to emphasize its dependency on k) is defined as the weighted average of the Gini concentration indexes of D_k and S_k :

$$\begin{aligned} K_2(k) &= q_k \int_0^1 [p - L_{S_k}(p)] dp + (1 - q_k) \int_0^1 [p - L_{D_k}(p)] dp \\ &= K_2^-(k) + K_2^+(k) . \end{aligned}$$

Notice that the two components of $K_2(k)$ can be rewritten in terms of Gini’s mean difference: after setting

$$\Delta_{S_k} = \int \int |x - y| dF_{S_k}(x) dF_{S_k}(y) \quad \text{and} \quad \Delta_{D_k} = \int \int |x - y| dF_{D_k}(x) dF_{D_k}(y),$$

one gets

$$K_2^-(k) = q_k \frac{\Delta_{S_k}}{2 \delta_{S_k}} \quad \text{and} \quad K_2^+(k) = (1 - q_k) \frac{\Delta_{D_k}}{2 \delta_{D_k}} .$$

The two ratios above evaluate the kurtosis of X on the left and on the right of k , by looking at the relative variability of S_k and D_k , respectively. By a similar logic, Zenga (1996) considered an alternative index based on the ratios between variances and second moments,

$$\frac{\text{Var}(S_k)}{\text{E}(S_k^2)} = 1 - \frac{\text{E}^2(S_k)}{\text{E}(S_k^2)} \quad \text{and} \quad \frac{\text{Var}(D_k)}{\text{E}(D_k^2)} = 1 - \frac{\text{E}^2(D_k)}{\text{E}(D_k^2)},$$

which give

$$\begin{aligned} K_1(k) &= q_k \left[1 - \frac{\text{E}^2(S_k)}{\text{E}(S_k^2)} \right] + (1 - q_k) \left[1 - \frac{\text{E}^2(D_k)}{\text{E}(D_k^2)} \right] \\ &= K_1^-(k) + K_1^+(k). \end{aligned}$$

When the cutting point k is chosen as μ or γ , some simplifications apply. For instance, it is easy to notice that $\delta_{S_\mu} = (2q_\mu)^{-1} \delta_\mu$ and $\delta_{D_\mu} = ((2(1 - q_\mu))^{-1} \delta_\mu$, so that

$$K_2(\mu) = \frac{\Delta}{\delta_\mu} - 1 \tag{4}$$

[see De Capitani and Poliscchio (2016)], where δ_μ and $\Delta = \int \int |x - y| \, dF \, dF$ are not actually referred to any conditional distributions. As emphasized in the following, $K_2(\mu)$ is, in effect, the only considered index which does not need to be separated into components to be evaluated. Other simplifications come after noticing that $q_\gamma = 1/2$ so that, for instance, $\delta_{D_\gamma} = 2\delta_\gamma - \delta_{S_\gamma}$.

Despite some possible simplifications, the occurrence of conditional distributions is the main problem in the determination of the asymptotic distributions of the sample estimators of the above-defined indexes. To ease such a process, in the following we will express our indexes in terms of some suitable non-conditional random variables. Consider the indicator function $\mathbb{1}(A)$ on the set A and define

$$h^-(x, k) = (k - x) \mathbb{1}(x \leq k) \quad \text{and} \quad h^+(x, k) = (x - k) \mathbb{1}(x > k).$$

Similarly, consider the functions

$$\begin{aligned} H^-(x, y, k) &= |(k - x) \mathbb{1}(x \leq k) - (k - y) \mathbb{1}(y \leq k)| = |h^-(x, k) - h^-(y, k)| \\ H^+(x, y, k) &= |(x - k) \mathbb{1}(x > k) - (y - k) \mathbb{1}(y > k)| = |h^+(x, k) - h^+(y, k)| \end{aligned}$$

and notice that

$$H^-(x, y, k) + H^+(x, y, k) = |x - y|.$$

After choosing $k = \gamma$ or $k = \mu$, the considered kurtosis indexes can be easily re-defined as functionals of the following non-conditional random variables:

$$X_k^- = h^-(X, k) \quad \text{and} \quad X_k^+ = h^+(X, k). \tag{5}$$

More specifically, begin by considering the expectation as a first functional and set

$$\delta^- = \text{E}\left(X_\gamma^-\right) \quad \text{and} \quad \delta^+ = \text{E}\left(X_\gamma^+\right).$$

It turns out that

$$\begin{aligned} \delta^- &= \int h^-(x, \gamma) \, dF(x) = \int_{-\infty}^{\gamma} (\gamma - x) \, dF(x) = \frac{1}{2} \delta_{S_{\gamma}} \\ \delta^+ &= \int h^+(x, \gamma) \, dF(x) = \int_{\gamma}^{\infty} (x - \gamma) \, dF(x) = \frac{1}{2} \delta_{D_{\gamma}}. \end{aligned}$$

Now consider Gini's mean difference as a second functional, that is, for every random variable X ,

$$\mathbf{A}(X) = E[|X - Y|] \quad \text{where } (X, Y) \text{ are i.i.d. } \sim X$$

and define

$$\Delta^- = \mathbf{A}(X_{\gamma}^-) \quad \text{and} \quad \Delta^+ = \mathbf{A}(X_{\gamma}^+).$$

One can easily notice that

$$\begin{aligned} \Delta^- &= \int \int |h^-(x, \gamma) - h^-(y, \gamma)| \, dF(y) \, dF(x) = \\ &= \int \int H^-(x, y, \gamma) \, dF(y) \, dF(x) = \\ &= \int_{-\infty}^{\gamma} \int_{-\infty}^{\gamma} |x - y| \, dF(y) \, dF(x) + \int_{-\infty}^{\gamma} \int_{\gamma}^{\infty} (\gamma - x) \, dF(y) \, dF(x) + \\ &\quad + \int_{\gamma}^{\infty} \int_{-\infty}^{\gamma} (\gamma - y) \, dF(y) \, dF(x) = \frac{1}{4} \Delta_{S_{\gamma}} + \delta^- \end{aligned}$$

so that $\Delta_{S_{\gamma}} = 4(\Delta^- - \delta^-)$. Similarly, $\Delta_{D_{\gamma}} = 4(\Delta^+ - \delta^+)$, which finally gives:

$$K_2^-(\gamma) = \frac{\Delta^-}{2\delta^-} - \frac{1}{2}, \quad K_2^+(\gamma) = \frac{\Delta^+}{2\delta^+} - \frac{1}{2} \quad \text{and} \quad K_2(\gamma) = \frac{\Delta^-}{2\delta^-} + \frac{\Delta^+}{2\delta^+} - 1. \tag{6}$$

When the cutting point is $k = \mu$, the two functionals above let one define

$$d^- = E(X_{\mu}^-), \quad d^+ = E(X_{\mu}^+), \quad D^- = \mathbf{A}(X_{\mu}^-) \quad \text{and} \quad D^+ = \mathbf{A}(X_{\mu}^+),$$

so that the needed summary measures of the conditional random variables in (3) can be computed as

$$\delta_{S_{\mu}} = \frac{d^-}{q_{\mu}} \quad \text{and} \quad \delta_{D_{\mu}} = \frac{d^+}{1 - q_{\mu}}$$

and as

$$\Delta_{S_{\mu}} = \frac{D^- - 2(1 - q_{\mu})d^-}{q_{\mu}^2} \quad \text{and} \quad \Delta_{D_{\mu}} = \frac{D^+ - 2q_{\mu}d^+}{(1 - q_{\mu})^2}.$$

Consequently, the components of the index $K_2(\mu)$ are written as

$$K_2^-(\mu) = \frac{D^-}{2d^-} - (1 - q_\mu) \quad \text{and} \quad K_2^+(\mu) = \frac{D^+}{2d^+} - q_\mu. \tag{7}$$

Now notice that $d^- = d^+ = \delta_\mu/2$ and $D^- + D^+ = \Delta$, so that (4) is actually re-obtained: once again, the definition of $K_2(\mu)$ does not need any conditional distributions. However, the decomposition provided by (7) is likely to be useful in those applications where the left and right kurtosis should be studied separately. For instance, in a statistical test aiming at comparing the two sides of kurtosis, one might need to know the joint sample distribution of the estimators of the two components.

Let us now turn to the indexes $K_1(\gamma)$ and $K_1(\mu)$. To re-define them in terms of non-conditional random variables, we need to consider the l -th moments as well ($l \geq 2$). Define first

$$\delta_l^- = E\left[\left(X_\gamma^-\right)^l\right] \quad \text{and} \quad \delta_l^+ = E\left[\left(X_\gamma^+\right)^l\right]$$

and consider, for now, δ_2^- and δ_2^+ : by definition,

$$\delta_2^- = \int_{-\infty}^{\gamma} (\gamma - x)^2 dF(x) = \frac{E(S_\gamma^2)}{2},$$

so that $E(S_\gamma^2) = 2\delta_2^-$ and, similarly, $E(D_\gamma^2) = 2\delta_2^+$. It is easily seen that

$$K_1^-(\gamma) = \frac{1}{2} - \frac{(\delta^-)^2}{\delta_2^-}, \quad K_1^+(\gamma) = \frac{1}{2} - \frac{(\delta^+)^2}{\delta_2^+}$$

and that

$$K_1(\gamma) = 1 - \left[\frac{(\delta^-)^2}{\delta_2^-} + \frac{(\delta^+)^2}{\delta_2^+} \right]. \tag{8}$$

When the cutting point $k = \mu$ is chosen and the l -th moments are considered,

$$d_l^- = E\left[\left(X_\mu^-\right)^l\right] \quad \text{and} \quad d_l^+ = E\left[\left(X_\mu^+\right)^l\right],$$

one can easily obtain that

$$K_1^-(\mu) = q_\mu - \frac{(d^-)^2}{d_2^-}, \quad K_1^+(\mu) = (1 - q_\mu) - \frac{(d^+)^2}{d_2^+}$$

and

$$K_1(\mu) = 1 - \left[\frac{(d^-)^2}{d_2^-} + \frac{(d^+)^2}{d_2^+} \right]. \tag{9}$$

On passing, notice that when the distribution of X is symmetric, $\gamma = \mu$ and $d_2^- = d_2^+ = \sigma^2/2$, where σ^2 denotes the variance of X . In this chance, $K_1(\mu) = K_1(\gamma) = 1 - (\delta_\mu/\sigma)^2$, which is closely related to the definition of Geary’s kurtosis measure $\alpha_G = \delta_\mu/\sigma$.

3 Asymptotics

3.1 Asymptotic theory for $K_2(\mu)$ and $K_2(\mathcal{Y})$

As above outlined, the first aim of this paper is to derive the asymptotic distribution of the natural estimators of the considered kurtosis indexes. Let us start from the “simplest” of them, $K_2(\mu)$. After easing notation by $q = q_\mu$ and $d = \delta_\mu$, recall that

$$K_2(\mu) = \left[\frac{D^-}{2d^-} - (1 - q) \right] + \left[\frac{D^+}{2d^+} - q \right] = \frac{\Delta}{d} - 1.$$

Now let (X_1, \dots, X_n) be a iid random sample from F and consider

$$\widehat{K}_{2n}(\widehat{\mu}_n) = \frac{\widehat{\Delta}_n}{\widehat{d}_n} - 1, \tag{10}$$

where $\widehat{\Delta}_n = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |X_i - X_j|$, $\widehat{d}_n = \frac{1}{n} \sum_{i=1}^n |X_i - \widehat{\mu}_n|$ and $\widehat{\mu}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Notice that, in the definition of the estimator $\widehat{K}_{2n}(\widehat{\mu}_n)$, the cutting point μ is estimated from data as well. When the integral representation of the latter estimators is recalled, it is not difficult to approximate them by suitable linear combinations of the sample. After denoting by \widehat{F}_n the empirical distribution function, for instance,

$$\begin{aligned} \widehat{\Delta}_n - \Delta &= \int \int |x - y| \, d\widehat{F}_n(y) \, d\widehat{F}_n(x) - \Delta \\ &= \int 2 \left(\int |x - y| \, dF(y) - \Delta \right) \, d\widehat{F}_n(x) + o_p(n^{-1/2}) \\ &= \frac{1}{n} \sum_{i=1}^n \left[2 \int |X_i - y| \, dF(y) - 2\Delta \right] + o_p(n^{-1/2}) \\ &= \frac{1}{n} \sum_{i=1}^n g_1(X_i) + o_p(n^{-1/2}) \end{aligned} \tag{11}$$

which is a well-known result by Hoeffding (1948). We will use a similar logic for the remaining component \widehat{d}_n in (10); the following lemma, whose proof is found in De Capitani and Pasquazzi (2015), is of use:

Lemma 1 *Let F be a continuous distribution function and let (X_1, X_2, \dots, X_n) be a iid random sample from F . Denote by \widehat{F}_n the empirical distribution function, and let $\widehat{z}_n = T(\widehat{F}_n)$ be an estimator of $z = T(F)$ such that $\widehat{z}_n = z + O_p(n^{-1/2})$ for $n \rightarrow \infty$. Now consider the estimators*

$$\hat{I}_{1n} = \int (\hat{z}_n - x) \mathbb{1}(x \leq \hat{z}_n) d\hat{F}_n(x) \quad \text{and} \quad \hat{I}_{2n} = \int (x - \hat{z}_n) \mathbb{1}(x > \hat{z}_n) d\hat{F}_n(x)$$

of the functionals

$$I_1 = \int (z - x) \mathbb{1}(x \leq z) dF(x) \quad \text{and} \quad I_2 = \int (x - z) \mathbb{1}(x > z) dF(x). \tag{12}$$

Then, the following asymptotic representations hold:

$$\begin{aligned} \hat{I}_{1n} &= \int (z - x) \mathbb{1}(x \leq z) d\hat{F}_n(x) + F(z)(\hat{z}_n - z) + o_P(n^{-1/2}) \\ \hat{I}_{2n} &= \int (x - z) \mathbb{1}(x > z) d\hat{F}_n(x) - (1 - F(z))(\hat{z}_n - z) + o_P(n^{-1/2}). \end{aligned}$$

Lemma 1 easily gives:

$$\begin{aligned} \hat{d}_n - d &= 2 \int_{-\infty}^{\hat{\mu}_n} (\hat{\mu}_n - x) d\hat{F}_n(x) - d \\ &= 2 \left(\int (\mu - x) \mathbb{1}(x \leq \mu) d\hat{F}_n(x) + q(\hat{\mu}_n - \mu) \right) - d + o_P(n^{-1/2}) \\ &= \frac{1}{n} \sum_{i=1}^n \{ 2[(\mu - X_i) \mathbb{1}(X_i \leq \mu) + q(X_i - \mu)] - d \} + o_P(n^{-1/2}) \\ &= \frac{1}{n} \sum_{i=1}^n g_2(X_i) + o_P(n^{-1/2}). \end{aligned} \tag{13}$$

Of course, the linearization of $\hat{\Delta}_n$ and \hat{d}_n is a useful step to derive both their marginal asymptotic distributions (a known result in the literature) and their joint distributions, which is the key for the needed conclusion in Theorem 1. Before stating it, let us define two further functionals for a given random variable $X \sim F$,

$$\begin{aligned} \mathcal{F}(X) &= \int \int \int |x - y| |x - z| dF(z) dF(y) dF(x) = \int \left(\int |x - y| dF(y) \right)^2 dF(x) \\ \mathcal{D}(X) &= \int \int x |x - y| dF(y) dF(x) \end{aligned}$$

and set, to ease notation,

$$\mathcal{F} = \mathcal{F}(X), \quad \mathcal{D} = \mathcal{D}(X) \quad \text{and} \quad \mathcal{D}_\mu^- = \mathcal{D}(X_\mu^-).$$

Theorem 1 *Let (X_1, X_2, \dots, X_n) be a iid random sample from a continuous distribution function F such that $E(X_1) = \mu < \infty$ and $E[(X_1 - E(X_1))^2] = \sigma^2 < \infty$. Then, the vector $\sqrt{n} [\hat{\Delta}_n - \Delta, \hat{d}_n - d]^T$ is asymptotically normally distributed with zero mean and variance matrix Σ , where*

$$\begin{aligned} \Sigma_{11} &= 4(\mathcal{F} - \Delta^2) \\ \Sigma_{22} &= 4q\sigma^2 - 4(1 - 2q)d_2^- - d^2 \\ \Sigma_{12} &= 4q(\mathcal{D} - \mu\Delta) - 4\left(\mathcal{D}_\mu^- + \frac{d^2}{4}\right) - 2d\Delta. \end{aligned}$$

As a consequence, $\sqrt{n}(\widehat{K}_{2n}(\widehat{\mu}_n) - K_2(\mu))$ is asymptotically normal with zero mean and variance

$$\sigma_{K_2(\mu)}^2 = \frac{1}{d^2}\Sigma_{11} + \frac{\Delta^2}{d^4}\Sigma_{22} - \frac{2\Delta}{d^3}\Sigma_{12}.$$

Proof Thanks to the linearizations above, the result is a simple application of the central limit theorem, the Cramer–Wold device and the delta method. The expressions for Σ_{11} and Σ_{22} can be found in Hoeffding (1948) and Gastwirth (1974), respectively. Concerning the asymptotic covariance Σ_{12} , note first that

$$\begin{aligned} \mathcal{D}_\mu^- &= \int \int h^-(x, \mu) H^-(x, y, \mu) dF(y) dF(x) \\ &= \int_{-\infty}^\mu \int_{-\infty}^\mu (\mu - x) |x - y| dF(y) dF(x) + \int_{-\infty}^\mu \int_\mu^\infty (\mu - x)^2 dF(y) dF(x) \\ &= \int_{-\infty}^\mu \int_{-\infty}^\mu (\mu - x) |x - y| dF(y) dF(x) + (1 - q)d_2^-. \end{aligned}$$

Hence,

$$\begin{aligned} &\text{Cov}[g_1(X), g_2(X)] \\ &= 4 \text{E} \left\{ \left(\int |X - y| dF(y) \right) [(X - \mu)q - (X - \mu) \mathbb{1}(X \leq \mu)] \right\} - 2d\Delta \\ &= 4q \text{E} \left[\left(\int |X - y| dF(y) \right) (X - \mu) \right] + \\ &\quad - 4 \text{E} \left[\left(\int |X - y| dF(y) \right) (X - \mu) \mathbb{1}(X \leq \mu) \right] - 2d\Delta \\ &= 4q \int \int (x - \mu) |x - y| dF(y) dF(x) + \\ &\quad - 4 \int_{-\infty}^\mu \int (x - \mu) |x - y| dF(y) dF(x) - 2\Delta d \\ &= 4q(\mathcal{D} - \mu\Delta) - 4 \int_{-\infty}^\mu \int_{-\infty}^\mu (x - \mu) |x - y| dF(y) dF(x) + \\ &\quad - 4 \int_{-\infty}^\mu \int_\mu^\infty (x - \mu) |x - y| dF(y) dF(x) - 2\Delta d \\ &= 4q(\mathcal{D} - \mu\Delta) - 4 \left(\mathcal{D}_\mu^- + \frac{d^2}{4} \right) - 2d\Delta. \end{aligned}$$

□

On passing, notice that asymptotic linearizations similar to (11) and (13) can be provided for the m -th sample centered moment $\hat{\mu}_{mn} = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu}_n)^m$, for every real m (see Serfling (1980), p. 72). Hence, the asymptotic distribution of the estimators of Geary's α_G and Pearson's β_2

$$\hat{\alpha}_{Gn} = \frac{\hat{d}_n}{\hat{\sigma}_n} \quad \text{and} \quad \hat{\beta}_{2n} = \frac{\hat{\mu}_{4n}}{\hat{\sigma}_n^4},$$

where $\hat{\sigma}_n^2 = \hat{\mu}_{2n}$, are easily derived as in the next theorem:

Theorem 2 *Let (X_1, X_2, \dots, X_n) be a iid random sample from a continuous distribution function F such that $E(X_1) = \mu < \infty$ and $E[(X_1 - E(X_1))^l] = \mu_l < \infty$ up to $l = 4$. Then, $\sqrt{n}(\hat{\alpha}_{Gn} - \alpha_G)$ is asymptotically normally distributed with zero mean and variance*

$$\sigma_{\alpha_G}^2 = \frac{1}{\sigma^2}A_{11} + \frac{d^2}{\sigma^4}A_{22} - \frac{2d}{\sigma^3}A_{12},$$

where

$$A_{11} = 4q\sigma^2 - 4(1 - 2q)d_2^- - d^2 \quad A_{22} = \frac{\mu_4 - \sigma^4}{4\sigma^4} \quad A_{12} = \frac{1}{\sigma} [\mu_3q - d_3^-] - \frac{\sigma d}{2}.$$

Moreover, under the assumption that $\mu_l < \infty$ up to $l = 8$, $\sqrt{n}(\hat{\beta}_{2n} - \beta_2)$ is asymptotically normal with zero mean and variance

$$\sigma_{\beta_2}^2 = \frac{1}{\sigma^8}B_{11} + \frac{\mu_4^2}{\sigma^{16}}B_{22} - \frac{2\mu_4}{\sigma^{12}}B_{12}$$

where

$$B_{11} = (\mu_8 - \mu_4^2) + 8\mu_3(2\mu_3\sigma^2 - \mu_5), \quad B_{22} = 4\sigma^4(\mu_4 - \sigma^4), \quad \text{and}$$

$$B_{12} = 2\sigma^2(\mu_6 - \mu_4\sigma^2 - 4\mu_3^2).$$

Proof The representation of \hat{d}_n in (13) can be used, along with

$$\hat{\sigma}_n - \sigma = \frac{1}{n} \sum_{i=1}^n \frac{(X - \mu)^2}{2\sigma} - \frac{\sigma}{2} + o_p(n^{-1/2}) = \frac{1}{n} \sum_{i=1}^n g_3(X_i) + o_p(n^{-1/2}),$$

to prove that the two statistics are asymptotically jointly normal. Now consider that, after some computations, $\text{Cov}[g_3(X), g_2(X)] = \frac{1}{\sigma} [\mu_3q - d_3^-] - \frac{\sigma d}{2}$, so that the delta method easily gives the result for $\hat{\alpha}_{Gn}$. Concerning $\hat{\beta}_{2n}$, one can use Theorem B in Serfling (1980), pag. 72 to show that the second and the fourth sample centered moments are jointly asymptotically normal, and again, the result follows from a simple application of the delta method. □

We now turn to the estimation of $K_2(\gamma)$. After recalling (6), a first step is to estimate δ^- and δ^+ by

$$\hat{\delta}_n^- = \int h^-(x, \hat{\gamma}_n) d\hat{F}_n(x) \quad \text{and} \quad \hat{\delta}_n^+ = \int h^+(x, \hat{\gamma}_n) d\hat{F}_n(x), \tag{14}$$

respectively. In the formula above, $\hat{\gamma}_n = \inf \left\{ x : \hat{F}_n(x) \geq 1/2 \right\}$ denotes the sample median, which, assuming that F has density f strictly positive and continuous in γ , is known to admit the following asymptotic representation (see Serfling (1980), p. 77):

$$\hat{\gamma}_n - \gamma = \frac{1}{f(\gamma)} \int \left(\frac{1}{2} - \mathbb{1}(x \leq \gamma) \right) d\hat{F}_n(x) + o_p(n^{-1/2}). \tag{15}$$

Moreover, Lemma 1 can be applied again to get

$$\begin{aligned} \hat{\delta}_n^- - \delta^- &= \int (\gamma - x) \mathbb{1}(x \leq \gamma) d\hat{F}_n(x) + \frac{1}{2}(\hat{\gamma}_n - \gamma) - \delta^- + o_p(n^{-1/2}) \\ &= \int \left[(\gamma - x) \mathbb{1}(x \leq \gamma) - \delta^- + \frac{1}{2f(\gamma)} \left(\frac{1}{2} - \mathbb{1}(x \leq \gamma) \right) \right] d\hat{F}_n(x) + o_p(n^{-1/2}) \\ &= \frac{1}{n} \sum_{i=1}^n \left[(\gamma - X_i) \mathbb{1}(X_i \leq \gamma) - \frac{\mathbb{1}(X_i \leq \gamma)}{2f(\gamma)} - \left(\delta^- - \frac{1}{4f(\gamma)} \right) \right] + o_p(n^{-1/2}) \\ &= \frac{1}{n} \sum_{i=1}^n l_3(X_i) + o_p(n^{-1/2}). \end{aligned} \tag{16}$$

Similarly, Lemma 1 gives $\hat{\delta}_n^+ - \delta^+ = \frac{1}{n} \sum_{i=1}^n l_4(X_i) + o_p(n^{-1/2})$, where

$$l_4(X_i) = (X_i - \gamma) \mathbb{1}(X_i > \gamma) - \frac{\mathbb{1}(X_i > \gamma)}{2f(\gamma)} - \left(\delta^+ - \frac{1}{4f(\gamma)} \right). \tag{17}$$

However, to get a linearization for the estimators of Δ^- and Δ^+ ,

$$\hat{\Delta}_n^- = \int \int H^-(x, y, \hat{\gamma}_n) d\hat{F}_n(y) d\hat{F}_n(x) \quad \text{and} \quad \hat{\Delta}_n^+ = \int \int H^+(x, y, \hat{\gamma}_n) d\hat{F}_n(y) d\hat{F}_n(x),$$

a different statement is needed:

Lemma 2 *Under the assumptions of Lemma 1, consider the estimators*

$$\begin{aligned} \hat{I}_{3n} &= \int \int |(\hat{z}_n - x) \mathbb{1}(x \leq \hat{z}_n) - (\hat{z}_n - y) \mathbb{1}(y \leq \hat{z}_n)| d\hat{F}_n(y) d\hat{F}_n(x) \\ \hat{I}_{4n} &= \int \int |(x - \hat{z}_n) \mathbb{1}(x > \hat{z}_n) - (y - \hat{z}_n) \mathbb{1}(y > \hat{z}_n)| d\hat{F}_n(y) d\hat{F}_n(x), \end{aligned}$$

respectively, of the functionals

$$I_3 = \int \int |(z - x) \mathbb{1}(x \leq z) - (z - y) \mathbb{1}(y \leq z)| dF(y) dF(x)$$

$$I_4 = \int \int |(x - z) \mathbb{1}(x > z) - (y - z) \mathbb{1}(y > z)| dF(y) dF(x).$$

Then, the following asymptotic representations hold:

$$\hat{I}_{3n} = \int \int |(z - x) \mathbb{1}(x \leq z) - (z - y) \mathbb{1}(y \leq z)| d\hat{F}_n(y) d\hat{F}_n(x) + 2F(z)(1 - F(z))(\hat{z}_n - z) + o_p(n^{-1/2})$$

$$\hat{I}_{4n} = \int \int |(x - z) \mathbb{1}(x > z) - (y - z) \mathbb{1}(y > z)| d\hat{F}_n(y) d\hat{F}_n(x) - 2F(z)(1 - F(z))(\hat{z}_n - z) + o_p(n^{-1/2}).$$

For the sake of readability, the proof of Lemma 2 is reported in Appendix. Its application easily gives:

$$\hat{\Delta}_n^- = \int \int H^-(x, y, \gamma) d\hat{F}_n(y) d\hat{F}_n(x) + \frac{1}{2}(\hat{\gamma}_n - \gamma) + o_p(n^{-1/2})$$

where the first summand can be regarded as the mean difference of the pseudo-sample $\{(\gamma - X_i) \mathbb{1}(X_i \leq \gamma)\}_{i=1}^n$. By the usual asymptotic representation, one thus obtains:

$$\begin{aligned} \hat{\Delta}_n^- - \Delta^- &= \int \left(2 \int H^-(x, y, \gamma) dF(y) - 2\Delta^- \right) d\hat{F}_n(x) \\ &\quad + \frac{1}{2f(\gamma)} \int \left(\frac{1}{2} - \mathbb{1}(x \leq \gamma) \right) d\hat{F}_n(x) + o_p(n^{-1/2}) \\ &= \frac{1}{n} \sum_{i=1}^n \left[2 \int H^-(X_i, y, \gamma) dF(y) - \frac{\mathbb{1}(X_i \leq \gamma)}{2f(\gamma)} - \left(2\Delta^- - \frac{1}{4f(\gamma)} \right) \right] \\ &\quad + o_p(n^{-1/2}) \\ &= \frac{1}{n} \sum_{i=1}^n l_1(X_i) + o_p(n^{-1/2}) \end{aligned}$$

Analogously, $\hat{\Delta}_n^+ - \Delta^+ = \frac{1}{n} \sum_{i=1}^n l_2(X_i) + o_p(n^{-1/2})$ where

$$l_2(X_i) = 2 \int H^+(X_i, y, \gamma) dF(y) - \frac{\mathbb{1}(X_i > \gamma)}{2f(\gamma)} - \left(2\Delta^+ - \frac{1}{4f(\gamma)} \right).$$

We are now ready to provide the asymptotic distribution of the natural estimator of $K_2(\gamma)$,

$$\hat{K}_{2n}(\hat{\gamma}_n) = \frac{\hat{\Delta}_n^-}{2\hat{\delta}_n^-} + \frac{\hat{\Delta}_n^+}{2\hat{\delta}_n^+} - 1$$

in the next theorem, whose statement requires to set

$$\mathcal{F}^- = \mathcal{F}(X_\gamma^-), \quad \mathcal{F}^+ = \mathcal{F}(X_\gamma^+), \quad \mathcal{D}^- = \mathcal{D}(X_\gamma^-), \quad \mathcal{D}^+ = \mathcal{D}(X_\gamma^+) :$$

Theorem 3 *Let (X_1, X_2, \dots, X_n) be a iid random sample from an absolutely continuous distribution function F whose density f is strictly positive and continuous at γ . Assume that $E(X_1) < \infty$ and $E[(X_1 - E(X_1))^2] < \infty$. Then, $\sqrt{n}(\widehat{K}_{2n}(\widehat{\gamma}_n) - K_2(\gamma))$ is asymptotically normally distributed with zero mean and variance $\sigma_{K_2(\gamma)}^2 = \mathbf{s}^\top \Sigma \mathbf{s}$, where $\mathbf{s}^\top = \left[\frac{1}{2\delta^-}, \frac{1}{2\delta^+}, -\frac{\Delta^-}{2(\delta^-)^2}, -\frac{\Delta^+}{2(\delta^+)^2} \right]$ and $\Sigma = (\Sigma_{ij})_{i,j=1,\dots,4}$ is a symmetric matrix with elements:*

$$\begin{aligned} \Sigma_{11} &= 4(\mathcal{F}^- - (\Delta^-)^2) - \frac{\Delta^- - \delta^-}{f(\gamma)} + \frac{1}{16f^2(\gamma)} \\ \Sigma_{22} &= 4(\mathcal{F}^+ - (\Delta^+)^2) - \frac{\Delta^+ - \delta^+}{f(\gamma)} + \frac{1}{16f^2(\gamma)} \\ \Sigma_{33} &= (\delta_2^- - (\delta^-)^2) - \frac{\delta^-}{2f(\gamma)} + \frac{1}{16f^2(\gamma)} \\ \Sigma_{44} &= (\delta_2^+ - (\delta^+)^2) - \frac{\delta^+}{2f(\gamma)} + \frac{1}{16f^2(\gamma)} \\ \Sigma_{12} &= 4\delta^+ \left(\Delta^- - \frac{\delta^-}{2} \right) + 4\delta^- \left(\Delta^+ - \frac{\delta^+}{2} \right) - \frac{\delta^+ + \delta^-}{2f(\gamma)} \\ &\quad - \left(2\Delta^- - \frac{1}{4f(\gamma)} \right) \left(2\Delta^+ - \frac{1}{4f(\gamma)} \right) \\ \Sigma_{13} &= 2(\mathcal{D}^- - \Delta^- \delta^-) + \frac{1}{16f^2(\gamma)} - \frac{2\Delta^- - \delta^-}{4f(\gamma)} \\ \Sigma_{14} &= 2\delta^- \delta^+ - \frac{\delta^-}{2f(\gamma)} - \left(2\Delta^- - \frac{1}{4f(\gamma)} \right) \left(\delta^+ - \frac{1}{4f(\gamma)} \right) \\ \Sigma_{23} &= 2\delta^- \delta^+ - \frac{\delta^+}{2f(\gamma)} - \left(2\Delta^+ - \frac{1}{4f(\gamma)} \right) \left(\delta^- - \frac{1}{4f(\gamma)} \right) \\ \Sigma_{24} &= 2(\mathcal{D}^+ - \Delta^+ \delta^+) + \frac{1}{16f^2(\gamma)} - \frac{2\Delta^+ - \delta^+}{4f(\gamma)} \\ \Sigma_{34} &= - \left(\delta^- - \frac{1}{4f(\gamma)} \right) \left(\delta^+ - \frac{1}{4f(\gamma)} \right) \end{aligned}$$

From the proof of the theorem, which is left in Appendix, one can also easily derive the asymptotic distributions of the two components of $\widehat{K}_{2n}(\widehat{\gamma}_n)$ or of their difference. Even if details are omitted here, notice that those results could be useful in many testing problems aiming at comparing the right and the left side of kurtosis. As above anticipated, despite no decomposition is needed to get (10), when those testing problems arise, one can apply a similar reasoning to the single components of $\widehat{K}_{2n}(\widehat{\mu}_n)$ as well. The evaluation of Zenga’s indexes as tools to compare different sides of kurtosis will be the object of a future research.

3.2 Asymptotic theory for $K_1(\mu)$ and $K_1(\gamma)$

After recalling (9), a natural estimator for $K_1(\mu)$ is easily obtained from

$$\widehat{d}_n^- = \int h^-(x, \widehat{\mu}_n) d\widehat{F}_n(x)$$

and

$$\widehat{d}_{2n}^- = \int (h^-(x, \widehat{\mu}_n))^2 d\widehat{F}_n(x), \quad \widehat{d}_{2n}^+ = \int (h^+(x, \widehat{\mu}_n))^2 d\widehat{F}_n(x).$$

Again, a linear approximation for d_n^- is provided by Lemma 1. For d_{2n}^- and d_{2n}^+ a new statement is needed (proof in Appendix):

Lemma 3 *Under the assumptions of Lemma 1, consider the estimators*

$$\widehat{I}_{5n} = \int (\widehat{z}_n - x)^2 \mathbb{1}(x \leq \widehat{z}_n) d\widehat{F}_n(x) \quad \text{and} \quad \widehat{I}_{6n} = \int (x - \widehat{z}_n)^2 \mathbb{1}(x > \widehat{z}_n) d\widehat{F}_n(x)$$

of the functionals

$$I_5 = \int (z - x)^2 \mathbb{1}(x \leq z) dF(x) \quad \text{and} \quad I_6 = \int (x - z)^2 \mathbb{1}(x > z) dF(x).$$

Then, the following asymptotic representations hold:

$$\begin{aligned} \widehat{I}_{5n} &= \int (z - x)^2 \mathbb{1}(x \leq z) d\widehat{F}_n(x) + 2I_1(\widehat{z}_n - z) + o_p(n^{-1/2}) \\ \widehat{I}_{6n} &= \int (x - z)^2 \mathbb{1}(x > z) d\widehat{F}_n(x) - 2I_2(\widehat{z}_n - z) + o_p(n^{-1/2}) \end{aligned}$$

where I_1 and I_2 are as in (12).

Now,

$$\widehat{d}_{2n}^- - d_{2n}^- = \frac{1}{n} \sum_{i=1}^n [(\mu - X_i)^2 \mathbb{1}(X_i \leq \mu) + 2d^-(X_i - \mu) - d_{2n}^-] + o_p(n^{-1/2})$$

and a similar result is obtained for \widehat{d}_{2n}^+ . The next theorem (stated without proof) follows:

Theorem 4 *Let (X_1, X_2, \dots, X_n) be a random sample from a continuous distribution function F such that $E(X_1) < \infty$ and $E[(X_1 - E(X_1))^4] < \infty$. After considering*

$$\widehat{K}_{1n}(\widehat{\mu}_n) = 1 - \left[\frac{(\widehat{d}_n^-)^2}{\widehat{d}_{2n}^-} + \frac{(\widehat{d}_n^+)^2}{\widehat{d}_{2n}^+} \right],$$

the random variable $\sqrt{n}(\widehat{K}_{1n}(\widehat{\mu}_n) - K_1(\mu))$ is asymptotically normally distributed with zero mean and variance $\sigma_{K_1(\mu)}^2 = \mathbf{s}^\top \Sigma \mathbf{s}$, where $\mathbf{s}^\top = \left[-\frac{2d^-}{d_2^-} - \frac{2d^+}{d_2^+}, \left(\frac{d^-}{d_2^-}\right)^2, \left(\frac{d^+}{d_2^+}\right)^2 \right]$ and $\Sigma = (\Sigma_{ij})_{i,j=1,2,3}$ is a symmetric matrix with elements:

$$\begin{aligned} \Sigma_{11} &= q^2 \sigma^2 + (1 - 2q)d_2^- - (d^-)^2 \\ \Sigma_{22} &= (d_4^- - (d_2^-)^2) - 4d^- (d_3^- - \sigma^2 d^-) \\ \Sigma_{33} &= (d_4^+ - (d_2^+)^2) - 4d^+ (d_3^+ - \sigma^2 d^+) \\ \Sigma_{12} &= (1 - q)d_3^- - d^- (3d_2^- - 2q\sigma^2) \\ \Sigma_{13} &= qd_3^+ + 2(1 - q)d^- \sigma^2 - 3d^- d_2^+ \\ \Sigma_{23} &= 2d^- (d_3^- + d_3^+) - 4(d^-)^2 \sigma^2 - d_2^- d_2^+ . \end{aligned}$$

Finally, to estimate $K_1(\gamma)$ as defined in (8), the tools in (14) can be used jointly with

$$\widehat{\delta}_{2n}^- = \int (h^-(x, \widehat{\gamma}_n))^2 d\widehat{F}_n(x) \quad \text{and} \quad \widehat{\delta}_{2n}^+ = \int (h^+(x, \widehat{\gamma}_n))^2 d\widehat{F}_n(x).$$

While an asymptotic representation for $\widehat{\delta}_n^-$ and $\widehat{\delta}_n^+$ has been already provided in (16) and (17), respectively, for $\widehat{\delta}_{2n}^-$ and $\widehat{\delta}_{2n}^+$ another application of Lemma 3, jointly with (15) is needed: $\widehat{\delta}_{2n}^- - \delta_2^- = \frac{1}{n} \sum_{i=1}^n r(X_i) + o_p(n^{-1/2})$ with

$$r(X_i) = (\gamma - X_i)^2 \mathbb{1}(X_i \leq \gamma) - \frac{2\delta^-}{f(\gamma)} \mathbb{1}(X_i \leq \gamma) - \left(\delta_2^- - \frac{\delta^-}{f(\gamma)} \right)$$

(and analogously for $\widehat{\delta}_{2n}^+$). The next theorem, whose proof is omitted as well, follows easily:

Theorem 5 Let (X_1, X_2, \dots, X_n) be a random sample from an absolutely continuous distribution F whose density f is strictly positive and continuous at $\gamma = F^{-1}(1/2)$. Assume that $E(X_1) < \infty$ and $E[(X_1 - E(X_1))^4] < \infty$. Then, the random variable $\sqrt{n}(\widehat{K}_{1n}(\widehat{\gamma}_n) - K_1(\gamma))$, with

$$\widehat{K}_{1n}(\widehat{\gamma}_n) = 1 - \left(\frac{(\widehat{\delta}_n^-)^2}{\widehat{\delta}_{2n}^-} + \frac{(\widehat{\delta}_n^+)^2}{\widehat{\delta}_{2n}^+} \right),$$

is asymptotically normal with zero mean and variance $\sigma_{K_1(\gamma)}^2 = \mathbf{s}^\top \Sigma \mathbf{s}$, where $\mathbf{s}^\top = \left[-\frac{2\delta^-}{\delta_2^-}, -\frac{2\delta^+}{\delta_2^+}, \left(\frac{\delta^-}{\delta_2^-}\right)^2, \left(\frac{\delta^+}{\delta_2^+}\right)^2 \right]$ and $\Sigma = (\Sigma_{ij})_{i,j=1,\dots,4}$ is a symmetric matrix whose elements are:

$$\begin{aligned} \Sigma_{11} &= \delta_2^- - (\delta^-)^2 - \frac{\delta^-}{2f(\gamma)} + \frac{1}{16f^2(\gamma)} \\ \Sigma_{22} &= \delta_2^+ - (\delta^+)^2 - \frac{\delta^+}{2f(\gamma)} + \frac{1}{16f^2(\gamma)} \\ \Sigma_{33} &= \left(\delta_4^- - (\delta_2^-)^2 \right) + \left(\frac{\delta^-}{f(\gamma)} \right)^2 - \frac{2\delta^- \delta_2^-}{f(\gamma)} \\ \Sigma_{44} &= \left(\delta_4^+ - (\delta_2^+)^2 \right) + \left(\frac{\delta^+}{f(\gamma)} \right)^2 - \frac{2\delta^+ \delta_2^+}{f(\gamma)} \\ \Sigma_{12} &= - \left(\delta^- - \frac{1}{4f(\gamma)} \right) \left(\delta^+ - \frac{1}{4f(\gamma)} \right) \\ \Sigma_{13} &= \left(\delta_3^- - \delta_2^- \delta^- \right) - \frac{(\delta^-)^2}{f(\gamma)} + \frac{\delta^-}{4f^2(\gamma)} - \frac{\delta_2^-}{4f(\gamma)} \\ \Sigma_{14} &= - \left(\delta^- - \frac{1}{4f(\gamma)} \right) \left(\delta_2^+ - \frac{\delta^+}{f(\gamma)} \right) \\ \Sigma_{23} &= - \left(\delta_2^- - \frac{\delta^-}{f(\gamma)} \right) \left(\delta^+ - \frac{1}{4f(\gamma)} \right) \\ \Sigma_{24} &= \left(\delta_3^+ - \delta_2^+ \delta^+ \right) - \frac{(\delta^+)^2}{f(\gamma)} + \frac{\delta^+}{4f^2(\gamma)} - \frac{\delta_2^+}{4f(\gamma)} \\ \Sigma_{34} &= - \left(\delta_2^- - \frac{\delta^-}{f(\gamma)} \right) \left(\delta_2^+ - \frac{\delta^+}{f(\gamma)} \right). \end{aligned}$$

4 Simulation study

The asymptotic theory for the considered kurtosis indexes provides deeper tools than simple point estimation, thanks to the evaluation of standard errors and the computation of confidence intervals. Despite that all the asymptotic distributions above share the same theoretical rate of convergence, however, the quality of their approximation may result differently good when the indexes are computed on samples with a fixed, though large, size. That is likely to depend both on the chosen index and on the characteristics of the sampled population. With this respect, notice that, even if one could think to improve results by suitable modifications in the definition of a given index (or sometimes in the computation of its standard error), such changes often depend on the choice of a specific model for the population. In many applications, however, the user just wonders which index to choose for kurtosis, without the possibility of making any preliminary assumption, because it is not yet clear whether the underlying population is affected by skewness or heavy tails.

To get further insight about the practical usage of the considered indexes on (large) samples, this section reports some results of a simulation study aimed at comparing the coverage rates of the related confidence intervals and, more generally, at evaluating their performance as estimators. To get coverage, data were

randomly drawn from distributions for which the exact value of each kurtosis index can be computed as a function of the model parameters. The latter, in addition, were varied to explore a wide set of situations of kurtosis, skewness and heavy-tailedness. The usual Wald 0.95-confidence interval $\hat{\theta}_n \mp 1.96 (\widehat{ASV}/n)^{1/2}$ was applied where, for each sample index $\hat{\theta}_n$, \widehat{ASV} denotes the estimated asymptotic variance according to the formulas provided in Theorems 1 to 5. On passing, notice that such formulas require to use simulated data to estimate several moments and functionals of the underlying population and of the variables in (5). At this aim, the ordered dataset $x_{(1)}, \dots, x_{(n)}$ proves to be a useful tool both where the median is the cutting point (for instance, δ_2^- can be estimated as $\frac{1}{n} \sum_{i=1}^{m-1} (x_{(m)} - x_{(i)})^2$, where $m = \lfloor 0.5(n + 1) \rfloor$) and to ease the computation of some sample functionals. Indeed, following Zenga et al. (2007), the estimate of \mathcal{F} can be computed as

$$s^2 + \frac{4}{n^3} \sum_{i=1}^n (ix_{(i)} - t_i)^2 - \frac{4}{n^2} \sum_{i=1}^n (x_{(i)} - \bar{x})(ix_{(i)} - t_i),$$

where \bar{x} and s^2 are the sample mean and variance and $t_j = \sum_{i=1}^j x_{(i)}$ ($j = 1, \dots, n$). Analogously, we obtain that a computationally easy formula for the estimate of \mathcal{D} is

$$\frac{1}{n^2} \sum_{i=1}^n x_{(i)} [2(ix_{(i)} - t_i) - n(x_{(i)} - \bar{x})].$$

Consider first a set of simulations from the normal model. Despite the regularity of the distribution, the estimation process may be still strongly affected by the quality of the sampled data, especially if it rests on the evaluation of some “deep” characteristics of the population, like moments of high order. This fact is fairly known for the estimation of Pearson’s β_2 : to evaluate asymptotic standard errors, for instance, one needs to estimate central moments up to the 8-th order as well. In addition, though the estimator $\hat{\beta}_{2n}$ is asymptotically unbiased, a not negligible bias can be observed even in large samples. Table 1 provides evidence of the two drawbacks: it reports the results obtained by computing $\hat{\beta}_{2n}$ on 10,000 simulated samples from the standard normal distribution and by comparing it with its true value 3. First of all, notice that a significant bias (0.1136) is shown when the sample size is fixed to $n = 50$ and that such a bias is not completely negligible even for $n = 2000$ (still 0.0040), especially if compared with that corresponding to other indexes (commented below). Of course, some corrections to get unbiased estimates from $\hat{\beta}_{2n}$ exist [see Joanes and Gill (1998)], but they rest on the assumption of normality of the population, which is suitable here but not generally and thus, as above commented, it is not safe for many applications. As a second issue, Table 1 underlines the problems arising in the evaluation of the standard error of $\hat{\beta}_{2n}$: the standard deviation of the 10,000 simulated values of $\hat{\beta}_{2n}$ is compared with the mean of $(\widehat{ASV}/n)^{1/2}$ over the same samples. When $n = 50$, a clear problem of underestimation of the standard error exists (0.4502 against 0.5933) and, again, this divergence is not completely negligible when n is raised to 2000 (0.1057 against 0.1093). More importantly for applications, the problems arising in bias and standard error of $\hat{\beta}_{2n}$ reflect on a bad performance in terms of coverage: Table 1 shows that, for 0.95-confidence intervals

Table 1 Simulation results when sampling from the standard normal model

		$\hat{\beta}_{2n}$	$\hat{\alpha}_{Gn}$	$\hat{K}_{1n}(\hat{\gamma}_n)$	$\hat{K}_{1n}(\hat{\mu}_n)$	$\hat{K}_{2n}(\hat{\gamma}_n)$	$\hat{K}_{2n}(\hat{\mu}_n)$
<i>n</i> = 50	True	3.0000	0.7979	0.3634	0.3634	0.4142	0.4142
	Bias	-0.1136	0.0038	-0.0027	-0.0125	-0.0033	-0.0126
	Sim. SE	0.5933	0.0294	0.0462	0.0463	0.0356	0.0356
	Mean est. SE	0.4502	0.0299	0.0491	0.0469	0.0415	0.0390
	Coverage	0.8065	0.9356	0.9501	0.9255	0.9657	0.9390
<i>n</i> = 250	Bias	-0.0221	0.0008	-0.0004	-0.0026	-0.0005	-0.0025
	Sim. SE	0.3030	0.0132	0.0210	0.0210	0.0158	0.0157
	Mean est. SE	0.2632	0.0134	0.0215	0.0212	0.0167	0.0164
	Coverage	0.8823	0.9480	0.9513	0.9470	0.9575	0.9536
<i>n</i> = 1000	Bias	-0.0060	0.0003	-0.0002	-0.0008	-0.0003	-0.0008
	Sim. SE	0.1532	0.0067	0.0106	0.0107	0.0080	0.0080
	Mean est. SE	0.1461	0.0067	0.0107	0.0107	0.0081	0.0081
	Coverage	0.9252	0.9488	0.9500	0.9477	0.9540	0.9502
<i>n</i> = 2000	Bias	-0.0040	0.0002	-0.0002	-0.0004	-0.0002	-0.0004
	Sim. SE	0.1093	0.0048	0.0076	0.0076	0.0057	0.0057
	Mean est. SE	0.1057	0.0047	0.0076	0.0076	0.0057	0.0057
	Coverage	0.9289	0.9484	0.9479	0.9480	0.9510	0.9497

The table gives the true value of the considered indexes (true), the simulated bias (bias), the simulated Standard Error (sim. SE), the mean estimated Standard Error (mean est. SE) that is the average root of $A\hat{S}V/n$ over the simulated samples and the simulated coverage for 95%-confidence intervals (coverage). Bias, sim. SE, mean est. SE and coverage are given for each considered sample size: $n = (50, 250, 1000, 2000)$

based on samples of size $n = 50$, the actual coverage (0.8065) is significantly far from the nominal value. One may check that no relevant changes can be induced by varying both the model parameters or the confidence level (results not shown) and that the coverage problem is not negligible even when $n = 2000$ (see Table 1).

Turning to the other considered kurtosis indexes, the problems above seem to have minor incidence, with some distinctions. Table 1 reports the true value of each kurtosis index for the normal model and the corresponding simulated bias on the 10,000 samples, which is considerably lower than the one for $\hat{\beta}_2$. Geary’s $\hat{\alpha}_{Gn}$ is almost as unbiased as Zenga’s indexes based on the median ($\hat{K}_{1n}(\hat{\gamma}_n)$ and $\hat{K}_{2n}(\hat{\gamma}_n)$), but its bias is of opposite sign: that is coherent with a general tendency to underestimate the real extent of kurtosis. Zenga’s indexes using the mean as a cutting point ($\hat{K}_{1n}(\hat{\mu}_n)$ and $\hat{K}_{2n}(\hat{\mu}_n)$) suffer from a slightly higher bias than the latter ones: perhaps that can be ascribed to the scarce robustness of the sample mean. Nonetheless, the effect clearly vanishes when the sample size increases, which is sensible if one considers that, in the normal model, the mean and the median are actually the same parameter. For all indexes other than $\hat{\beta}_{2n}$, the mean estimated asymptotic standard error is almost coincident with the simulated value. In this aspect (and conversely to the problems in bias), $\hat{K}_{1n}(\hat{\mu}_n)$ and $\hat{K}_{2n}(\hat{\mu}_n)$ behave better than $\hat{K}_{1n}(\hat{\gamma}_n)$ and $\hat{K}_{2n}(\hat{\gamma}_n)$ but, again, that effect is not significant when the sample size increases. Finally, the most relevant strength of the

competitors of $\hat{\beta}_{2n}$ in Table 1 is their ability to reach coverage rates very close to the nominal value. This fact seems to be particularly true for $\hat{K}_{1n}(\hat{\gamma}_n)$ (and partly for the related $\hat{K}_{2n}(\hat{\gamma}_n)$), even if the values in the table should be judged with some caution, due to the simulation error which can be roughly assessed as ± 0.0065 for 10,000 replications. Anyway, practical differences from the nominal 0.95 are observed just for the case $n = 50$ and they are surely of minor relevance, if compared with those commented above for $\hat{\beta}_{2n}$.

A second set of simulations was conducted by using the T-distribution, to get insight about the sensitivity of the considered indexes to heavy-tailedness. A new problem arises because, to guarantee the existence of the r -th moment, the distribution needs to have at least $r + 1$ degrees of freedom (df). Thus, few df may prevent the computation of the true value of a given kurtosis index as population parameter, despite that its estimate can be still obtained from the simulated data. Obviously, under this chance (denoted as “scenario A”), the performance of the estimator cannot be even evaluated. More interestingly, another scenario (B) arises when the df are enough to compute a given population parameter, but too low to provide the existence of the asymptotic variance of its estimator, which is likely to depend on high-order moments. Being that the unaware user may still use the dataset to compute standard errors and confidence intervals, simulations are quite useful here, to evaluate the potential error occurred when using not properly defined tools. Naturally, each index can be evaluated under a third scenario C, where there are sufficient df both to guarantee its existence as a parameter and for the applicability of the related asymptotic theory.

Clearly, the index which mostly depends on the existence of moments is β_2 . Table 2 reports a first set of simulations where $n = 50$ and the df are set to 3, which means that β_2 falls under scenario A and that its estimator cannot be valued. Notice that there are consequences also for α_G , $K_1(\mu)$, and $K_1(\gamma)$, which are all under scenario B: while the mean asymptotic standard error (which is theoretically undefined) departs just partially from the simulated value, the actual coverage is, indeed, always remarkably distant from the nominal 0.95. This problem seems to be more severe for the K_1 -indexes than for α_G , which shows also a lower bias. On the contrary, the K_2 -indexes, which fall under scenario C, perform quite well in this critical situation, both from the point of view of bias and from that of coverage. If the df are raised to 5 (see Table 2, again with $n = 50$), Pearson's β_2 falls under scenario B: the bias of its estimator, which can now be evaluated, reveals to be huge. As expected, this fact combines with a very bad coverage rate of the related confidence interval. The remaining indexes are now under scenario C, but, in spite of that, the estimators based on second moments ($\hat{\alpha}_{Gn}$, $\hat{K}_{1n}(\hat{\mu}_n)$, and $\hat{K}_{1n}(\hat{\gamma}_n)$) behave worse than those based just on the first moment ($\hat{K}_{2n}(\hat{\mu}_n)$ and $\hat{K}_{2n}(\hat{\gamma}_n)$), which show practically no bias and almost exact coverage. Starting from 9 df, all the considered indexes are in scenario C, but notice that the performance of $\hat{\beta}_{2n}$ is still far from being optimal, with no justification based on the existence of moments. Even when the df are raised to 15, the coverage rate for $\hat{\beta}_{2n}$ reaches just 0.6626 when $n = 50$ (see details in Table 2). On the contrary, in this chance, the coverage for all other indexes gets quite close to the nominal value.

Table 2 Simulation results when sampling from the T distribution with $n = 50$

		$\hat{\beta}_{2n}$	$\hat{\alpha}_{Gn}$	$\hat{K}_{1n}(\hat{\gamma}_n)$	$\hat{K}_{1n}(\hat{\mu}_n)$	$\hat{K}_{2n}(\hat{\gamma}_n)$	$\hat{K}_{2n}(\hat{\mu}_n)$
3 df	True	—	0.6366	0.5947	0.5947	0.5000	0.5000
	Bias	—	0.0651	-0.1105	-0.1274	-0.0173	-0.0306
	Sim. SE	—	0.0771	0.0798	0.0760	0.0480	0.0460
	Mean est. SE	—	0.0416	0.0504	0.0550	0.0449	0.0458
	Coverage	—	0.4597	0.4084	0.3736	0.8973	0.8493
5 df	True	9.0000	0.7351	0.4596	0.4596	0.4583	0.4583
	Bias	-4.3251	0.0173	-0.0328	-0.0441	-0.0091	-0.0193
	Sim. SE	2.8916	0.0508	0.0642	0.0636	0.0412	0.0410
	Mean est. SE	0.4427	0.0165	0.0228	0.0227	0.0193	0.0189
	Coverage	0.1480	0.7900	0.8078	0.7631	0.9407	0.9019
9 df	True	4.2000	0.7700	0.4071	0.4071	0.4364	0.4364
	Bias	-0.5982	0.0086	-0.0137	-0.0240	-0.0060	-0.0158
	Sim. SE	1.3845	0.0372	0.0536	0.0536	0.0376	0.0376
	Mean est. SE	0.1451	0.0075	0.0113	0.0110	0.0095	0.0091
	Coverage	0.5157	0.8917	0.9104	0.8729	0.9548	0.9259
15 df	True	3.5455	0.7827	0.3874	0.3874	0.4269	0.4269
	Bias	-0.2982	0.0068	-0.0091	-0.0191	-0.0057	-0.0151
	Sim. SE	0.9499	0.0330	0.0495	0.0500	0.0363	0.0366
	Mean est. SE	0.0871	0.0050	0.0079	0.0076	0.0066	0.0063
	Coverage	0.6626	0.9172	0.9367	0.9031	0.9621	0.9333

For each considered value of the degrees of freedom (df = 3, 5, 9, 15), the table gives the true value of the considered indexes (true), the simulated bias (bias), the simulated Standard Error (sim. SE), the mean estimated Standard Error (mean est. SE) that is the average root of $A\hat{S}V/n$ over the simulated samples and the simulated coverage for 95%-confidence intervals (coverage)

Overall, Table 2 shows that, when heavy-tailedness prevents the existence of moments, $\hat{K}_{2n}(\hat{\mu}_n)$ and $\hat{K}_{2n}(\hat{\gamma}_n)$ seem to be the most accurate choice, both because the related indexes are always defined (and thus their estimates are meaningful) and because a good inference is provided in a wide range of situations. It is interesting to underline that, differently from what happened with the normal model, the real criterion to distinguish among Zenga’s indexes seems not to be the cutting point (mean or median), but the index definition in terms of moments. To appreciate possible effects of the sample size, while Table 2 just considers cases where $n = 50$, Table 3 concentrates on two relevant situations (5 df and 15 df) with raising values of n (250, 1000 and 2000). Large sample sizes obviously tend to even out differences among the indexes; however, the conclusions of Table 2 are substantially confirmed. Notice that, despite the high sample size, $\hat{\beta}_{2n}$ cannot recover from its disadvantage. In addition, a certain tendency for over-coverage seems to characterize $\hat{K}_{2n}(\hat{\gamma}_n)$.

Tables 2 and 3 show the effects of heavy tails conjugated with that of the non-existence of moments. To get further insight, a new set of simulations was considered under the the Generalized Error distribution [see Subbotin (1923)]. This

Table 3 Simulations results when sampling from the T-distribution with 5 and 15 degrees of freedom (df)

		$\hat{\beta}_{2n}$	$\hat{\alpha}_{Gn}$	$\hat{K}_{1n}(\hat{\gamma}_n)$	$\hat{K}_{1n}(\hat{\mu}_n)$	$\hat{K}_{2n}(\hat{\gamma}_n)$	$\hat{K}_{2n}(\hat{\mu}_n)$
5 df	True	9.0000	0.7351	0.4596	0.4596	0.4583	0.4583
	$n = 250$ Bias	-2.8133	0.0050	-0.0108	-0.0131	-0.0018	-0.0039
	Sim. SE	5.0339	0.0291	0.0381	0.0379	0.0191	0.0191
	Mean est. SE	1.3184	0.5987	0.6919	0.6902	0.5488	0.5473
$n = 1000$	Coverage	0.2560	0.8459	0.8325	0.8194	0.9451	0.9371
	Bias	-1.5766	0.0014	-0.0035	-0.0041	-0.0006	-0.0011
	Sim. SE	7.5739	0.0175	0.0237	0.0237	0.0097	0.0097
	Mean est. SE	1.8301	0.0285	0.0391	0.0390	0.0193	0.0192
$n = 2000$	Coverage	0.3580	0.8828	0.8774	0.8706	0.9487	0.9469
	Bias	-1.0158	0.0005	-0.0016	-0.0019	-0.0002	-0.0005
	Sim. SE	13.3304	0.0135	0.0182	0.0181	0.0069	0.0069
	Mean est. SE	1.9984	0.0154	0.0215	0.0215	0.0097	0.0097
15 df	Coverage	0.3986	0.8969	0.8934	0.8903	0.9465	0.9454
	True	3.5455	0.7827	0.3874	0.3874	0.4269	0.4269
	$n = 250$ Bias	-0.0674	0.0013	-0.0017	-0.0039	-0.0009	-0.0029
	Sim. SE	0.6153	0.0155	0.0238	0.0239	0.0165	0.0165
$n = 1000$	Mean est. SE	0.4227	0.0054	0.0083	0.0082	0.0061	0.0060
	Coverage	0.7899	0.9367	0.9402	0.9319	0.9554	0.9480
	Bias	-0.0158	0.0004	-0.0006	-0.0011	-0.0003	-0.0008
	Sim. SE	0.3520	0.0078	0.0122	0.0122	0.0083	0.0083
$n = 2000$	Mean est. SE	0.2785	0.0156	0.0243	0.0243	0.0168	0.0167
	Coverage	0.8517	0.9426	0.9440	0.9414	0.9531	0.9508
	Bias	-0.0082	0.0002	-0.0003	-0.0006	-0.0002	-0.0004
	Sim. SE	0.2688	0.0057	0.0089	0.0089	0.0060	0.0060
$n = 1000$	Mean est. SE	0.2118	0.0078	0.0122	0.0122	0.0084	0.0083
	Coverage	0.8844	0.9437	0.9431	0.9419	0.9505	0.9495

The table gives the true value of the considered indexes (true), the simulated bias (bias), the simulated Standard Error (sim. SE), the mean estimated Standard Error (mean est. SE) that is the average root of $A\hat{S}V/n$ over the simulated samples and the simulated coverage for 95%-confidence intervals (coverage). Bias, sim SE, mean est. SE and coverage are given for each considered sample size $n = (50, 250, 1000)$

model has a shape parameter ν providing normality when $\nu = 2$ and heavy or light tails when ν is, respectively, less or greater than 2, but, in contrast with the T-distribution, it possesses all moments. Table 4 shows some results obtained by a combination of three sample sizes ($n = 50;250;1000$) and two values of the shape parameter ($\nu = 0.5;4$), while the location and the scale parameters of the model were always set to get zero mean and unit variance. The first part of the table, the one with heavy tails, resembles what faced for the T-distribution with few df: all indexes except $\hat{K}_{2n}(\hat{\gamma}_n)$ and $\hat{K}_{2n}(\hat{\mu}_n)$ suffer from the irregularity of the distribution and, while $\hat{\alpha}_{Gn}$ and the K_1 -indexes recover fast with the increase in the sample size, $\hat{\beta}_{2n}$ is still quite inaccurate even for $n = 1000$. On the contrary, in the

Table 4 Simulation results when sampling from the Generalized Error Distribution with shape parameter $\nu = (0.5, 4)$

		$\hat{\beta}_{2n}$	$\hat{\alpha}_{Gn}$	$\hat{K}_{1n}(\hat{\gamma}_n)$	$\hat{K}_{1n}(\hat{\mu}_n)$	$\hat{K}_{2n}(\hat{\gamma}_n)$	$\hat{K}_{2n}(\hat{\mu}_n)$
$\nu = 0.5$	True	25.2000	0.5477	0.7000	0.7000	0.6250	0.6250
$n = 50$	Bias	-15.3803	0.0588	-0.0802	-0.1211	-0.0310	-0.0742
	Sim. SE	32.9222	0.0048	0.0051	0.0072	0.0023	0.0038
	Mean est. SE	3.3602	0.0481	0.0474	0.0752	0.0439	0.0728
	Coverage	0.1428	0.6346	0.5601	0.6058	0.8569	0.7962
$n = 250$	Bias	-8.3927	0.0163	-0.0241	-0.0345	-0.0064	-0.0199
	Sim. SE	11.4192	0.0425	0.0424	0.0446	0.0227	0.0268
	Mean est. SE	3.7552	0.0315	0.0298	0.0343	0.0215	0.0294
	Coverage	0.2784	0.7717	0.7340	0.7269	0.9143	0.9016
$n = 1000$	Bias	-3.4837	0.0048	-0.0076	-0.0104	-0.0017	-0.0058
	Sim. SE	14.5569	0.0247	0.0254	0.0256	0.0115	0.0123
	Mean est. SE	5.3612	0.0207	0.0206	0.0211	0.0113	0.0131
	Coverage	0.4668	0.8518	0.8372	0.8260	0.9388	0.9379
$\nu = 4$	True	2.1884	0.8409	0.2929	0.2929	0.3662	0.3662
$n = 50$	Bias	0.0095	0.0001	0.0064	-0.0050	0.0012	-0.0089
	Sim. SE	0.2821	0.0249	0.0412	0.0414	0.0346	0.0347
	Mean est. SE	0.2731	0.0263	0.0480	0.0432	0.0418	0.0372
	Coverage	0.9326	0.9511	0.9682	0.9422	0.9709	0.9385
$n = 250$	Bias	0.0021	0.0001	0.0012	-0.0012	0.0002	-0.0020
	Ssim. SE	0.1219	0.0109	0.0183	0.0184	0.0154	0.0154
	Mean est. SE	0.1201	0.0111	0.0192	0.0186	0.0163	0.0157
	Coverage	0.9442	0.9492	0.9549	0.9474	0.9573	0.9464
$n = 1000$	Bias	0.0008	0.0000	0.0004	-0.0002	0.0001	-0.0004
	Sim. SE	0.0598	0.0055	0.0092	0.0092	0.0077	0.0077
	Mean est. SE	0.0601	0.0055	0.0093	0.0092	0.0078	0.0077
	Coverage	0.9488	0.9508	0.9535	0.9503	0.9533	0.9510

The table gives the true value of the considered indexes (true), the simulated bias (bias), the simulated Standard Error (sim. SE), the mean estimated Standard Error (mean est. SE) that is the average root of $A\hat{S}V/n$ over the simulated samples and the simulated coverage for 95%-confidence intervals (coverage). Bias, sim SE, mean est. SE and coverage are given for each considered sample size $n = (50, 250, 1000)$

second part of Table 4 ($\nu = 4$ and light tails), the situation is similar to the one of Table 1 (normal model): the performance of all indexes is far from being critical, with some limitations just for $\hat{\beta}_{2n}$. Heavy-tailedness, then, confirms to be a major inconvenience for the performance of many kurtosis indexes, beyond the problem of the existence of moments.

All the above considered simulations were based on symmetric models. To check for cases where the user faces skew data, Table 5 reports some results obtained with the standard skew-normal model [for a recent review, see Azzalini and Capitanio (2014)]. In this family, the standard normal distribution is obtained by setting the shape parameter (denoted here as ξ) to zero. Large positive values

Table 5 Simulation results when sampling from the skew normal distribution with shape parameter $\xi = (1, 4)$

		$\hat{\beta}_{2n}$	$\hat{\alpha}_{Gn}$	$\hat{K}_{1n}(\hat{\gamma}_n)$	$\hat{K}_{1n}(\hat{\mu}_n)$	$\hat{K}_{2n}(\hat{\gamma}_n)$	$\hat{K}_{2n}(\hat{\mu}_n)$
$\xi = 1$	True	3.0617	0.7962	0.3654	0.3651	0.4154	0.4152
$n = 50$	Bias	-0.1253	0.0036	-0.0024	-0.0121	-0.0030	-0.0121
	Sim. SE	0.4321	0.0009	0.0021	0.0022	0.0013	0.0013
	Mean est. SE	0.2088	0.4621	0.5894	0.5767	0.5420	0.5258
	Coverage	0.7977	0.9320	0.9485	0.9250	0.9630	0.9391
$n = 250$	Bias	-0.0299	0.0010	-0.0008	-0.0029	-0.0007	-0.0027
	Sim. SE	0.3332	0.0136	0.0214	0.0214	0.0160	0.0160
	Mean est. SE	0.1410	0.0136	0.0216	0.0213	0.0167	0.0164
	Coverage	0.8633	0.9461	0.9487	0.9447	0.9561	0.9504
$n = 1000$	Bias	-0.0057	0.0002	-0.0001	-0.0006	-0.0002	-0.0007
	Sim. SE	0.1695	0.0068	0.0107	0.0107	0.0080	0.0080
	Mean est. SE	0.1131	0.0068	0.0108	0.0108	0.0081	0.0081
	Coverage	0.9146	0.9470	0.9485	0.9465	0.9530	0.9500
$\xi = 4$	True	3.6828	0.7948	0.3495	0.3367	0.4041	0.3945
$n = 50$	Bias	-0.3113	0.0053	-0.0017	-0.0110	-0.0029	-0.0116
	Sim. SE	1.0857	0.0337	0.0453	0.0446	0.0353	0.0350
	Mean est. SE	0.3011	0.0051	0.0078	0.0073	0.0066	0.0061
	Coverage	0.6708	0.9132	0.9542	0.9281	0.9656	0.9374
$n = 250$	Bias	-0.0609	0.0010	0.0000	-0.0021	-0.0003	-0.0023
	Sim. SE	0.6523	0.0157	0.0210	0.0207	0.0160	0.0159
	Mean est. SE	0.2478	0.0154	0.0213	0.0206	0.0167	0.0162
	Coverage	0.8059	0.9362	0.9487	0.9425	0.9555	0.9488
$n = 1000$	Bias	-0.0211	0.0002	0.0002	-0.0004	0.0001	-0.0004
	Sim. SE	0.3387	0.0078	0.0104	0.0102	0.0079	0.0078
	Mean est. SE	0.2149	0.0079	0.0106	0.0103	0.0081	0.0080
	Coverage	0.8765	0.9480	0.9565	0.9529	0.9587	0.9537

The table gives the true value of the considered indexes (true), the simulated bias (bias), the simulated Standard Error (sim. SE), the mean estimated Standard Error (mean est. SE) that is the average root of $A\hat{S}V/n$ over the simulated samples and the simulated coverage for 95%-confidence intervals (coverage). Bias, sim. SE, mean est. SE and coverage are given for each considered sample size $n = (50, 250, 1000)$

of ξ increase right skewness of the distribution, while negative values make it skewed to the left. For our purposes, the direction of skewness is not relevant; thus, just two positive values (1 and 4) are set; such values are combined with three sample sizes (50, 250 and 1000). Looking at Table 5, one can notice that $\hat{\beta}_{2n}$ is also strongly affected by skewness, differently from the other considered indexes. However, when the degree of skewness increases, $\hat{\alpha}_{Gn}$ reveals some problems as well, performing quite worse than in the case of symmetric distributions with light tails. On the contrary, Zenga's indexes show a good robustness to skewness, with a dominance of those using the median as a cutting point (conjugated with a slight tendency to over-coverage).

Of course, skewness and heavy tails might jointly affect data and, to understand the effect of that, a final set of simulations was conducted by using the skew-T model [see Azzalini and Capitanio (2014)]. Not surprisingly, the index which best opposes the related “double” violation to normality is $\hat{K}_{2n}(\hat{\gamma}_n)$. Indeed, as noticed from Table 2 on, the effect of heavy tails tends to favor indexes based on moments of first order. In addition, skewed distributions and a generic attitude to robustness lead to the dominance of the median as a cutting point (see Tables 1 and 5, for instance). The conclusions above are clearly depicted in Table 6 which considers, for the sake of brevity, just the case $n = 50$ and combines two values of the shape parameter ($\xi = 1;4$) with two levels of df (5 and 15). Notice that, like for the T-distribution, the existence of some moments depends on the latter parameter: in our settings, however, the only index under scenario B is β_2 when there are 5 df, while all other indexes are under scenario in C in all situations considered in Table 6. Even after discarding cases where the application of the asymptotic theory is shaky, however, the performance of $\hat{\beta}_{2n}$ results to be strongly affected by both skewness and heavy-taildeness. The same is partially true for $\hat{\alpha}_{Gn}$ and mildly for the K_1 -indexes.

Table 6 Simulation results when sampling from the Skew T-Distribution with $n = 50$

		$\hat{\beta}_{2n}$	$\hat{\alpha}_{Gn}$	$\hat{K}_{1n}(\hat{\gamma}_n)$	$\hat{K}_{1n}(\hat{\mu}_n)$	$\hat{K}_{2n}(\hat{\gamma}_n)$	$\hat{K}_{2n}(\hat{\mu}_n)$
$\xi = 1$	True	11.9208	0.7306	0.4538	0.4463	0.4555	0.4510
5 df	Bias	-6.9916	0.0214	-0.0326	-0.0427	-0.0094	-0.0193
	Sim. SE	3.2644	0.0524	0.0617	0.0599	0.0401	0.0395
	Mean est. SE	1.1239	0.0374	0.0510	0.0511	0.0432	0.0420
	Coverage	0.0950	0.7633	0.8264	0.7917	0.9459	0.9135
$\xi = 1$	True	3.7063	0.7809	0.3879	0.3866	0.4271	0.4263
15 df	Bias	-0.3660	0.0064	-0.0080	-0.0178	-0.0047	-0.0139
	Sim. SE	1.0784	0.0341	0.0497	0.0500	0.0362	0.0364
	Mean est. SE	0.2625	0.0144	0.0223	0.0216	0.0188	0.0178
	Coverage	0.6170	0.9097	0.9338	0.9005	0.9636	0.9343
$\xi = 4$	True	20.6763	0.7202	0.4114	0.3749	0.4309	0.4067
5 df	Bias	-14.6303	0.0301	-0.0232	-0.0309	-0.0064	-0.0151
	Sim. SE	4.4112	0.0622	0.0570	0.0503	0.0390	0.0365
	Mean est. SE	0.3894	0.0091	0.0113	0.0110	0.0097	0.0090
	Coverage	0.0634	0.6848	0.8668	0.8606	0.9528	0.9272
$\xi = 4$	True	4.8169	0.7778	0.3651	0.3472	0.4118	0.3987
15 df	Bias	-0.8830	0.0096	-0.0063	-0.0158	-0.0043	-0.0133
	Sim. SE	1.8370	0.0402	0.0482	0.0467	0.0362	0.0357
	Mean est. SE	0.1382	0.0055	0.0079	0.0074	0.0067	0.0062
	Coverage	0.4867	0.8773	0.9418	0.9153	0.9616	0.9355

The table gives the true value of the considered indexes (true), the simulated bias (bias), the simulated Standard Error (sim. SE), the mean estimated Standard Error (mean est. SE) that is the average root of $A\hat{S}V/n$ over the simulated samples and the simulated coverage for 95%-confidence intervals (coverage). Bias, sim. SE, mean est. SE and coverage are given for each combination of degrees of freedom $df = (5, 15)$ and shape parameter $\xi = (1, 4)$

In addition, notice that a distinguishing feature of $\widehat{K}_{2n}(\widehat{\gamma}_n)$ seems to be its definite tendency to over-coverage: this fact can be an advantage, because it leaves room to seek for shorter, thus more informative, confidence intervals. As a final comment, let us consider that Table 6 depicts some very realistic situations, resulted from the sum of many deviations from regularity. Therefore, one should be at least aware that, under many circumstances, the use of some classical indexes of kurtosis (markedly of β_2), or at least the use of a limited set of them, may lead to inaccurate results.

5 Conclusions

This paper gave indications about the usage of several kurtosis measures on large datasets. The need to estimate kurtosis is found in many fields of applications, basically in all those circumstances where departure from normality may cause the unreliability of classical tools and, more generally, where the shape of the distribution is of interest. The empirical finance is surely one of those fields as demonstrated by the ubiquitous fat-tailedness of daily stock returns, the widespread use of GARCH models and the attempts made in the literature to improve the Black–Scholes formula [see Corrado and Su (1996) and Carnero et al. (2004)]. At the same time, those examples show that kurtosis is too often simply identified with the centered fourth moments, while this paper shows that more informative measures can be considered. Actually, some practitioners may find it difficult to compute indexes other than the well-established β_2 , mainly because they are unsure on how to interpret the obtained results. However, the meaning of those alternative indexes is quite similar to the one for β_2 , because it suffices to compare an observed value with its theoretical counterpart in the case of the normal distribution (which, for the indexes considered in this paper, is provided in the first row of Table 1). We want to highlight, in addition, that computing a different kurtosis index, or at least a range of alternative measures, can provide a series of advantages.

First of all, researchers should consider that, despite its size, a dataset is often a sample from a given population. Thus, the simple point estimation is meaningless unless accompanied by the corresponding standard error. Our simulations reveal that, while there is scarcely a problem of bias, all considered kurtosis measure are often characterized by high standard errors and, as a consequence, the real extent of kurtosis is likely to be different from the estimated value. To ease the interpretation, the user should then revert to confidence intervals which, however, must also provide reliable results. In this sense, the choice of a kurtosis index must match the need to guarantee the desired coverage of confidence intervals, against possible choices of the sample size or possible irregularities of the sampled distribution.

Something which is often neglected is that, even when the sampled population is quite regular, too low sample sizes can still lead to unreliable results, if some indexes are applied. As expected, such problems arise when the indexes are based on unrobust summaries of the sample. Specifically, when n is as low as 50, our simulations show that β_2 cannot be properly estimated even for samples from the normal distribution: some problems arise both in bias both in the coverage

of confidence intervals. Being based on the 4-th power of deviations from the mean, namely, β_2 is likely to be strongly affected by the sample variability which, first, reflects into high standard errors. The real problem, however, is that even those standard errors cannot be easily estimated because, in turn, they depend on the 8-th centered moment. As a consequence, confidence intervals for β_2 are quite large, thus not informative, and unreliable. The same problem, of course, affects other indexes, but to a lesser extent: among the considered indexes, α_G , $K_1(\mu)$ and $K_1(\gamma)$ are defined upon moments up to the second order, so that the related standard errors need moments up to the fourth. A further improvement, of course, can be obtained by Zenga's K_2 -indexes, whose definition rests on first moments, while standard errors just depend on the second order. In our opinion, then, when the sample size is too low, the researcher cannot forget to include $K_2(\mu)$ and (markedly) $K_2(\gamma)$ in the set of computed kurtosis measures and to check whether they provide contrasting evidence with respect to the classical indexes.

High sample sizes, of course, make the choice of a specific index redundant. Our simulation show that, when n is around 200, all considered indexes provide similarly reliable information. However, this judgment fits just those situations (like the one depicted in Table 1) where the underlying distribution is quite regular, while it is not uncommon to face heavy-tailed and/or skewed populations. The former kind of irregularity is, in our opinion, the one which deserves most attention when a kurtosis index is chosen. First, heavy-tailed distributions often do not possess some moments, a fact which can make the point estimate meaningless, because the related parameter does not even exist. Sometimes this problem affects just the estimation of standard errors, but it can be considered almost as severe as before because, in this case, confidence intervals are not properly defined. Of course, a possible indication is to use preliminary statistical testing to check for the existence of the needed moments in the underlying distribution. The literature provides both graphical and parametric methods based on tail-index estimation [Hill (1975)] and, more recently, some nonparametric approaches using bootstrap [see Ng and Yau (2018)]. However, it is clear that the process to get a final assessment of kurtosis could result too cumbersome, at least if compared with the possibility of starting with indexes which are quickly robust against the nonexistence of moments, like $K_2(\mu)$ and $K_2(\gamma)$. In addition, our simulations (see Table 4) show that heavy-tailedness itself may undermine the performance of some indexes, beyond the problem of the existence of moments, so that just robust indexes are actually a safe choice. The researcher must be aware, however, that even those indexes cannot always cope with too low sample sizes, as evidenced in Tables 2 and 4.

Overall, $K_2(\mu)$ and $K_2(\gamma)$ react to heavy-tailedness quite similarly. A distinction among them could be appreciated in the case of skewness. The use of the median as a cutting point seems to be the strength of $K_2(\gamma)$ over $K_2(\mu)$ (see Table 5). Nonetheless, under a mild skewness, the real answer is just a high sample size: excluding β_2 , no relevant differences can be reported over 200 observations, similar to the case of a fully regular distribution. Of course, things change when skewness is conjugated with heavy tails: our simulation show that, in those cases, the only reliable choice is $K_2(\gamma)$, even if n is as high as 200.

As a final remark, we would like to underline that the conclusions above can be useful, beyond the process of the estimation of kurtosis, when researchers need to compare the levels of kurtosis of two populations or of two sides of a distribution. Indeed, suitable testing procedures can be built upon the asymptotic distributions derived in this paper, so that it is important to have knowledge of their sensitivity to the level of the sample size and to the characteristics of the sampled population. This study will be the object of a future research.

Concerning other topics of research, the above-reported need to use estimated kurtosis in many fields of application leaves, in our opinion, a relevant problem unexplored: the dependence of data across time. Such a characteristic is frequently encountered, for instance, in financial applications based on the analysis of returns. Beyond a specific framework, however, we think that the search for reliable inferential instruments for kurtosis under serial dependence is an urgent problem which deserves attention in the nearest future. That issue could be the starting point for the development of further generalizations of the presented methodology. In the literature, for instance, a certain need is exhibited to conjugate the study of kurtosis with the measurement of dependence among variables, to cope with the so-called co-kurtosis [see Martellini and Ziemann (2010)]. Again, we think that, even in the multivariate framework, some efforts could be paid to go well beyond the usual identification with the fourth centered moment, so that kurtosis can be measured by means of other logics.

Appendix

Proof of Lemma 2

First, recall that, under the assumptions of the lemma, $\widehat{F}_n(\widehat{z}_n) - \widehat{F}_n(z) = o_p(1)$ and, consequently, $\widehat{F}_n(\widehat{z}_n) = F(z) + o_p(1)$ [see De Capitani and Pasquazzi (2015)]. Now, simplify notation in \widehat{I}_{3n} and decompose it as:

$$\begin{aligned} \widehat{I}_{3n} &= \int \int H^-(x, y, \widehat{z}_n) d\widehat{F}_n(y) d\widehat{F}_n(x) \\ &= \int \int [H^-(x, y, \widehat{z}_n) - H^-(x, y, z)] d\widehat{F}_n(y) d\widehat{F}_n(x) \\ &\quad + \int \int H^-(x, y, z) d\widehat{F}_n(y) d\widehat{F}_n(x). \end{aligned}$$

After assuming that $\widehat{z}_n \leq z$, it is easily seen that

$$H^-(x, y, \widehat{z}_n) - H^-(x, y, z) = \begin{cases} \widehat{z}_n - z & (x, y) \in (-\infty, \widehat{z}_n] \times [z, \infty) \cup [z, \infty) \times (-\infty, \widehat{z}_n] \\ x - z & (x, y) \in [\widehat{z}_n, z] \times [z, \infty) \\ y - z & (x, y) \in [z, \infty) \times [\widehat{z}_n, z] \\ \widehat{z}_n - x & (x, y) \in [\widehat{z}_n, z] \times (-\infty, \widehat{z}_n] \\ \widehat{z}_n - y & (x, y) \in (-\infty, \widehat{z}_n] \times [\widehat{z}_n, z] \\ -|x - y| & (x, y) \in [\widehat{z}_n, z] \times [\widehat{z}_n, z] \\ 0 & \text{otherwise .} \end{cases}$$

Moreover, note that

$$\begin{aligned}
 & \int_{-\infty}^{\hat{z}_n} \int_z^{\infty} (\hat{z}_n - z) d\hat{F}_n(y) d\hat{F}_n(x) \\
 & \quad + \int_z^{\infty} \int_{-\infty}^{\hat{z}_n} (\hat{z}_n - z) d\hat{F}_n(y) d\hat{F}_n(x) \\
 & = 2 \int_{-\infty}^{\hat{z}_n} \int_z^{\infty} (\hat{z}_n - z) d\hat{F}_n(y) d\hat{F}_n(x) \\
 & = 2(\hat{z}_n - z) \hat{F}_n(\hat{z}_n)(1 - \hat{F}_n(z)) \\
 & = 2(\hat{z}_n - z) (F(z) + o_p(1))(1 - F(z) - O_p(n^{-1/2})) \\
 & = 2(\hat{z}_n - z) F(z) (1 - F(z)) + o_p(n^{-1/2})
 \end{aligned} \tag{18}$$

$$\begin{aligned}
 & \int_{\hat{z}_n}^z \int_z^{\infty} (x - z) d\hat{F}_n(y) d\hat{F}_n(x) + \int_z^{\infty} \int_{\hat{z}_n}^z (y - z) d\hat{F}_n(y) d\hat{F}_n(x) \\
 & = 2 \int_{\hat{z}_n}^z \int_z^{\infty} (x - z) d\hat{F}_n(y) d\hat{F}_n(x) \\
 & = 2(1 - \hat{F}_n(z)) \int_{\hat{z}_n}^z (x - z) d\hat{F}_n(x) \\
 & = o_p(n^{-1/2})
 \end{aligned} \tag{19}$$

$$\begin{aligned}
 & \int_{\hat{z}_n}^z \int_{-\infty}^{\hat{z}_n} (\hat{z}_n - x) d\hat{F}_n(y) d\hat{F}_n(x) \\
 & \quad + \int_{-\infty}^{\hat{z}_n} \int_{\hat{z}_n}^z (\hat{z}_n - y) d\hat{F}_n(y) d\hat{F}_n(x) \\
 & = 2 \int_{\hat{z}_n}^z \int_{-\infty}^{\hat{z}_n} (\hat{z}_n - x) d\hat{F}_n(y) d\hat{F}_n(x) \\
 & = 2\hat{F}_n(\hat{z}_n) \int_{\hat{z}_n}^z (\hat{z}_n - x) d\hat{F}_n(x) \\
 & = o_p(n^{-1/2}).
 \end{aligned} \tag{20}$$

Finally, being that

$$\left| - \int_{\hat{z}_n}^z \int_{\hat{z}_n}^z |x - y| d\hat{F}_n(y) d\hat{F}_n(x) \right| < |z - \hat{z}_n| (\hat{F}_n(z) - \hat{F}_n(\hat{z}_n))^2 = o_p(n^{-1/2}),$$

one has

$$- \int_{\hat{z}_n}^z \int_{\hat{z}_n}^z |x - y| d\hat{F}_n(y) d\hat{F}_n(x) = o_p(n^{-1/2}). \tag{21}$$

Now, expressions from (18) to (21) allow to write, at least when $\hat{z}_n \leq z$,

$$\hat{I}_{3n} = 2(\hat{z}_n - z)F(z)(1 - F(z)) + \int \int H^-(x, y, z) d\hat{F}_n(y) d\hat{F}_n(x) + o_P(n^{-1/2}).$$

However, similar steps can be followed under the assumption that $\hat{z}_n > z$, which shows that the statement above is a general result. The proof for \hat{I}_{3n} is obtained consistently.

Proof of Lemma 3

As in the proof of Lemma 2, notice first that

$$\int_{\hat{z}_n}^z (\hat{z}_n - x)^l d\hat{F}_n(x) = o_P(n^{-1/2}) \quad \text{and} \quad \int_{\hat{z}_n}^z (z - x)^l d\hat{F}_n(x) = o_P(n^{-1/2})$$

with $l = 1;2$. Now decompose \hat{I}_{5n} as

$$\begin{aligned} \hat{I}_{5n} &= \int (\hat{z}_n - x)^2 \mathbb{1}(x \leq \hat{z}_n) d\hat{F}_n(x) \\ &= \int_{-\infty}^z (\hat{z}_n - x)^2 d\hat{F}_n(x) + \int_z^{\hat{z}_n} (\hat{z}_n - x)^2 d\hat{F}_n(x) \\ &= \int_{-\infty}^z [(\hat{z}_n - z) + (z - x)]^2 d\hat{F}_n(x) + o_P(n^{-1/2}) \\ &= \hat{F}_n(z)(\hat{z}_n - z)^2 + \int_{-\infty}^z (z - x)^2 d\hat{F}_n(x) \\ &\quad + 2(\hat{z}_n - z) \int_{-\infty}^z (z - x) d\hat{F}_n(x) + o_P(n^{-1/2}) \\ &= \int_{-\infty}^z (z - x)^2 d\hat{F}_n(x) \\ &\quad + 2(\hat{z}_n - z) \left[\left(\int_{-\infty}^z (z - x) d\hat{F}_n(x) - I_1 \right) + I_1 \right] + o_P(n^{-1/2}) \\ &= \int (z - x)^2 \mathbb{1}(x \leq z) d\hat{F}_n(x) + 2I_1 (\hat{z}_n - z) + o_P(n^{-1/2}). \end{aligned}$$

The asymptotic representation for \hat{I}_{6n} can be obtained similarly.

Proof of Theorem 3

Clearly, Σ is the asymptotic variance matrix of the vector

$$\sqrt{n} \begin{bmatrix} \widehat{\Delta}_n^- - \Delta^- \\ \widehat{\Delta}_n^+ - \Delta^+ \\ \widehat{\delta}_n^- - \delta^- \\ \widehat{\delta}_n^+ - \delta^+ \end{bmatrix}$$

whose asymptotic normality follows from the linearizations provided in Section 3.1, the central limit theorem and the Cramer–Wold device. Then, the result about $\widehat{K}_{2n}(\widehat{\gamma}_n)$ is an application of the delta method. To get into details about the the derivation of the elements of Σ , consider first the following expectations:

$$\begin{aligned} & \int \int H^-(x, y, \gamma) \mathbb{1}(x \leq \gamma) dF(y) dF(x) \\ &= \int \int H^-(x, y, \gamma) (1 - \mathbb{1}(x > \gamma)) dF(y) dF(x) \\ &= \Delta^- - \int_{\gamma}^{\infty} \int_{-\infty}^{\gamma} (\gamma - y) dF(y) dF(x) \\ &= \Delta^- - \frac{\delta^-}{2} \end{aligned}$$

Similarly, $\int \int H^+(x, y, \gamma) \mathbb{1}(x > \gamma) dF(y) dF(x) = \Delta^+ - \frac{\delta^+}{2}$. In addition,

$$\begin{aligned} & \int \int H^-(x, y, \gamma) \mathbb{1}(x > \gamma) dF(y) dF(x) \\ &= \int_{\gamma}^{\infty} \int H^-(x, y, \gamma) dF(y) dF(x) \\ &= \int_{\gamma}^{\infty} \int_{-\infty}^{\gamma} (\gamma - y) dF(y) dF(x) \\ &= \frac{\delta^-}{2} \end{aligned}$$

and $\int \int H^+(x, y, \gamma) \mathbb{1}(x \leq \gamma) dF(y) dF(x) = \frac{\delta^+}{2}$. Moreover,

$$\begin{aligned} & \int \int H^-(x, y, \gamma) (x - \gamma) \mathbb{1}(x > \gamma) dF(y) dF(x) \\ &= \int_{\gamma}^{\infty} \int_{-\infty}^{\gamma} (\gamma - y)(x - \gamma) dF(y) dF(x) \\ &= \int_{\gamma}^{\infty} (x - \gamma) dF(x) \int_{-\infty}^{\gamma} (\gamma - y) dF(y) \\ &= \delta^+ \delta^- \end{aligned}$$

and analogously $\int \int H^+(x, y, \gamma) (\gamma - x) \mathbb{1}(x \leq \gamma) dF(y) dF(x) = \delta^+ \delta^-$. Now notice that

$$\int H^-(x, y, \gamma) dF(y) = \begin{cases} \int_{-\infty}^{\gamma} |x - y| dF(y) + \frac{\gamma - x}{2} & x \leq \gamma \\ \delta^- & x > \gamma \end{cases}$$

and that, equivalently,

$$\int H^+(x, y, \gamma) dF(y) = \begin{cases} \delta^+ & x \leq \gamma \\ \int_{\gamma}^{\infty} |x - y| dF(y) + \frac{x - \gamma}{2} & x > \gamma. \end{cases}$$

Thus, a last relevant expectation can be computed as:

$$\begin{aligned} & \int \left(\int H^+(x, y, \gamma) dF(y) \int H^-(x, y, \gamma) dF(y) \right) dF(x) \\ &= \int_{-\infty}^{\gamma} \delta^+ \left(\int_{-\infty}^{\gamma} |x - y| dF(y) + \frac{\gamma - x}{2} \right) dF(x) \\ & \quad + \int_{\gamma}^{\infty} \delta^- \left(\int_{\gamma}^{\infty} |x - y| dF(y) + \frac{x - \gamma}{2} \right) dF(x) \\ &= \delta^+ \left(\int_{-\infty}^{\gamma} \int_{-\infty}^{\gamma} |x - y| dF(y) dF(x) + \int_{-\infty}^{\gamma} \frac{\gamma - x}{2} dF(x) \right) \\ & \quad + \delta^- \left(\int_{\gamma}^{\infty} \int_{\gamma}^{\infty} |x - y| dF(y) dF(x) + \int_{\gamma}^{\infty} \frac{x - \gamma}{2} dF(x) \right) \\ &= \delta^+ \left(\Delta^- - \frac{\delta^-}{2} \right) + \delta^- \left(\Delta^+ - \frac{\delta^+}{2} \right). \end{aligned}$$

The elements of Σ can now be obtained:

$$\begin{aligned} \Sigma_{11} &= \text{Var}(I_1(X)) \\ &= \mathbb{E} \left[4 \left(\int H^-(X, y, \gamma) dF(y) \right)^2 + \frac{\mathbb{1}(X \leq \gamma)}{4f^2(\gamma)} - 2 \frac{\mathbb{1}(X \leq \gamma)}{f(\gamma)} \int H^-(X, y, \gamma) dF(y) \right] + \\ & \quad - \left(2\Delta^- - \frac{1}{4f(\gamma)} \right)^2 \\ &= 4\mathcal{F}^- + \frac{1}{8f^2(\gamma)} - \frac{2}{f(\gamma)} \left(\Delta^- - \frac{\delta^-}{2} \right) - \left(2\Delta^- - \frac{1}{4f(\gamma)} \right)^2 \end{aligned}$$

(and similarly for Σ_{22}).

$$\begin{aligned} \Sigma_{33} &= \text{Var}(l_3(X)) \\ &= \mathbb{E} \left[(\gamma - X)^2 \mathbb{1}(X \leq \gamma) + \frac{\mathbb{1}(X \leq \gamma)}{4f^2(\gamma)} - \frac{(\gamma - X) \mathbb{1}(X_i \leq \gamma)}{f(\gamma)} \right] + \\ &\quad - \left(\delta^- - \frac{1}{4f(\gamma)} \right)^2 \\ &= \delta_2^- + \frac{1}{8f^2(\gamma)} - \frac{\delta^-}{f(\gamma)} - \left(\delta^- - \frac{1}{4f(\gamma)} \right)^2 \end{aligned}$$

(and similarly for Σ_{44}).

$$\begin{aligned} \Sigma_{12} &= \text{Cov}(l_1(X), l_2(X)) \\ &= \mathbb{E} \left[\left(2 \int H^-(X, y, \gamma) dF(y) - \frac{\mathbb{1}(X \leq \gamma)}{2f(\gamma)} \right) \left(2 \int H^+(X, y, \gamma) dF(y) - \frac{\mathbb{1}(X > \gamma)}{2f(\gamma)} \right) \right] + \\ &\quad - \left(2\Delta^- - \frac{1}{4f(\gamma)} \right) \left(2\Delta^+ - \frac{1}{4f(\gamma)} \right) \\ &= 4\delta^+ \left(\Delta^- - \frac{\delta^-}{2} \right) + 4\delta^- \left(\Delta^+ - \frac{\delta^+}{2} \right) - \frac{1}{f(\gamma)} \frac{\delta^+}{2} - \frac{1}{f(\gamma)} \frac{\delta^-}{2} \\ &\quad - \left(2\Delta^- - \frac{1}{4f(\gamma)} \right) \left(2\Delta^+ - \frac{1}{4f(\gamma)} \right); \end{aligned}$$

$$\begin{aligned} \Sigma_{13} &= \text{Cov}(l_1(X), l_3(X)) \\ &= \mathbb{E} \left[\left(2 \int H^-(X, y, \gamma) dF(y) - \frac{\mathbb{1}(X \leq \gamma)}{2f(\gamma)} \right) \left((\gamma - X) \mathbb{1}(X \leq \gamma) - \frac{\mathbb{1}(X \leq \gamma)}{2f(\gamma)} \right) \right] + \\ &\quad - \left(2\Delta^- - \frac{1}{4f(\gamma)} \right) \left(\delta^- - \frac{1}{4f(\gamma)} \right) \\ &= 2\mathcal{D}^- - \frac{1}{f(\gamma)} \left(\Delta^- - \frac{\delta^-}{2} \right) + \frac{\delta^-}{2f(\gamma)} + \frac{1}{8f^2(\gamma)} - \left(2\Delta^- - \frac{1}{4f(\gamma)} \right) \left(\delta^- - \frac{1}{4f(\gamma)} \right) \end{aligned}$$

(and similarly for Σ_{24}).

$$\begin{aligned} \Sigma_{14} &= \text{Cov}(l_1(X), l_4(X)) \\ &= \mathbb{E} \left[\left(2 \int H^-(X, y, \gamma) dF(y) - \frac{\mathbb{1}(X \leq \gamma)}{2f(\gamma)} \right) \left((X - \gamma) \mathbb{1}(X > \gamma) - \frac{\mathbb{1}(X > \gamma)}{2f(\gamma)} \right) \right] + \\ &\quad - \left(2\Delta^- - \frac{1}{4f(\gamma)} \right) \left(\delta^+ - \frac{1}{4f(\gamma)} \right) \\ &= 2\delta^- \delta^+ - \frac{1}{f(\gamma)} \frac{\delta^-}{2} - \left(2\Delta^- - \frac{1}{4f(\gamma)} \right) \left(\delta^+ - \frac{1}{4f(\gamma)} \right) \end{aligned}$$

(and similarly for Σ_{23}). Finally,

$$\begin{aligned}
\Sigma_{34} &= \text{Cov}(I_3(X), I_4(X)) \\
&= E \left[\left((\gamma - X) \mathbb{1}(X \leq \gamma) - \frac{\mathbb{1}(X \leq \gamma)}{2f(\gamma)} \right) \left((X - \gamma) \mathbb{1}(X > \gamma) - \frac{\mathbb{1}(X > \gamma)}{2f(\gamma)} \right) \right] + \\
&\quad - \left(\delta^- - \frac{1}{4f(\gamma)} \right) \left(\delta^+ - \frac{1}{4f(\gamma)} \right) \\
&= - \left(\delta^- - \frac{1}{4f(\gamma)} \right) \left(\delta^+ - \frac{1}{4f(\gamma)} \right).
\end{aligned}$$

Funding Open access funding provided by Università degli Studi di Milano - Bicocca within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Azzalini, A., Capitanio, A.: The Skew-normal and related families. Cambridge University Press, Cambridge (2014)
- Balanda, K.P., MacGillivray, H.L.: Kurtosis: a critical review. *Am Stat* **42**, 111–119 (1988)
- Balanda, K.P., MacGillivray, H.L.: Kurtosis and spread. *Can J Stat* **18**, 17–30 (1990)
- Blest, D.C.: A new measure of kurtosis adjusted for skewness. *Aust N Z J Stat* **45**, 175–179 (2003)
- Carnero, M.A., Peña, D., Ruiz, E.: Persistence and kurtosis in GARCH and stochastic volatility models. *J Financ Econom* **2**, 319–342 (2004)
- Corrado, C., Su, T.: Skewness and kurtosis in S&P 500 index returns implied by option prices. *J Financ Res* **19**, 175–192 (1996)
- Darlington, R.B.: Is kurtosis really “peakedness”? *Am Stat* **24**, 19–22 (1970)
- De Capitani, L., Pasquazzi, L.: Inference for performance measures for financial assets. *Metron* **73**, 73–98 (2015)
- De Capitani, L., Poliscicchio, M.: Some remarks on Zenga’s approach to Kurtosis. *Statistica Applicazioni* **14**, 159–195 (2016)
- Domma, F.: Kurtosis diagram for the Log-Dagum distribution. *Statistica Applicazioni* **2**, 3–23 (2004)
- Finucan, H.M.: A note on kurtosis. *J R Stat Soc* **26**, 111–112 (1964)
- Fiori, A.M.: Kurtosis measures: tradition, contradictions, alternatives. *Statistica Applicazioni* **5**, 185–203 (2007)
- Fiori, A.M.: Measuring kurtosis by right and left inequality orders. *Commun Stat* **37**, 2665–2680 (2008)
- Fiori, A.M., Zenga, M.M.: Karl Pearson and the origin of kurtosis. *Int Stat Rev* **77**, 40–50 (2009)
- Gastwirth, J.: Large sample theory of some measures of income inequality. *Econometrica* **42**, 191–196 (1974)
- Geary, R.C.: Moments of the ratio of the mean deviation to the standard deviation for normal samples. *Biometrika* **28**, 295–307 (1936)

- Hill, B.: A simple general approach to inference about the tail of a distribution. *Ann Stat* **3**, 1163–1174 (1975)
- Hoeffding, W.: A class of statistics with asymptotically normal distribution. *Ann Math Stat* **19**, 293–325 (1948)
- Joanes, D.N., Gill, C.A.: Comparing measures of sample skewness and kurtosis. *J R Stat Soc* **47**, 183–189 (1998)
- Jones, M.C., Rosco, J.F., Pewsey, A.: Skewness-Invariant measures of kurtosis. *Am Stat* **65**, 89–95 (2011)
- Kaplansky, I.: A common error concerning kurtosis. *J Am Stat Assoc* **40**, 259 (1945)
- Martellini, L., Ziemann, V.: Improved estimates of higher-order comoments and implications for portfolio selection. *Rev Financ Stud* **23**, 1467–1502 (2010)
- Ng, W.L., Yau, C.Y.: Test for the existence of finite moments via bootstrap. *J Nonparametr Stat* **30**, 28–48 (2018)
- Pearson, K.: Contributions to the mathematical theory of evolution. *Philos Trans Roy Soc Lond* **A185**, 71–114 (1894)
- Pearson, K.: Skew variation, a rejoinder. *Biometrika* **4**, 169–212 (1905)
- Serfling, R.J.: *Approximation theorems of mathematical statistics*. Wiley (1980)
- Subbotin, M.T.: On the law of frequency of error. *Mat Sb* **31**, 296–301 (1923)
- Zenga, M.: La curtosi. *Statistica* **56**, 87–102 (1996)
- Zenga, M.: Kurtosis. In: Kotz, S., Read, C.B., Balakrishna, N., Vidakovic, B. (eds.) *Encyclopedia of statistical sciences* (second edition). Wiley, New York (2006)
- Zenga, M., Poliscchio, M., Greselin, F.: The variance of the Gini's mean difference and its estimators. *Statistica* **64**, 455–475 (2007)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.