Annals of the Institute of Statistical Mathematics (2022) 74:713–735
https://doi.org/10.1007/s10463-021-00813-3

Empirical tail conditional allocation and its consistency under minimal assumptions

N. V. Gribkova[1,2] · J. Su[3,4] · R. Zitikis[4,5]

Received: 7 January 2021 / Revised: 16 October 2021 / Accepted: 19 October 2021 /
Published online: 26 November 2021

Abstract

Under minimal assumptions, we prove that an empirical estimator of the tail conditional allocation (TCA), also known as the marginal expected shortfall, is consistent. Examples are provided to confirm the minimality of the assumptions. A simulation study illustrates the performance of the estimator in the context of developing confidence intervals for the TCA. The philosophy adopted in the present paper relies on three principles: easiness of practical use, mathematical rigor, and practical justifiability and verifiability of assumptions.

Keywords Tail conditional allocation · Marginal expected shortfall · Inference · Order statistic · Concomitant

1 Introduction

Let (X, Y) be a random pair whose joint cumulative distribution function (cdf) we denote by H. Furthermore, let F and G denote the marginal cdf's of X and Y, respectively. For any $p \in [0, 1)$, the tail conditional allocation is defined by

$$\mathrm{TCA}(p) = \frac{1}{1-p}\mathbb{E}(X\mathbb{1}_{[p,1]}(G(Y))), \tag{1}$$

We are indebted to Chief Editor Hironori Fujisawa, an Associate Editor, and two referees for constructive criticism and generous suggestions. This research has been supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada, and the national research organization Mathematics of Information Technology and Complex Systems (MITACS) of Canada.

The original version of this article was revised to update Reference Citation corrections.

J. Su
jianxi@purdue.edu

Extended author information available on the last page of the article

where $\mathbb{1}_{[p,1]}(t)$ is equal to 1 when $t \in [p, 1]$ and 0 otherwise. It has played a prominent role in a number of optimization and allocation problems in economics, finance, and insurance (e.g., Bauer & Zanjani, 2016; Furman et al., 2021a; Guo et al., 2021; Schechtman et al., 2008; Shalit & Yitzhaki, 1994; Yitzhaki & Schechtman, 2013).

The following examples serve concrete illustrations of where TCA can naturally be used, with X being any of the two specified random variables R_1 and R_2, with $Y = R_1 + R_2$:

- In an auto insurance context, suppose that R_1 is the indemnity payment per insurance claim, and R_2 is the loss adjustment expense per claim (e.g., Frees & Valdez, 1998). Note that in the current insurance practice, it is not uncommon for insurance companies to have the records of a very large number n of claims (e.g., $n = 10^5$, which we shall use in our simulations). Similarly, in home insurance products, R_1 might mean the structural damage per claim, and R_2 the additional living expense per claim (e.g. Insurance Information Institute, 2021). In both cases, Y is the total insurance payment. The TCA can be used to determine the risk driver in the large insurance payment scenario for insurance risk pricing and management purposes.
- In the case of simulated climate scenarios provided by catastrophe software, third-party companies, or regulators, R_1 might mean the total loss of property insurance business line per policy period (e.g., one year) based on a given catastrophe scenario, and R_2 might mean the total loss of life-insurance business line per policy period based on the same catastrophe scenario. In this case, Y would be the total insurance loss. The TCA can be used to determine the insurance portfolio's minimal capital requirement and capital allocation.
- In the case of Economic Scenario Generator (ESG) (Pedersen et al., 2016), R_1 could mean the loss/profit for an investment, and R_2 could mean the loss/profit for another investment, in which case Y would be the total investment loss/profit. The TCA can be used to determine the optimal asset allocation and measure investment performance (e.g., Tasche, 2004).

Coming back to equation (1), when $p = 0$, we obviously have $\mathrm{TCA}(0) = \mathbb{E}(X)$, which is often called the net premium in insurance, with X carrying the meaning of a loss variable. Naturally, throughout the paper we make the following assumption.

Assumption 1 The moment $\mathbb{E}(|X|)$ is finite.

Note the equation

$$\mathrm{TCA}(p) = \frac{1}{1-p}\mathbb{E}(X\mathbb{1}_{[y_p,\infty)}(Y)),$$

where $y_p = G^{-1}(p)$, with the generalized inverse G^{-1}, also called the quantile function, defined by the equation $G^{-1}(p) = \inf\{y \in \mathbb{R} : G(y) \geq p\}$. Usually in the finance and insurance literature, $G^{-1}(p)$ is called the p^{th} value-at-risk of the random variable Y and denoted by $\mathrm{VaR}_p(Y)$. When $X = Y$, $\mathrm{TCA}(p)$ turns into the expected shortfall

$$\mathrm{ES}(p) = \frac{1}{1-p}\mathbb{E}(X\mathbb{1}_{[p,1]}(F(X))) = \frac{1}{1-p}\mathbb{E}(X\mathbb{1}_{[x_p,\infty)}(X)),$$

where $x_p = F^{-1}(p)$. We note that the expected shortfall is also known in the literature under other names, such as conditional tail expectation (CTE), tail conditional expectation (TCE), and tail value at risk (TVaR). We refer to Wang & Zitikis (2020), Embrechts et al. (2021), and Wang et al. (2021) for axiomatic foundations that distinguish the expected shortfall from other risk measures.

Empirical estimation of $\mathbb{E}(X)$ and thus of ES(0) and TCA(0) is of course a well-understood and developed area of classical statistics. In practice, when $\mathrm{ES}(p)$ and $\mathrm{TCA}(p)$ are of importance, the values of $p \in [0, 1)$ are, however, quite close to 1, such as $p = 0.975$ and $p = 0.99$ (BCBS, 2016, 2019). In such cases, statistical inference for $\mathrm{ES}(p)$ and $\mathrm{TCA}(p)$ becomes involved. To see the reason, we start with the expected shortfall $\mathrm{ES}(p)$.

Let $X_1, X_2, \ldots, X_n$ be identically distributed random variables, each with the same cdf F. Let F_n be the empirical cdf based on these random variables, that is,

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{1}\{X_k \le x\} = \frac{1}{n}\sum_{k=1}^{n}\mathbb{1}\{X_{k:n} \le x\},$$

where $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ are the order statistics of $X_1, X_2, \ldots, X_n$ (e.g., Ahsanullah et al., 2013; Arnold et al., 2008; David & Nagaraja, 2003). A non-parametric estimator of $\mathrm{ES}(p)$ can now be defined as follows:

$$\begin{aligned}\mathrm{ES}_n(p) &= \frac{1}{n}\sum_{i=1}^{n} X_i \frac{1}{1-p}\mathbb{1}_{[p,1]}\left(\frac{n}{n+1}F_n(X_i)\right)\\ &= \frac{1}{n}\sum_{i=1}^{n} X_{i:n} \frac{1}{1-p}\mathbb{1}_{[p,1]}\left(\frac{n}{n+1}F_n(X_{i:n})\right)\\ &= \frac{1}{n}\sum_{i=1}^{n} X_{i:n} \frac{1}{1-p}\mathbb{1}_{[p,1]}\left(\frac{i}{n+1}\right),\end{aligned} \tag{2}$$

where to get the final equation, we assumed continuity of the cdf F and thus, without loss of generality, the strict ordering $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$. (In what follows, when analyzing the tail conditional allocation, we shall drop the continuity of F but impose the continuity on the cdf G; see Assumption 2.) The factor $n/(n+1)$ in front of the empirical cdf F_n in the above equations can of course be deleted, but it is useful to keep it there for a number of reasons, not least for the consistency of our following discussion.

Hence, $\mathrm{ES}_n(p)$ is a linear combination of order statistics, which are of course dependent and non-identically distributed random variables, thus explaining the leap in complexity if compared with the non-parametric estimator $\overline{X} = n^{-1}\sum_{i=1}^{n} X_i$ of the mean $\mathbb{E}(X)$. Nevertheless, the case has been successfully tackled even at a much higher level of generality than that on the right-hand side of equation (2). Namely, various statistical properties of the L-statistic

$$L_n(w) = \frac{1}{n}\sum_{i=1}^{n} X_{i:n} w\left(\frac{i}{n+1}\right) \tag{3}$$

have been thoroughly explored in the immense body of the literature, under various sets of conditions and levels of generality. Note that having $n+1$ and not n in the denominator on the right-hand side of equation (3) allows us to use weight functions $w : [0,1] \to \mathbb{R} \cup \{\pm\infty\}$ that take infinite values at any of the two end-points of the interval [0, 1], thus explaining one of the reasons behind the factor $n/(n+1)$ in front of the empirical cdf F_n in equations (2).

The complexity of statistical inference for TCA(p) is even higher, as we shall see in the next section.

2 The main result, its proof and performance

Let each pair $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ follow the same joint cdf H, thus making them independent copies of the earlier introduced pair (X, Y). A natural empirical estimator of $H(x, y)$ is

$$H_n(x, y) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}\{X_k \le x, Y_k \le y\} = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}\{X_{k,n} \le x, Y_{k:n} \le y\},$$

where $Y_{1:n} \le Y_{2:n} \le \cdots \le Y_{n:n}$ are the order statistics of $Y_1, Y_2, \dots, Y_n$, with $X_{1,n}, X_{2,n}, \dots, X_{n,n}$ denoting the corresponding concomitants (e.g., David & Nagaraja, 2003, Section 6.8), which are defined by

$$X_{k,n} = \sum_{i=1}^{n} X_i \mathbb{1}\{Y_i = Y_{k:n}\}$$

for every $k = 1, \dots, n$. The concomitants are uniquely defined when the cdf G is continuous, in which case we have $Y_{1:n} < Y_{2:n} < \cdots < Y_{n:n}$ almost surely. Therefore, from now on, unless noted otherwise, we work under the following assumption.

Assumption 2 The cdf G of Y is a continuous function.

An empirical estimator of TCA(p) can now be defined as follows:

$$\begin{aligned}\mathrm{TCA}_n(p) &= \frac{1}{n}\sum_{i=1}^{n} X_i \frac{1}{1-p}\mathbb{1}_{[p,1]}\left(\frac{n}{n+1}G_n(Y_i)\right)\\ &= \frac{1}{n}\sum_{i=1}^{n} X_{i,n} \frac{1}{1-p}\mathbb{1}_{[p,1]}\left(\frac{n}{n+1}G_n(Y_{i:n})\right)\\ &= \frac{1}{n}\sum_{i=1}^{n} X_{i,n} \frac{1}{1-p}\mathbb{1}_{[p,1]}\left(\frac{i}{n+1}\right),\end{aligned} \tag{4}$$

where G_n is the empirical cdf based on $Y_1, Y_2, \dots, Y_n$. More generally than the above estimator, we might be interested in considering

$$\Lambda_n(w) = \frac{1}{n}\sum_{i=1}^{n} X_{i,n} w\left(\frac{i}{n+1}\right)$$

for various weight functions $w : [0,1] \to \mathbb{R} \cup \{\pm\infty\}$. If we compare $\Lambda_n(w)$ with $L_n(w)$, the appearance of concomitants explains the added level of complexity. Nevertheless, a number of results have appeared in the literature, and we shall discuss them next.

In particular, we see two trends in developing non-parametric statistical inference for

$$\mathrm{TCA}(w) = \mathbb{E}(Xw(G(Y))).$$

One of them is based on the desire to cover as large a class of weight functions w as possible, and then to specify classes of distributions of (X, Y), which often happen to be sub-optimal. For example, Gribkova & Zitikis (2017, 2019), and Dudkina & Gribkova (2020) require higher finite moments than the first moment of X in order to establish weak or strong consistency of estimators. Although not being an impediment in a number of applications, such conditions nevertheless need to be relaxed to accommodate other applications. There are situations when even the first moment is not finite (e.g., Nešlehová et al., 2006).

The other trend in the development of statistical inference for $\mathrm{TCA}(w)$ is based on fixing a weight function w of some particular interest and then seeking for (nearly) optimal classes of distributions that ensure desired inference results, such as consistency or asymptotic normality of $\Lambda_n(w)$. For example, when the weight function is $w(t) = t$ for all $t \in [0,1]$, which gives rise to the Gini covariance and correlation (e.g., Yitzhaki & Schechtman, 2013), statistical inference results have been thoroughly developed by Kattumannil et al. (2020), where extensive references to earlier works can also be found. Departures from iid observations to those driven by time series models with accompanying statistical inference results for the Gini covariance and related quantities can be found in Shelef (2013), Carcea & Serfling (2015), Shelef (2016), and Shelef & Schechtman (2019).

The present paper is in the spirit of the latter trend, albeit in the iid case, and is devoted to establishing consistency of the estimator $\mathrm{TCA}_n(p)$, which is $\Lambda_n(w_p)$ with

$$w_p(t) = \frac{1}{1-p}\mathbb{1}_{[p,1]}(t), \quad t \in [0,1]. \tag{5}$$

This weight function w_p plays a pivotal role when developing ES-based capital allocations (e.g., Furman & Zitikis, 2008) and, by now, has been widely employed when tackling a myriad of insurance- and finance-related problems (e.g., Bauer & Zanjani, 2016; Furman et al., 2021a; Guo et al., 2021, and references therein).

Having thus mentioned insurance and finance, we note that the earlier noted regulatory documents (BCBS, 2016, 2019) stipulate the use of $p = 0.975$ and $p = 0.99$ when using the expected shortfall $\mathrm{ES}(p)$; see Wang & Zitikis (2020) for an in-depth discussion on the matter. We therefore also use these p values in the current context of the tail conditional allocation $\mathrm{TCA}(w_p)$, which is also known in the literature under the name of marginal expected shortfall and written as $\mathrm{MES}(p) = \mathbb{E}(X \mid Y \geq G^{-1}(p))$ (e.g., Cai et al., 2015; Cai & Musta, 2020; Kulik & Tong, 2019). For related results in the case $Y = X$, that is, when dealing with the expected shortfall $\mathrm{ES}(p)$, we refer to Necir et al. (2010), Laidi et al. (2020), and references therein. The aforementioned studies employ the extreme value theory (e.g., Beirlant et al., 2004; Castillo et al., 2004; de Haan & Ferreira, 2006) that facilitates the development of ES and MES estimation even in situations when there are very few observations in the far right-hand tail of the distribution. In particular, in Cai et al. (2015), Eq. (5), and Kulik & Tong (2019), Eq. (11), we find consistency and asymptotic normality of an empirical estimator, which is akin to the herein employed $\mathrm{TCA}_n(p)$.

These state-of-the-art results give rise to asymptotically precise statistical inference results, including minimal length confidence intervals and asymptotically precise coverage proportions. But these achievements come at a heavy price, which is the assumptions inevitably imposed on the underlying population (e.g., Kulik & Tong, 2019, Section 3.3). Although highly natural from the theoretical point of view, the assumptions are challenging to verify in practice and thus give rise to unsurmountable difficulties when convincing practitioners in the worthiness of adopting the results. Clearly, a compromise is warranted, and as we shall argue next, it could be achieved via properly delivered (weak) consistency results under minimal assumptions on the population. In this sense, the following theorem is the main result of the present paper. Its connection to the classical standard-error-based approach, with additional clarifying details and a numerical illustration, will be provided after the theorem has been proved.

Theorem 1 *Let the pairs* $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ *be independent copies of* (X, Y), *and let Assumptions 1 and 2 be satisfied. Then* $\mathrm{TCA}_n(p)$ *is a consistent estimator of* $\mathrm{TCA}(p)$ *for every* $p \in [0, 1)$. *That is, for any pre-specified margin of error* $\epsilon > 0$, *the statement*

$$\mathrm{TCA}(p) \in \left[\mathrm{TCA}_n(p) - \epsilon, \mathrm{TCA}_n(p) + \epsilon\right] \tag{6}$$

holds with the asymptotically perfect (*i.e., equal to* 1) *probability when the sample size n indefinitely increases.*

Proof When $p = 0$, the theorem is obviously true. Hence, from now on, we consider only $p \in (0, 1)$. With a parameter $\delta > 0$, let $w_{p,\delta} : \mathbb{R} \to [0, 1]$ be the function defined by

$$w_{p,\delta}(t) = \begin{cases} 0 & \text{when} \quad t \le p - \delta, \\ \frac{1}{2}\left(1 + \frac{t-p}{\delta} + \frac{1}{\pi}\sin\left(\frac{t-p}{\delta}\pi\right)\right) & \text{when} \quad t \in (p-\delta, p+\delta), \\ 1 & \text{when} \quad t \ge p + \delta. \end{cases}$$

It is a non-decreasing and continuously differentiable function, known in the statistical literature as the raised cosine cdf, whose first derivative (i.e., probability density function) is

$$w'_{p,\delta}(t) = \begin{cases} \frac{1}{2\delta}\left(1 + \cos\left(\frac{t-p}{\delta}\pi\right)\right) & \text{when} \quad t \in (p-\delta, p+\delta), \\ 0 & \text{otherwise.} \end{cases}$$

Obviously, $w'_{p,\delta}(t)$ is always non-negative and never exceeds $1/\delta$. We have

$$\begin{aligned} (1-p)|\mathrm{TCA}_n(p) - \mathrm{TCA}(p)| &\le \left| \frac{1}{n}\sum_{i=1}^n X_{i,n}\mathbb{1}_{[p,1]}\left(\frac{i}{n+1}\right) - \frac{1}{n}\sum_{i=1}^n X_{i,n} w_{p,\delta}\left(\frac{i}{n+1}\right)\right| \\ &\quad + \left| \frac{1}{n}\sum_{i=1}^n X_{i,n} w_{p,\delta}\left(\frac{i}{n+1}\right) - \mathbb{E}(X w_{p,\delta}(G(Y)))\right| \\ &\quad + \left| \mathbb{E}(X w_{p,\delta}(G(Y))) - \mathbb{E}(X\mathbb{1}_{[p,1]}(G(Y)))\right| \\ &=: \alpha_n(\delta) + \beta_n(\delta) + \gamma(\delta). \end{aligned}$$

We shall next show that no matter what $\delta > 0$ is, as long as it does not depend on n, the quantity $\beta_n(\delta)$ converges in probability to zero when $n \to \infty$. We start as follows:

$$\begin{aligned} \beta_n(\delta) &= \left| \frac{1}{n}\sum_{i=1}^n X_{i,n} w_{p,\delta}\left(\frac{n}{n+1}G_n(Y_{i:n})\right) - \mathbb{E}(X w_{p,\delta}(G(Y)))\right| \\ &= \left| \frac{1}{n}\sum_{i=1}^n X_i w_{p,\delta}\left(\frac{n}{n+1}G_n(Y_i)\right) - \mathbb{E}(X w_{p,\delta}(G(Y)))\right| \\ &\le \left| \frac{1}{n}\sum_{i=1}^n X_i w_{p,\delta}\left(\frac{n}{n+1}G_n(Y_i)\right) - \frac{1}{n}\sum_{i=1}^n X_i w_{p,\delta}\big(G(Y_i)\big)\right| \\ &\quad + \left| \frac{1}{n}\sum_{i=1}^n X_i w_{p,\delta}\big(G(Y_i)\big) - \mathbb{E}(X w_{p,\delta}(G(Y)))\right| \\ &=: \beta_n^*(\delta) + \beta^{**}(\delta). \end{aligned} \tag{7}$$

By the law of large numbers, the right-most term $\beta^{**}(\delta)$ converges in probability to zero, and even almost surely, irrespective of the value of $\delta > 0$. As to $\beta_n^*(\delta)$, we utilize the fact that $w'_{p,\delta}(t) \in [0, 1/\delta]$ for all $t \in \mathbb{R}$ and have

$$\begin{aligned}
\beta_n^*(\delta) &\le \frac{1}{\delta n}\sum_{i=1}^n |X_i|\left|\frac{n}{n+1}G_n(Y_i) - G(Y_i)\right| \\
&= \frac{1}{\delta n}\sum_{i=1}^n |X_i|\left|\frac{-1}{n+1}G_n(Y_i) + G_n(Y_i) - G(Y_i)\right| \\
&\le \frac{1}{\delta n}\sum_{i=1}^n |X_i|\left(\frac{1}{n+1} + \sup_{y\in\mathbb{R}}|G_n(y) - G(y)|\right) \qquad (8)\\
&= \frac{1}{\delta}\left(\frac{1}{n}\sum_{i=1}^n |X_i| - \mathbb{E}(|X|)\right)\left(\frac{1}{n+1} + \sup_{y\in\mathbb{R}}|G_n(y) - G(y)|\right) \\
&\quad + \frac{\mathbb{E}(|X|)}{\delta}\left(\frac{1}{n+1} + \sup_{y\in\mathbb{R}}|G_n(y) - G(y)|\right).
\end{aligned}$$

By the Glivenko-Cantelli theorem, $\sup_{y\in\mathbb{R}}|G_n(y) - G(y)|$ converges in probability to zero, and even almost surely. By the law of large numbers, $n^{-1}\sum_{i=1}^n |X_i| - \mathbb{E}(|X|)$ converges in probability to zero, and even almost surely. Hence, no matter what $\delta > 0$ is, the entire right-hand side of bound (8) converges in probability to zero when $n \to \infty$, and so does $\beta_n(\delta)$ in view of bound (7).

Consequently, to complete the proof of Theorem 1, we need to show that no matter what $\varepsilon > 0$ is, we can choose $\delta = \delta(\varepsilon) > 0$ such that

$$\mathbb{P}\big(\alpha_n(\delta) + \gamma(\delta) \ge \varepsilon\big) \to 0 \qquad (9)$$

when $n \to \infty$. We start with the bounds

$$\begin{aligned}
\gamma(\delta) &\le \mathbb{E}\big(|X||w_{p,\delta}(G(Y)) - \mathbb{1}_{[p,1]}(G(Y))|\big) \\
&\le \mathbb{E}\big(|X|\mathbb{1}_{[p-\delta,p+\delta]}(G(Y))\big).
\end{aligned}$$

The expectation on the right-hand side converges to 0 when $\delta \downarrow 0$ because $\mathbb{E}(|X|) < \infty$ and $\mathbb{P}((G(Y) \in [p-\delta, p+\delta]) = 2\delta \downarrow 0$ when $\delta \downarrow 0$. (Note that continuity of the cdf G implies that $G(Y)$ follows the uniform on [0, 1] distribution.) Hence, there is $\delta^* = \delta^*(\varepsilon) > 0$ such that

$$\gamma(\delta) \le \frac{\varepsilon}{2}$$

for every $\delta \in (0, \delta^*]$. Consequently, from now on we work with only such δ's, and thus statement (9) holds provided that we show the convergence

$$\mathbb{P}\left(\alpha_n(\delta) \ge \frac{\varepsilon}{2}\right) \to 0 \qquad (10)$$

when $n \to \infty$. That is, we need to show that for every $\varepsilon^* > 0$, there is n^* such that

$$\mathbb{P}\left(\alpha_n(\delta) \ge \frac{\varepsilon}{2}\right) \le \varepsilon^* \qquad (11)$$

for all $n \geq n^*$. We start with the bounds

$$\begin{aligned}
\alpha_n(\delta) &\leq \frac{1}{n}\sum_{i=1}^{n}|X_{i,n}|\left|\mathbb{1}_{[p,1]}\left(\frac{i}{n+1}\right)-w_{p,\delta}\left(\frac{i}{n+1}\right)\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}|X_{i,n}|\mathbb{1}_{[p-\delta,p+\delta]}\left(\frac{i}{n+1}\right) \\
&\leq \frac{1}{n}\sum_{i=1}^{n}|X_{i,n}|\mathbb{1}_{[0,M]}(|X_{i,n}|)\mathbb{1}_{[p-\delta,p+\delta]}\left(\frac{i}{n+1}\right)+\frac{1}{n}\sum_{i=1}^{n}|X_{i,n}|\mathbb{1}_{(M,\infty)}(|X_{i,n}|),
\end{aligned}$$

where $M \geq 0$ can be arbitrary at the moment but it will soon need to be chosen sufficiently large. We have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}|X_{i,n}|\mathbb{1}_{[0,M]}(|X_{i,n}|)\mathbb{1}_{[p-\delta,p+\delta]}\left(\frac{i}{n+1}\right) &\leq \frac{M}{n}\sum_{i=1}^{n}\mathbb{1}_{[p-\delta,p+\delta]}\left(\frac{i}{n+1}\right) \\
&\leq \frac{M}{n}\big(2\delta(n+1)+1\big).
\end{aligned} \tag{12}$$

We see that no matter what $\varepsilon > 0$ and $M < \infty$ are, we can always choose $\delta \in (0, \delta^*]$ so small that the right-hand side of bound (12) becomes smaller than $\varepsilon/4$ for all sufficiently large n, that is, for all $n \geq n_0$ for some n_0. This implies statement (11) provided that

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}|X_{i,n}|\mathbb{1}_{(M,\infty)}(|X_{i,n}|) \geq \frac{\varepsilon}{4}\right) \leq \varepsilon^* \tag{13}$$

for all $n \geq n^*$, where n^* is an integer not smaller than n_0. Using Markov's inequality, we have

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}|X_{i,n}|\mathbb{1}_{(M,\infty)}(|X_{i,n}|) \geq \frac{\varepsilon}{4}\right) &\leq \frac{4}{n\varepsilon}\sum_{i=1}^{n}\mathbb{E}\big(|X_{i,n}|\mathbb{1}_{(M,\infty)}(|X_{i,n}|)\big) \\
&= \frac{4}{n\varepsilon}\sum_{i=1}^{n}\mathbb{E}\big(|X_i|\mathbb{1}_{(M,\infty)}(|X_i|)\big) \\
&= \frac{4}{\varepsilon}\mathbb{E}\big(|X|\mathbb{1}_{(M,\infty)}(|X|)\big).
\end{aligned}$$

Since $\mathbb{E}(|X|) < \infty$, the right-hand side of the above bound can be made smaller than ε^* by choosing a sufficiently large $M = M(\varepsilon, \varepsilon^*)$. This completes the proof of Theorem 1. □

We next discuss several aspects related to the use of $\mathrm{TCA}_n(p)$ in practice. Specifically, in Section 2.1 we provide a large sample analysis of $\mathrm{TCA}(p)$ within the context of ESG (Pedersen et al., 2016), which gives the risk analyst the ability to generate practice-relevant samples of as large sizes as desired. In Section 2.2 we deal with the traditional case of fixed-size samples, although with a non-traditional but highly practical twist: we are interested in confidence intervals that have fixed margins of error and increasing (when n grows) coverage probabilities.

2.1 Statistical analysis under unlimited sample sizes

Consider a scenario in which the risk analyst has the ability to generate data of any sample size. This situation is not uncommon in the finance and insurance practice, due to the aforementioned ESG. The data generating process that underlines the ESG is, however, rather complex and therefore, admittedly, can be viewed as unknown. In this case, the estimator $\mathrm{TCA}_n(p)$ can be used to determine the risk capital and its allocation to individual business lines, since the sample size n can be made as large as needed to obtain an accurate estimate of $\mathrm{TCA}(p)$.

To imitate such a process, instead of using the complex ESG, we consider an insurance portfolio consisting of two dependent risks R_1 and R_2 that follow Mardia's bivariate Pareto distribution (Mardia, 1962). Its joint survival function is given by the formula

$$S(r_1, r_2) = \left(1 + r_1/\theta_1 + r_2/\theta_2\right)^{-\alpha}, \qquad r_1 > 0,\ r_2 > 0, \tag{14}$$

where $\theta_1 > 0$ and $\theta_2 > 0$ are the scale parameters, and $\alpha > 0$ is the shape parameter. It is one of the most popular multivariate Pareto distributions (Arnold et al., 1999; Arnold, 2015) that has recently been shown to be particularly useful when modelling insurance and financial risks under the background risk model (Asimit et al., 2016; Furman et al., 2021b). Both R_1 and R_2 have finite first moments when $\alpha > 1$ and finite second moments only when $\alpha > 2$.

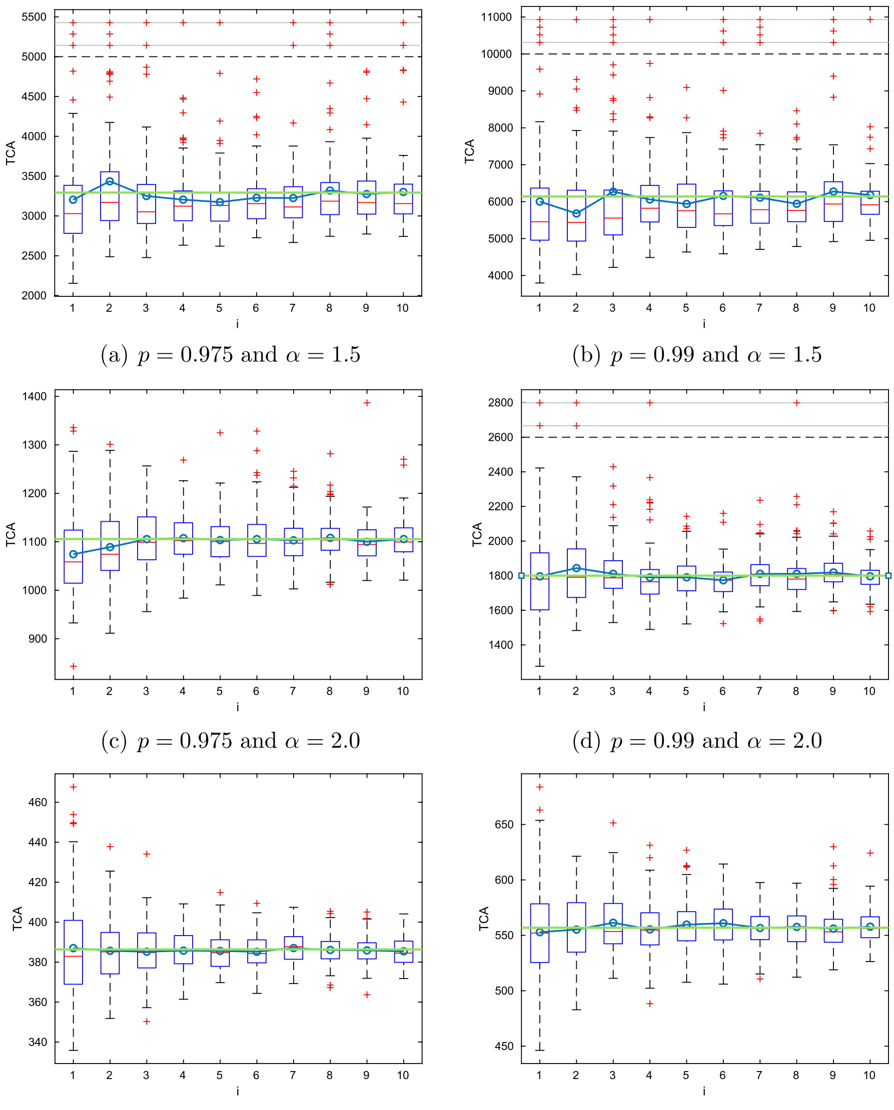
The associated with the pair (R_1, R_2) aggregate risk is $Y = R_1 + R_2$, and suppose that we are interested in $\mathrm{TCA}(p)$ when $X = R_1$. Hence, we simulate n independent pairs $(r_{1,k}, r_{2,k})_{k=1}^{n}$ based on distribution (14), from which we obtain the pairs

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \tag{15}$$

via the equations $x_k = r_{1,k}$ and $y_k = r_{1,k} + r_{2,k}$. We make the following parameter choices:

(A.1) $\theta_1 = 100$, $\theta_2 = 50$, and $\alpha \in \{1.5, 2.0, 3.0\}$ as motivated by Furman et al. (2021b);
(A.2) $p = 0.975$ and 0.99 as stipulated by BCBS (2016, 2019);
(A.3) $n = i \times 10,000$ when $i = 1, \dots, 10$.

We obtain one value of $\mathrm{TCA}_n(p)$ in each simulation for every given n. Then, for each n, we repeat the procedure 100 times. Figure 1 depicts the box plots of the obtained values of $\mathrm{TCA}_n(p)$. We observe that the averaged values of the TCA estimates follow closely the corresponding theoretical TCA values, which have been computed using a matrix-analytic method described by Furman et al. (2021b).



(a) $p = 0.975$ and $\alpha = 1.5$

(b) $p = 0.99$ and $\alpha = 1.5$

(c) $p = 0.975$ and $\alpha = 2.0$

(d) $p = 0.99$ and $\alpha = 2.0$

(e) $p = 0.975$ and $\alpha = 3.0$

(f) $p = 0.99$ and $\alpha = 3.0$

Fig. 1 Box plots of TCA estimates $\mathrm{TCA}_n(p)$ for $n = i \times 10,000$ when $i = 1, \dots, 10$, with their averaged (blue lines with circles) and theoretical (horizontal green lines) values; the parameter α refers to distribution (14)

2.2 Statistical analysis under limited sample sizes

In this section, we consider another practical situation in which there is no possibility to generate data sets of arbitrary size, and thus the risk analyst can only estimate the TCA based on one set of historical data. In this case, we resort to bootstrap methodology in order to understand the performance of the estimator.

Specifically, we assume the availability of $n = 10^5$ observed pairs (15), which we call the parent data set, which we simulate from distribution (14) with parameters (A.1). Then we use bootstrap to re-sample different samples of size $m \in \{10\,000, 20\,000, \ldots, 100\,000\}$, which mimic our analysis in Section 2.1.

To facilitate readability, the rest of the current section is divided into three parts: Sect. 2.2.1 develops understanding of the standard errors of $\mathrm{TCA}_n(p)$, which can be used, e.g., for constructing shrinking-margin-of-error and fixed-coverage-probability confidence intervals for the population TCA. Sect. 2.2.2 transitions our (traditional) thinking into fixed-margin-of-error and increasing-coverage-probability confidence intervals, which are discussed in Sect. 2.2.3.

2.2.1 Estimated standard errors

The standard error of $\mathrm{TCA}_n(p)$ can conveniently and speedily be assessed using bootstrap (e.g., Efron & Tibshirani, 1993; Davison & Hinkley, 1997; DasGupta, 2008; Hall, 1992; Shao & Tu, 1995):

$$\widehat{\mathrm{se}}_{n,m,\mathrm{boot}} = \sqrt{s^2_{n,m,\mathrm{boot}}}\,, \tag{16}$$

where

$$s^2_{n,m,\mathrm{boot}} = \frac{1}{L}\sum_{\ell=1}^{L}\left(\mathrm{TCA}^*_{n,m,\ell}(p) - \frac{1}{L}\sum_{i=1}^{L}\mathrm{TCA}^*_{n,m,i}(p)\right)^2$$

with $\mathrm{TCA}^*_{n,m,\ell}(p)$ obtained as follows: Given pairs (15), we select (with replacement) m pairs for various choices of the parameter m. Given the just obtained pairs $(x^*_1, y^*_1), (x^*_2, y^*_2), \ldots, (x^*_m, y^*_m)$, we calculate the empirical TCA using formula (4) and denote the obtained value by $\mathrm{TCA}^*_{n,m,1}(p)$. We repeat the procedure L times and obtain $\mathrm{TCA}^*_{n,m,1}(p), \ldots, \mathrm{TCA}^*_{n,m,L}(p)$, which we use to calculate $s^2_{n,m,\mathrm{boot}}$.

As an illustration, we use Mardia's bivariate Pareto distribution (14) and from it arising pairs (15). We use the same parameters θ_1, θ_2, α and p as in specifications (A.1) and (A.2), with the other specifications as follows:

(B.1) $n = 100,000$, which reflects what we have observed in real-world insurance data;
(B.2) $m = i \times 10,000$ with $i = 1, \ldots, 10$ (we shall discuss "$m < n$ vs $m = n$" in Sect. 2.2.3);
(B.3) $L = 5,000$.

Table 1 summarizes simulation results in the case of one parent data set, which we call I, whose size is $n = 10^5$, simulated from distribution (14) with parameters (A.1).

Table 1 Estimated standard errors based on the parent data set I, with α referring to distribution (14)

	$\alpha = 1.5$		$\alpha = 2$		$\alpha = 3$	
	p =97.5%	p =99%	p =97.5%	p =99%	p =97.5%	p =99%
$m = 1 \times 10^4$	496	1230	70.8	167	11.2	22.7
$m = 2 \times 10^4$	319	788	49	115	8	16.1
$m = 3 \times 10^4$	254	626	40	93.7	6.5	13.1
$m = 4 \times 10^4$	220	542	34.3	80.5	5.6	11.3
$m = 5 \times 10^4$	195	479	31	72.4	6	9.9
$m = 6 \times 10^4$	178	438	28.4	66.8	4.5	9
$m = 7 \times 10^4$	165	408	26.3	62.1	4.2	8.4
$m = 8 \times 10^4$	154	379	24.5	57.6	3.9	7.9
$m = 9 \times 10^4$	145	357	23.2	54.4	3.7	7.4
$m = n = 10^5$	136	335	21.8	51.1	3.5	7
Theoretical TCA	3290	6140	1100	1800	386	557

To show the effect of parent-data variability, we have also produced another parent data set, called II, using the same data generating process. The simulation results are summarized in Table 2.

The theoretical TCA values reported in the bottom rows of the two tables have been calculated using a matrix-analytic method discussed by Furman et al. (2021b).

Note 1 Looking at Tables 1 and 2, the monotonicity of values in each column would suggest that bootstrap works even when $\alpha = 1.5$, in which case X, being equal to R_1, lacks finite second moment, and thus the TCA estimator lacks finite asymptotic variance. This may not be an issue when working with data and bootstrap, because large

Table 2 Estimated standard errors based on the parent data set II, with α referring to distribution (14)

	$\alpha = 1.5$		$\alpha = 2$		$\alpha = 3$	
	p =97.5%	p =99%	p =97.5%	p =99%	p =97.5%	p =99%
$m = 1 \times 10^4$	420	824	85.7	191	20	40.3
$m = 2 \times 10^4$	296	602	60.1	135	13.9	28.7
$m = 3 \times 10^4$	241	481	49.1	110	11.4	23.2
$m = 4 \times 10^4$	208	413	43	95.2	9.9	20.3
$m = 5 \times 10^4$	187	386	39	85.2	8.9	17.7
$m = 6 \times 10^4$	174	343	34.3	78	8.1	16.4
$m = 7 \times 10^4$	161	317	31.8	72.1	7.5	15.2
$m = 8 \times 10^4$	149	303	30.3	67	7.1	14.3
$m = 9 \times 10^4$	139	279	28.3	64.1	6.6	13.3
$m = n = 10^5$	136	268	21.8	60.4	6.3	12.8
Theoretical TCA	3290	6140	1110	1800	386	557

deviations of the estimator from the true TCA value may compensate increasing values of the variance estimate. Nevertheless, to rigorously explore this phenomenon from the theoretical point of view is a worthwhile problem, well beyond the scope of the present paper.

2.2.2 A pivotal practical consideration

From the traditional point of view of looking at confidence intervals, the standard error determines the shrinking to 0 (when n grows) margin of error while asymptotically maintaining the pre-specified level of confidence. The view that we take next is in a sense the opposite: the confidence interval has a pre-specified (fixed) margin of error ϵ and an asymptotically growing to 1 coverage probability

$$C_n := \mathbb{P}\big(|\mathrm{TCA}_n(p) - \mathrm{TCA}(p)| \le \epsilon\big).$$

The established consistency result ensures that $C_n \to 1$ when $n \to \infty$. Note that the value of ϵ may not be small. It could, for example, be a certain percentage (e.g., 10%) of the average TCA values obtained beforehand from historical data, and such values can be, or at least look, large.

To assess the rate of convergence of C_n to 1, lower bounds for C_n can be established using, e.g., Markov's bound, or—assuming finite second moment—Chebyshev's bound

$$C_n \ge 1 - \frac{1}{\epsilon^2}\mathbb{E}\big((\mathrm{TCA}_n(p) - \mathrm{TCA}(p))^2\big), \tag{17}$$

or perhaps using some other bound such as Chernoff's, which would require finiteness of the moment generating function. The verification can also be done in an asymptotically precise manner—although naturally under more stringent conditions—via the CLT-type result

$$C_n = \mathbb{P}\left(\chi_1^2 \le \frac{n\epsilon^2}{\sigma_{\mathrm{TCA}}^2}\right) + o(1) \tag{18}$$

when $n \to \infty$, where χ_1^2 is the χ^2-random variable with 1 degree of freedom, and σ_{TCA}^2 is the asymptotic variance of $\mathrm{TCA}_n(p)$ whose magnitude we assessed in Sect. 2.2.1 using bootstrap.

The validity of such results requires considerably stronger assumptions than those of the present paper (e.g., Gribkova & Zitikis, 2017, 2019). For example, statements (17) and (18) require at least finite second moment of X. The extreme value theory (EVT) would enable us to circumvent the need for finite second or larger-order moments, but it relies on a plethora of other conditions whose validity might be challenging to establish in practice (e.g., Beirlant et al., 2004; Castillo et al., 2004; de Haan & Ferreira, 2006). Hence, we next adopt a practical bootstrap-based approach for assessing convergence of the error probabilities

$$\pi_n(p, \epsilon) = 1 - C_n$$

to 0. The approach is akin to the classical bootstrap approach for assessing standard errors, and it will hint at the sample sizes n needed to achieve prescribed confidence levels in practice.

2.2.3 Estimated error probabilities

Given pairs (15), we first calculate $\mathrm{TCA}_n(p)$ using formula (4), and then obtain the bootstrap values $\mathrm{TCA}^*_{n,m,1}(p), \dots, \mathrm{TCA}^*_{n,m,L}(p)$ following the procedure described after definition (16). Using these values, we then calculate the proportion

$$\pi^*_{n,m}(p,\epsilon) := \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{1}\left\{ \left|\mathrm{TCA}^*_{n,m,\ell}(p) - \mathrm{TCA}_n(p)\right| > \epsilon \right\}.$$

Asymptotically, this is a proxy to $\pi_m(p,\epsilon)$, which tends to be larger than $\pi_n(p,\epsilon)$ due to $n \geq m$ and the established consistency of the estimator. In other words, $\pi^*_{n,m}(p,\epsilon)$ asymptotically gives a conservative upper bound for the probability $\pi_n(p,\epsilon)$, which we do not want to exceed the prescribed significance level α, say 0.05. Hence, if we find m such that $\pi^*_{n,m}(p,\epsilon)$ dips below α, then we conclude that for the given sample size n, the probability $\pi_n(p,\epsilon)$ also dips below α, and thus the coverage probability C_n becomes (asymptotically) at least $1-\alpha$.

If, however, we cannot find m for which $\pi^*_{n,m}(p,\epsilon)$ dips below α, then we conclude that the sample size n is not sufficiently large to suggest the coverage probability of at least $1-\alpha$, and thus additional data need to be obtained.

To illustrate the proximity of $\pi^*_{n,m}(p,\epsilon)$ to 0 depending on m, we use the same experiment as described earlier under parameter specifications (A.1)–(A.2) and (B.1)–(B.3). We set ϵ to be 10% of the corresponding theoretical $\mathrm{TCA}(p)$ value, which is unknown in practice but, of course, is known in our experiment. Tables 3

Table 3 Error proportions (in %) based on the parent data set I, with α referring to distribution (14)

	$\alpha = 1.5$		$\alpha = 2$		$\alpha = 3$	
	p =97.5%	p =99%	p =97.5%	p =99%	p =97.5%	p =99%
$m = 1 \times 10^4$	56.8	67	30.7	48.2	3.8	14
$m = 2 \times 10^4$	48.6	60.3	15.6	33.3	0.4	3.7
$m = 3 \times 10^4$	39.9	54	8.7	23.4	0	1
$m = 4 \times 10^4$	33.8	48.3	5	16.5	0	0.4
$m = 5 \times 10^4$	28.2	42.6	3.1	12.3	0	0
$m = 6 \times 10^4$	24.4	38.7	2	9.2	0	0
$m = 7 \times 10^4$	21.2	35.4	1.3	7.2	0	0
$m = 8 \times 10^4$	17.8	31.8	0.9	5.7	0	0
$m = 9 \times 10^4$	15	28.6	0.6	4.7	0	0
$m = n = 10^5$	13	25.6	0.3	3.5	0	0

Table 4 Error proportions (in %) based on the parent data set II, with α referring to distribution (14)

	$\alpha = 1.5$		$\alpha = 2$		$\alpha = 3$	
	p =97.5%	p =99%	p =97.5%	p =99%	p =97.5%	p =99%
$m = 1 \times 10^4$	43.1	46.1	19.2	34.1	4.8	16.6
$m = 2 \times 10^4$	25.8	30.8	6.6	17.6	0.6	4.9
$m = 3 \times 10^4$	16.3	20.2	2.2	9.7	0.1	1.8
$m = 4 \times 10^4$	11.1	13.6	1	6	0	0.7
$m = 5 \times 10^4$	7.5	11	0.5	3.5	0	0.3
$m = 6 \times 10^4$	5.3	7.5	0.1	2.1	0	0.1
$m = 7 \times 10^4$	4.2	4.8	0	1.4	0	0.1
$m = 8 \times 10^4$	2.7	3.9	0	0.7	0	0
$m = 9 \times 10^4$	2.2	2.8	0	0.4	0	0
$m = n = 10^5$	1.3	2.3	0	0.2	0	0

and 4 summarize the results in the case of the same parent data sets I and II that we used to produce Tables 1 and 2. A few notes conclude this section.

Note 2 It is natural to ask why $m < n$ could be better than $m = n$, given that $\pi_n(p, \epsilon)$ is asymptotically dominated by $\pi_m(p, \epsilon)$ whenever $m < n$. The reason is that the "m out of n" bootstrap tends to outperform the "n out of n" bootstrap when the latter lacks consistency. For details, we refer to, e.g., Bickel et al. (1997), Bickel & Sakov (2008), DasGupta (2008), Gribkova & Helmers (2007, 2011). Hence, the two competing forces at play in the current context can be conciliated by appropriately choosing m, and this leads to the above suggested recipe for deciding whether or not the sample size n of the original data set is sufficiently large. Nevertheless, we do not see $m < n$ outperforming $m = n$ in Tables 1 and 2, which, in our opinion, suggests that bootstrap is consistent. Establishing this fact rigorously, however, presents yet another interesting theoretical problem, well beyond the scope of the present paper.

Note 3 By comparing Tables 3 and 4, we see that there are more entries below 5% in the latter table than in the former, and thus Table 4 is more likely to suggest that the sample size $n = 10^5$ is sufficient for statistical purposes than Table 3. This is of course natural, given the inevitable data variability, but it reminds us that some data sets are more telling about the population than other ones.

3 The role of continuity

The continuity assumption on the cdf's—whether F or G depending on the context—has been natural throughout the above considerations, but this naturalness does not serve a rigorous basis for claiming that the assumption is truly necessary for consistency. For this reason, the current section is devoted to a detailed analysis of a discrete case. Naturally, to show that consistency fails for the tail conditional allocation, it is sufficient to show that it fails for the expected shortfall because the latter is a special case of the former: we just need to set $Y = X$. Hence, we concentrate on the latter risk measure, dropping the factor $1/(1-p)$ for typographical simplicity, as it does not change any of the following arguments.

Fix any $p \in (0,1)$, and let the distribution of X be

$$X = \begin{cases} x_1 & \text{with probability} \quad \alpha \in (0,1), \\ x_2 & \text{otherwise,} \end{cases}$$

for some $x_1 < x_2$ be chosen later. Then

$$L(p) := \mathbb{E}(X \mathbb{1}_{[p,1]}(F(X))) = x_1 \alpha \mathbb{1}_{[p,1]}(\alpha) + x_2(1-\alpha).$$

Note that in the discontinuous case, such as the one we consider now, equations (2) give two empirical "estimators" of $L(p)$: one is in the bottom (third) line, and another one is in any of the two preceding lines. For convenience, we denote them as

$$L_n(p) := \frac{1}{n} \sum_{i=1}^{n} X_{i:n} \mathbb{1}_{[p,1]}\left(\frac{i}{n+1}\right)$$

and

$$L_n^*(p) := \frac{1}{n} \sum_{i=1}^{n} X_i \mathbb{1}_{[p,1]}\left(\frac{n}{n+1} F_n(X_i)\right),$$

respectively. The rest of the section is subdivided into two parts.

3.1 Failure of $L_n(p)$

Let $x_1 = 0$ and $x_2 = 1$. Then

$$L(p) = 1 - \alpha.$$

Furthermore, we have

$$L_n(p) = \min\left\{ \frac{1}{n} \sum_{i=1}^{n} X_i, \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{[p,1]}\left(\frac{i}{n+1}\right) \right\} \xrightarrow{\mathbb{P}} \min\{1-\alpha, 1-p\}$$

because $n^{-1} \sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} 1 - \alpha$ by the law of large numbers. Hence, when $\alpha < p$, the "estimator" $L_n(p)$ fails to consistently estimate $L(p)$.

3.2 Failure of $L_n^*(p)$

Let $x_1 = 1$ and $x_2 = 2$. Then

$$L(p) = \begin{cases} \alpha + 2(1-\alpha) & \text{when} \quad \alpha \geq p, \\ 2(1-\alpha) & \text{when} \quad \alpha < p. \end{cases}$$

Furthermore, we have

$$\begin{aligned} L_n^*(p) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{1\}}(X_i)\mathbb{1}_{[p,1]}\left(\frac{n}{n+1}F_n(1)\right) + 2\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{2\}}(X_i)\mathbb{1}_{[p,1]}\left(\frac{n}{n+1}F_n(2)\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{1\}}(X_i)\mathbb{1}_{[p,1]}\left(\frac{n}{n+1}F_n(1)\right) + 2\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{2\}}(X_i)\mathbb{1}_{[p,1]}\left(\frac{n}{n+1}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{1\}}(X_i)\mathbb{1}_{[p,1]}\left(\frac{n}{n+1}F_n(1)\right) + 2\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{2\}}(X_i) \end{aligned}$$

for all sufficiently large n, so that $n/(n+1) \geq p \in (0,1)$. Hence,

$$\begin{aligned} L_n^*(p) &= F_n(1)\mathbb{1}_{[p,1]}\left(\frac{n}{n+1}F_n(1)\right) + 2(1-F_n(1)) \\ &= \alpha\mathbb{1}_{[p,1]}\left(\frac{n}{n+1}F_n(1)\right) + 2(1-\alpha) + o_{\mathbb{P}}(1) \end{aligned}$$

when $n \to \infty$, because $F_n(1) \xrightarrow{\mathbb{P}} \alpha$ by the law of large numbers. This implies that

$$L_n^*(p) \xrightarrow{\mathbb{P}} L(p)$$

if and only if

$$\mathbb{1}_{[p,1]}\left(\frac{n}{n+1}F_n(1)\right) \xrightarrow{\mathbb{P}} \begin{cases} 1 & \text{when} \quad \alpha \geq p, \\ 0 & \text{when} \quad \alpha < p. \end{cases} \tag{19}$$

We shall next check whether of not the latter statement holds.

When $\alpha < p$, for every $\varepsilon \in (0,1)$ we have

$$\begin{aligned} \mathbb{P}\left(\mathbb{1}_{[p,1]}\left(\frac{n}{n+1}F_n(1)\right) > \varepsilon\right) &= \mathbb{P}\left(\frac{n}{n+1}F_n(1) \geq p\right) \\ &= \mathbb{P}\left(\frac{n}{n+1}(F_n(1)-\alpha) - \frac{\alpha}{n+1} \geq p - \alpha\right) \to 0 \end{aligned}$$

when $n \to \infty$, because $p - \alpha > 0$ and $F_n(1) - \alpha \xrightarrow{\mathbb{P}} 0$ by the law of large numbers.

When $\alpha \geq p$, for every $\varepsilon \in (0,1)$ we have

$$\begin{aligned} \mathbb{P}\left(\left|\mathbb{1}_{[p,1]}\left(\frac{n}{n+1}F_n(1)\right) - 1\right| > \varepsilon\right) &= \mathbb{P}\left(\frac{n}{n+1}F_n(1) < p\right) \\ &= \mathbb{P}\left(\frac{n}{n+1}(F_n(1)-\alpha) - \frac{\alpha}{n+1} < p - \alpha\right). \end{aligned}$$

Hence, when $\alpha > p$, the right-hand side converges to 0 when $n \to \infty$, because $p - \alpha < 0$ and $F_n(1) - \alpha \xrightarrow{\mathbb{P}} 0$ by the law of large numbers.

When $\alpha = p$, the above equations give

$$\begin{aligned}\mathbb{P}\left(\left|\mathbb{1}_{[p,1]}\left(\frac{n}{n+1}F_n(1)\right) - 1\right| > \varepsilon\right) &= \mathbb{P}\left(\frac{n}{n+1}(F_n(1) - \alpha) < \frac{\alpha}{n+1}\right)\\ &= \mathbb{P}\left(\frac{\sqrt{n}(F_n(1) - \alpha)}{\sqrt{\alpha(1-\alpha)}} < \frac{\sqrt{\alpha}}{\sqrt{n}\sqrt{1-\alpha}}\right)\\ &\geq \mathbb{P}\left(\frac{\sqrt{n}(F_n(1) - \alpha)}{\sqrt{\alpha(1-\alpha)}} \leq 0\right)\\ &= \frac{1}{2} + o(1)\end{aligned}$$

when $n \to \infty$, due to the central limit theorem and the symmetry of the standard Gaussian distribution. Hence, when $\alpha = p$, statement (19) fails, which in turn implies that the "estimator" $L_n^*(p)$ fails to consistently estimate $L(p)$.

4 Conclusion

In response to a concluding remark by Dudkina & Gribkova (2020), we have shown under minimal assumptions that the empirical estimator of the tail conditional allocation $\mathrm{TCA}(p)$ is (weakly) consistent. This, however, only partially solves the problem of Dudkina & Gribkova (2020), as they consider a *strong* law of large numbers and thus obviously wish a *strong* consistency result, which we next formulate as a conjecture.

Conjecture 1 *Let Assumptions 1 and 2 be satisfied. Then* $\mathrm{TCA}_n(p)$ *is a strongly consistent estimator of* $\mathrm{TCA}(p)$ *for every* $p \in [0, 1)$.

The minimality of Assumptions 1 and 2 is, of course, meant in some general sense. For example, the necessity of the finite first moment $\mathbb{E}(X)$ can, irrespective of $p \in [0, 1)$, be concluded by considering independent X and Y, in which case we have

$$\mathrm{TCA}(p) = \frac{1}{1-p}\mathbb{E}(X)\mathbb{E}(\mathbb{1}_{[p,1]}(G(Y))) = \mathbb{E}(X),$$

due to the equation $\mathbb{E}(\mathbb{1}_{[p,1]}(G(Y))) = 1 - p$ that follows from the continuity of the cdf G and thus from the uniformity of the distribution of $G(Y)$ on the interval $[0, 1]$. This establishes the necessity of $\mathbb{E}(X) \in \mathbb{R}$. Nevertheless, the other extreme case (in the sense of dependence) $X = Y$ gives the equations

$$\mathrm{TCA}(p) = \frac{1}{1-p}\mathbb{E}(Y\mathbb{1}_{[p,1]}(G(Y))) = \frac{1}{1-p}\int_p^1 G^{-1}(u)\mathrm{d}u,$$

thus suggesting that only $\mathbb{E}(Y_+) \in \mathbb{R}$ (or, equivalently, $\mathbb{E}(X_+) \in \mathbb{R}$) is necessary. Hence, next is our second conjecture.

Conjecture 2 *Let Assumption 2 be satisfied. Given any* $p \in [0, 1)$, *if the moment* $\mathbb{E}(X\mathbb{1}_{[p,1]}(G(Y)))$ *is finite, then* $\mathrm{TCA}_n(p)$ *is a (weakly or strongly) consistent estimator of* $\mathrm{TCA}(p)$.

Solving these conjectures are primarily of theoretical interest. The two bootstrap-related problems that we raised in Notes 1 and 2 are of practical value. Given our intuition for the technicalities needed to attack these problems, their solutions will require considerable space to establish. That said, we can give the following expression

$$\sigma_p^2 = \frac{1}{(1-p)^2}\left\{ \int_p^1 g_2(G^{-1}(t))\mathrm{d}t - \left(\int_p^1 g(G^{-1}(t))\mathrm{d}t \right)^2 \right.$$
$$\left. + \gamma_p^2 p(1-p) - 2p\gamma_p \int_p^1 g(G^{-1}(t))\mathrm{d}t \right\}$$

of the asymptotic variance of the estimator $\mathrm{TCA}_n(p)$, where $g_2(y) = \mathbb{E}\big(X^2 \mid Y = y\big)$, $g(y) = \mathbb{E}\big(X \mid Y = y\big)$, and $\gamma_p = g(G^{-1}(p))$.

Conjecture 3 *Let Assumption 2 be satisfied. If the asymptotic variance* σ_p^2 *is finite, then*

$$\sqrt{n}\big(\mathrm{TCA}_n(p) - \mathrm{TCA}(p)\big) \xrightarrow{d} \mathcal{N}(0, \sigma_p^2),$$

where $\xrightarrow{d}$ *denotes convergence in distribution.*

To compare the performance of various estimation procedures based on confidence intervals with prescribed margins of error under minimal assumptions will be a very interesting problem for future research, but for doing so, we first need to establish Conjecture 3 in the stated or a modified form, depending on the inevitability of technicalities.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

References

Ahsanullah, M., Nevzorov, V. B., Shakil, M. (2013). *An introduction to order statistics*. Paris: Atlantis Press.

Arnold, B. C. (2015). *Pareto distributions*, 2nd ed. Boca Raton: Chapman and Hall/CRC.

Arnold, B. C., Balakrishnan, N., Nagaraja, H. N. (2008). *A first course in order statistics*. Philadelphia: Society for Industrial and Applied Mathematics.
Arnold, B. C., Castillo, E., Sarabia, J. M. (1999). *Conditional specification of statistical models*. New York: Springer.
Asimit, A. V., Vernic, R., Zitikis, R. (2016). Background risk models and stepwise portfolio construction. *Methodology and Computing in Applied Probability, 18,* 805–827.
Bauer, D., Zanjani, G. (2016). The marginal cost of risk, risk measures, and capital allocation. *Management Science, 62,* 1431–1457.
BCBS (2016). *Minimum capital requirements for market risk.* (2016). Basel committee on banking supervision. Bank for international settlements, Basel. https://www.bis.org/bcbs/publ/d352.htm.
BCBS (2019). *Minimum capital requirements for market risk.* (2019). Basel committee on banking supervision. Bank for international settlements, Basel. https://www.bis.org/bcbs/publ/d457.htm.
Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J. (2004). *Statistics of extremes*: *Theory and applications*. Chichester: Wiley.
Bickel, P. J., Sakov, A. (2008). On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica, 18,* 967–985.
Bickel, P. J., Götze, F., van Zwet, W. R. (1997). Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statistica Sinica, 7,* 1–31.
Cai, J.-J., Musta, E. (2020). Estimation of the marginal expected shortfall under asymptotic independence. *Scandinavian Journal of Statistics, 47,* 56–83.
Cai, J.-J., Einmahl, J. H. J., de Haan, L., Zhou, C. (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme. *Journal of the Royal Statistical Society*: *Series B* (*Statistical Methodology*), *77,* 417–442.
Carcea, M., Serfling, R. (2015). A Gini autocovariance function for time series modelling. *Journal of Time Series Analysis, 36,* 817–838.
Castillo, E., Hadi, A. S., Balakrishnan, N., Sarabia, J. M. (2004). *Extreme value and related models with applications in engineering and science*. Chichester: Wiley.
DasGupta, A. (2008). *Asymptotic theory of statistics and probability*. New York: Springer.
David, H. A., Nagaraja, H. N. (2003). *Order statistics*, 3rd ed. Hoboken: Wiley.
Davison, A. C., Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge: Cambridge University Press.
de Haan, L., Ferreira, A. (2006). *Extreme value theory*: *An introduction*. New York: Springer.
Dudkina, O. I., Gribkova, N. V. (2020). On the strong law of large numbers for linear combinations of concomitants. Vestnik St. *Petersburg University, Mathematics, 53,* 282–286.
Efron, B., Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Boca Raton: Chapman and Hall/CRC.
Embrechts, P., Mao, T., Wang, Q., Wang, R. (2021). Bayes risk, elicitability, and the expected shortfall. *Mathematical Finance* (in press). https://onlinelibrary.wiley.com/doi/abs/10.1111/mafi.12313.
Frees, E. W., Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal, 2,* 1–25.
Furman, E., Kye, Y., Su, J. (2021a). A reconciliation of the top-down and bottom-up approaches to risk capital allocations: Proportional allocations revisited. *North American Actuarial Journal, 25,* 395–416.
Furman, E., Kye, Y., Su, J. (2021b). Multiplicative background risk models: Setting a course for the idiosyncratic risk factors distributed phase-type. *Insurance*: Mathematics and Economics, 96, 153–167.
Furman, E., Zitikis, R. (2008). Weighted risk capital allocations. *Insurance*: *Mathematics and Economics*, 43, 263–269.
Gribkova, N. V., Helmers, R. (2007). On the Edgeworth expansion and the M out of N bootstrap accuracy for a Studentized trimmed mean. *Mathematical Methods of Statistics, 16,* 142–176.
Gribkova, N. V., Helmers, R. (2011). On the consistency of the $M \ll N$ bootstrap approximation for a trimmed mean. *Theory of Probability and Its Applications, 55,* 42–53.
Gribkova, N., Zitikis, R. (2017). Statistical foundations for assessing the difference between the classical and weighted-Gini betas. *Mathematical Methods of Statistics, 26,* 267–281.
Gribkova, N., Zitikis, R. (2019). Weighted allocations, their concomitant-based estimators, and asymptotics. *Annals of the Institute of Statistical Mathematics, 71,* 811–835.
Guo, Q., Bauer, D., Zanjani, G. (2021). Capital allocation techniques: Review and comparison. *Variance* (in press).
Hall, P. (1992). *The bootstrap and Edgeworth expansion*. New York: Springer.

Insurance Information Institute (2021). *What Is Covered by Standard Homeowners Insurance?* Accessed online on September 17, 2021, at the address https://www.iii.org/article/what-covered-standard-homeowners-policy.
Kattumannil, S. K., Sreelakshmi, N., Balakrishnan, N. (2020). Non-parametric inference for Gini covariance and its variants. *Sankhyā A* . https://doi.org/10.1007/s13171-020-00218-z.
Kulik, R., Tong, Z. (2019). Estimation of the expected shortfall given an extreme component under conditional extreme value model. *Extremes, 22,* 29–70.
Laidi, M., Rassoul, A., Ould Rouis, H. (2020). Improved estimator of the conditional tail expectation in the case of heavy-tailed losses. *Statistics, Optimization and Information Computing, 8,* 98–109.
Mardia, K. V. (1962). Multivariate Pareto distributions. *Annals of Mathematical Statistics, 33,* 1008–1015.
Necir, A., Rassoul, A., Zitikis, R. (2010). Estimating the conditional tail expectation in the case of heavy-tailed losses. *Journal of Probability and Statistics, 596839,* 1–17 (Special issue on "Actuarial and Financial Risks: Models, Statistical Inference, and Case Studies").
Nešlehová, J., Embrechts, P., Chavez-Demoulin, V. (2006). Infinite-mean models and the LDA for operational risk. *Journal of Operational Risk, 1,* 3–25.
Pedersen, H., Campbell, M. P., Christiansen, S. L., Cox, S. H., Finn, D., Griffin, K., et al. (2016). *Economic scenario generators*: *A practical guide*. Schaumburg: The Society of Actuaries.
Schechtman, E., Shelef, A., Yitzhaki, S., Zitikis, R. (2008). Testing hypotheses about absolute concentration curves and marginal conditional stochastic dominance. *Econometric Theory, 24,* 1044–1062.
Shalit, H., Yitzhaki, S. (1994). Marginal conditional stochastic dominance. *Management Science, 40,* 549–684.
Shao, J., Tu, D. (1995). *The jackknife and bootstrap*. New York: Springer.
Shelef, A. (2013). Statistical analyses based on Gini for time series data. PhD Dissertation, Ben-Gurion University of the Negev, Beer Sheva.
Shelef, A. (2016). A Gini-based unit root test. *Computational Statistics and Data Analysis, 100,* 763–772.
Shelef, A., Schechtman, E. (2019). A Gini-based time series analysis and test for reversibility. *Statistical Papers, 60,* 687–716.
Tasche, D. (2004). Allocating portfolio economic capital to sub-portfolios. In A. Dev (Ed.), *Economic capital*: *A practitioner guide* (pp. 275–302). London: Risk Books.
Wang, R., Zitikis, R. (2020). An axiomatic foundation for the expected shortfall. *Management Science, 67,* 1413–1429.
Wang, Q., Wang, R. Zitikis, R. (2021). *Risk Measures Induced by Efficient Insurance Contracts.* Available at arXiv:org/abs/2109.00314.
Yitzhaki, S., Schechtman, E. (2013). *The Gini methodology*. New York: Springer.

Authors and Affiliations

N. V. Gribkova[1,2] · J. Su[3,4] · R. Zitikis[4,5]

N. V. Gribkova
n.gribkova@spbu.ru

R. Zitikis
rzitikis@uwo.ca

1 Faculty of Mathematics and Mechanics, Saint Petersburg State University, 7/9 Universitetskaya Embankment, Saint Petersburg 199034, Russia

2 Department of Mathematics, Emperor Alexander I St. Petersburg State Transport University, 9 Moskovsky Prospekt, Saint Petersburg 199034, Russia

3 Department of Statistics, Purdue University, 150 N. University Street, West Lafayette, IN 47907, USA

[4] Risk and Insurance Studies Centre, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada

[5] School of Mathematical and Statistical Sciences, Western University, 1151 Richmond Street North, London, ON N6A 5B7, Canada