SIAM J. OPTIM.
Vol. 28, No. 2, pp. 1337–1366

A CENTRAL LIMIT THEOREM AND HYPOTHESES TESTING FOR RISK-AVERSE STOCHASTIC PROGRAMS*

VINCENT GUIGUES†, VOLKER KRÄTSCHMER‡, AND ALEXANDER SHAPIRO§

Abstract. We study statistical properties of the optimal value and optimal solutions of the sample average approximation of risk-averse stochastic problems. Central limit theorem-type results are derived for the optimal value when the stochastic program is expressed in terms of a law invariant coherent risk measure having a discrete Kusuoka representation. The obtained results are applied to hypotheses testing problems aiming at comparing the optimal values of several risk-averse convex stochastic programs on the basis of samples of the underlying random vectors. We also consider nonasymptotic tests based on confidence intervals on the optimal values of the stochastic programs obtained using the stochastic mirror descent algorithm. Numerical simulations show how to use our developments to choose among different distributions and on the considered class of risk-averse stochastic programs the asymptotic tests show better results.

Key words. stochastic optimization, sample average approximation, hypotheses testing, coherent risk measures, statistical inference, central limit theorem

AMS subject classifications. 90C15, 90C90, 90C30

DOI. 10.1137/16M1104639

1. Introduction.

Consider the following risk-averse stochastic program:

$$\min_{x\in\mathcal{X}} \left\{ g(x) := \mathcal{R}(G_x) \right\}. \tag{1}$$

Here $\mathcal{X}$ is a nonempty compact subset of $\mathbb{R}^m$, G_x is a random variable depending on $x \in \mathcal{X}$ and $\mathcal{R}$ is a risk measure. We assume that G_x is given in the form $G_x(\omega) = G(x, \xi(\omega))$, where $G : \mathcal{X} \times \mathbb{R}^d \to \mathbb{R}$ and $\xi : \Omega \to \mathbb{R}^d$ is a random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ whose distribution is supported on set $\Xi \subset \mathbb{R}^d$. We assume that the functional $\mathcal{R}$, defined on a space of random variables, is law invariant (we will give precise definitions in section 2).

Let $\xi_j = \xi_j(\omega)$, $j = 1, \dots, N$, be an independently and identically distributed (i.i.d.) sample of the random vector ξ defined on the same probability space. Then the respective sample estimate of $g(x)$, denoted $\hat{g}_N(x)$, is obtained by replacing the "true" distribution of the random vector ξ with its empirical estimate. Consequently the true optimization problem (1) is approximated by the problem

$$\operatorname*{Min}_{x\in\mathcal{X}} \hat{g}_N(x), \tag{2}$$

*Received by the editors November 22, 2016; accepted for publication (in revised form) January 22, 2018; published electronically May 8, 2018.
http://www.siam.org/journals/siopt/28-2/M110463.html
Funding: Research of the first author was partly supported by an FGV grant, CNPq grant 307287/2013-0, and FAPERJ grants E-26/110.313/2014 and E-26/201.599/2014. Research of the third author was partly supported by NSF grant 1633196 and DARPA EQUiPS program, grant SNL 014150709.
†School of Applied Mathematics, Fundação Getulio Vargas, 190 Praia de Botafogo, Rio de Janeiro, Brazil (vguigues@fgv.br).
‡Faculty of Mathematics, University of Duisburg-Essen, Essen, Germany (volker.kraetschmer@uni-due.de).
§School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205 (ashapiro@isye.gatech.edu).

referred to as the sample average approximation (SAA) problem. Note that $\hat{g}_N(x) = \hat{g}_N(x,\omega)$ is a random function; sometimes we suppress dependence on ω in the notation. In particular, if $\mathcal{R}$ is the expectation operator, i.e., $g(x) = \mathbb{E}[G_x]$, then $\hat{g}_N(x) = N^{-1}\sum_{j=1}^{N} G(x,\xi_j)$.

We denote by ϑ_* and $\hat{\vartheta}_N$ the optimal values of problems (1) and (2), respectively, and study statistical properties of $\hat{\vartheta}_N$. The random sample can be given by collected data or can be generated by Monte Carlo sampling techniques in the goal of solving the true problem by the SAA method. Although conceptually different, both situations lead to the same statistical inference.

The statistical analysis allows us to address the following question of asymptotic tests of hypotheses. Suppose that we are given $V \geq 2$ optimization problems of the form (1) with ξ, G, and $\mathcal{X}$ respectively replaced by ξ^v, G_v, and $\mathcal{X}_v$ for problem $v \in \{1,\dots,V\}$. On the basis of samples $\xi_1^v,\dots,\xi_N^v$, of size N, of $\xi^v, v = 1,\dots,V$, and denoting by ϑ_*^v the optimal value of problem v, we study statistical tests of the null hypotheses

$$
\begin{array}{ll}
\text{(a)} & H_0 :\ \vartheta_*^1 = \vartheta_*^2 = \cdots = \vartheta_*^V, \\
\text{(b)} & H_0 :\ \vartheta_*^p \leq \vartheta_*^q \text{ for } p \text{ fixed and all } 1 \leq q \leq V, \\
\text{(c)} & H_0 :\ \vartheta_*^1 \leq \vartheta_*^2 \leq \cdots \leq \vartheta_*^V,
\end{array}
\tag{3}
$$

against the corresponding unrestricted alternatives. As a special case, if the feasibility sets of the V optimizations problems are singletons, say $\{x_*^v\}$ for problem v, the above tests aim at comparing the risks $\mathcal{R}(G_{x_*^1}),\dots,\mathcal{R}(G_{x_*^V})$. These tests are useful when we want to choose among V candidate solutions $x_*^1,\dots,x_*^V$ of problem (1) the one with the smallest risk measure value, using risk measure $\mathcal{R}$ to rank the distributions $G_{x_*^v}$, $v = 1,\dots,V$, e.g., to decide about the preference of one set of assets over another. In this situation, if the risk measure $\mathcal{R}$ is polyhedral [5] or extended polyhedral [9], then it can be expressed as the optimal value of a risk-neutral optimization problem and tests on the equality of risk measure values $\mathcal{R}(G_{x_*^1}),\dots,\mathcal{R}(G_{x_*^V})$ are of the form (3)(a).

Setting $\theta := (\vartheta_*^1,\dots,\vartheta_*^V)$, we also consider the following extension of tests (3):

$$
H_0 : \theta \in \Theta_0 \quad \text{against} \quad H_1 : \theta \in \mathbb{R}^V,
\tag{4}
$$

with $\Theta_0 \subset \mathbb{R}^V$ being a linear space or a convex cone. Tests (3) will also be studied in a nonasymptotic setting.

The paper is organized as follows. In section 2 we specify the type of objective functions in the optimization problem (1). We introduce the class of so-called law invariant convex risk measures and point out the specific subclass of risk measures which we use. In section 3 we study the asymptotics of the SAA estimator for the optimal value of problem 1. Besides consistency its asymptotic distribution is given in Theorem 2. As a by product we may derive a result on asymptotic distributions of sample estimators for the law invariant convex risk measures which we consider, and it will turn out that it improves already known general results. The proof of Theorem 2 is the subject of section 4. This theorem allows us to derive in section 5.1 asymptotic rejection regions for tests (3) and (4). In section 5.2 we derive nonasymptotic rejection regions for tests (3). This analysis is first conducted in a risk-neutral setting (when $\mathcal{R} = \mathbb{E}$ is the expectation) and is then extended to risk-averse problems. In particular, in this latter case, we obtain nonasymptotic confidence intervals for the optimal value of (1) for a larger class of risk measures than the class considered in [11], where

$\mathcal{R} = \mathsf{AVaR}$ (the *average value-at-risk*; see section 2) was considered. Also, when $\mathcal{R} = \mathsf{AVaR}$, our bounds are slightly refined versions of the bounds from [11]. Finally, section 6 presents numerical simulations that illustrate our results: we show how to use our developments to choose, using tests (3), among different distributions. We also use these tests to compare the optimal value of several risk-averse stochastic programs. It is shown that the normal (Gaussian) distribution already approximates well the distribution of $\hat{\vartheta}_N$ for $N = 20$ and problem sizes (dimension of decision variables) up to $m = 10\,000$, and that the asymptotic tests yield much smaller type II errors than the considered nonasymptotic tests for small to moderate sample size (N up to 10^5) and problem size (m up to 500).

We use the following notation throughout the paper. By $F_Z(z) := \mathbb{P}(Z \le z)$ we denote the cumulative distribution function (c.d.f.) of a random variable $Z : \Omega \to \mathbb{R}$. By $F^{-1}(\alpha) = \inf\{t : F(t) \ge \alpha\}$ we denote the left-side α-quantile of the c.d.f. F. By $\mathfrak{Q}_F(\alpha)$ we denote the interval of α-quantiles of c.d.f. F, i.e.,

$$\mathfrak{Q}_F(\alpha) = [a, b], \quad \text{where } a := F^{-1}(\alpha),\ b := \sup\{t : F(t) \le \alpha\}. \tag{5}$$

By $\mathbf{1}_A(\cdot)$ we denote the indicator function of set A. For $p \in [1, \infty)$ we consider the space $\mathcal{Z} := L_p(\Omega, \mathcal{F}, \mathbb{P})$ of random variables $Z : \Omega \to \mathbb{R}$ having finite pth order moments. The dual of space $\mathcal{Z}$ is the space $\mathcal{Z}^* = L_q(\Omega, \mathcal{F}, P)$, where $q \in (1, \infty]$ is such that $1/p + 1/q = 1$. The notation $Z \succeq Z'$ means that $Z(\omega) \ge Z'(\omega)$ for a.e. $\omega \in \Omega$. By $\delta(a)$ we denote the measure of mass one at a.

2. Preliminary discussion. Let us turn to specifying the functional (risk measure) $\mathcal{R}$ in the goal of problem (1). It is defined as a mapping $\mathcal{R} : \mathcal{Z} \to \mathbb{R}$ on a linear space $\mathcal{Z}$ consisting of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Specifically we assume that $\mathcal{Z} := L_p(\Omega, \mathcal{F}, \mathbb{P})$, $p \in [1, \infty)$. Note that we consider here real-valued risk measures, i.e., we do not allow $\mathcal{R}(Z)$ to have an infinite value. It is said that risk measure $\mathcal{R}(Z)$ is *law invariant* if it depends only on the distribution of Z, i.e., if $Z, Z' \in \mathcal{Z}$ and $F_Z = F_{Z'}$, then $\mathcal{R}(Z) = \mathcal{R}(Z')$.

In the influential paper of Artzner et al. [2] it was suggested that a "good" risk measure should satisfy the following conditions (axioms).

(i) *Monotonicity:* If $Z, Z' \in \mathcal{Z}$ and $Z \succeq Z'$, then $\mathcal{R}(Z) \ge \mathcal{R}(Z')$.
(ii) *Subadditivity:* $\mathcal{R}(Z + Z') \le \mathcal{R}(Z) + \mathcal{R}(Z')$ for all $Z, Z' \in \mathcal{Z}$.
(iii) *Translation equivariance:* If $a \in \mathbb{R}$ and $Z \in \mathcal{Z}$, then $\mathcal{R}(Z + a) = \mathcal{R}(Z) + a$.
(iv) *Positive homogeneity:* If $t \ge 0$ and $Z \in \mathcal{Z}$, then $\mathcal{R}(tZ) = t\mathcal{R}(Z)$.

Conditions (ii) and (iv) imply that $\mathcal{R}$ is convex, i.e.,

$$\mathcal{R}(tZ + (1-t)Z') \le t\mathcal{R}(Z) + (1-t)\mathcal{R}(Z')$$

for all $Z, Z' \in \mathcal{Z}$ and all $t \in [0, 1]$.

In [2] such risk measures were called *coherent* and suggested as a mathematical tool to assess the risks of financial positions. Unless stated otherwise we deal in this paper with law invariant coherent risk measures. Systematic accounts of this class of risk measures can be found in the monographs [29, Chapter 6] and [6, Chapter 4].

An important example of law invariant coherent risk measure is the so-called *average value-at-risk* (also called conditional value-at-risk, expected shortfall, and expected tail loss)

$$\mathsf{AVaR}_\alpha(Z) := \frac{1}{1-\alpha}\int_\alpha^1 F_Z^{-1}(t)\, dt, \quad \alpha \in [0, 1). \tag{6}$$

It is naturally defined, and is finite valued, on the space $\mathcal{Z} = L_1(\Omega, \mathcal{F}, \mathbb{P})$, and has the following useful representation (cf. [23]):

$$\mathsf{AVaR}_\alpha(Z) = \inf_{t\in\mathbb{R}} \left\{ t + (1-\alpha)^{-1}\mathbb{E}[(Z-t)_+] \right\}. \tag{7}$$

Note that $\mathsf{AVaR}_0(\cdot) = \mathbb{E}[\cdot]$.

The average value-at-risk $\mathsf{AVaR}_\alpha(Z)$ is an index to describe the tail behavior of the distribution function F_Z on the interval $(F_Z^{-1}(\alpha), \infty)$. If we want to take into account different regions of tail behavior, we may choose different levels $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_k < 1$ and then weight the average value-at-risk at the respective levels. That is, consider

$$\mathcal{R}(Z) := \sup_{w\in\mathfrak{W}} \left\{ w_0\mathbb{E}[Z] + \sum_{i=1}^k w_i \mathsf{AVaR}_{\alpha_i}(Z) \right\}, \tag{8}$$

where $\mathfrak{W}$ is a nonempty subset of $\Delta_{k+1} := \{w \in \mathbb{R}_+^{k+1} : w_0 + \cdots + w_k = 1\}$. This is a law invariant coherent risk measure defined on the space $\mathcal{Z} = L_1(\Omega, \mathcal{F}, \mathbb{P})$. Note that $\mathcal{R}$ is not changed if $\mathfrak{W}$ is replaced by the topological closure of its convex hull. Note also that the set Δ_{k+1} and hence the set $\mathfrak{W}$ are bounded. Therefore if $\mathfrak{W}$ is closed, then it is compact. In view of (7) we can write this risk measure in the following minimax form:

$$\mathcal{R}(Z) = \sup_{w\in\mathfrak{W}} \inf_{\tau\in\mathbb{R}^k} \mathbb{E}[\phi(Z, w, \tau)], \tag{9}$$

where

$$\phi(z, w, \tau) := w_0 z + \sum_{i=1}^k w_i \left(\tau_i + (1-\alpha_i)^{-1}[z-\tau_i]_+ \right). \tag{10}$$

Assuming that the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is *nonatomic*, any law invariant coherent risk measure $\mathcal{R} : \mathcal{Z} \to \mathbb{R}$ has the following so-called Kusuoka representation (cf. [10]):

$$\mathcal{R}(Z) = \sup_{\mu\in\mathfrak{M}} \int_0^1 \mathsf{AVaR}_\alpha(Z) d\mu(\alpha), \tag{11}$$

where $\mathfrak{M}$ is a set of probability measures on the interval [0,1). We can view risk measure (8) as a discretized version of Kusuoka representation where probability measures $\mu \in \mathfrak{M}$ are restricted to have finite support $\{\alpha_0, \ldots, \alpha_k\}$.

3. Asymptotics of the optimization problem. Since a law invariant risk measure $\mathcal{R}$ can be considered as a function of its c.d.f. $F(\cdot) = F_Z(\cdot)$, we also write $\mathcal{R}(F)$ to denote the corresponding value $\mathcal{R}(Z)$. Let $Z_1, \ldots, Z_N$ be an i.i.d. sample of Z and $\widehat{F}_N = N^{-1}\sum_{j=1}^N \mathbf{1}_{[Z_j,\infty)}$ be the corresponding empirical estimate of the c.d.f. F. By replacing F with its empirical estimate $\widehat{F}_N$, we obtain the estimate $\mathcal{R}(\widehat{F}_N)$ to which we refer as the *sample* or *empirical* estimate of $\mathcal{R}(F)$. We assume that for every $x \in \mathcal{X}$ the random variable G_x belongs to the space $\mathcal{Z}$, and hence $g(x) = \mathcal{R}(G_x)$ is well defined for every $x \in \mathcal{X}$. Let F_x be the c.d.f. of random variable G_x, $x \in \mathcal{X}$, and $\widehat{F}_{x,N}$ be the empirical c.d.f. associated with the sample $G(x, \xi_1), \ldots, G(x, \xi_N)$. Then we can write $g(x) = \mathcal{R}(F_x)$ and $\hat{g}_N(x) = \mathcal{R}(\widehat{F}_{x,N})$.

We have the following result about the convergence of the optimal value and optimal solutions of the SAA problem (2) to their counterparts of the "true" problem (1) (cf. [26, Theorem 3.3]).

THEOREM 1. *Let $\mathcal{R} : \mathcal{Z} \to \mathbb{R}$ be a law invariant risk measure satisfying the axioms of monotonicity, convexity, and translation equivariance. Suppose that the set $\mathcal{X}$ is nonempty and compact and the following conditions hold:* (i) *the function $G_x(\omega)$ is random lower semicontinuous, i.e., the epigraphical multifunction $\omega \mapsto \{(x,t) \in \mathbb{R}^{n+1} : G_x(\omega) \le t\}$ is closed valued and measurable;* (ii) *for every $\bar{x} \in \mathbb{R}^n$ there is a neighborhood $\mathcal{V}_{\bar{x}}$ of $\bar{x}$ and a function $h \in \mathcal{Z}$ such that $G_x(\cdot) \ge h(\cdot)$ for all $x \in \mathcal{V}_{\bar{x}}$.*

Then the optimal value $\hat{\vartheta}_N$ of problem (2) *converges with probability one (w.p.1) to the optimal value ϑ_* of the "true" problem* (1)*, and the distance from an optimal solution $\hat{x}_N$ of* (2) *to the set of optimal solutions of* (1) *converges w.p.1 to zero as $N \to \infty$.*

We derive first order asymptotics of the SAA optimal value for risk measures $\mathcal{R}$ of the form (8), i.e., having discretized Kusuoka representation. We assume that the set $\mathcal{X}$ is nonempty convex compact, $G(x,\xi)$ is convex in x for all $\xi \in \Xi$, and $\mathbb{E}|G_x| < +\infty$ for all $x \in \mathcal{X}$. It follows that functions $g(x)$ and $\hat{g}_N(x)$ are convex and finite valued, and hence the respective optimization problems (1) and (2) are convex.

Since $\mathcal{R}$ is of the form (8), the optimal value ϑ_* of problem (1) can be written as

$$\vartheta_* = \inf_{x\in\mathcal{X}} \sup_{w\in\mathfrak{W}} \left\{ w_0\mathbb{E}[G_x] + \sum_{i=1}^{k} w_i \mathsf{AVaR}_{\alpha_i}(G_x) \right\}. \tag{12}$$

As it was pointed before we can assume that the set $\mathfrak{W} \subset \Delta_{k+1}$ is convex and closed. Note that the objective function in the right-hand side of (12) is convex in x and linear in w. Therefore, since $\mathfrak{W}$ and $\mathcal{X}$ are convex compact, the "min" and "max" operators can be interchanged, i.e.,

$$\vartheta_* = \sup_{w\in\mathfrak{W}} \inf_{x\in\mathcal{X}} \left\{ w_0\mathbb{E}[G_x] + \sum_{i=1}^{k} w_i \mathsf{AVaR}_{\alpha_i}(G_x) \right\}, \tag{13}$$

and both problems (12) and (13) have nonempty sets of optimal solutions, denoted respectively as $\overline{\mathcal{X}}$ and $\overline{\mathfrak{W}}$. We make the following assumption.

(A) For every $i \in \{1,\dots,k\}$ there exists $w \in \mathfrak{W}$ such that $w_i \ne 0$.

This is a natural condition. Otherwise there is $i \in \{1,\dots,k\}$ such that $w_i = 0$ for all $w \in \mathfrak{W}$. In that case we can reduce the considered set $\{\alpha_0, \alpha_1, \dots, \alpha_k\}$ by removing the corresponding point α_i.

We also can write

$$\vartheta_* = \inf_{(x,\tau)\in\mathcal{X}\times\mathbb{R}^k} \sup_{w\in\mathfrak{W}} \mathbb{E}[\phi(G_x, w, \tau)] \tag{14}$$

$$= \sup_{w\in\mathfrak{W}} \inf_{(x,\tau)\in\mathcal{X}\times\mathbb{R}^k} \mathbb{E}[\phi(G_x, w, \tau)], \tag{15}$$

where the function $\phi(z,w,\tau)$ is defined in (10). Define $\mathcal{Y} := \mathcal{X} \times \mathbb{R}^k$ and let $\overline{\mathcal{Y}} \subset \mathcal{Y}$ be the set of optimal solutions of problem (14). Note that under assumption (A), the set $\overline{\mathcal{Y}}$ consists of points $(\bar{x}, \bar{\tau})$ such that $\bar{x} \in \overline{\mathcal{X}}$ and $\bar{\tau}_i$ belongs to the α_i-quantile interval of the c.d.f. of $G_{\bar{x}}$, $i = 1,\dots,k$. It follows that the set $\overline{\mathcal{Y}}$ is nonempty, convex, and compact. The set of optimal solutions of problem (15) is $\overline{\mathfrak{W}}$, the same as the one of problem (13). The minimax problem (14)–(15) is convex in $(x,\tau) \in \mathcal{Y}$ and concave (linear) in $w \in \mathbb{R}^k$. The set of saddle points of this minimax problem is $\overline{\mathfrak{W}} \times \overline{\mathcal{Y}}$. The SAA problem for (14) is written

$$\hat{\vartheta}_N = \inf_{(x,\tau)\in\mathcal{X}\times\mathbb{R}^k} \sup_{w\in\mathfrak{W}} \frac{1}{N}\sum_{j=1}^{N} \phi\left(G(x,\xi_j), w, \tau\right). \tag{16}$$

The following theorem is the main result of this section. Its proof is presented in section 4. The main tools in the derivation of this result are the minimax representations (14)–(16) and a minimax functional central limit theorem (cf. [25]).

THEOREM 2. *Suppose that* (i) $\mathcal{R}$ *is of the form* (8) *with the set* $\mathfrak{W} \subset \Delta_{k+1}$ *being convex and closed,* (ii) *the set* $\mathcal{X}$ *is nonempty, convex, and compact and* $G(x,\xi)$ *is convex in* x, (iii) *condition* (A) *holds,* (iv) $\mathbb{E}[G_{x^*}^2]$ *is finite for some* $x^* \in \mathcal{X}$, (v) *there is a measurable function* $C(\xi)$ *such that* $\mathbb{E}[C(\xi)^2]$ *is finite and*

$$|G(x,\xi) - G(x',\xi)| \le C(\xi)\|x - x'\| \quad \forall x, x' \in \mathcal{X},\ \forall \xi \in \Xi. \tag{17}$$

Then

$$\begin{aligned} \hat{\vartheta}_N = \quad & \inf_{(x,\tau)\in\overline{\mathcal{Y}}} \sup_{w\in\overline{\mathfrak{W}}} \Big\{ \tfrac{w_0}{N} \textstyle\sum_{j=1}^N G(x,\xi_j) \\ & + \textstyle\sum_{i=1}^k w_i \Big(\tau_i + \tfrac{1}{N(1-\alpha_i)} \textstyle\sum_{j=1}^N [G(x,\xi_j) - \tau_i]_+\Big)\Big\} + o_p(N^{-1/2}), \end{aligned} \tag{18}$$

and

$$N^{1/2}\big(\hat{\vartheta}_N - \hat{\vartheta}_*\big) \xrightarrow{\mathcal{D}} \sup_{w\in\overline{\mathfrak{W}}} \inf_{(x,\tau)\in\overline{\mathcal{Y}}} \mathbb{Y}(x,\tau,w), \tag{19}$$

where $\mathbb{Y}(w,\tau)$ *is a Gaussian process with mean zero and covariances*

$$\begin{gathered} \mathbb{E}[\mathbb{Y}(x,\tau,w)\mathbb{Y}(x',\tau',w')] = \\ \mathrm{Cov}\Big(w_0 G_x + \sum_{i=1}^k \tfrac{w_i}{1-\alpha_i}\,[G_x - \tau_i]_+\,, w_0' G_{x'} + \sum_{i=1}^k \tfrac{w_i'}{1-\alpha_i}\,[G_{x'} - \tau_i']_+\Big). \end{gathered} \tag{20}$$

Moreover, if the sets $\overline{\mathfrak{W}} = \{\bar{w}\}$ *and* $\overline{\mathcal{Y}} = \{(\bar{x},\bar{\tau})\}$ *are singletons, then* $N^{1/2}\big(\hat{\vartheta}_N - \vartheta_*\big)$ *converges in distribution to normal* $\mathcal{N}(0,\nu_*^2)$ *with variance*

$$\nu_*^2 := \mathrm{Var}\left[\phi(G_{\bar{x}}, \bar{w}, \bar{\tau})\right] = \mathrm{Var}\left\{\bar{w}_0 G_{\bar{x}} + \sum_{i=1}^k \frac{\bar{w}_i}{1-\alpha_i}\big[G_{\bar{x}} - \bar{\tau}_i\big]_+\right\}. \tag{21}$$

Remark 1. It is assumed in the above theorem that the set $\mathcal{X}$ is compact. Actually it is possible to push the proof through by relaxing this assumption to the respective set $\overline{\mathcal{X}}$ of optimal solutions being nonempty and compact.

Remark 2. For further calculation of the covariance structure (20) we may invoke Hoeffding's covariance formula (e.g., Lemma 5.24 in [15]) to obtain, for $t, s \in \mathbb{R}$,

$$\mathrm{Cov}\big([G_x - t]_+, [G_{x'} - s]_+\big) = \int_t^\infty \int_s^\infty \big(F_{x,x'}(u,v) - F_x(u)F_{x'}(v)\big)\, du\, dv, \tag{22}$$

where $F_{x,x'}$ denotes the joint distribution function of G_x and $G_{x'}$ and F_x and $F_{x'}$ denote their marginal distribution functions, respectively.

Let us discuss now estimation of the variance ν_*^2 given in (21). Let $(\hat{x}_N, \hat{\tau}_N, \hat{w}_N)$ be a saddle point of the SAA problem (16). Suppose that the sets $\overline{\mathfrak{W}} = \{\bar{w}\}$ and $\overline{\mathcal{Y}} = \{(\bar{x}, \bar{\tau})\}$ are singletons. Since the sets $\mathcal{Y}$ and $\mathfrak{W}$ are convex and the function $\phi\left(G(x,\xi), w, \tau\right)$ is convex in (x, τ) and concave (linear) in w, it follows that $(\hat{x}_N, \hat{\tau}_N)$ converges w.p.1 to $(\bar{x}, \bar{\tau})$ and $\hat{w}_N$ converges w.p.1 to $\bar{w}$ as $N \to \infty$ (see, e.g., [29, Theorem 5.4]). It follows that the variance ν_*^2 can be consistently estimated by its sample counterpart, i.e., the estimator

$$\hat{\nu}_N^2 = \frac{1}{N-1} \sum_{j=1}^{N} \left[\phi(G(\hat{x}_N, \xi_j), \hat{w}_N, \hat{\tau}_N) - \frac{1}{N} \sum_{j=1}^{N} \phi(G(\hat{x}_N, \xi_j), \hat{w}_N, \hat{\tau}_N) \right]^2 \tag{23}$$

converges w.p.1 to ν_*^2. Then employing Slutsky's theorem we obtain that, under the assumptions of Theorem 2, it follows that

$$\frac{N^{1/2}\left(\hat{\vartheta}_N - \vartheta_*\right)}{\hat{\nu}_N} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{24}$$

In particular, Theorem 2 provides an interesting asymptotic result concerning empirical estimates of risk measures. We again assume that $\mathcal{R}$ has representation as in (8) with $\mathfrak{W}$ being convex and compact. Let Z be an integrable random variable having c.d.f. F, and $\hat{F}_N$ be the empirical c.d.f. based on an i.i.d. sample $Z_1, \dots, Z_N \sim F$, and hence

$$\mathcal{R}(\hat{F}_N) = \sup_{w \in \mathfrak{W}} \inf_{\tau \in \mathbb{R}^k} \frac{1}{N} \sum_{j=1}^{N} \phi(Z_j, w, \tau).$$

Consider the sets

$$\overline{\mathfrak{W}}(F) := \arg\max_{w \in \mathfrak{W}} \left\{ w_0 \mathbb{E}[Z] + \sum_{i=1}^{k} w_i \mathsf{AVaR}_{\alpha_i}(Z) \right\} = \arg\max_{w \in \mathfrak{W}} \left\{ \inf_{\tau \in \mathbb{R}^k} \mathbb{E}[\phi(Z, w, \tau)] \right\},$$
$$\overline{\mathfrak{T}}(F) := \mathfrak{Q}_F(\alpha_1) \times \cdots \times \mathfrak{Q}_F(\alpha_k)$$

associated with c.d.f. F of random variable Z. Note that under assumption (A), the set $\overline{\mathfrak{T}}(F)$ gives the set of minimizers of $\mathbb{E}[\phi(Z, w, \tau)]$ for any $w \in \mathfrak{W}$. We have that $\overline{\mathfrak{W}}(F) \times \overline{\mathfrak{T}}(F)$ is the set of saddle points of the respective minimax problem associated with $\mathcal{R}$. Then an application of Theorem 2 to the sample estimate $\mathcal{R}(\hat{F}_N)$ reads as follows.

COROLLARY 3. *Suppose that $\mathcal{R}$ is of the form* (8) *with $\mathfrak{W}$ being convex and closed, condition* (A) *holds and $\mathbb{E}_F[Z^2] < +\infty$. Then*

$$\mathcal{R}(\hat{F}_N) = \sup_{w \in \overline{\mathfrak{W}}(F)} \inf_{\tau \in \overline{\mathfrak{T}}(F)} \left\{ \frac{w_0}{N} \sum_{j=1}^{N} Z_j + \sum_{i=1}^{k} w_i \Big(\tau_i + \frac{1}{N(1-\alpha_i)} \sum_{j=1}^{N} [Z_j - \tau_i]_+ \Big) \right\} + o_p(N^{-1/2})$$

and

$$N^{1/2}\left[\mathcal{R}(\widehat{F}_N) - \mathcal{R}(F)\right] \xrightarrow{\mathcal{D}} \sup_{w \in \overline{\mathfrak{W}}(F)} \inf_{\tau \in \overline{\mathfrak{T}}(F)} \mathbb{Y}(w, \tau), \tag{25}$$

where $\mathbb{Y}(w,\tau)$ is a Gaussian process with mean zero and covariances

$$\begin{gathered} \mathbb{E}_F[\mathbb{Y}(w,\tau)\mathbb{Y}(w',\tau')] = \\ \mathrm{Cov}_F\Big(w_0 Z + \textstyle\sum_{i=1}^{k} \frac{w_i}{1-\alpha_i} \left[Z - \tau_i\right]_+, w_0' Z + \sum_{i=1}^{k} \frac{w_i'}{1-\alpha_i} \left[Z - \tau_i'\right]_+ \Big). \end{gathered} \tag{26}$$

Moreover, if the sets $\overline{\mathfrak{W}}(F) = \{\bar{w}\}$ and $\overline{\mathfrak{T}}(F) = \{\bar{\tau}\}$ are singletons, then $N^{1/2}\big[\mathcal{R}(\widehat{F}_N) - \mathcal{R}(F)\big]$ converges in distribution to normal $\mathcal{N}(0, \nu^2)$ with variance

$$\nu^2 = \mathrm{Var}_F \left\{ \bar{w}_0 Z + \sum_{i=1}^{k} \frac{\bar{w}_i}{1-\alpha_i} \left[Z - \bar{\tau}_i\right]_+ \right\}. \tag{27}$$

Remark 3. Corollary 3 provides an alternative representation of the asymptotic distribution of the estimator $\mathcal{R}(\hat{F}_N)$ in comparison with the already known ones from [20] and [3]. The results there are formulated for general law invariant coherent risk measures, with, however, additional assumptions about tail behavior of the distribution F. In particular, in [20] F is required to have a polynomial tail; more precisely,

$$\sup_{\alpha \in]0,1[} \big(F^{-1}(\alpha)\ \alpha^{d_1}\ (1-\alpha^{d_2})\big) < \infty \quad \text{for some } d_1, d_2 \in (0, 1/2)$$

(cf. [20, Theorem 3.7]). For law invariant coherent risk measures on $L_1(\Omega, \mathcal{F}, \mathbb{P})$ this condition was relaxed in [3] by

$$\int_{-\infty}^{\infty} \sqrt{F(u)[1-F(u)]}\ du < \infty \tag{28}$$

(cf. [3, Theorem 3.1]). It is well known that condition (28) is fulfilled if the random variable Z has absolute moments of order q for some $q > 2$, that property (28) implies that Z has absolute moments of order 2 (see, e.g., [13, p. 10]). Moreover, if Z has absolute moments of second order, it does not satisfy (28) necessarily. Hence in the case of risk measures with representation of the form (8), Corollary 3 improves existing results as it only assumes the existence of the second order moments.

Remark 4. Theorem 2 and Corollary 3 give quite a complete description of the asymptotics in the case in which the risk measure $\mathcal{R}$ has the discrete Kusuoka representation (8). It would be natural to try to extend this analysis to the general case of the Kusuoka representation (11) by writing the corresponding risk measure in the respective minimax form. It turned out to be surprisingly difficult to handle such a general setting in a rigorous way. The following examples demonstrate that asymptotics of empirical estimates of law invariant coherent risk measures could behave in quite a weird way; some specific conditions are required in order for the empirical estimates to have asymptotically normal distributions.

Example 1 (absolute semideviation risk measure). Consider the risk measure

$$\mathcal{R}_c(F) := \mathbb{E}_F[Z] + c\,\mathbb{E}_F[Z - \mathbb{E}_F(Z)]_+, \quad c \in (0, 1]. \tag{29}$$

We assume that the c.d.f. F has finite first order moment. This risk measure has the following representation (cf. [27]):

$$\mathcal{R}_c(F) = \sup_{\gamma \in [0,1]} \big\{(1-c\gamma)\mathbb{E}_F(Z) + c\gamma \mathsf{AVaR}_{1-\gamma}(F)\big\} \tag{30}$$

$$= \sup_{\gamma \in [0,1]} \inf_{t \in \mathbb{R}} \mathbb{E}_F \big\{(1-c\gamma)Z + c\gamma t + c[Z-t]_+\big\} \tag{31}$$

$$= \inf_{t \in \mathbb{R}} \sup_{\gamma \in [0,1]} \mathbb{E}_F \big\{(1-c\gamma)Z + c\gamma t + c[Z-t]_+\big\}. \tag{32}$$

Representation (30) is the (minimal) Kusuoka representation (11) of $\mathcal{R}_c$ with the corresponding set $\mathfrak{M} = \bigcup_{\gamma \in [0,1]} \{(1-c\gamma)\delta(0) + c\gamma\delta(1-\gamma)\}$. Since

$$\sup_{\gamma \in [0,1]} \mathbb{E}_F \big\{(1-c\gamma)Z + c\gamma t + c[Z-t]_+\big\} = \mathbb{E}_F[Z] + c \max\left\{\mathbb{E}_F[Z-t]_+, \mathbb{E}_F[t-Z]_+\right\},$$

it follows that problem (32) has unique optimal solution $t^* = \mathfrak{m}$, where $\mathfrak{m} := \mathbb{E}_F[Z]$, i.e., $\mathcal{R}_c(F) = \mathbb{E}_F[Z] + c\max\{\mathbb{E}_F[Z - \mathbb{E}_F[Z]]_+, \mathbb{E}_F[\mathbb{E}_F[Z] - Z]_+\}$, which is consistent with definition (29) of $\mathcal{R}_c$ because $\mathbb{E}_F[Z - \mathbb{E}_F[Z]]_+ = \mathbb{E}_F[\mathbb{E}_F[Z] - Z]_+$.

Now the set of minimizers of $\gamma t + \mathbb{E}[Z - t]_+$, over $t \in \mathbb{R}$, is defined by the equation $F(t) = 1 - \gamma$. It follows that the set of saddle points of the minimax representation (31) is $[\underline{\gamma}, \overline{\gamma}] \times \{\mathfrak{m}\}$, where

$$\underline{\gamma} := 1 - \mathsf{Pr}(Z \le \mathfrak{m}), \quad \overline{\gamma} := 1 - \mathsf{Pr}(Z < \mathfrak{m})$$

(cf. [29, section 6.6.2]). In other words, here the set of maximizers of measures $\mu \in \mathfrak{M}$ in the Kusuoka representation is

$$\bar{\mathfrak{M}}(F) = \bigcup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \{(1 - c\gamma)\delta(0) + c\gamma\delta(1 - \gamma)\},$$

and the respective set $\bar{\mathfrak{T}}(F) = \{\bar{\tau}(\alpha)\}$ is the singleton with $\bar{\tau}(\alpha) = \mathbb{E}_F[Z]$ for all $\alpha \in [0, 1)$.

The minimax representation (31) leads to the following asymptotics. Suppose that $\mathbb{E}_F[Z^2] < +\infty$. Then by a finite-dimensional minimax asymptotics theorem (cf. [25]),

$$\mathcal{R}_c(\widehat{F}_N) = \sup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \left\{ c\gamma\mathfrak{m} + (1 - c\gamma)\bar{Z} + cN^{-1} \sum_{j=1}^{N} \left[Z_j - \mathfrak{m}\right]_+ \right\} + o_p(N^{-1/2}), \tag{33}$$

where $\bar{Z} := N^{-1} \sum_{j=1}^N Z_j$. We have here that a condition which is *required* for asymptotic normality of the corresponding empirical estimate is that $\underline{\gamma} = \overline{\gamma}$, i.e., that $F(\cdot)$ should be continuous at $\mathfrak{m} = \mathbb{E}_F[Z]$. If the c.d.f. $F(\cdot)$ is continuous at $\mathfrak{m} = \mathbb{E}_F[Z]$, then $N^{1/2}\big[\mathcal{R}_c(\widehat{F}_N) - \mathcal{R}_c(F)\big]$ converges in distribution to normal $\mathcal{N}(0, \nu^2)$ with variance

$$\nu^2 = \mathrm{Var}_F\big\{(1 - c\gamma^*)Z + c[Z - \mathfrak{m}]_+\big\}, \tag{34}$$

where $\gamma^* := 1 - F(\mathfrak{m}) = \bar{F}(\mathfrak{m})$.

Example 2 (mean-semideviation risk measure). Consider the following risk measure:

$$\mathcal{R}_c(F) := \mathbb{E}_F[Z] + c\left(\mathbb{E}_F[Z - \mathbb{E}_F(Z)]_+^2\right)^{1/2}, \quad c \in (0, 1]. \tag{35}$$

Asymptotics of empirical estimates of such risk measures were discussed in [4]. If $F(\cdot)$ is continuous at $\mathfrak{m} := \mathbb{E}_F[Z]$, then $\mathcal{R}_c(\cdot)$ is Gâteaux differentiable at F and the corresponding influence function is

$$IF(z) = z + c(2\theta)^{-1}\left([z - \mathfrak{m}]_+^2 - \theta^2 + 2\kappa(1 - F(\mathfrak{m}))(z - \mathfrak{m})\right), \tag{36}$$

where $\theta := \left(\mathbb{E}_F[Z - \mathbb{E}_F[Z]]_+^2\right)^{1/2}$ and $\kappa := \mathbb{E}_F[Z - \mathfrak{m}]_+$ (see, e.g., [29, p. 345] for a more detailed discussion of this example). This indicates that continuity of $F(\cdot)$ at $\mathfrak{m}$ is a necessary condition for $\mathcal{R}_c(\cdot)$ to be Gâteaux differentiable at F. Here again continuity of $F(\cdot)$ at $\mathfrak{m}$ is a required condition for $\mathcal{R}(\widehat{F}_N)$ to be asymptotically normal.

4. Proof of Theorem 2. Throughout this section, we shall use notation and assumptions from Theorem 2. Moreover, let us define for $x \in \mathcal{X}$, $\tau \in \mathbb{R}^k$, and $w \in \mathfrak{W}$ the function

$$f_{x,\tau,w} : \mathbb{R}^d \to \mathbb{R},\ z \mapsto w_0 G(x,z) + \sum_{i=1}^{k} w_i \left(\tau_i + \frac{1}{1-\alpha_i}[G(x,z) - \tau_i]_+ \right).$$

The idea to show Theorem 2 is to apply asymptotic results from empirical process theory to the class of the functions $f_{x,\tau,w}$, and then to invoke a minimax delta theorem. In preparation to make use of mentioned results from empirical process theory, we shall verify first that the functions $f_{x,\tau,w}$ satisfy pointwise some certain Lipschitz continuity w.r.t. their parameters.

LEMMA 4. *For any $n \in \mathbb{N}$, there is a Borel-measurable function $C_n : \mathbb{R}^d \to \mathbb{R}$ such that $\mathbb{E}[C_n(\xi)^2] < \infty$ holds, and*

$$|f_{x,\tau,w}(z) - f_{\overline{x},\overline{\tau},\overline{w}}(z)| \le C_n(z)\big(\|x - \overline{x}\|_{m,2} + \|\tau - \overline{\tau}\|_{k,2} + \|w - \overline{w}\|_{k+1,2}\big)$$

is valid for any $z \in \mathbb{R}^d$, $x, \overline{x} \in \mathcal{X}$, $\tau, \overline{\tau} \in [-n,n]^k$ as well as $w, \overline{w} \in \mathfrak{W}$. Here $\|\cdot\|_{m,2}$, $\|\cdot\|_{k,2}$, and $\|\cdot\|_{k+1,2}$ respectively denote the Euclidean norms on $\mathbb{R}^m$, $\mathbb{R}^k$, and $\mathbb{R}^{k+1}$.

Proof. Let $x, \overline{x} \in \mathcal{X}$, $\tau, \overline{\tau} \in [-n,n]^k$ and $w, \overline{w} \in \mathfrak{W}$. Furthermore, let $x^* \in \mathcal{X}$ as in assumption (iv) of Theorem 2. Then using the triangle inequality several times we may observe, for $z \in \mathbb{R}^d$, that

$$\begin{aligned}
& |f_{x,\tau,w}(z) - f_{\overline{x},\overline{\tau},\overline{w}}(z)| \\
\le\ & w_0|G(x,z) - G(\overline{x},z)| + |(w_0 - \overline{w}_0)(G(\overline{x},z) - G(x^*,z))| + |w_0 - \overline{w}_0||G(x^*,z)| \\
& + \sum_{i=1}^{k} w_i \left|(\tau_i - \overline{\tau}_i) + (1-\alpha_i)^{-1}\big[(G(x,z) - \tau_i)_+ - (G(\overline{x},z) - \overline{\tau}_i)_+\big]\right| \\
& + \sum_{i=1}^{k} \frac{|w_i - \overline{w}_i|}{1-\alpha_i} \cdot \Big(|(G(x^*,z) - \overline{\tau}_i)_+ - (G(\overline{x},z) - \overline{\tau}_i)_+| \\
& \qquad\qquad + |(1-\alpha_i)\overline{\tau_i} + (G(x^*,z) - \overline{\tau}_i)_+| \Big) \\
\le\ & w_0|G(x,z) - G(\overline{x},z)| + |(w_0 - \overline{w}_0)(G(\overline{x},z) - G(x^*,z))| + |w_0 - \overline{w}_0||G(x^*,z)| \\
& + \sum_{i=1}^{k} \frac{w_i}{1-\alpha_i} \cdot ((2-\alpha_i)|\tau_i - \overline{\tau}_i| + |G(x,z) - G(\overline{x},z)|) \\
& + \sum_{i=1}^{k} \frac{|w_i - \overline{w}_i|}{1-\alpha_i} \cdot \Big(|G(\overline{x},z) - G(x^*,z)| + (2-\alpha_i)|\overline{\tau}_i| + |G(x^*,z)| \Big).
\end{aligned}$$

Then invoking the Cauchy–Schwarz inequality we obtain

$$\begin{aligned}
& |f_{x,\tau,w}(z) - f_{\overline{x},\overline{\tau},\overline{w}}(z)| \\
\le\ & \max_{i=1,\ldots,k} (1-\alpha_i)^{-1} \cdot \|w\|_{k+1,2}\big(\sqrt{k+1}|G(x,z) - G(\overline{x},z)| + 2\|\tau - \overline{\tau}\|_{k,2}\big) \\
& + \max_{i=1,\ldots,k} (1-\alpha_i)^{-1} \\
& \qquad \cdot \|w - \overline{w}\|_{k+1,2}\big(\sqrt{k+1}(|G(x^*,z) - G(\overline{x},z)| + |G(x^*,z)|) + 2\|\overline{\tau}\|_{k,2}\big).
\end{aligned}$$

By assumption the inequalities

$$|G(x^*, z) - G(\overline{x}, z)| \leq C(z)\|x^* - \overline{x}\|_{m,2} \leq C(z) \ \mathrm{diam}(\mathcal{X})$$

and $|G(x, z) - G(\overline{x}, z)| \leq C(z)\|x - \overline{x}\|_{m,2}$ hold, where C denotes the nonnegative Borel-measurable function C as in assumption (v) of Theorem 2, and $\mathrm{diam}(\mathcal{X})$ stands for the diameter of the compact set $\mathcal{X}$ w.r.t. the Euclidean norm on $\mathbb{R}^n$. Hence

$$\begin{aligned}
& |f_{x,\tau,w}(z) - f_{\overline{x},\overline{\tau},\overline{w}}(z)| \\
\leq \ & \max_{i=1,\ldots,k}(1-\alpha_i)^{-1} \cdot \|w\|_{k+1,2}\big(\sqrt{k+1}C(z)\|x - \overline{x}\|_{m,2} + 2\|\tau - \overline{\tau}\|_{k,2}\big) \\
& + \max_{i=1,\ldots,k}(1-\alpha_i)^{-1} \cdot \|w - \overline{w}\|_{k+1,2}\big(\sqrt{k+1}(C(z) \ \mathrm{diam}(\mathcal{X}) + |G(x^*, z)|) + 2\|\overline{\tau}\|_{k,2}\big) \\
\leq \ & \max_{i=1,\ldots,k}(1-\alpha_i)^{-1} \cdot \big(\sqrt{k+1}C(z)\|x - \overline{x}\|_{m,2} + 2\|\tau - \overline{\tau}\|_{k,2}\big) \\
& + \max_{i=1,\ldots,k}(1-\alpha_i)^{-1} \cdot \|w - \overline{w}\|_{k+1,2}\big(\sqrt{k+1}(C(z) \ \mathrm{diam}(\mathcal{X}) + |G(x^*, z)|) + 2\sqrt{k}n\big).
\end{aligned}$$

Now the function $C_n : \mathbb{R}^d \to \mathbb{R}$, defined by

$$C_n(z) := \max_{i=1,\ldots,k}(1-\alpha_i)^{-1} \cdot \max\left[\sqrt{k+1}C(z), 2\sqrt{k}n + \sqrt{k+1}(C(z) \ \mathrm{diam}(\mathcal{X}) + |G(x^*, z)|)\right],$$

is as required due to assumptions (iv) and (v) of Theorem 2. ∎

In the next step we want to show that from an asymptotic view point we may replace the estimator $\hat{\vartheta}_N$ with the estimator

$$\tilde{\vartheta} := \inf_{(x,\tau)\in\mathcal{K}} \sup_{w\in\mathfrak{W}} \left\{ \frac{w_0}{N}\sum_{j=1}^{N} G(x, \xi_j) + \sum_{i=1}^{k} w_i\Big(\tau_i + \frac{1}{N(1-\alpha_i)}\sum_{j=1}^{N}[G(x, \xi_j) - \tau_i]_+\Big)\right\}$$

for some compact subset $\mathcal{K}$ of $\mathcal{X} \times \mathbb{R}^k$. This estimator is more convenient as it allows us to apply the minimax functional central limit theorem from [25].

Lemma 5. *Let $n_0 \in \mathbb{N}$ such that $\overline{\mathcal{Y}} \subseteq \mathcal{X} \times (-n_0, n_0)^k$. Then*

$$\begin{aligned}
\hat{\vartheta}_N = \ & \inf_{(x,\tau)\in\mathcal{K}} \sup_{w\in\mathfrak{W}} \Big\{ \tfrac{w_0}{N}\textstyle\sum_{j=1}^{N} G(x, \xi_j) \\
& + \textstyle\sum_{i=1}^{k} w_i\Big(\tau_i + \tfrac{1}{N(1-\alpha_i)}\sum_{j=1}^{N}[G(x, \xi_j) - \tau_i]_+\Big)\Big\} + o_p(N^{-1/2}),
\end{aligned}$$

where $\mathcal{K} := \mathcal{X} \times [-n_0, n_0]^k$.

Proof. Let ϕ denote the mapping as defined in (10). In particular,

$$f_{x,\tau,w}(z) = \phi(G(x, z), w, \tau) \quad \text{for } (x, \tau, w, z) \in \mathcal{X} \times \mathbb{R}^k \times \mathfrak{W} \times \mathbb{R}^d. \tag{37}$$

According to Lemma 4, we may draw on [32, Example 19.7 and Theorem 19.4] to find for every $n \in \mathbb{N}$ some $A_n \in \mathcal{F}$ with $\mathbb{P}(A_n) = 1$ such that

$$\sup_{(x,\tau)\in\mathcal{X}\times[-n,n]^k} \sup_{w\in\mathfrak{W}} \Big|\frac{1}{N}\sum_{j=1}^{N}\phi\big(G(x, \xi_j(\omega)), w, \tau\big) - \mathbb{E}\big[\phi\big(G(x, \xi), w, \tau\big)\big]\Big| \to 0 \tag{38}$$

for $\omega \in A_n$. Then $A := \bigcap_{n=1}^{\infty} A_n$ satisfies $\mathbb{P}(A) = 1$. Moreover, in view of Lemma 4 along with (38),

$$h_N : \mathcal{X} \times \mathbb{R}^k \times \Omega \to \mathbb{R}, \ (x, \tau, \omega) \mapsto \sup_{w\in\mathfrak{W}} \frac{1}{N}\sum_{j=1}^{N}\phi\big(G(x, \xi_j(\omega)), w, \tau\big)$$

defines a sequence $(h_N)_{N\in\mathbb{N}}$ of mappings such that, for any $\omega \in A$, the mapping $h_N(\cdot,\cdot,\omega)$ is lower-semicontinuous and the sequence $(h_N(\cdot,\cdot,\omega))_{N\in\mathbb{N}}$ converges uniformly on compact subsets to the function $\sup_{w\in\mathfrak{W}} \mathbb{E}[\phi(G(\cdot,\xi),w,\cdot)]$ with

$$\inf_{(x,\tau)\in\mathcal{K}} h_N(x,\tau,\omega) \to \inf_{(x,\tau)\in\mathcal{K}} \sup_{w\in\mathfrak{W}} \mathbb{E}[\phi(G(\cdot,\xi),w,\cdot)] = \vartheta_* \quad \text{for } N\to\infty.$$

Since $\mathcal{K}$ is compact, we may find for every $\omega \in A$ a sequence $\big((\hat{x}_N(\omega),\hat{\tau}_N(\omega))\big)_{N\in\mathbb{N}}$ in $\mathcal{K}$ such that $(\hat{x}_N(\omega),\hat{\tau}_N(\omega))$ minimizes $h_N(\cdot,\cdot,\omega)|_{\mathcal{K}}$ for any $N\in\mathbb{N}$. By compactness of $\mathcal{K}$ for any $\omega\in A$, the sequence $\big((\hat{x}_N(\omega),\hat{\tau}_N(\omega))\big)_{N\in\mathbb{N}}$ has cluster points which all belong to the set $\mathfrak{S}$ of minimizers of the function $\sup_{w\in\mathfrak{W}} \mathbb{E}[\phi(G(\cdot,\xi),w,\cdot)]|_{\mathcal{K}}$ because $(h_N(\cdot,\cdot,\omega))_{N\in\mathbb{N}}$ converges uniformly on $\mathcal{K}$ to $\sup_{w\in\mathfrak{W}} \mathbb{E}[\phi(G(\cdot,\xi),w,\cdot)]$ (cf. [24, Theorem 7.31]). In particular, the distance of $(\hat{x}_N(\omega),\hat{\tau}_N(\omega))$ to $\mathfrak{S}$ tends to zero as $N\to\infty$ for every $\omega\in A$. Note $\mathfrak{S}=\overline{\mathcal{Y}}$ so that for every $\omega\in A$ there is some $N(\omega)\in\mathbb{N}$ such that

$$(\hat{x}_N(\omega),\hat{\tau}_N(\omega)) \in \mathcal{X}\times(-n_0,n_0)^k \quad \text{for arbitrary } N\in\mathbb{N} \text{ with } N\geq N(\omega). \tag{39}$$

In view of assumption (ii) of Theorem 2, the mapping $\phi(G(\cdot,z),w,\cdot)$ is convex for every $z\in\mathbb{R}^d$ and any $w\in\mathfrak{W}$. This implies that $h_N(\cdot,\cdot,\omega)$ is convex for $N\in\mathbb{N}$ and $\omega\in A$, and thus

$$\min_{\lambda\in(0,1)} h_N\big(\lambda(x,\tau)+(1-\lambda)(\hat{x}_N(\omega),\hat{\tau}_N(\omega)),\omega\big) \leq \min\big\{h_N(x,\tau,\omega), h_N\big(\hat{x}_N(\omega),\hat{\tau}_N(\omega),\omega\big)\big\}$$

holds for $(x,\tau)\in\mathcal{X}\times\mathbb{R}^k$. Then by (39), we obtain, for any $\omega\in A$ and every $N\in\mathbb{N}$ with $N\geq N(\omega)$,

$$\hat{\vartheta}_N(\omega) = \inf_{(x,\tau)\in\mathbb{R}^k} h_N(x,\tau,\omega) = \inf_{(x,\tau)\in\mathcal{K}} h_N(x,\tau,\omega),$$

and then

$$\sqrt{N}\Big[\inf_{(x,\tau)\in\mathcal{X}\times\mathbb{R}^k} \sup_{w\in\mathfrak{W}} \quad \Big\{\tfrac{1}{N}\textstyle\sum_{j=1}^N \phi\big(G(x,\xi_j(\omega)),w,\tau\big) \\ - \inf_{(x,\tau)\in\mathcal{K}} \sup_{w\in\mathfrak{W}} \tfrac{1}{N}\textstyle\sum_{j=1}^N \phi\big(G(x,\xi_j(\omega)),w,\tau\big)\Big\}\Big] = 0$$

for $N\in\mathbb{N}$ with $N\geq N(\omega)$. Hence

$$\sqrt{N}\Big[\inf_{(x,\tau)\in\mathcal{X}\times\mathbb{R}^k} \sup_{w\in\mathfrak{W}} \quad \Big\{\frac{1}{N}\sum_{j=1}^N \phi\big(G(x,\xi_j(\omega)),w,\tau\big) \\ - \inf_{(x,\tau)\in\mathcal{K}} \sup_{w\in\mathfrak{W}} \frac{1}{N}\sum_{j=1}^N \phi\big(G(x,\xi_j(\omega)),w,\tau\big)\Big\}\Big] \to 0 \quad \mathbb{P}\text{-a.s.},$$

implying

$$\sqrt{N}\Big[\inf_{(x,\tau)\in\mathcal{X}\times\mathbb{R}^k} \sup_{w\in\mathfrak{W}} \quad \Big\{\frac{1}{N}\sum_{j=1}^N \phi\big(G(x,\xi_j(\omega)),w,\tau\big) \\ - \inf_{(x,\tau)\in\mathcal{K}} \sup_{w\in\mathfrak{W}} \frac{1}{N}\sum_{j=1}^N \phi\big(G(x,\xi_j(\omega)),w,\tau\big)\Big\}\Big] \to 0$$

in probability. This completes the proof. $\square$

Now we are ready to prove Theorem 2.

Proof of Theorem 2. Let ϕ denote the function defined in (10). By Lemma 5 we may find some $n_0 \in \mathbb{N}$ such that $\overline{\mathcal{Y}} \subseteq \mathcal{X} \times [-n_0, n_0]^k$,

$$\vartheta_* = \inf_{(x,\tau)\in\mathcal{X}\times[-n_0,n_0]^k} \sup_{w\in\mathfrak{W}} \mathbb{E}[\phi(G(x,\xi), w, \tau)] \tag{40}$$

and

$$\hat{\vartheta}_N = \inf_{(x,\tau)\in\mathcal{X}\times[-n_0,n_0]^k} \sup_{w\in\mathfrak{W}} \frac{1}{N}\sum_{j=1}^{N} \phi(G(x,\xi_j), w, \tau) + o_p(N^{-1/2}). \tag{41}$$

Set $\mathcal{K} := \mathcal{X} \times [-n_0, n_0]^k$ and

$$\overline{\vartheta}_N = \inf_{(x,\tau)\in\mathcal{K}} \sup_{w\in\mathfrak{W}} \frac{1}{N}\sum_{j=1}^{N} \phi(G(x,\xi_j), w, \tau).$$

The idea now is to apply Theorem 2.1 from [25], a minimax delta theorem, to $(\overline{\vartheta}_N)_{N\in\mathbb{N}}$ and ϑ_*. For this purpose consider the stochastic process $(V^N_{x,\tau,w})_{(x,\tau,w)\in\mathcal{K}\times\mathfrak{W}}$, defined by

$$V^N_{x,\tau,w} = \frac{1}{N}\sum_{j=1}^{N} \phi(G(x,\xi_j), w, \tau) \quad \text{for } (x,\tau,w) \in \mathcal{K}\times\mathfrak{W}.$$

Using Lemma 4 and recalling (37) it may be viewed as a Borel random element V^N of the space $\mathcal{C}(\mathcal{K}\times\mathfrak{W})$ of continuous real-valued mappings on $\mathcal{K}\times\mathfrak{W}$ which is endowed with the uniform metric. In the same way the mapping

$$V : \mathcal{K}\times\mathfrak{W} \to \mathbb{R},\ (x,\tau,w) \mapsto \mathbb{E}[\phi(G(x,\xi), w, \tau)]$$

may be verified as a member of $\mathcal{C}(\mathcal{K}\times\mathfrak{W})$. Drawing on Lemma 4 again, we may apply Example 19.7 from [32] to conclude that the sequence $N^{-1/2}(V^N - V)_{N\in\mathbb{N}}$ converges in law to some centered Gaussian random element $\mathbb{Y}$ of $\mathcal{C}(\mathcal{K}\times\mathfrak{W})$ with covariances

$$\mathbb{E}[\mathbb{Y}(x,\tau,w)\cdot\mathbb{Y}(x',\tau',w')] = \mathrm{Cov}\big(\phi(G(x,\xi), w, \tau), \phi(G(x',\xi), w', \tau')\big). \tag{42}$$

By assumption (ii) of Theorem 2, the mapping $\phi(G(\cdot,z), w, \cdot)$ is convex for every $z \in \mathbb{R}^d$ and any $w \in \mathfrak{W}$. Hence the stochastic process $(V^N_{x,\tau,w})_{(x,\tau)\in\mathcal{K}}$ has convex paths and $V(\cdot,\cdot,w)$ is convex for any $w \in \mathfrak{W}$. Moreover, the mapping $\phi(G(x,z),\cdot,\tau)$ is concave for every $(x,\tau)\in\mathcal{K}$, which implies that the stochastic process $(V^N_{x,\tau,w})_{w\in\mathfrak{W}}$ has concave paths and $V(x,\tau,\cdot)$ is concave for arbitrary $(x,\tau)\in\mathcal{K}$. Now the statement of Theorem 2 follows immediately from [25, Theorem 2.1] along with (40), (41), and (42). ∎

5. Hypotheses testing. Using the results of the previous sections, we now propose asymptotic rejection regions for tests (3) and (4) (in section 5.1) on the basis of samples $\xi^{N,v} = (\xi^v_1, \ldots, \xi^v_N)$ of ξ^v for $v = 1, \ldots, V$. We will also study tests (3) in a nonasymptotic framework (in section 5.2) deriving nonasymptotic confidence intervals on the optimal value of (1). We will denote by $0 < \beta < 1$ the maximal probability of type I error.

5.1. Asymptotic tests. Tests (3) and (4). Let us consider $V > 1$ optimization problems of the form (1) with ξ, $g(x)$, and $\mathcal{X}$ respectively replaced by ξ^v, $g_v(x) = \mathcal{R}(G_{vx})$, and $\mathcal{X}_v$ for problem v. In the above definition of g_v, G_{vx} satisfies $G_{vx}(\omega) = G_v(x, \xi^v(\omega))$. For $v = 1, \ldots, V$, let $(\xi_1^v, \ldots, \xi_N^v)$ be a sample from the distribution of ξ^v, let ϑ_*^v be the optimal value of problem v and let $z_*^v = (x_*^v, \tau_*^v, w_*^v)$ be an optimal solution of the problem, written under form (14), in variables $z = (x, \tau, w)$. Let $\hat{\vartheta}_N^v$ be the SAA estimator of the optimal value for problem $v = 1, \ldots, V$. Defining the function $H_v(z, \xi^v) = \phi(G_v(x, \xi^v), w, \tau)$ in variables $z = (x, \tau, w)$ with ϕ given by (10), we also denote by $\hat{\nu}_N^v$ the empirical estimator of the variance $\mathrm{Var}[H_v(z_*^v, \xi^v)]$ based on the sample for problem v. We assume that the samples are i.i.d. and that $\xi^{N,1}, \ldots, \xi^{N,V}$ are independent. Under the assumptions of Theorem 2 for N large we can approximate the distribution of $N^{1/2}\big(\hat{\vartheta}_N^v - \vartheta_*^v\big)/\hat{\nu}_N^v$ by the standard normal $\mathcal{N}(0,1)$.

Let us first consider the statistical tests (3)(a) and (3)(b) with $V = 2$:

$$\begin{array}{l} H_0 : \vartheta_*^1 = \vartheta_*^2 \text{ against } H_1 : \vartheta_*^1 \neq \vartheta_*^2, \\ H_0 : \vartheta_*^1 \leq \vartheta_*^2 \text{ against } H_1 : \vartheta_*^1 > \vartheta_*^2. \end{array}$$

For N large, we approximate the distribution of

$$\frac{(\hat{\vartheta}_N^1 - \hat{\vartheta}_N^2) - (\vartheta_*^1 - \vartheta_*^2)}{\sqrt{\frac{(\hat{\nu}_N^1)^2}{N} + \frac{(\hat{\nu}_N^2)^2}{N}}}$$

by the standard normal $\mathcal{N}(0,1)$ and we obtain the rejection regions

$$\begin{array}{l} \left\{(\xi_{N,1}, \xi_{N,2}) \;:\; |\hat{\vartheta}_N^1 - \hat{\vartheta}_N^2| > \sqrt{\frac{(\hat{\nu}_N^1)^2}{N} + \frac{(\hat{\nu}_N^2)^2}{N}}\,\Phi^{-1}\left(1 - \frac{\beta}{2}\right)\right\} \text{ for test (3)(a) with } V = 2, \\ \left\{(\xi_{N,1}, \xi_{N,2}) \;:\; \hat{\vartheta}_N^1 > \hat{\vartheta}_N^2 + \sqrt{\frac{(\hat{\nu}_N^1)^2}{N} + \frac{(\hat{\nu}_N^2)^2}{N}}\,\Phi^{-1}(1 - \beta)\right\} \text{ for test (3)(b) with } V = 2. \end{array} \tag{43}$$

Let us now consider test (4):

$$H_0 : \theta \in \Theta_0 \quad \text{against} \quad H_1 : \theta \in \mathbb{R}^V$$

for $\theta = (\vartheta_*^1, \ldots, \vartheta_*^V)^\top$ with Θ_0 a linear space or a closed convex cone.

Let Θ_0 be the subspace

$$\Theta_0 = \{\theta \in \mathbb{R}^V : A\theta = 0\}, \tag{44}$$

where A is a $k_0 \times V$ matrix of full rank k_0. Note that test (3)(a) can be written in this form with A a $(V-1) \times V$ matrix of rank $V - 1$. We have for θ the estimator $\hat{\theta}_N = (\hat{\vartheta}_N^1, \ldots, \hat{\vartheta}_N^V)^T$. Fixing N large, since $\xi^{N,1}, \ldots, \xi^{N,V}$ are independent, using the fact that $N^{1/2}(\hat{\vartheta}_N^v - \vartheta_*^v)/\hat{\nu}_N^v \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$, the distribution of $\hat{\theta}_N$ can be approximated by the Gaussian $\mathcal{N}(\theta, \Sigma)$ distribution with Σ the diagonal matrix $\Sigma = (1/N)\mathrm{diag}(\mathrm{Var}(H_1(z_*^1, \xi^1)), \ldots, \mathrm{Var}(H_V(z_*^V, \xi^V)))$. The log-likelihood ratio statistic for test (4) is $\Lambda = \sup_{\theta \in \Theta_0, \Sigma \succ 0} \mathcal{L}(\theta, \Sigma)/\sup_{\theta, \Sigma \succ 0} \mathcal{L}(\theta, \Sigma)$, where $\mathcal{L}(\theta, \Sigma)$ is the likelihood function for a Gaussian multivariate model. For a sample $(\tilde{\theta}_1, \ldots, \tilde{\theta}_M)$

of $\hat{\theta}_N$[1], introducing the estimators

$$\hat{\theta} = \frac{1}{M}\sum_{i=1}^{M}\tilde{\theta}_i \quad \text{and} \quad \hat{\Sigma} = \frac{1}{M-1}\sum_{i=1}^{M}\Big(\tilde{\theta}_i - \hat{\theta}\Big)\Big(\tilde{\theta}_i - \hat{\theta}\Big)^T$$

of θ and Σ, respectively, we have

$$-2\ln\Lambda = V\ln\Big(1+\frac{T^2}{M-1}\Big), \quad \text{where} \quad T^2 = V\min_{\theta\in\Theta_0}\Big(\hat{\theta}-\theta\Big)^T\hat{\Sigma}^{-1}\Big(\hat{\theta}-\theta\Big) \tag{45}$$

and when Θ_0 is of the form (44), under H_0, we have that Hotelling's T^2 squared statistic approximately has distribution $\frac{k_0(M-1)}{M-k_0}F_{k_0,M-k_0}$ (see, e.g., [16]), where $F_{p,q}$ is the Fisher–Snedecor distribution with degrees of freedom p and q. For asymptotic test (4) at confidence level β with Θ_0 given by (44), we then reject H_0 if $T^2 \geq \frac{k_0(M-1)}{M-k_0}F^{-1}_{k_0,M-k_0}(1-\beta)$, where $F^{-1}_{p,q}(\beta)$ is the β-quantile of the Fisher–Snedecor distribution.

Now take for Θ_0 the convex cone $\Theta_0 = \{\theta \in \mathbb{R}^V : A\theta \leq 0\}$, where A is a $k_0\times V$ matrix of full rank k_0 (tests (3)(b), (c) are special cases) and assume that $M \geq V+1$. Since the corresponding null hypothesis is θ *belongs to a one-sided cone*, on the basis of the sample $(\tilde{\theta}_1, \ldots, \tilde{\theta}_M)$ of $\hat{\theta}_N$, we can use [18] and we reject H_0 for large values of the statistic

$$\mathcal{U}(\Theta_0) = \|\hat{\theta}\|_S^2 - \|\Pi_S(\hat{\theta}|\Theta_0)\|_S^2 = \|\hat{\theta} - \Pi_S(\hat{\theta}|\Theta_0)\|_S^2,$$

where $S = \frac{M-1}{M}\hat{\Sigma}$, $\|x\|_S = \sqrt{x^T S^{-1}x}$, and $\Pi_S(x|A)$ is any point in A minimizing $\|y-x\|_S$ among all $y \in A$. For a type I error of at most $0 < \beta < 1$, knowing that [18]

$$\sup_{\theta\in\Theta_0,\Sigma\succ 0}\mathbb{P}\Big(\mathcal{U}(\Theta_0)\geq u\Big) \leq \texttt{Err}(u) := \frac{1}{2}\Big[\mathbb{P}\Big(G_{V-1,M-V-1}\geq u\Big) + \mathbb{P}\Big(G_{V,M-V}\geq u\Big)\Big], \tag{46}$$

where $G_{m,n} = (m/n)F_{m,n}$, we reject H_0 if $\mathcal{U}(\Theta_0) \geq u_\beta$, where u_β satisfies $\beta = \texttt{Err}(u_\beta)$ with $\texttt{Err}(\cdot)$ given by (46).

5.2. Nonasymptotic tests.

5.2.1. Risk-neutral case. Let us consider $V \geq 2$ optimization problems of the form (1) with $\mathcal{R} := \mathbb{E}$ the expectation. In this situation, several papers have derived nonasymptotic confidence intervals on the optimal value of (1): [19], using the Talagrand inequality (see [30, 31]); [28, 8], using large-deviation-type results; [17, 11, 7], using robust stochastic approximation (RSA) [21, 22], stochastic mirror descent (SMD) [17], and variants of SMD; see also [33]. In all cases, the confidence interval depends on a sample $\xi^N = (\xi_1, \ldots, \xi_N)$ of ξ and of parameters. For instance, the confidence interval $[\texttt{Low}(\Theta_2,\Theta_3,N), \texttt{Up}(\Theta_1,N)]$ with confidence level $1-\beta$ from [7] obtained using RSA depends on parameters $\Theta_1 = 2\sqrt{\ln(2/\beta)}$, $\Theta_3 = 2\sqrt{\ln(4/\beta)}$, Θ_2 satisfying $e^{1-\Theta_2^2} + e^{-\Theta_2^2/4} = \frac{\beta}{4}$, and L, M_1, M_2, $D(\mathcal{X})$ with $D(\mathcal{X})$ the maximal Euclidean distance in $\mathcal{X}$ to x_1 (the initial point of the RSA algorithm), L a uniform upper bound on $\mathcal{X}$ on the $\|\cdot\|_2$-norm of some selection (say, selection $g'(x) \in \partial g(x)$ at x) of subgradients of g, and $M_1, M_2 < +\infty$ such that for all $x \in \mathcal{X}$ it holds that

$$\begin{array}{lrcl} \text{(a)} & \mathbb{E}\Big[(G(x,\xi)-g(x))^2\Big] & \leq & M_1^2, \\ \text{(b)} & \mathbb{E}\Big[\|G'_x(x,\xi)-\mathbb{E}[G'_x(x,\xi)]\|_2^2\Big] & \leq & M_2^2 \end{array} \tag{47}$$

[1]This sample is obtained from independent samples $\xi^{N,m,v}$ of size N of ξ^v for $m = 1, \ldots, M, v = 1, \ldots, V$. More precisely, the vth component of $\tilde{\theta}_m$ is the optimal value of the SAA of problem v obtained taking sample $\xi^{N,m,v}$ of ξ^v.

for some selection $G'_x(x,\xi)$ belonging to the subdifferential $\partial_x G(x,\xi)$.

With this notation, on the basis of a sample $\xi^N=(\xi_1,\ldots,\xi_N)$ of size N of ξ and of the trajectory $x_1,\ldots,x_N$ of the RSA algorithm, setting

$$a(\Theta,N)=\frac{\Theta M_1}{\sqrt{N}} \quad \text{and} \quad b(\Theta,\mathcal{X},N)=\frac{K_1(\mathcal{X})+\Theta(K_2(\mathcal{X})-M_1)}{\sqrt{N}}, \tag{48}$$

where the constants $K_1(\mathcal{X})$ and $K_2(\mathcal{X})$ are given by

$$K_1(\mathcal{X})=\frac{D(\mathcal{X})(M_2^2+2L^2)}{\sqrt{2(M_2^2+L^2)}} \quad \text{and} \quad K_2(\mathcal{X})=\frac{D(\mathcal{X})M_2^2}{\sqrt{2(M_2^2+L^2)}}+2D(\mathcal{X})M_2+M_1,$$

the lower bound $\texttt{Low}(\Theta_2,\Theta_3,N)$ is

$$\texttt{Low}(\Theta_2,\Theta_3,N)=\frac{1}{N}\sum_{t=1}^{N}G(x_t,\xi_t)-b(\Theta_2,\mathcal{X},N)-a(\Theta_3,N) \tag{49}$$

and the upper bound $\texttt{Up}(\Theta_1,N)$ is

$$\texttt{Up}(\Theta_1,N)=\frac{1}{N}\sum_{t=1}^{N}G(x_t,\xi_t)+a(\Theta_1,N). \tag{50}$$

More precisely, we have $\mathbb{P}(\vartheta_*<\texttt{Low}(\Theta_2,\Theta_3,N))\le\beta/2$ and $\mathbb{P}(\vartheta_*>\texttt{Up}(\Theta_1,N)\le\beta/2$.

Test (3)(a). Using the bounds Low and Up or one of the aforementioned cited procedures, we can determine for the optimization problem $v\in\{1,\ldots,V\}$ (stochastic) lower and upper bounds on ϑ_*^v which we will respectively denote by $\texttt{Low}_v$ and $\texttt{Up}_v$ for short, such that $\mathcal{P}(\vartheta_*^v<\texttt{Low}_v)\le\frac{\beta}{2V}$ and $\mathcal{P}(\vartheta_*^v>\texttt{Up}_v)\le\frac{\beta}{2V}$.

We define for test (3)(a) the rejection region $\mathcal{W}_{(3)(\mathrm{a})}$ to be the set of samples such that the realizations of the confidence intervals $[\texttt{Low}_v,\texttt{Up}_v]$, $v=1,\ldots,V$, on the optimal values have no intersection, i.e.,

$$\begin{aligned}\mathcal{W}_{(3)(\mathrm{a})} &= \left\{(\xi^{N,1},\ldots,\xi^{N,V}) \;:\; \bigcap_{v=1}^{V}\Big[\texttt{Low}_v,\texttt{Up}_v\Big]=\emptyset\right\}\\ &= \left\{(\xi^{N,1},\ldots,\xi^{N,V}) \;:\; \max_{v=1,\ldots,V}\ \texttt{Low}_v>\min_{v=1,\ldots,V}\ \texttt{Up}_v\right\}.\end{aligned}$$

If H_0 holds, writing $\vartheta_*=\vartheta_*^1=\vartheta_*^2=\cdots=\vartheta_*^V$, we have

$$\begin{aligned}&\mathbb{P}\Big(\max_{v=1,\ldots,V}\ \texttt{Low}_v>\min_{v=1,\ldots,V}\ \texttt{Up}_v\Big)=\mathbb{P}\left(\max_{v=1,\ldots,V}\ \Big[\texttt{Low}_v-\vartheta_*\Big]+\max_{v=1,\ldots,V}\ \Big[\vartheta_*-\texttt{Up}_v\Big]>0\right)\\ &\le\textstyle\sum_{v=1}^{V}\Big[\mathbb{P}\Big(\texttt{Low}_v-\vartheta_*^v>0\Big)+\mathbb{P}\Big(\vartheta_*^v-\texttt{Up}_v>0\Big)\Big]\le\beta\end{aligned}$$

and $\mathcal{W}_{(3)(\mathrm{a})}$ is a rejection region for (3)(a) yielding a probability of type I error of at most β. Moreover, as stated in the following lemma, if H_0 does not hold and if two optimal values are sufficiently distant then the probability of accepting H_0 will be small.

LEMMA 6. *Consider test* (3)(a) *with rejection region* $\mathcal{W}_{(3)(\mathrm{a})}$. *If for some* $p,q\in\{1,\ldots,V\}$ *with* $p\neq q$ *we have almost surely* $\vartheta_*^p>\vartheta_*^q+\texttt{Up}_p-\texttt{Low}_p+\texttt{Up}_q-\texttt{Low}_q$, *then the probability of accepting* H_0 *is not larger than* $\frac{\beta}{V}$.

Proof. We first check that

$$
\left.\begin{array}{ll}
\vartheta_*^p > \vartheta_*^q + \mathtt{Up}_q - \mathtt{Low}_q + \mathtt{Up}_p - \mathtt{Low}_p & \text{(a)} \\
\mathtt{Low}_q \le \vartheta_*^q & \text{(b)} \\
\vartheta_*^p \le \mathtt{Up}_p & \text{(c)}
\end{array}\right\} \Rightarrow \mathtt{Up}_q < \mathtt{Low}_p. \tag{51}
$$

Indeed, if (51)(a), (b), and (c) hold, then

$$
\mathtt{Up}_q = \mathtt{Low}_q + \mathtt{Up}_q - \mathtt{Low}_q \overset{(51)(b)}{\le} \vartheta_*^q + \mathtt{Up}_q - \mathtt{Low}_q \overset{(51)(a)}{<} \vartheta_*^p + \mathtt{Low}_p - \mathtt{Up}_p \overset{(51)(c)}{\le} \mathtt{Low}_p.
$$

Assume now that $\vartheta_*^p > \vartheta_*^q + \mathtt{Up}_q - \mathtt{Low}_q + \mathtt{Up}_p - \mathtt{Low}_p$. Since $\mathtt{Up}_q < \mathtt{Low}_p$ implies that H_0 is rejected, we get

$$
\begin{array}{rcl}
\mathbb{P}\Big(\text{reject } H_0\Big) & \ge & \mathbb{P}\Big(\mathtt{Up}_q < \mathtt{Low}_p\Big) \overset{(51)}{\ge} \mathbb{P}\Big(\Big\{\mathtt{Low}_q \le \vartheta_*^q\Big\} \bigcap \Big\{\vartheta_*^p \le \mathtt{Up}_p\Big\}\Big) \\
& \ge & \mathbb{P}\Big(\mathtt{Low}_q \le \vartheta_*^q\Big) + \mathbb{P}\Big(\vartheta_*^p \le \mathtt{Up}_p\Big) - 1 \ge 1 - \frac{\beta}{V},
\end{array}
$$

which achieves the proof of the lemma. ∎

Simlarly, for tests (3)(b) and (c), we respectively define the rejection regions $\mathcal{W}_{(3)(\mathrm{b})}$ and $\mathcal{W}_{(3)(\mathrm{c})}$ by

$$
\begin{array}{rcl}
\mathcal{W}_{(3)(\mathrm{b})} & = & \left\{(\xi^{N,1}, \ldots, \xi^{N,V}) \;:\; \exists 1 \le q \ne p \le V \text{ such that } \mathtt{Low}_p > \mathtt{Up}_q\right\}, \\
\mathcal{W}_{(3)(\mathrm{c})} & = & \left\{(\xi^{N,1}, \ldots, \xi^{N,V}) \;:\; \exists v \in \{1, \ldots, V-1\} \text{ such that } \mathtt{Low}_v > \mathtt{Up}_{v+1}\right\},
\end{array}
$$

yielding a probability of type I error of at most β provided $[\mathtt{Low}_v, \mathtt{Up}_v]$ is a confidence interval with confidence level at least $1 - \beta/2(V-1)$ for problem v:

$$
\mathbb{P}(\vartheta_*^v < \mathtt{Low}_v) \le \beta/2(V-1) \quad \text{and} \quad \mathbb{P}(\vartheta_*^v > \mathtt{Up}_v) \le \beta/2(V-1). \tag{52}
$$

Similarly to Lemma 6, we can bound from above the probability of type I error for test (3)(b) if $\vartheta_*^p > \vartheta_*^q + \mathtt{Up}_p - \mathtt{Low}_p + \mathtt{Up}_q - \mathtt{Low}_q$ almost surely and for test (3)(c) if $\vartheta_*^v > \vartheta_*^{v+1} + \mathtt{Up}_v - \mathtt{Low}_v + \mathtt{Up}_{v+1} - \mathtt{Low}_{v+1}$ almost surely.

Remark 5. Though $\mathtt{Low}$ and $\mathtt{Up}$ are stochastic, for bounds (49) and (50), the difference $\mathtt{Up} - \mathtt{Low} = a(\Theta_1, N) + b(\Theta_2, \mathcal{X}, N) + a(\Theta_3, N)$ is deterministic and inequalities $\vartheta_*^p > \vartheta_*^q + \mathtt{Up}_p - \mathtt{Low}_p + \mathtt{Up}_q - \mathtt{Low}_q$ in Lemma 6 and $\vartheta_*^v > \vartheta_*^{v+1} + \mathtt{Up}_v - \mathtt{Low}_v + \mathtt{Up}_{v+1} - \mathtt{Low}_{v+1}$ are deterministic too.

5.2.2. Risk-averse case. Consider $K \ge 2$ optimization problems of the form (1). For such problems, nonasymptotic confidence intervals $[\mathtt{Low}, \mathtt{Up}]$ on the optimal value ϑ_* were derived in [7] and [11] using RSA and SMD, taking for $\mathcal{R}$ an extended polyhedral risk measure (introduced in [9]) in [7] and $\mathcal{R} = \mathsf{AVaR}_\alpha$ and $G(x, \xi) = \xi^\top x$ in [11]. With such confidence intervals at hand, we can use the developments of the previous section for testing hypotheses (3). However, the analysis in [7] assumes boundedness of the feasible set of the optimization problem defining the risk measure; an assumption that can be enforced for risk measure $\mathcal{R}$ given by (53). We provide in this situation formulas for the constants L, M_1, and M_2 defined in the previous section, necessary to compute the bounds from [7]. These constants are slightly refined versions of the constants given in section 4.2 of [11] for the special case in which $\mathcal{R} = \mathsf{AVaR}_\alpha$ and $G(x, \xi) = \xi^\top x$.

We assume here that the set Ξ is compact, $G(\cdot,\cdot)$ is continuous, for every $x \in \mathcal{X}$ the distribution of G_x is continuous, and that the set $\mathfrak{W} = \{w\}$ is a singleton, i.e.,

$$\mathcal{R}(Z) = w_0 \mathbb{E}[Z] + \sum_{i=1}^{k} w_i \mathsf{AVaR}_{\alpha_i}(Z) \tag{53}$$

for some $w \in \Delta_{k+1}$. Consequently problem (1) can be written as

$$\vartheta_* = \inf_{(x,\tau)\in\mathcal{X}\times\mathbb{R}^k} \big\{\mathbb{E}[\phi(G_x,\tau)] = \mathbb{E}[H(x,\tau,\xi)]\big\}, \tag{54}$$

where $\phi(G_x,\tau)$ is defined in (10), with vector w omitted, and

$$H(x,\tau,\xi) := w_0 G(x,\xi) + \sum_{i=1}^{k} w_i \left(\tau_i + \frac{1}{1-\alpha_i}[G(x,\xi)-\tau_i]_+\right).$$

For a given $x \in \mathcal{X}$ the minimum in (54) is attained at $\tau_i = F_x^{-1}(\alpha_i)$, $i = 1,\dots,k$, where F_x is the c.d.f. of G_x. Therefore, using the lower and upper bounds from [11] for the quantile of a continuous distribution with finite mean and variance, we can restrict τ to the compact set $\mathcal{T} = [\underline{\tau},\bar{\tau}] \subset \mathbb{R}^k$, where

$$\begin{array}{rcl} \underline{\tau}_i &=& \min_{x\in\mathcal{X}} \mathbb{E}[G_x] - \sqrt{\frac{1-\alpha_i}{\alpha_i}}\sqrt{\max_{x\in\mathcal{X}} \mathrm{Var}(G_x)}, \\ \bar{\tau}_i &=& \max_{x\in\mathcal{X}} \mathbb{E}[G_x] + \sqrt{\frac{\alpha_i}{1-\alpha_i}}\sqrt{\max_{x\in\mathcal{X}} \mathrm{Var}(G_x)} \end{array} \tag{55}$$

for $i = 1,\dots,k$. This implies that we can take $D(\mathcal{X}\times\mathcal{T}) = \sqrt{D(\mathcal{X})^2 + \|\bar{\tau}-\underline{\tau}\|_2^2}$.

Computation of M_1. Setting

$$M_0 := \max_{(x,\xi)\in\mathcal{X}\times\Xi} G(x,\xi) \text{ and } m_0 := \min_{(x,\xi)\in\mathcal{X}\times\Xi} G(x,\xi),$$

we have for $(x,\tau) \in \mathcal{X}\times\mathcal{T}$ that $|G_x - \mathbb{E}[G_x]| \le M_0 - m_0$ and $|[G_x-\tau_i]_+ - \mathbb{E}[G_x-\tau_i]_+| \le M_0 - \underline{\tau}_i$, which implies that almost surely

$$|\phi(G_x,\tau) - \mathbb{E}[\phi(G_x,\tau)]| \le M_1 := w_0(M_0-m_0) + \sum_{i=1}^{k} \frac{w_i}{1-\alpha_i}(M_0 - \underline{\tau}_i).$$

Computation of M_2 and L. We have $H'_{x,\tau}(x,\tau,\xi) = [H'_x(x,\tau,\xi); H'_\tau(x,\tau,\xi)]$ with

$$\begin{array}{rcl} H'_x(x,\tau,\xi) &=& w_0 G'_x(x,\xi) + \sum_{i=1}^k \frac{w_i}{1-\alpha_i} G'_x(x,\xi)\mathbf{1}_{G(x,\xi)\ge\tau_i}, \\ H'_\tau(x,\tau,\xi) &=& (w_i(1-\frac{1}{1-\alpha_i}\mathbf{1}_{G(x,\xi)\ge\tau_i}))_{i=1,\dots,k}. \end{array}$$

We assume that for every $x \in \mathcal{X}$ the stochastic subgradients $G'_x(x,\xi)$ are almost surely bounded and we denote by $\underline{m}$ and $\overline{M}$ vectors such that almost surely $\underline{m} \le G'_x(x,\xi) \le \overline{M}$. Then, for $(x,\tau) \in \mathcal{X}\times\mathcal{T}$, setting $b_i = \max(w_0\overline{M}_i, (w_0 + \sum_{j=1}^k \frac{w_j}{1-\alpha_j})\overline{M}_i)$ and $a_i = \min(w_0\underline{m}_i, (w_0 + \sum_{j=1}^k \frac{w_j}{1-\alpha_j})\underline{m}_i)$, we have

$$\begin{array}{l} \|\mathbb{E}[H'_{x,\tau}(x,\tau,\xi)]\|_2^2 \le L^2 := \sum_{i=1}^m \max(a_i^2,b_i^2) + \sum_{i=1}^k w_i^2 \max\left(1, \frac{\alpha_i^2}{(1-\alpha_i)^2}\right), \\ \mathbb{E}\|H'_{x,\tau}(x,\tau,\xi) - \mathbb{E}[H'_{x,\tau}(x,\tau,\xi)]\|_2^2 \le M_2^2 := \sum_{i=1}^m (a_i-b_i)^2 + \sum_{i=1}^k \left(\frac{w_i}{1-\alpha_i}\right)^2. \end{array}$$

In some cases, the above formulas for $\bar{\tau}$, $\underline{\tau}$, L, M_1, and M_2 can be simplified, as is shown by the following example.

Example 3. Let $k = 1$ in (53) and $G(x, \xi) = \xi^T x$, where ξ is a random vector with mean μ and covariance matrix Σ. In this case $\min_{x\in\mathcal{X}} \mathbb{E}[G_x]$ and $\max_{x\in\mathcal{X}} \mathbb{E}[G_x]$ are convex optimization problems with linear objective functions and, denoting by U_1 the quantity $\max_{x\in\mathcal{X}} \|x\|_1$ or an upper bound on this quantity, we can replace $\max_{x\in\mathcal{X}} \mathrm{Var}(G_x)$ by $U_1^2 \max_i \Sigma(i,i)$ in the expressions of $\underline{\tau}_i$ and $\bar{\tau}_i$. Computing M_0 and m_0 also amounts to solving convex optimization problems with linear objective. Assume also that almost surely $\|\xi\|_\infty \le U_2$ for some $0 < U_2 < +\infty$. We have $|G_x - \mathbb{E}[G_x]| \le 2U_1U_2$ and $|[G_x - \tau]_+ - \mathbb{E}[G_x - \tau]_+| \le U_1U_2 - \underline{\tau}$, which shows that we can take $M_1 = 2w_0U_1U_2 + \frac{w_1}{1-\alpha_1}(U_1U_2 - \underline{\tau})$. We have $\mathbb{E}[H'_\tau(x,\tau,\xi)] = w_1(1 - \frac{\mathbb{P}(\xi^T x \ge \tau)}{1-\alpha_1})$ so that $|\mathbb{E}[H'_\tau(x,\tau,\xi)]| \le w_1 \max(1, \frac{\alpha_1}{1-\alpha_1})$ and $\|\mathbb{E}[H'_x(x,\tau,\xi)]\|_2^2 \le m(w_0 + \frac{w_1}{1-\alpha_1})^2 U_2^2$, i.e., we can take

$$L^2 = w_1^2 \max\left(1, \frac{\alpha_1^2}{(1-\alpha_1)^2}\right) + m\left(w_0 + \frac{w_1}{1-\alpha_1}\right)^2 U_2^2.$$

Next, for all $\xi_0 \in \Xi$ we have

$$\begin{array}{rcl} |H'_\tau(x,\tau,\xi_0) - \mathbb{E}[H'_\tau(x,\tau,\xi)]| & = & \frac{w_1(1-\mathbb{P}(\xi^T x \ge \tau))}{1-\alpha_1} \text{ if } \xi_0^T x \ge \tau, \\ & = & \frac{w_1 \mathbb{P}(\xi^T x \ge \tau)}{1-\alpha_1} \text{ otherwise,} \end{array}$$

implying that $|H'_\tau(x,\tau,\xi_0) - \mathbb{E}[H'_\tau(x,\tau,\xi)]| \le \frac{w_1}{1-\alpha_1}$.

Since $\|H'_x(x,\tau,\xi_0) - \mathbb{E}[H'_x(x,\tau,\xi)]\|_\infty$ is bounded from above by $2(w_0 + \frac{w_1}{1-\alpha_1})U_2$, we can take

$$M_2^2 = \frac{w_1^2}{(1-\alpha_1)^2} + 4m\left(w_0 + \frac{w_1}{1-\alpha_1}\right)^2 U_2^2.$$

In the special case in which $\mathcal{X} = \{x_*\}$ is a singleton, defining $\eta = \xi^T x_*$, we have $\vartheta_* = \mathcal{R}(\eta)$, $H(x,\tau,\xi) = H(x_*,\tau,\xi)$, $H'_x(x,\tau,\xi) = 0$ almost surely and the above computations show that we can take

$$L = w_1 \max(1, \tfrac{\alpha_1}{1-\alpha_1}),\ M_1 = w_0(b_0 - a_0) + \tfrac{w_1}{1-\alpha_1}(b_0 - \underline{\tau}),\ \text{and } M_2 = \tfrac{w_1}{1-\alpha_1}, \tag{56}$$

where $\underline{\tau} = \mathbb{E}[\eta] - \sqrt{\frac{1-\alpha_1}{\alpha_1}}\sqrt{\mathrm{Var}(\eta)}$ with a_0, b_0 satisfying $a_0 \le \eta \le b_0$ almost surely.

Discussion: Asymptotic versus nonasymptotic tests and confidence intervals for the optimal value of (1). The nonasymptotic tests of this and the previous section do not require the independence of $\xi^{N,1}, \dots, \xi^{N,V}$ and are valid for any sample size N. On the contrary, the asymptotic tests are valid as the sample size N goes to infinity and theory does not tell us for which values of N the Gaussian distribution "approximates well" the optimal value of SAA (2) of (1). Moreover, experiments in [8] and in the next section show that this value of N depends on dimension m of x.

A (known) drawback of nonasymptotic confidence bounds is their conservativeness. On the one hand, this conservativeness allows us, when the sample size N is not much larger than problem dimension m, to provide confidence sets of the prescribed risk, which asymptotic confidence intervals (based on the CLT of section 3) fail to do; see [8]. On the other hand, for testing problems (3), (4), nonasymptotic rejection regions can lead to large probabilities of type II errors. Even if the asymptotic tests of section 5.1 are valid as the sample size tends to infinity, they can work well in practice for small sample sizes ($N = 20$) and problems of small to moderate size (m up to 500); see the numerical simulations of section 6. The derivation of less conservative nonasymptotic confidence sets (especially the lower bound) is an interesting future research goal.

6. Numerical experiments.

6.1. Comparing the risk of two distributions: Test (3) with a singleton for $\mathcal{X}$. We consider test (3) with $V = 2$ and $\mathcal{X}$ a singleton. We use the rejection regions given in section 5.2 (resp., given by (43)) in the nonasymptotic (resp., asymptotic) case. In this situation, the test aims at comparing the risk of two distributions. We use the notation $\mathcal{N}(m_0, \sigma^2; a_0, b_0)$ for the normal distribution with mean m_0 and variance σ^2 conditional on this random variable being in $[a_0, b_0]$ (truncated normal distribution with support $[a_0, b_0]$). More precisely, we compare the risks $\mathcal{R}(\xi_1)$ and $\mathcal{R}(\xi_2)$ of two truncated normal (loss) distributions ξ_1 and ξ_2 with support $[a_0, b_0] = [0, 30]$ in three cases: (I) $\xi_1 \sim \mathcal{N}(10, 1; 0, 30)$, $\xi_2 \sim \mathcal{N}(20, 1; 0, 30)$, (II) $\xi_1 \sim \mathcal{N}(5, 1; 0, 30)$, $\xi_2 \sim \mathcal{N}(10, 25; 0, 30)$, and (III) $\xi_1 \sim \mathcal{N}(10, 49; 0, 30)$, $\xi_2 \sim \mathcal{N}(14, 0.25; 0, 30)$. For these three cases, the densities of ξ_1 and ξ_2 are represented in Figure 1 (left for (I), middle for (II), right for (III)).

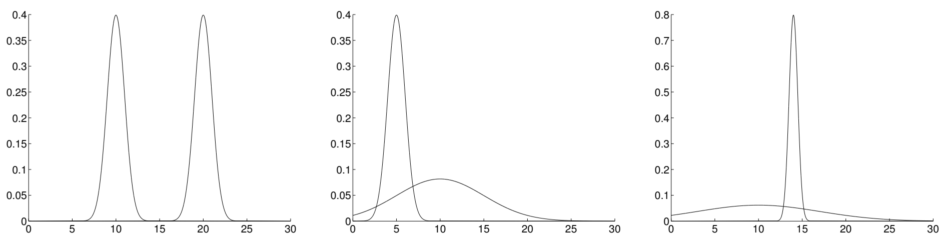

FIG. 1. *Densities of truncated normal loss distributions ξ_1 and ξ_2. Left plot: $\xi_1 \sim \mathcal{N}(10, 1; 0, 30)$ and $\xi_2 \sim \mathcal{N}(20, 1; 0, 30)$. Middle plot: $\xi_1 \sim \mathcal{N}(5, 1; 0, 30)$ and $\xi_2 \sim \mathcal{N}(10, 25; 0, 30)$. Right plot: $\xi_1 \sim \mathcal{N}(10, 49; 0, 30)$ and $\xi_2 \sim \mathcal{N}(14, 0.25; 0, 30)$.*

We take for $\mathcal{R}$ the risk measure $\mathcal{R}(\xi) = w_0\mathbb{E}[\xi] + w_1 \mathsf{AVaR}_\alpha(\xi)$ for $0 < \alpha < 1$, where $w_0, w_1 \geq 0$ with $w_0 + w_1 = 1$. We assume that only the support $[a_0, b_0]$ of ξ_1 and ξ_2 and two samples ξ_1^N and ξ_2^N of size N of ξ_1 and ξ_2, respectively, are known. Since the distribution of ξ has support $[a_0, b_0]$, we can write

$$\mathcal{R}(\xi) = \min_{\tau \in [a_0, b_0]} w_0\mathbb{E}[\xi] + w_1 \left(\tau + \frac{1}{1-\alpha}\mathbb{E}[\xi - \tau]_+ \right), \tag{57}$$

which is of the form (1) with a risk-neutral objective function, $G(\tau, \xi) = w_0\xi + w_1\tau + \frac{w_1}{1-\alpha}[\xi - \tau]_+$, and $\mathcal{X}$ the compact set $\mathcal{X} = [a_0, b_0] = [0, 30]$.

It follows that the RSA algorithm can be used to estimate $\mathcal{R}(\xi_1)$ and $\mathcal{R}(\xi_2)$ and to compute the confidence bounds (49) and (50) with L, M_1, and M_2 given by (56). In these formulas, we replace $\underline{\tau}$ by its lower bound 0 since we do not assume that the mean and standard deviation of ξ_1 and ξ_2 are known. We obtain $L = w_1 \max(1, \frac{\alpha}{1-\alpha})$, $M_2 = \frac{w_1}{1-\alpha}$, and $M_1 = 30(w_0 + \frac{w_1}{1-\alpha})$.

Let us first consider **Case (I)**. We illustrate Corollary 3 computing the empirical estimation $\mathcal{R}(\hat{F}_{N,1})$ of $\mathcal{R}(\xi_1)$ on 200 samples of size N of $\xi_1 \sim \mathcal{N}(10, 1; 0, 30)$ for $w_0 = 0.1$, $w_1 = 0.9$, and various values of α and of the sample size N. For this experiment, the QQ-plots of the empirical distribution of $\mathcal{R}(\hat{F}_{N,1})$ versus the normal distribution with parameters the empirical mean and standard deviation of this empirical distribution are reported in the supplementary materials of this article. We see that even for small values of $1-\alpha$ and N as small as 20, the distribution of $\mathcal{R}(\hat{F}_{N,1})$ is well approximated by a Gaussian distribution: for $N = 20$ the Jarque–Bera test accepts the hypothesis of normality at the significance level 0.05 for $1-\alpha = 0.01$ and $1-\alpha = 0.5$.

TABLE 1
Estimation of the risk measure value $\mathcal{R}(\xi_1)$ for $\xi_1 \sim \mathcal{N}(10, 1; 0, 30)$ using SAA and RSA for various values of (w_0, w_1, α) and various sample sizes N.

			Sample size N						
(w_0, w_1)	$1-\alpha$	Method	20	50	10^2	10^3	10^4	10^5	10^6
(0.1, 0.9)	10^{-2}	SAA	11.71	12.00	12.21	12.37	12.40	12.40	12.40
(0.1, 0.9)	10^{-2}	RSA	14.35	14.26	14.16	13.46	12.75	12.51	12.43
(0.1, 0.9)	0.1	SAA	11.51	11.50	11.54	11.58	11.58	11.58	11.58
(0.1, 0.9)	0.1	RSA	20.50	16.78	15.10	12.61	11.90	11.68	11.61
(0.1, 0.9)	0.5	SAA	10.71	10.69	10.72	10.72	10.72	10.72	10.72
(0.1, 0.9)	0.5	RSA	11.42	11.12	11.02	10.81	10.75	10.73	10.72
(0.9, 0.1)	10^{-2}	SAA	10.19	10.23	10.25	10.26	10.27	10.27	10.27
(0.9, 0.1)	10^{-2}	RSA	10.49	10.48	10.47	10.38	10.31	10.28	10.27
(0.9, 0.1)	0.1	SAA	10.17	10.16	10.19	10.18	10.18	10.18	10.18
(0.9, 0.1)	0.1	RSA	10.34	10.28	10.27	10.20	10.18	10.18	10.18
(0.9, 0.1)	0.5	SAA	10.09	10.07	10.08	10.08	10.08	10.08	10.08
(0.9, 0.1)	0.5	RSA	10.17	10.11	10.12	10.09	10.08	10.08	10.08

We fix again the distribution $\xi_1 \sim \mathcal{N}(10, 1; 0, 30)$ and approximately compute $\mathcal{R}(\xi_1)$ for various values of (w_0, w_1, α, N) using the RSA and SAA methods on samples ξ_1^N of size N of ξ_1. For a sample of size N of ξ_1, let $\hat{\mathcal{R}}_{N,\mathtt{RSA}}(\xi_1)$ and $\hat{\mathcal{R}}_{N,\mathtt{SAA}}(\xi_1) = \mathcal{R}(\hat{F}_{N,1})$ be these estimations using RSA and SAA, respectively. For fixed (w_0, w_1, α, N), we generate 200 samples of size N of ξ_1 and for each sample we compute $\hat{\mathcal{R}}_{N,\mathtt{RSA}}(\xi_1)$ and $\hat{\mathcal{R}}_{N,\mathtt{SAA}}(\xi_1)$ and report in Table 1 the average of these values for $N \in \{20, 50, 100, 10^3, 10^4, 10^5, 10^6\}$. Considering that $\mathcal{R}(\xi_1)$ is the value obtained using SAA for $N = 10^6$, we observe that RSA correctly approximates $\mathcal{R}(\xi_1)$ as N grows and that the estimation of $\mathbb{E}[\hat{\mathcal{R}}_{N,\mathtt{SAA}}(\xi_1)]$ (resp., $\mathbb{E}[\hat{\mathcal{R}}_{N,\mathtt{RSA}}(\xi_1)]$) increases (resp., decreases) with the sample size N, as expected. We also naturally observe that the more weight is given to the AVaR and the smaller $1-\alpha$, the more difficult it is to estimate the risk measure, i.e., the more distant the expectation of the approximation is to the optimal value and the larger the sample size needs to be to obtain an expected approximation with given accuracy.

We now study for Case (I) the test

$$H_0 : \mathcal{R}(\xi_1) = \mathcal{R}(\xi_2) \quad \text{against} \quad H_1 : \mathcal{R}(\xi_1) \neq \mathcal{R}(\xi_2). \tag{58}$$

We first fix $(w_0, w_1) = (0.1, 0.9)$ and report in Tables 2 and 3 for various values of the pair (α, N) the average nonasymptotic and asymptotic confidence bounds for $\mathcal{R}(\xi_1)$ and $\mathcal{R}(\xi_2)$ when $\xi_1 \sim \mathcal{N}(10, 1; 0, 30)$ and $\xi_2 \sim \mathcal{N}(20, 1; 0, 30)$.[2] We observe that even for small values of the sample size and of the confidence level $1-\alpha$, the asymptotic confidence interval is of small width and its bounds are close to the risk measure value. For RSA, a large sample is needed to obtain a confidence interval of small width, especially when $1-\alpha$ is small.

[2]The nonasymptotic confidence interval is given by (49), (50). Recalling that $\mathcal{R}(\xi)$ is the optimal value of optimization problem (57), which is of the form (1), we compute for $\mathcal{R}(\xi)$ the asymptotic confidence interval $[\hat{\vartheta}_N - \Phi^{-1}(1-\beta/2)\frac{\hat{\nu}_N}{\sqrt{N}}, \hat{\vartheta}_N + \Phi^{-1}(1-\beta/2)\frac{\hat{\nu}_N}{\sqrt{N}}]$, where $\hat{\vartheta}_N$ is the optimal value of the SAA of (57). See also [12, 14] for the computation of asymptotic confidence intervals on the optimal value of a risk-neutral optimization problem (observe that for reformulation (57) of the risk measure, the objective is risk-neutral). Note that in this case the optimal value $\hat{\tau}_N$ of the SAA problem is the α-quantile of the distribution of ξ (no optimization step is necessary to solve the SAA problem).

Table 2

Average values of the asymptotic and nonasymptotic confidence bounds for $\mathcal{R}(\xi_1)$ and $\mathcal{R}(\xi_2)$ when $\xi_1 \sim \mathcal{N}(10, 1; 0, 30)$ and $\xi_2 \sim \mathcal{N}(20, 1; 0, 30)$, $1-\alpha = 0.1$. For $\mathcal{R}(\xi_i)$, the average asymptotic confidence interval is [`L-Ai`, `U-Ai`] *and the average nonasymptotic confidence interval is* [`L-NAi`, `U-NAi`].

N	`L-A1`	`U-A1`	`L-NA1`	`U-NA1`	`L-A2`	`U-A2`	`L-NA2`	`U-NA2`
50	11.20	11.94	-347	146	21.30	21.97	-335	158
10^3	11.49	11.67	-68.65	41.71	21.49	21.67	-58.37	51.99
10^4	11.55	11.61	-13.79	21.11	21.55	21.61	-3.71	31.19
10^5	11.57	11.59	3.56	14.59	21.57	21.59	13.58	24.62
1.5×10^5	11.57	11.59	5.03	14.04	21.57	21.59	15.05	24.06

Table 3

Average values of the asymptotic and nonasymptotic confidence bounds for $\mathcal{R}(\xi_1)$ and $\mathcal{R}(\xi_2)$ when $\xi_1 \sim \mathcal{N}(10, 1; 0, 30)$ and $\xi_2 \sim \mathcal{N}(20, 1; 0, 30)$, $1-\alpha = 0.5$. For $\mathcal{R}(\xi_i)$, the average asymptotic confidence interval is [`L-Ai`, `U-Ai`] *and the average nonasymptotic confidence interval is* [`L-NAi`, `U-NAi`].

N	`L-A1`	`U-A1`	`L-NA1`	`U-NA1`	`L-A2`	`U-A2`	`L-NA2`	`U-NA2`
50	10.45	10.98	-230.29	40.16	20.47	21.00	-220.25	50.21
10^3	10.65	10.77	-43.18	17.30	20.65	20.77	-33.18	27.29
10^4	10.70	10.74	-6.33	12.80	20.70	20.74	3.68	22.80
10^5	10.71	10.72	5.33	11.38	20.71	20.72	15.33	21.38
1.5×10^5	10.71	10.72	6.32	11.26	20.71	20.72	16.32	21.25

For all the remaining tests of this section, we choose $\beta = 0.1$ for the maximal type I error and $1-\alpha = 0.1$. Since in Case (I) we have $\mathcal{R}(\xi_1) \neq \mathcal{R}(\xi_2)$ (see Figure 1), from this experiment we expect to obtain a large probability of type II error using the nonasymptotic tests of section 5.2 based on the confidence intervals computed using RSA, unless the sample size is very large. More precisely, we compute the probability of type II error[3] for (58) considering asymptotic and nonasymptotic rejection regions using various sample sizes $N \in \{20, 50, 100, 1\,000, 5\,000, 10\,000, 20\,000, 50\,000, 100\,000, 130\,000, 150\,000\}$, taking $1-\alpha = 0.1$ and $(w_0, w_1) \in \{(0, 1), (0.1, 0.9), (0.2, 0.8), (0.3, 0.7), (0.4, 0.6), (0.5, 0.5), (0.6, 0.4), (0.7, 0.3), (0.8, 0.2), (0.9, 0.1)\}$. For fixed N, the probability of type II error is estimated using 100 samples of size N of ξ_1 and ξ_2. Using the asymptotic rejection region, we reject H_0 for all realizations and all parameter combinations, meaning that the probability of type II error is null (since H_1 holds for all parameter combinations). For the nonasymptotic test, the probabilities of type II errors are reported in Table 4. For sample sizes less than 5 000, the probability of type II error is always 1 (the nonasymptotic test always takes the wrong decision) and the larger w_1 the larger the sample size N needs to be to obtain a probability of type II error of zero. In particular, if $w_1 = 1$ (we estimate the AVaR_α of the distribution) as much as 150 000 observations are needed to obtain a null probability of type II error. However, if the sample size is sufficiently large, both tests always take the correct decision $\mathcal{R}(\xi_1) \neq \mathcal{R}(\xi_2)$.

Given (possibly small) samples of size N of ξ_1 and ξ_2, to know which of the two risks $\mathcal{R}(\xi_1)$ and $\mathcal{R}(\xi_2)$ is the smallest, we now consider the test

$$H_0 : \mathcal{R}(\xi_1) \geq \mathcal{R}(\xi_2) \quad \text{against} \quad H_1 : \mathcal{R}(\xi_1) < \mathcal{R}(\xi_2). \tag{59}$$

Computing $\mathcal{R}(\xi_1)$ and $\mathcal{R}(\xi_2)$ with a very large sample (of size 10^6) of ξ_1 and ξ_2

[3]All computed probabilities of type II error are empirical probabilities. However, for short, we will use in what follows the term probabilities of type II error.

Table 4
Empirical probabilities of type II *error for tests* (58) *and* (59) *using a nonasymptotic rejection region when* $\xi_1 \sim \mathcal{N}(10, 1; 0, 30)$, $\xi_2 \sim \mathcal{N}(20, 1; 0, 30)$, *and* $1-\alpha = 0.1$.

	Sample size N						
(w_0, w_1)	5 000	10 000	20 000	50 000	100 000	130 000	150 000
(0.0, 1.0)	1	1	1	1	1	1	0
(0.1, 0.9)	1	1	1	1	1	0	0
(0.2, 0.8)	1	1	1	1	0	0	0
(0.3, 0.7)	1	1	1	1	0	0	0
(0.4, 0.6)	1	1	1	1	0	0	0
(0.5, 0.5)	1	1	1	0	0	0	0
(0.6, 0.4)	1	1	1	0	0	0	0
(0.7, 0.3)	1	1	0	0	0	0	0
(0.8, 0.2)	1	0	0	0	0	0	0
(0.9, 0.1)	0	0	0	0	0	0	0

either with SAA or RSA or looking at Figure 1, we know that $\mathcal{R}(\xi_1) < \mathcal{R}(\xi_2)$. We again analyze the probability of type II error using the asymptotic and nonasymptotic rejection regions when the decision is taken on the basis of a much smaller sample. For the nonasymptotic test, the empirical probabilities of type II error for various sample sizes (estimated, for fixed N, using 100 samples of size N of ξ_1 and ξ_2) are exactly those obtained for test (58) and are given in Table 4. The asymptotic test again always takes the correct decision $\mathcal{R}(\xi_1) < \mathcal{R}(\xi_2)$ while a large sample size is needed to always take the correct decision using the nonasymptotic test (as large as 150 000 for $w_1 = 1$).

We now consider tests (58) and (59) for **Case (II)**. In this case, there is a larger overlap between the distributions of ξ_1 and ξ_2. However, from Figure 1 and computing $\mathcal{R}(\xi_1)$ and $\mathcal{R}(\xi_2)$ with a very large sample (say of size 10^6) of ξ_1 and ξ_2 either using SAA or RSA, we check that we have again $\mathcal{R}(\xi_2) > \mathcal{R}(\xi_1)$ for all values of (w_0, w_1). The empirical probabilities of type II error are null for the asymptotic test for all sample sizes N tested while, for the nonasymptotic test, the probabilities of type II error are given in Table 5 for both tests (58) and (59). As a result, here again, the asymptotic test always takes the correct decision $\mathcal{R}(\xi_1) < \mathcal{R}(\xi_2)$ while a large sample size is needed to always take the correct decision using the nonasymptotic test (as large as 110 000 for $w_1 = 1$). For sample sizes less than 10 000, the empirical probability of type II error with the nonasymptotic test is 1. We see that for fixed (w_0, w_1), in most cases, we need a larger sample size than in Case (I) to have a null probability of type II error, due the overlap of the two distributions.

We finally consider **Case (III)**, where the choice between ξ_1 and ξ_2 is more delicate and depends on the pair (w_0, w_1). In this case, we have (see Figure 1) $\mathbb{E}[\xi_2] > \mathbb{E}[\xi_1]$ and $\mathsf{AVaR}_\alpha(\xi_2) < \mathsf{AVaR}_\alpha(\xi_1)$ for $1-\alpha = 0.1$. It follows that for pairs (w_0, w_1) summing to one, when

$$0 \le w_0 < w_{\mathtt{Crit}} = \frac{\mathsf{AVaR}_\alpha(\xi_1) - \mathsf{AVaR}_\alpha(\xi_2)}{\mathbb{E}[\xi_2] - \mathbb{E}[\xi_1] + \mathsf{AVaR}_\alpha(\xi_1) - \mathsf{AVaR}_\alpha(\xi_2)},$$

we have $\mathcal{R}(\xi_2) < \mathcal{R}(\xi_1)$ and, for $w_0 > w_{\mathtt{Crit}}$, we have $\mathcal{R}(\xi_2) > \mathcal{R}(\xi_1)$. The empirical estimation of $w_{\mathtt{Crit}}$ (estimated using a sample of size 10^6) is 0.71. For w_0 close to $w_{\mathtt{Crit}}$, $\mathcal{R}(\xi_1)$ and $\mathcal{R}(\xi_2)$ are close and the probability of type II error for test (58) can be large even for the asymptotic test if the sample size is not sufficiently large. More precisely,

TABLE 5

Empirical probabilities of type II *error for tests* (58) *and* (59) *using a nonasymptotic rejection region when* $\xi_1 \sim \mathcal{N}(5, 1; 0, 30)$, $\xi_2 \sim \mathcal{N}(10, 25; 0, 30)$, *and* $1 - \alpha = 0.1$.

	Sample size N				
(w_0, w_1)	10 000	20 000	50 000	100 000	110 000
(0.0, 1.0)	1	1	1	1	0
(0.1, 0.9)	1	1	1	0	0
(0.2, 0.8)	1	1	1	0	0
(0.3, 0.7)	1	1	1	0	0
(0.4, 0.6)	1	1	1	0	0
(0.5, 0.5)	1	1	1	0	0
(0.6, 0.4)	1	1	0	0	0
(0.7, 0.3)	1	1	0	0	0
(0.8, 0.2)	1	1	0	0	0
(0.9, 0.1)	0.06	0	0	0	0

TABLE 6

Empirical probabilities of type II *error for test* (58) *using an asymptotic rejection region when* $\xi_1 \sim \mathcal{N}(10, 49; 0, 30)$, $\xi_2 \sim \mathcal{N}(14, 0.25; 0, 30)$, *and* $1 - \alpha = 0.1$.

	Sample size N							
(w_0, w_1)	20	50	100	200	500	1000	2000	5000
(0.0, 1.0)	0.13	0.01	0	0	0	0	0	0
(0.1, 0.9)	0.24	0	0	0	0	0	0	0
(0.2, 0.8)	0.32	0.03	0	0	0	0	0	0
(0.3, 0.7)	0.50	0.07	0	0	0	0	0	0
(0.4, 0.6)	0.61	0.11	0	0	0	0	0	0
(0.5, 0.5)	0.71	0.46	0.11	0.01	0	0	0	0
(0.6, 0.4)	0.86	0.69	0.50	0.28	0.01	0	0	0
(0.7, 0.3)	0.83	0.85	0.90	0.91	0.87	0.89	0.69	0.53
(0.8, 0.2)	0.71	0.71	0.65	0.29	0.07	0	0	0
(0.9, 0.1)	0.57	0.34	0.09	0	0	0	0	0

for the asymptotic test, when $(w_0, w_1) = (0.7, 0.3)$, the empirical probabilities of type II error are given in Table 6 for $N \in \{20, 50, 100, 200, 500, 1\,000, 2\,000, 5000\}$, and are $0.28, 0.11, 0.01$, and 0 for $N = 10\,000, 20\,000, 40\,000$, and $45\,000$, respectively. For the remaining values of w_0 the empirical probabilities of type II error are given in Table 6 for the asymptotic test. For the nonasymptotic test, the empirical probabilities of type II error for test (58) are given in Table 7. It is seen that much larger sample sizes are needed in this case to obtain a small probability of type II error. However, for the sample size $N = 5{\times}10^6$, the nonasymptotic test still always takes the wrong decision for the difficult case $w_0 = 0.7$.

For $w_0 < w_{\mathtt{Crit}}$ with $w_0 \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$, we are interested in the probability of type II error of the test

$$H_0 : \mathcal{R}(\xi_2) \geq \mathcal{R}(\xi_2) \quad \text{against} \quad H_1 : \mathcal{R}(\xi_2) < \mathcal{R}(\xi_1) \tag{60}$$

since H_1 holds in this case. Using the asymptotic rejection region, except for the difficult case $w_0 = 0.7$ where the probability of type II error is still positive for $N = 30\,000$, the empirical probability of type II error is null for small to moderate (at most 1 000) sample sizes; see Table 8. Using the nonasymptotic rejection region, much larger sample sizes are necessary to obtain a small probability of type II error;

TABLE 7

Empirical probabilities of type II *error for test* (58) *using a nonasymptotic rejection region when* $\xi_1 \sim \mathcal{N}(10, 49; 0, 30)$, $\xi_2 \sim \mathcal{N}(14, 0.25; 0, 30)$, *and* $1 - \alpha = 0.1$.

	Sample size N					
(w_0, w_1)	100 000	300 000	500 000	700 000	10^6	$5{\times}10^6$
(0.0, 1.0)	1	0.83	0	0	0	0
(0.1, 0.9)	1	1	0	0	0	0
(0.2, 0.8)	1	1	0	0	0	0
(0.3, 0.7)	1	1	0	0	0	0
(0.4, 0.6)	1	1	1	0	0	0
(0.5, 0.5)	1	1	1	1	0	0
(0.6, 0.4)	1	1	1	1	1	0
(0.7, 0.3)	1	1	1	1	1	1
(0.8, 0.2)	1	1	1	1	1	0
(0.9, 0.1)	0	0	0	0	0	0

TABLE 8

Empirical probabilities of type II *error for test* (60) *using an asymptotic rejection region when* $\xi_1 \sim \mathcal{N}(10, 49; 0, 30)$, $\xi_2 \sim \mathcal{N}(14, 0.25; 0, 30)$, *and* $1 - \alpha = 0.1$.

	Sample size N							
(w_0, w_1)	20	100	200	1 000	5 000	10 000	30 000	50 000
(0.0, 1.0)	0.11	0	0	0	0	0	0	0
(0.1, 0.9)	0.26	0	0	0	0	0	0	0
(0.2, 0.8)	0.28	0	0	0	0	0	0	0
(0.3, 0.7)	0.35	0	0	0	0	0	0	0
(0.4, 0.6)	0.51	0	0	0	0	0	0	0
(0.5, 0.5)	0.66	0.2	0.01	0	0	0	0	0
(0.6, 0.4)	0.83	0.53	0.22	0	0	0	0	0
(0.7, 0.3)	0.87	0.88	0.90	0.81	0.61	0.39	0.05	0

see Table 9.

For $w_0 > w_{\mathtt{Crit}}$ with $w_0 \in \{0.8, 0.9\}$, we are interested in the probability of type II error for test (59) since H_1 holds in this case. The probability of type II error for this test using the nonasymptotic rejection region is 1 (resp., 0) for $(N, w_0, w_1) = (10^6, 0.8, 0.2)$ (resp., $(N, w_0, w_1) = (10^6, 0.9, 0.1)$), and null for $(N, w_0, w_1) = (5{\times}10^6, 0.8, 0.2), (5{\times}10^6, 0.9, 0.1)$, meaning that we always take the correct decision $\mathcal{R}(\xi_1) < \mathcal{R}(\xi_2)$ for $N = 5{\times}10^6$ and $(w_0, w_1) = (0.8, 0.2), (0.9, 0.1)$. Using the asymptotic rejection region, the probabilities of type II errors are null already for $N = 1\,000$. For $N = 100$, we get probabilities of type II error of 0.09 and 0.42 for $(w_0, w_1) = (0.8, 0.2)$ and $(w_0, w_1) = (0.9, 0.1)$, respectively.

6.2. Tests on the optimal value of two risk-averse stochastic programs. We illustrate the results of section 3 on the risk-averse problem

(61)
$$\begin{array}{l} \min w_0 \mathbb{E}[\sum_{i=1}^m \xi_i x_i] + w_1 \left(x_0 + \mathbb{E}\left[\frac{1}{1-\alpha} [\sum_{i=1}^m \xi_i x_i - x_0]_+ \right] \right) + \lambda_0 \|[x_0; x_1; \cdots ; x_m]\|_2^2 + c_0 \\ -1 \le x_0 \le 1, \sum_{i=1}^m x_i = 1,\ x_i \ge 0, i = 1, \ldots, m, \end{array}$$

Table 9
Empirical probabilities of type II *error for test* (60) *using a nonasymptotic rejection region when* $\xi_1 \sim \mathcal{N}(10, 49; 0, 30)$*,* $\xi_2 \sim \mathcal{N}(14, 0.25; 0, 30)$*, and* $1 - \alpha = 0.1$*.*

	Sample size N						
(w_0, w_1)	300 000	400 000	500 000	700 000	900 000	2×10^6	5×10^6
(0.0, 1.0)	0	0	0	0	0	0	0
(0.1, 0.9)	0.85	0	0	0	0	0	0
(0.2, 0.8)	1	0	0	0	0	0	0
(0.3, 0.7)	1	1	0	0	0	0	0
(0.4, 0.6)	1	1	1	0	0	0	0
(0.5, 0.5)	1	1	1	1	0	0	0
(0.6, 0.4)	1	1	1	1	1	0.75	0
(0.7, 0.3)	1	1	1	1	1	1	1

Table 10
Definition of instances $\mathcal{I}_1$*,* $\mathcal{I}_2$*,* $\mathcal{I}_3$*,* $\mathcal{I}_4$*,* $\mathcal{I}_5$*, and* $\mathcal{I}_6$ *of problem* (61) *(*Ψ_1 *and* Ψ_2 *are vectors with entries drawn independently and randomly over* $[0,1]$*).*

`Instance`	$(w_0, w_1, 1-\alpha, \lambda_0)$	c_0	m	$(\mathbb{P}(\xi_i = 1))_i$
$\mathcal{I}_1$	$(0.9, 0.1, 0.1, 2)$	0	100	Ψ_1
$\mathcal{I}_2$	$(0.9, 0.1, 0.1, 2)$	0	100	$0.8\Psi_1$
$\mathcal{I}_3$	$(0.9, 0.1, 0.1, 2)$	-3	100	$0.8\Psi_1$
$\mathcal{I}_4$	$(0.9, 0.1, 0.1, 2)$	0	500	Ψ_2
$\mathcal{I}_5$	$(0.9, 0.1, 0.1, 2)$	0	500	$0.8\Psi_2$
$\mathcal{I}_6$	$(0.9, 0.1, 0.1, 2)$	-3	500	$0.8\Psi_2$

where ξ is a random vector with i.i.d. Bernoulli entries: $\mathbb{P}(\xi_i = 1) = \Psi_i$, $\mathbb{P}(\xi_i = -1) = 1 - \Psi_i$, with Ψ_i randomly drawn over $[0,1]$.[4] This problem amounts to minimizing a linear combination of the expectation and the AVaR_α of $\sum_{i=1}^m \xi_i x_i$ plus a penalty obtained taking $\lambda_0 > 0$. Therefore, it has a unique optimal solution. SAA formulation of this problem as well as the quadratic problems of each iteration of RSA were solved numerically using the MOSEK optimization toolbox [1]. We will again use the rejection regions given in section 5.2 (resp., given by (43)) in the nonasymptotic (resp., asymptotic) case.

To illustrate Theorem 2, for several instances of this problem, we report in the supplementary materials of this article the QQ-plots of the empirical distribution of the SAA optimal value for problem (61) versus the normal distribution with parameters the empirical mean and standard deviation of this empirical distribution for various sample sizes N. We observe again that this distribution is well approximated by a Gaussian distribution even when the sample size is small ($N = 20$): For all problem sizes ($m = 100$, $m = 500$, $m = 10^3$, and $m = 10^4$) and the smallest sample size tested ($N = 20$), the Jarque–Bera test accepts the null hypothesis (the data comes from a normal distribution with unknown mean and variance) at the 5% significance level.

We now define in Table 10 six instances $\mathcal{I}_1$, $\mathcal{I}_2$, $\mathcal{I}_3$, $\mathcal{I}_4$, $\mathcal{I}_5$, and $\mathcal{I}_6$ of problem (61).

[4]Of course c_0 can be ignored to solve the problem. However, it will be used to define several instances and test the equality about their optimal values.

TABLE 11
Average approximate optimal value of instance $\mathcal{I}_2$ (computed using 100 samples of ξ^N) using SAA and RSA for various sample sizes N.

Method	$N = 20$	$N = 50$	$N = 10^2$	$N = 10^3$	$N = 10^4$	$N = 10^5$
SAA	-0.7205	-0.6965	-0.6883	-0.6799	-0.6791	-0.6791
RSA	-0.4615	-0.5274	-0.5646	-0.6389	-0.6654	-0.6738

We first compare the estimation of the optimal value of $\mathcal{I}_2$ using RSA and SAA. For the RSA algorithm, we take $\|\cdot\| = \|\cdot\|_2 = \|\cdot\|_*$ and (see [7])

$$\begin{aligned} M_1 &= 2\left(w_0 + \frac{w_1}{1-\alpha}\right), \\ L &= \sqrt{\left(\frac{w_1\alpha}{1-\alpha}\right)^2 + m\left(w_0 + \frac{w_1}{1-\alpha}\right)^2} + 2\lambda_0, \\ M_2 &= \sqrt{\left(\frac{w_1}{1-\alpha}\right)^2 + 4m\left(w_0 + \frac{w_1}{1-\alpha}\right)^2}. \end{aligned}$$

The average approximate optimal value of instance $\mathcal{I}_2$ (averaging taking 100 samples of ξ^N) using RSA and SAA is given in Table 11 for various sample sizes N. These values increase (resp., decrease) with the sample size for SAA (resp., RSA). With SAA, the optimal value is already well approximated with small sample sizes while large samples are needed to obtain a good approximation with RSA. We also report in Table 12 the average values of the asymptotic and nonasymptotic confidence bounds (computed using 100 samples of ξ^N) on the optimal values of instances $\mathcal{I}_1$ and $\mathcal{I}_2$ and various sample sizes.[5] Knowing that the optimal values of $\mathcal{I}_1$ and $\mathcal{I}_2$, estimated using SAA with a sample of size 10^6, are, respectively, $\vartheta_1 = -0.6515$ and $\vartheta_2 = -0.6791$, we observe that the asymptotic confidence interval is in mean much closer to the optimal value and of small width while large samples are needed to obtain a nonasymptotic confidence interval of small width. However, the confidence bounds on the optimal value obtained using RSA are almost independent of the problem size and, as for the one-dimensional problem of the previous section, the sample size $N = 10^5$ provides confidence intervals of small width and allows us to have small probabilities of type I and type II errors for nonasymptotic tests on the optimal value of two instances of (61) if their optimal values are sufficiently distant (see Lemma 6). To check that and the superiority of the asymptotic tests for problems of moderate sizes ($m = 100$ and $m = 500$), we compare the empirical probabilities of type II error of several tests of form (3) with $K = 2$ for which H_1 holds and where ϑ_i is the optimal value of instance $\mathcal{I}_i$.

More precisely, the empirical probabilities of type II error of asymptotic and nonasymptotic tests of form

$$H_0 : \vartheta_p = \vartheta_q \quad \text{against} \quad H_1 : \vartheta_p \neq \vartheta_q, \tag{62}$$

are reported in Table 13 (for all these tests, we check that H_1 holds computing ϑ_v

[5] The nonasymptotic confidence interval is $[\mathtt{Low}(\Theta_2, \Theta_3, N), \mathtt{Up}(\Theta_1, N)]$ with $\mathtt{Low}(\Theta_2, \Theta_3, N)$, $\mathtt{Up}(\Theta_1, N)$ given by (49), (50) and $\Theta_1 = 2\sqrt{\ln(2/\beta)}$, $\Theta_3 = 2\sqrt{\ln(4/\beta)}$, and Θ_2 satisfying $e^{1-\Theta_2^2} + e^{-\Theta_2^2/4} = \beta/4$. The asymptotic confidence interval for (61) is $[\hat{\vartheta}_N - \Phi^{-1}(1-\beta/2)\frac{\hat{\nu}_N}{\sqrt{N}}\hat{\vartheta}_N + \Phi^{-1}(1-\beta/2)\frac{\hat{\nu}_N}{\sqrt{N}}]$.

Table 12

Average values of the asymptotic and nonasymptotic confidence bounds (computed using 100 *samples of* ξ^N*) for instances* $\mathcal{I}_1$ *and* $\mathcal{I}_2$ *and various sample sizes. For instance* $\mathcal{I}_i$*, the average asymptotic confidence interval is* [L-Ai, U-Ai] *and the average nonasymptotic confidence interval is* [L-NAi, U-NAi].

N	L-A1	U-A1	L-NA1	U-NA1	L-A2	U-A2	L-NA2	U-NA2
20	-0.7207	-0.6666	-95.7926	2.5227	-0.7443	-0.6967	-95.8354	2.4799
50	-0.6888	-0.6475	-60.8057	1.3743	-0.7148	-0.6781	-60.8472	1.3329
10^2	-0.6752	-0.6444	-43.1779	0.7900	-0.7019	-0.6746	-43.2171	0.7508
10^3	-0.6573	-0.6474	-14.0952	-0.1913	-0.6843	-0.6755	-14.1269	-0.2230
10^4	-0.6532	-0.6501	-4.9019	-0.5051	-0.6805	-0.6777	-4.9307	-0.5339
10^5	-0.6520	-0.6510	-1.9947	-0.6043	-0.6796	-0.6787	-2.0226	-0.6322

Table 13

Empirical probabilities of type II *error for tests of form* (62).

			Sample size N					
H_0	H_1	Test type	20	50	10^2	10^3	10^4	10^5
$\vartheta_1 = \vartheta_2$	$\vartheta_1 \neq \vartheta_2$	Asymptotic	0.72	0.45	0.29	0	0	0
$\vartheta_1 = \vartheta_2$	$\vartheta_1 \neq \vartheta_2$	Nonasymptotic	1	1	1	1	1	1
$\vartheta_1 = \vartheta_3$	$\vartheta_1 \neq \vartheta_3$	Asymptotic	0	0	0	0	0	0
$\vartheta_1 = \vartheta_3$	$\vartheta_1 \neq \vartheta_3$	Nonasymptotic	1	1	1	1	1	0
$\vartheta_4 = \vartheta_5$	$\vartheta_4 \neq \vartheta_5$	Asymptotic	0.33	0.36	0.21	0	0	0
$\vartheta_4 = \vartheta_5$	$\vartheta_4 \neq \vartheta_5$	Nonasymptotic	1	1	1	1	1	1
$\vartheta_4 = \vartheta_6$	$\vartheta_4 \neq \vartheta_6$	Asymptotic	0	0	0	0	0	0
$\vartheta_4 = \vartheta_6$	$\vartheta_4 \neq \vartheta_6$	Nonasymptotic	1	1	1	1	1	0

solving the SAA problem of instance $\mathcal{I}_v$ with a sample of ξ of size 10^6: $\vartheta_1 = -0.6515$, $\vartheta_2 = -0.6791$, $\vartheta_3 = -3.6791$, $\vartheta_4 = -0.7725$, $\vartheta_5 = -0.7868$, and $\vartheta_6 = -3.7868$).

Though it was observed in [7, 8] that for sample sizes that are not much larger than the problem size the coverage probability of the asymptotic confidence interval is much lower than the coverage probability of the nonasymptotic confidence interval and than the target coverage probability, the asymptotic confidence bounds are much closer to each other and much closer to the optimal value than the nonasymptotic confidence bounds. This explains why the probability of type II error of the asymptotic test is much less than the probability of type II error of the nonasymptotic test, even for small sample sizes, and a smaller sample is needed to always take the correct decision H_1 with the asymptotic test, i.e., to obtain a null probability of type II error. Of course, in both cases, for fixed N, the empirical probability of type II error depends on the distance between ϑ_p and ϑ_q.

Similar conclusions can be drawn from Table 14, which reports the empirical probability of type II error for various tests of form

$$H_0 : \vartheta_p \leq \vartheta_q \quad \text{against} \quad H_1 : \vartheta_q < \vartheta_p. \tag{63}$$

In particular, from these results, we see that we always take the correct decision H_1 with the asymptotic test for sample sizes above $N = 100$.

Table 14
Empirical probabilities of type II *error for tests of form* (63).

			Sample size N					
H_0	H_1	Test type	20	50	10^2	10^3	10^4	10^5
$\vartheta_1 \le \vartheta_2$	$\vartheta_1 > \vartheta_2$	`Asymptotic`	0.54	0.38	0.16	0	0	0
$\vartheta_1 \le \vartheta_2$	$\vartheta_1 > \vartheta_2$	`Nonasymptotic`	1	1	1	1	1	1
$\vartheta_1 \le \vartheta_3$	$\vartheta_1 > \vartheta_3$	`Asymptotic`	0	0	0	0	0	0
$\vartheta_1 \le \vartheta_3$	$\vartheta_1 > \vartheta_3$	`Nonasymptotic`	1	1	1	1	1	0
$\vartheta_4 \le \vartheta_5$	$\vartheta_4 > \vartheta_5$	`Asymptotic`	0.29	0.26	0.15	0	0	0
$\vartheta_4 \le \vartheta_5$	$\vartheta_4 > \vartheta_5$	`Nonasymptotic`	1	1	1	1	1	1
$\vartheta_4 \le \vartheta_6$	$\vartheta_4 > \vartheta_6$	`Asymptotic`	0	0	0	0	0	0
$\vartheta_4 \le \vartheta_6$	$\vartheta_4 > \vartheta_6$	`Nonasymptotic`	1	1	1	1	1	0

REFERENCES

[1] E. D. Andersen and K. D. Andersen, *The MOSEK optimization toolbox for MATLAB* (manual), version 7.0, 2013; available online from https://www.mosek.com/.

[2] P. Artzner, F. Delbaen, J. Eber, and D. Heath, *Coherent measures of risk*, Math. Finance, 9 (1999), pp. 203–228.

[3] D. Belomestny and V. Krätschmer, *Central limit theorems for law-invariant coherent risk measures*, J. Appl. Prob., 49 (2012), pp. 1–21.

[4] D. Dentcheva, S. Penev, and A. Ruszczyński, *Statistical estimation of composite risk functionals and risk optimization problems*, Ann. Inst. Statist. Math., (2016), pp. 1–24.

[5] A. Eichhorn and W. Römisch, *Polyhedral risk measures in stochastic programming*, SIAM J. Optim., 16 (2005), pp. 69–95.

[6] H. Föllmer and A. Schied, *Stochastic Finance: An Introduction in Discrete Time*, 3rd ed., Walter de Gruyter, Berlin, 2011.

[7] V. Guigues, *Multistep stochastic mirror descent for risk averse convex stochastic programs based on extended polyhedral risk measures*, Math. Program., 163 (2017), pp. 169–212.

[8] V. Guigues, A. Juditsky, and A. Nemirovski, *Non-Asymptotic Confidence Bounds for the Optimal Value of a Stochastic Program*, preprint, http://arxiv.org/abs/1601.07592, 2016.

[9] V. Guigues and W. Römisch, *Sampling-based decomposition methods for multistage stochastic programs based on extended polyhedral risk measures*, SIAM J. Optim., 22 (2012), pp. 286–312.

[10] S. Kusuoka, *On law-invariant coherent risk measures*, in Adv. Math. Econ. 3, S. Kusuoka and T. Maruyama, eds., Springer, Tokyo, 2001, pp. 83–95.

[11] G. Lan, A. Nemirovski, and A. Shapiro, *Validation analysis of mirror descent stochastic approximation method*, Math. Program., 134 (2012), pp. 425–458.

[12] J. Linderoth, A. Shapiro, and S. Wright, *The empirical behavior of sampling methods for stochastic programming*, Ann. Oper. Res., 142 (2006), pp. 215–241.

[13] M. M. Ledoux and M. Talagrand, *Probability in Banach Space*, Springer, Berlin, 1991.

[14] W. Mak, D. Morton, and R. Wood, *Monte Carlo bounding techniques for determining solution quality in stochastic programs*, Oper. Res. Lett., 24 (1999), pp. 47–56.

[15] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Managements*, Princeton University Press, Princeton, NJ, 2005.

[16] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley Ser. Probab. Stat., John Wiley, Hoboken, NJ, 1982.

[17] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, *Robust stochastic approximation approach to stochastic programming*, SIAM J. Optim., 19 (2009), pp. 1574–1609.

[18] M. Perlman, *One-sided testing problems in multivariate analysis*, Ann. Math. Stat., 40 (1969), pp. 549–567.

[19] G. Pflug, *Stochastic programs and statistical data*, Ann. Oper. Res., 85 (1999), pp. 59–78.

[20] G. Pflug and N. Wozabal, *Asymptotic distribution of law-invariant risk functionals*, Finance Stoch., 14 (2010), pp. 397–418.

[21] B. Polyak, *New stochastic approximation type procedures*, Autom. Remote Control, 51 (1990), pp. 937–948.

[22] B. Polyak and A. Juditsky, *Acceleration of stochastic approximation by averaging*, SIAM J. Contr. and Optim., 30 (1992), pp. 838–855.

[23] R. Rockafellar and S. Uryasev, *Conditional value-at-risk for general loss distributions*, J. Bank. Financ., 26 (2002), pp. 1443–1471.
[24] R. Rockafellar and R.-B. Wets, *Variational Analysis*, Springer, New York, 1998.
[25] A. Shapiro, *Asymptotics of minimax stochastic programs*, Statist. Probab. Lett., 78 (2008), pp. 150–157.
[26] A. Shapiro, *Consistency of sample estimates of risk averse stochastic programs*, J. Appl. Probab., 50 (2013), pp. 533–541.
[27] A. Shapiro, *On Kusuoka representation of law invariant risk measures*, Math. Oper. Res., 38 (2013), pp. 142–152.
[28] A. Shapiro and T. H. de Mello, *On rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs*, SIAM J. Optim., 11 (2000), pp. 70–86.
[29] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, 2nd ed., SIAM, Philadelphia, 2014.
[30] M. Talagrand, *Sharper bounds for Gaussian and empirical processes*, Ann. Probab., 22 (1994), pp. 28–76.
[31] M. Talagrand, *The Glivenko–Cantelli problem, ten years later*, J. Theoret. Probab., 9 (1996), pp. 371–384.
[32] A. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, New York, 1998.
[33] S. Vogel, *Universal confidence sets for solutions of optimization problems*, SIAM J. Optim., 19 (2008), pp. 1467–1488.