Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Nonasymptotic Convergence Rates for the Plug-in Estimation of Risk Measures

Daniel Bartl, Ludovic Tangpi

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 48, No. 4, November 2023, pp. 2129–2155
ISSN 0364-765X (print), ISSN 1526-5471 (online)
https://pubsonline.informs.org/journal/moor

Nonasymptotic Convergence Rates for the Plug-in Estimation of Risk Measures

Daniel Bartl,[a] **Ludovic Tangpi**[b,*]

[a] Department of Mathematics, Vienna University, 1010 Vienna, Austria; [b] Operations Research and Financial Engineering, Bendheim Center for Finance, Princeton University, Princeton, New Jersey 08544

*Corresponding author

Contact: daniel.bartl@univie.ac.at, https://orcid.org/0000-0003-2245-7346 (DB); ludovic.tangpi@princeton.edu, https://orcid.org/0000-0003-3830-1418 (LT)

Received: June 18, 2021
Revised: May 18, 2022; September 21, 2022
Accepted: October 29, 2022
Published Online in Articles in Advance: November 22, 2022

MSC2020 Subject Classification: Primary: 91B82, 91B30, 91B16

https://doi.org/10.1287/moor.2022.1333

Abstract. Let ρ be a general law-invariant convex risk measure, for instance, the average value at risk, and let X be a financial loss, that is, a real random variable. In practice, either the true distribution μ of X is unknown, or the numerical computation of $\rho(\mu)$ is not possible. In both cases, either relying on historical data or using a Monte Carlo approach, one can resort to an independent and identically distributed sample of μ to approximate $\rho(\mu)$ by the finite sample estimator $\rho(\mu_N)$ (μ_N denotes the empirical measure of μ). In this article, we investigate convergence rates of $\rho(\mu_N)$ to $\rho(\mu)$. We provide nonasymptotic convergence rates for both the deviation probability and the expectation of the estimation error. The sharpness of these convergence rates is analyzed. Our framework further allows for hedging, and the convergence rates we obtain depend on neither the dimension of the underlying assets nor the number of options available for trading.

Funding: Daniel Bartl is grateful for financial support through the Vienna Science and Technology Fund [Grant MA16-021] and the Austrian Science Fund [Grants ESP-31 and P34743]. Ludovic Tangpi is supported by the National Science Foundation [Grant DMS-2005832] and CAREER award [Grant DMS-2143861].

Keywords: decision analysis • risk • statistics • estimation • decision analysis • approximations

1. Introduction

Risk is a pervasive aspect of the financial industry as every single financial decision carries a certain amount of risk. Correctly quantifying riskiness is, therefore, of central importance for financial institutions. The idea is often to consider the profit and loss $F(S)$ resulting from an investment in assets S. A fundamental innovation (that can be traced back to the work of Markovitz [41] in the 1950s) allowing us to quantify the risk of $F(S)$ was the introduction of the concept of risk measures, which allows us to assign a numerical value $\rho(F(S))$ to the profit and loss $F(S)$ depending on the agent's risk aversion. In other words, one can focus on a single number to make decisions rather than on the whole distribution of the loss. As a consequence, computing $\rho(F(S))$ becomes an essential task for the risk manager.

For a long time, the value at risk (VaR) has been the industry standard for risk management. As a result, the numerical simulation of VaR (i.e., of quantiles) is well-understood, and various methods can be found in Glasserman et al. [24], Hong [27], Jin et al. [29], and Jorion [30] and in their references. However, there are many criticisms for the VaR,[1] so much so that the Basel Committee on Banking Supervision [5], which oversees risk management for financial institutions, has recommended since 2013 to use expected shortfall (also known as average value at risk (AVaR) or conditional value at risk) as the benchmark risk measure.

Intuitively, the AVaR at level $u \in (0,1)$ can be understood as the average of all VaR_v over $v \in (u,1)$. Thus, it does not only take into account occurrence of large losses, but also their size. The estimation of AVaR (in the context of portfolio optimization) is, for instance, considered by Rockafellar and Uryasev [48], who fundamentally use the fact that the AVaR at level u of the loss $F(S)$ can be written as

$$\mathrm{AVaR}_u(F(S)) = \inf_{m\in\mathbb{R}} \left(\frac{1}{1-u} E[(F(S)-m)^+] + m \right). \tag{1.1}$$

This representation shows, in particular, that AVaR is almost risk neutral for very large losses as it is linear in the tails. This prompted the generalization to convex risk measures that behave nonlinearly in both the tail and the center of the distribution, including, for instance, the optimized certainty equivalent (OCE) obtained by replacing the function $x \longmapsto \frac{1}{1-u}x^+$ in (1.1) by a convex loss function $l: \mathbb{R} \to \mathbb{R}$ (see Ben-Tal and Teboulle [7, 8]) or the shortfall (SF) risk measure defined in a similar spirit (see Föllmer and Schied [18] and Section 2.2 for details).

More generally, a rigorous unifying approach to risk management was initiated by Artzner et. al. [2] and matured into an impressive theory of risk measures. We refer for instance to the monographs of Föllmer and Schied [19] and McNeil et al. [42] for excellent expositions. A general convex risk measure is defined as follows.

Definition 1.1 (Convex Risk Measure). A functional $\rho: L^\infty \to \mathbb{R}$ over a standard probability space is a convex risk measure[2] if

a. $\rho(X + m) = \rho(X) + m$ for all X and $m \in \mathbb{R}$ and $\rho(0) = 0$.
b. $\rho(X) \le \rho(Y)$ if $X \le Y$ almost surely (a.s.).
c. $\rho(\lambda X + (1-\lambda)Y) \le \lambda\rho(X) + (1-\lambda)\rho(Y)$ for $\lambda \in [0,1]$.

The monotonicity condition (b) is natural and models preference for more profits. The condition (a), known as cash invariance or translation invariance, stems from the desire to interpret $\rho(X)$ as a capital requirement, that is, the minimal cash value, which, if added to the position X, would make it acceptable for regulators. The convexity property (c) means that more diversified positions should be less risky. In practice, for numerical simulation, it is more convenient to work with the distribution of the loss rather than its observed realization. Therefore, it is often assumed that risk measures are law-invariant (or law-determined),[3] meaning that

d. $\rho(X) = \rho(Y)$ if $X \sim Y$, that is, if X and Y have the same distribution.

Observe that most examples of risk measures fulfill this condition. We make the convention that, throughout this paper, the term "risk measure" always refers to a convex law-invariant risk measure. Consequently, we use the shorthand notation

$$\rho^\mu(F) := \rho(F(S)) \quad \text{where} \quad S \sim \mu,$$

that is, $\rho^\mu(F)$ is the risk of $F(S)$ computed according to the risk measure ρ when S has the distribution μ.

The numerical computation of a law-invariant risk measure depends on the probability distribution μ of S. For AVaR, for instance, one issue is to efficiently approximate the integrals $\int_{\mathbb{R}}(F(x) - m)^+ \mu(dx)$ (assuming S is real-valued). In some cases, this integral operation is computationally costly. Moreover, in many practical applications, the distribution μ is not precisely known. A natural idea is to approximate the integral by the sample average $\frac{1}{N}\sum_{n=1}^N (F(S_n) - m)^+$, where $S_1, \ldots, S_N$ are independent random variables with distribution μ and the minimization over m (see (1.1)) can be reduced to linear programming; see Rockafellar and Uryasev [48, section 3]. When μ is not known, this Monte Carlo simulation can be carried out on historical data.

Example 1.1. For instance, in the context of portfolio optimization, S is a vector of d stock returns $S := (S_1^i - S_0^i)_{i=1,\ldots,d}$, and $F(S)$ takes the form $F(S) := \sum_{i=1}^d g_i(S_1^i - S_0^i)$, where $(g_1, \ldots, g_d)$ are portfolio weights. Strictly speaking, in practice, the time series formed by historical returns is of course not independent and identically distributed (i.i.d.). It shows patterns of changing volatility. Some workarounds in the literature include working with longer interval returns series (see McNeil et al. [42, section 4.1]) or using a semiparametric approach to estimate returns as $\Delta S_i := X_i Z_i$, where X_i is a diagonal volatility matrix modeled by a generalized autoregressive conditional heteroskedasticity model and the innovation processes Z_i are i.i.d. (see McNeil et al. [42, chapter 2]). Thus, we can make the simplifying assumption that we are working with i.i.d. observations. In both cases, denoting by

$$\mu_N := \frac{1}{N}\sum_{n\le N}\delta_{S_n} \quad \text{where} \quad S_1, \ldots, S_N \sim S \quad \text{i.i.d.}$$

the empirical distribution of the N observations $S_1, \ldots, S_N$, $\mathrm{AVaR}^{\mu_N}(F)$ is a nonparametric finite sample estimator of $\mathrm{AVaR}^\mu(F)$.

This idea extends to general risk measures and arbitrary functions F.

Definition 1.2 (Plug-in Estimator). For every $N \ge 1$, denote by

$$\rho^{\mu_N}(F) = \rho(F(\hat{S})) \quad \text{where } \hat{S} \sim \mu_N$$

the plug-in estimator of $\rho^\mu(F)$.

As we observe, this estimator[4] is consistent (see Corollary 2.1) but typically underestimates the true risk $\rho^\mu(F)$ (see Remark 2.3). The latter observation is consequential from the practical standpoint. In fact, the idea of risk measures is precisely to protect oneself from risky investments; thus, underestimating the risk of an asset is precisely what risk managers want to avoid. Because the finite sample estimator $\rho^{\mu_N}(F)$ is the most natural in nonparametric estimations and is widely used in practice, it is, therefore, essential to understand just how much it underestimates the true risk.

Thus, the question for the risk manager is

$$\text{How } \textit{far} \text{ is } \rho^{\mu_N}(F) \text{ from } \rho^{\mu}(F) \text{ for a fixed sample size } N?$$

This is an essential question because its answer gives theoretical insights allowing risk managers to parsimoniously use data. To make the question rigorous, one, of course, needs to give a meaning to "far" as the estimation error $|\rho^{\mu_N}(F) - \rho^{\mu}(F)|$ is random (it depends on the observations from S).

The goal of this article is to answer the question by providing nonasymptotic convergence rates on the expected estimation error and the probability that the estimation error exceeds some prescribed threshold. The difference between asymptotic and nonasymptotic rates should be underscored: whereas there are instances in which the asymptotic rates suggest a much faster convergence, this is only true within the asymptotic regime. In particular, relying on asymptotic rates, for example, for the computation of the needed sample size N to guarantee that the estimator preforms well with a certain confidence, might give a far too optimistic number. Nonasymptotic rates, however, hold for every N and give an order of magnitude of the sample size N needed to achieve a desired estimation accuracy.

Our main results show the following: for a general risk measure, the usual $1/\sqrt{N}$ convergence rate dictated by the central limit theorem needs not hold true. We introduce a simple and tractable notion of regularity for risk measures (quantified by a parameter $q \in (1, \infty)$) and show that this notion of regularity governs the convergence rates. More precisely, if a risk measure is regular with parameter q, then

$$E^*[|\rho^{\mu_N}(F) - \rho^{\mu}(F)|] \le \frac{C}{\sqrt{N}^q}$$
$$P^*[|\rho^{\mu_N}(F) - \rho^{\mu}(F)| \ge \varepsilon] \le C\exp(-cN\varepsilon^{2q}) \tag{1.2}$$

for all $N \ge 1$ and $\varepsilon > 0$, where c, C are two positive constants depending on ρ and (the L^∞-norm of) F; see Theorem 2.1. Here, E^* and P^* denote the outer expectation and outer probability, respectively (see, e.g., van der Vaart and Wellner [50]). They are used because ρ^{μ_N} does not necessarily need to be measurable. Notably, we show that these rates are sharp (at least up to a factor of two); see Proposition 2.1. Whereas this represents the major part of this work, we also consider three explicit risk measures separately (the average value at risk, the optimized certainty equivalent, and the shortfall risk) and show that, then, the usual $1/\sqrt{N}$ convergence rate dictated by the central limit theorem can be recovered in a nonasymptotic fashion; namely, (1.2) holds true for all $N \ge 1$ and $\varepsilon > 0$ with $q = 1$.

In practice, there is much more to risk management than computing the numerical value $\rho(F(S))$ for a given loss $F(S)$: in many situations, risk managers additionally hedge their exposure to $F(S)$ by investing in the stock market, resulting in a risk based (super)hedging problem or a utility maximization problem. Another important part of our work focuses on the investigation of these type of problems, and we show that the same rates of convergence remain valid in this more general situation.

The reader familiar with the theory of sample average approximation likely recognizes the average value at risk (1.1) as a stochastic programming problem. Finite sample approximation of such problems is extensively studied, at least as far as asymptotic convergence rates are concerned. We give a few references. Note, however, that, in the generality of Definition 1.1, risk measures cannot typically be written as the value of a stochastic programming problem (using Kusuoka's theorem, we can only write them as a maximum of more and more irregular stochastic programs). This substantially complicates the analysis. The main contribution of the paper is to consider this general situation by balancing the irregularity of the stochastic programs by means of the regularity of the risk measure. To obtain finite sample rates of convergence, we employ techniques from empirical process theory, specifically Dudley's theorem. This allows us to consider the aforementioned case of additional hedging and yields rates that do not depend on the dimension/number of hedging instruments. Finally, let us stress that F (as well as the hedging instruments) are not subject to any continuity condition.

1.1. Related Literature

The estimation of risk measures is an essential question in quantitative finance and, as such, has received a lot of attention; we refer, for instance, to the monograph of McNeil et al. [42] for an in-depth treatment. See also the book of Glasserman [23, chapter 9] for the case of (average) value at risk. In mathematical finance, there is a growing interest on statistical aspects of quantitative risk management. See Embrechts and Hofert [16] for an excellent review of the main lines of research in this direction. Concerning statistical estimation of risk measures, one of the earliest works is that of Weber [51], who considers the problem of estimating $\rho^{\mu}(F)$ in an asymptotic fashion as $N \to \infty$. By means of the theory of large deviations, he shows that, if ρ is sufficiently regular, then

$\rho^{\mu_N}(F)$ satisfies a large deviation principle. Along the same lines, Belomestny and Krätschmer [6], Beutner and Zähle [10], and Chen [11] obtain central limit theorems for $\rho^{\mu_N}(F)$; see also Shapiro et al. [49, chapter 6].

Aside from large deviation and central limit theorem results, several authors investigate estimation of specific risk measures and (super)hedging functionals. These include Pal [45, 46], who analyzes hedging under risk measures that can be written as the finite maxima of expectations. Let us further refer to Guigues et al. [25], Holzmann and Zwingmann [26], Krätschmer et al. [38], and Pitera and Schmidt [47] for other (asymptotic) estimation results, mostly for the average value at risk and expectiles and, under some assumptions, on the distribution μ (see, e.g., Hong et al. [28] for a review). A deviation-type inequality for the value at risk is proposed by Jin et al. [29]. The problem of strict superhedging is recently considered by Obłój and Wiesel [44] and the problem of portfolio optimization under heavy tails by Bartl and Mendelson [3].

When the estimation of $\rho^{\mu}(F)$ is performed repeatedly or periodically, it is important that the estimator $\rho^{\mu_N}(F)$ be stable, that is, insensitive to small changes of μ_N. Such insensitivity is often referred to as robustness of the risk measure and is first analyzed by Cont et al. [14], who investigate a concept of robustness essentially equivalent to continuity of ρ with respect to (w.r.t.) weak convergence of measures. Alternative approaches to robustness are later proposed and analyzed by Claus et al. [13], Krätschmer and Zähle [35], and Krätschmer et al. [36, 37, 39]. Along the same lines, some authors investigate risk measures (and other stochastic maximization problems) under model uncertainty to account for the effect of possible misspecification of the estimated model; see, for example, Bartl et al. [4], Eckstein et al. [15], and Esfahani and Kuhn [17], in which it is often assumed that the true model belongs to a Wasserstein ball.

Beyond estimation of risk measures, a rich literature in operations research is devoted to the estimation of the value of stochastic optimization problems similar to OCE through the empirical distribution of the underlying probability measure. This technique goes under the name sample average approximation as mentioned earlier. The bulk of the literature in this direction is concerned with convergence issues and questions related to computational complexity of the estimators; see, for example, Bertsimas et al. [9], Kleywegt et al. [33], and the book chapter Kim et al. [32] for a recent overview. We also refer to the recent preprint Krätschmer [34] for asymptotic estimation results as well as error bound estimations using empirical process theory.

1.2. Organization of the Rest of the Paper

We start by presenting the main results of this article in the next section. The proofs of the moment bounds are given in Section 3 for special cases of risk measures, and in Section 4, the main part of the paper, convergence rates for general risk measures, are proved. In these sections, we also state generalizations of our results to unbounded cases. The deviation inequalities are proved in Section 5. Sharpness of the rates for general risk measures is discussed in Section 6, and all remaining proofs are presented in Section 7. The paper ends with an appendix on the theory of empirical processes.

2. Main Results

Before presenting our main results, let us generalize the setting of the introduction to the more practically relevant situation in which the risk manager can offset the risk from $F(S)$ by trading. Henceforth, μ denotes the distribution of S, which is a probability measure on a Polish space $\mathcal{X}$, and μ_N denotes the empirical measure of μ built from an i.i.d. sample $(S_n)_{n\le N}$ defined on some abstract probability space $(\Omega, \mathcal{F}, P)$. Moreover, $F:\mathcal{X}\to\mathbb{R}$ is a measurable function. In fact, we can additionally consider (measurable) options $G_1,\dots,G_e:\mathcal{X}\to\mathbb{R}$ available for trading without loss of generality at price zero (at which $e\in\mathbb{N}$). Trading according to a strategy $g\in\mathbb{R}^e$ then yields the outcome $F+\sum_{i=1}^e g_iG_i$. Thus, assuming the interest rate to be zero throughout, the risk manager's task is to estimate the minimal risk incurred when trading in the option market, that is, to compute

$$\pi^{\mu}(F) := \inf_{g\in\mathcal{G}} \rho^{\mu}\left(F+\sum_{i=1}^{e} g_iG_i\right),$$

where $\mathcal{G}\subset\mathbb{R}^e$ is the set of all admissible trading strategies. Loosely speaking, the goal here is to "absorb" extreme outcomes of F by trading. For instance, $\mathcal{G}=\{g\in[0,1]^e : g_1+\cdots+g_e=1\}$ corresponds to portfolio optimization; see Shapiro et al. [49] for some background. Notice that, if zero is the only admissible trading strategy, that is, $\mathcal{G}=\{0\}$, then we have $\pi^{\mu}=\rho^{\mu}$, and hence, all results obtained for π translate to ρ as introduced in the previous section.

In an effort to simplify the presentation in this section, we state the results for risk measures defined on bounded random variables and assume throughout this section that F and G are all bounded functions. In the later sections, we partially replace boundedness by integrability assumptions at the cost of more involved notation. Moreover, $\mathcal{G}\subset\mathbb{R}^e$ is assumed to be a bounded[5] set throughout this article, and this assumption can quickly be checked to be necessary (see Proposition A.1). In order to avoid discussions regarding measurability issues, we assume throughout this article that $\mathcal{G}$ is a countable set. As explained in Remark A.1, this assumption can actually be made without loss of generality.

2.1. Results for General Convex Risk Measures

Let us now present our main results pertaining to the estimation of law-invariant risk measures in the generality of Definition 1.1. As becomes more and more apparent throughout this article, it is necessary to impose some form of continuity assumption on the risk measure in order to derive nonasymptotic convergence rates; see Proposition 2.1 and Remark 6.2.

In order to start with a positive result, we define the notion of continuity that we require right away and discuss its rationale and necessity afterward.

Definition 2.1 (q-Regularity). For $q \in (1, \infty)$, a convex risk measure is said to be q-regular if it satisfies

$$\sup_{n\in\mathbb{N}} \rho(X \wedge n) < \infty$$

for all random variables X following a Pareto distribution with shape parameter q.

Recall here that a random variable X has Pareto distribution with scale parameter $x > 0$ and shape parameter $q > 0$ if

$$P[X \geq t] = \begin{cases} (x/t)^q & \text{if } t \geq x, \\ 1 & \text{if } t < x. \end{cases}$$

As the name suggests, q-regularity reflects a certain notion of continuity, and the familiar reader may recognize it to be stronger than the two classic notions of regularity for risk measures, namely, the Fatou and Lebesgue properties. The latter two are, however, not enough to guarantee any convergence rates; see Section 6.

The following is the main result of this article.

Theorem 2.1 (Rates for General Risk Measures). *Let $q \in (1, \infty)$ and let $\rho : L^\infty \to \mathbb{R}$ be a q-regular risk measure. Then, there are constants $c, C > 0$ such that the following hold:*

i. *We have the moment bound*

$$E^*[|\pi^{\mu}(F) - \pi^{\mu_N}(F)|] \leq \frac{C}{N^{1/2q}}$$

for all $N \geq 1$.

ii. *We have the matching deviation inequality*

$$P^*[|\pi^{\mu}(F) - \pi^{\mu_N}(F)| \geq \varepsilon] \leq C \exp(-cN\varepsilon^{2q})$$

for all $N \geq 1$ and all $\varepsilon > 0$.

An important observation is that, throughout this paper, the rates never depend on the number e of options or on the "dimension" of the underlying space $\mathcal{X}$. The constants c and C depend on ρ; the maximal range of F, G; the number of options e; and the maximal Euclidean norm in $\mathcal{G}$.

As an immediate consequence of Theorem 2.1, part (ii), and the Borel–Cantelli lemma, we obtain that $\pi^{\mu_N}(F)$ is a strongly consistent estimator of $\pi^{\mu}(F)$. Note that, under the assumption that F and G are additionally continuous, this is a trivial consequence of weak continuity of $\nu \longmapsto \pi^{\nu}(F)$ and weak convergence of the empirical measure to the true one. However, this reasoning does not apply in the present setting as F and G are merely measurable, and thus, $\nu \longmapsto \pi^{\nu}(F)$ can be discontinuous.

Corollary 2.1 (Consistency). *In the setting of Theorem 2.1 we have that*

$$\lim_{N\to\infty} \pi^{\mu_N}(F) = \pi^{\mu}(F)$$

P^*-almost surely.

Remark 2.1. An interesting by-product of the deviation inequality given in Theorem 2.1 is that it allows us to give a nonasymptotic estimation of the error in the L^p norm. Indeed, using the tail integration $E^*[|X|^p] = p\int_0^\infty x^{p-1} P^*[|X| \geq x]\,dx$, it follows from part (ii) in Theorem 2.1 that

$$E^*[|\pi^{\mu}(F) - \pi^{\mu_N}(F)|^p]^{1/p} \leq C\frac{\sqrt{p}}{N^{1/2q}}.$$

This is for every $p \geq 1$.

Let us now come back to the notion of regularity in Definition 2.1 and explain both its rationale and necessity. This is easiest done with the example of following two risk measures: $\rho_{\max}(X) := \text{ess.sup}X$ and $\rho_{\text{mean}}(X) := E[X]$. Then, $\rho^{\mu_N}_{\text{mean}}(F)$ is just the empirical mean of $F(S)$, and thus, convergence happens at the usual rate $1/\sqrt{N}$. On the other extreme of the spectrum, $\rho^{\mu_N}_{\max}(F)$ equals the empirical 1-quantile, and it is well-known that, without very specific assumptions, convergence may happen at arbitrarily slow speed; the unfamiliar reader may skip to Remark 6.1. A simple observation pertaining to the source of this different behavior is that small changes of X result in small changes of ρ_{mean}, whereas this is not the case for $\rho_{\max}$. Indeed, changes of X on almost negligible sets can result in significant changes of $\rho_{\max}(X)$, and (unfortunately) a random sample cannot properly exhibit almost negligible events. From this perspective, ρ_{mean} is very regular (and, indeed, it is q-regular for every $q \in (1,\infty)$), whereas $\rho_{\max}$ is not regular at all (and, indeed, it lacks q-regularity for any $q \in (1,\infty)$).

Whereas the preceding discussion focuses only on two very extreme risk measures, it happens that Definition 2.1 actually interpolates between these two examples. Indeed, the following proposition shows that the rates obtained in Theorem 2.1 are optimal, at least up to a factor of two.

Proposition 2.1 (Sharpness of Rates). *Let $q \in (1,\infty)$ and assume that F takes (at least) two distinct values. Then, there is a coherent law-invariant risk measure $\rho : L^\infty \to \mathbb{R}$ that is $(q+\varepsilon)$-regular for every $\varepsilon > 0$ and a constant $c > 0$ such that*

$$\sup_{\mu} E[|\rho^{\mu}(F) - \rho^{\mu_N}(F)|] \geq \frac{c}{N^{1/q}}$$

for all $N \geq 1$.

Currently, the authors do not know whether Proposition 2.1 can be improved to show that the rates obtained in Theorem 2.1 are actually sharp (i.e., whether Proposition 2.1 holds with $N^{-1/2q}$ instead of $N^{-1/q}$). One indication that this might be true is the following: for $q \approx 1$, the lower bound of Proposition 2.1 is approximately $1/N$, but we already know that the actual best possible rate is $1/\sqrt{N}$ as is dictated by the central limit theorem; see Section 6 for a short discussion. That is, for $q \approx 1$, the lower bound is off exactly by the factor of two.

Let us conclude this section with a comment regarding the proof of Theorem 2.1. As already mentioned, it builds upon the empirical processes theory; specifically Dudley's entropy integral theorem (see Appendix A.2). One could, however, wonder whether the statements of Theorem 2.1 (and Theorem 2.2) follow from some rather simple-to-obtain continuity in Wasserstein distance of $\mu \mapsto \rho^{\mu}(F)$ in combination with convergence rates of empirical measure in Wasserstein distance—at least if $\mathcal{X} = \mathbb{R}^d$ and F, G are Lipschitz continuous. Whereas this technique certainly works for dimension $d = 1$, in the present general, multidimensional setting, this approach forces the convergence rates to be significantly worse: in dimension $d \geq 3$, the Wasserstein distance converges with rate $N^{-1/d}$; see Fournier and Guillin [20]. Thus, even for $q \approx 1$, these arguments give the rate $N^{-1/d}$ in Theorem 2.1 instead of $N^{-1/2}$.

2.2. Results for AVaR, OCE, and SF Risk Measures

It turns out that, for all the specific risk measures discussed in the introduction, the optimal rate $N^{-1/2}$ can be obtained and with easier arguments. We, therefore, state the results for these risk measures separately. For any measurable $F : \mathcal{X} \to \mathbb{R}$, recall that the shortfall risk measure (Föllmer and Schied [19]) is defined as

$$\text{SF}^{\mu}(F) := \inf\{m \in \mathbb{R} : E[l(F(S) - m)] \leq 0\}.$$

Here, $l : \mathbb{R} \to [-1,\infty)$ is a loss function, meaning that l is increasing and convex such that $1 \in \partial l(0)$ (the subdifferential at point zero) and $l(0) = 0$. In other words, $\text{SF}^{\mu}(F)$ is the smallest capital m by which we should reduce the loss F to make it acceptable, meaning that the expected loss $E[l(F(S) - m)]$ is below the threshold $l(0) = 0$.

In a similar spirit, the OCE of Ben-Tal and Teboulle [7, 8] is defined via

$$\text{OCE}^{\mu}(F) := \inf_{m \in \mathbb{R}} (E[l(F(S) - m)] + m). \tag{2.1}$$

Again, l is a loss function, and the interpretation is similar to that of shortfall risk. Importantly, OCEs cover popular risk measures, such as the average value at risk obtained with $l(x) = x^+/(1-u)$ or the entropic risk measure obtained with $l(x) = e^x - 1$. The following result gives the convergence rate for these particular examples of risk measures.

Theorem 2.2 (Rates for AVaR, OCE, and SF). *Let $\rho = \text{OCE}$ or $\rho = \text{SF}$, and in the latter case, assume that l is strictly increasing. There are constants $c, C > 0$ such that the following hold:*

i. *We have the moment bound*

$$E[|\pi^{\mu}(F) - \pi^{\mu_N}(F)|] \le \frac{C}{\sqrt{N}}$$

for all $N \ge 1$.

ii. *We have the matching deviation inequality*

$$P[|\pi^{\mu}(F) - \pi^{\mu_N}(F)| \ge \varepsilon] \le C\exp(-cN\varepsilon^2)$$

for all $N \ge 1$ *and all* $\varepsilon > 0$.

The constants c and C depend on l; the maximal range of F, G; the number of options e; and the largest Euclidean norm in $\mathcal{G}$. The rates obtained in both parts of Theorem 2.2 are the usual rates dictated by the central limit theorem and, in particular, are optimal; see Section 6.

Remark 2.2. Note that $\pi^{\mu_N}(F)$ is readily checked to be measurable, and we do not need to resort to the outer expectation and probability in Theorem 2.2. In fact, if, for instance, $\rho = \mathrm{OCE}$, then, by continuity of the function l, we can write

$$\pi^{\mu_N}(F) := \inf_{g\in\mathcal{G},\, m\in\mathbb{Q}} \frac{1}{N}\sum_{n\le N} l\left(F(S_n) + \sum_{i=1}^{e} g_i G_i(S_n) - m\right) + m.$$

Recalling that $\mathcal{G}$ is countable, this shows that the random variable $\pi^{\mu_N}(F)$ is measurable.

As before, Theorem 2.2 implies strong consistency.

Corollary 2.2 (Consistency). *In the setting of Theorem 2.2,* $\lim_{N\to\infty}\pi^{\mu_N}(F) = \pi^{\mu}(F)$ *P-almost surely.*

Let us conclude this section with a short discussion on the biasedness of $\rho^{\mu_N}(F)$ claimed in the introduction.

Remark 2.3 (Biasedness). For typical risk measures, $\rho^{\mu_N}(F)$ underestimates $\rho(F)$. This is easiest explained by considering the optimized certainty equivalents. In fact, we have

$$\begin{aligned} E[\mathrm{OCE}^{\mu_N}(F)] &= E\left[\inf_{m\in\mathbb{R}} \int_{\mathcal{X}} l(F(x) - m) + m\,\mu_N(dx)\right] \\ &\le \inf_{m\in\mathbb{R}} E\left[\int_{\mathcal{X}} l(F(x) - m) + m\,\mu_N(dx)\right] \\ &= \mathrm{OCE}^{\mu}(F), \end{aligned}$$

where the last equality follows as $E[\int f\, d\mu_N] = \int f\, d\mu$ for every μ-integrable function f. The same applies in the presence of trading, namely, $E[\pi^{\mu_N}(F)] \le \pi^{\mu}(F)$.

More generally, a quick inspection of OCE and SF reveals that both are concave considered as mappings of μ. As a matter of fact, this very concavity is the reason for the bias. Indeed, concavity and lower semicontinuity of $\mu \longmapsto \rho^{\mu}(F)$ implies, thanks to Jensen's inequality for infinite dimensional random variables (see, e.g., Nonnenmacher and Zagst [43, theorem 3.1]), that

$$E[\rho^{\mu_N}(F)] \le \rho^{E[\mu_N]}(F) = \rho^{\mu}(F),$$

where we use that the measure-valued random variable μ_N has mean μ. Whereas it should be noted that not all law-invariant risk measures are concave in μ, this is often the case;[6] see Acciaio and Svindland [1].

We further refer to Pitera and Schmidt [47] for a more in-depth discussion on the issue of biasedness and some empirical evidence.

2.3. Utility Maximization

It is conceivable that most of the results and methods of the present article extend beyond the estimation of risk measures. Other issues that seem to fit to our framework and method include the estimation of risk premium principles in insurance, (see, e.g., Young [52] or Furman and Zitikis [21] for an overview), or estimation of the value of some stochastic optimization problems.

To illustrate the latter, let us consider another popular approach for quantifying the riskiness of a position, namely, utility maximization: let $U : \mathbb{R} \to \mathbb{R}$ be a concave increasing function and set $u^{\mu}(F) := E^{\mu}[U(F(S))]$. Similar to before, allowing the agent to invest in a market (with stock returns $G_1, \ldots, G_e$), one obtains the utility maximization problem

$$u^{\mu}_{\max}(F) := \sup_{g\in\mathcal{G}} u^{\mu}\left(F + \sum_{i=1}^{e} g_i G_i\right).$$

In this case, we have the following.

Proposition 2.2 (Utility Maximization). *There are constants $c, C > 0$ such that*

$$E[|u^{\mu}_{\max}(F) - u^{\mu_N}_{\max}(F)|] \le \frac{C}{\sqrt{N}},$$
$$P[|u^{\mu}_{\max}(F) - u^{\mu_N}_{\max}(F)| \ge \varepsilon] \le C\exp(-cN\varepsilon^2)$$

for all $N \ge 1$ and $\varepsilon > 0$.

Again, note that the rates are optimal and do not depend on the dimension of the underlying nor the number e of available options and $u^{\mu_N}_{\max}(F)$ is a strongly consistent estimator that typically overestimates its true value (as we deal with maximization instead of minimization this time).

3. Rates for Average Value at Risk and Optimized Certainty Equivalents

Let us briefly fix our notation: throughout this paper, we make the important convention that $C > 0$ is a generic constant. This means that C may depend on all kinds of parameters (such as some L^p norms of F and G; features of the risk measure, such as growth of the loss function l in the OCE/SF case, etc.) but not on N. Moreover, the value of C is allowed to increase from line to line; for instance, $\sup_y(xy - y^2) = Cx^2 \le Cx^2/2$ or $C\sqrt{e+1} \le C\sqrt{e}$ for all $e \in \mathbb{N}$, but not $N \le C$ or $\sqrt{e+1} \le \sqrt{e}/C$.

To keep the distinction between the analysis for bounded random variables (i.e., random variables in L^p for $p = \infty$) and unbounded random variables ($p < \infty$) as light as possible, we use the following conventions: we put $1/\infty := 0$, and for $x > 0$, $x^0 := 1$ and $x^\infty := \infty$.

For a metric space (Λ, d_Λ) and $\varepsilon > 0$, denote by $\mathcal{N}(\Lambda, d_\Lambda, \varepsilon)$ the covering numbers at scale ε; that is, $\mathcal{N}(\Lambda, d_\Lambda, \varepsilon)$ is the smallest number for which there is a subset $\tilde{\Lambda}$ with that cardinality satisfying that, for every $s \in \Lambda$, there is $\tilde{s} \in \tilde{\Lambda}$ with $d_\Lambda(s, \tilde{s}) \le \varepsilon$. In other words, $\mathcal{N}(\Lambda, d_\Lambda, \varepsilon)$, the smallest number of balls of radius ε that covers Λ. The latter suggests this to be some measurement of compactness, and in fact, it is an important tool in understanding the behavior of empirical processes; see van der Vaart and Wellner [50].

Recall that $e \in \mathbb{N}$ is a fixed number and $F, G_1, \ldots, G_e : \mathcal{X} \to \mathbb{R}$ are measurable functions. For shorthand notation, write $g \cdot G := \sum_{i=1}^e g_i G_i$ for $g \in \mathcal{G}$ and $|G| := \sum_{i=1}^e |G_i|$. Recall that, throughout this article, the set $\mathcal{G} \subset \mathbb{R}^e$ is assumed to be countable and bounded. The former assumption is without loss of generality (see Remark A.1), and the latter is shown to be necessary in Proposition A.1.

The average value at risk also goes under several different names, such as expected shortfall, conditional value at risk, and expected tail loss, and has equally many different (equivalent) definitions, for instance, as the value at risk integrated over different levels; see Föllmer and Schied [19, section 4.3] for an overview. We only use the definition of AVaR given in (1.1). Given a loss function l, also recall the definition of OCE given in (2.1). We additionally assume that $\liminf_{x\to\infty} l(x)/x > 1$, which, by convexity and $1 \in \partial l(0)$, is equivalent to the fact that $l(x) > x$ for some $x \ge 0$. This assumption is there because F and G are possibly not bounded (in contrast to Section 2) but not needed if this is the case.

We often work under the assumption that l' (the right-continuous derivative of the convex function l) has polynomial growth of degree $p-1$, which means that $l'(x) \le C(1 + |x|^{p-1})$ for all $x \in \mathbb{R}$. In particular, recalling the convention $|x|^\infty := \infty$ for $x \ne 0$, we see that polynomial growth of degree ∞ is no restriction at all; for instance, the exponential function $l = \exp$ satisfies this assumption (only) for $p = \infty$.

The goal of this section is to prove Theorem 2.2, part (i), or, rather, the following generalization thereof.

Theorem 3.1. *Let $p \in [1, \infty]$ and assume that l' has polynomial growth of degree $p-1$ and $\|F\|_{L^{2p}(\mu)}$ and $\|G\|_{L^{2p}(\mu)}$ are finite. Then,*

$$E\left[\sup_{g\in\mathcal{G}} |\mathrm{OCE}^{\mu}(F + g\cdot G) - \mathrm{OCE}^{\mu_N}(F + g\cdot G)|\right] \le \frac{C}{\sqrt{N}}$$

for all $N \ge 1$. The constant C depends on μ only through the size of the $L^{2p}(\mu)$-norms of F and G on e, p, and the diameter of $\mathcal{G}$.

Observe that the measurability of $\mathrm{OCE}^{\mu_N}(F + g\cdot G)$ is readily checked; see, for instance, Remark 2.2.

Before presenting the proof of Theorem 3.1, let us shortly elaborate on the integrability conditions therein. These assumptions cannot be improved in general. In fact, consider the trivial case $\mathcal{G} = \{0\}$ and $u = 0$ for which we have $\mathrm{AVaR}_0(F(S)) = E[F(S)]$. Thus, we have $p = 1$, and it is well-known that

$$E\left[\left|\frac{1}{N}\sum_{n\le N} F(S_n) - E[F(S)]\right|\right] \le \frac{C}{\sqrt{N}}$$

requires that $F(S)$ has a finite second moment, that is, that $\|F\|_{L^2(\mu)}$ is finite. We now turn to the proof of Theorem 3.1. In fact, looking at the definition of the optimized certainty equivalent, the reader familiar with the theory of empirical processes recognizes this as a standard problem covered within this theory. Thus, at some point, an estimate of the covering numbers with respect to the random $L^2(\mu_N)$ norm must be computed. Fortunately, no geometric arguments are needed, and all randomness can be controlled by some estimates involving moments only. For this reason, it is useful to keep track of the following quantities:

$$J := 1 + |F| + |G|; \quad M := \|J\|_{L^p(\mu)} \quad \text{and} \quad M_N := \|J\|_{L^p(\mu_N)}. \tag{3.1}$$

The first result in this spirit follows.

Lemma 3.1. *Assume that l' has polynomial growth of degree $p-1$. Then, we have that*

$$|\mathrm{OCE}^{\mu}(F + g\cdot G)| \le CM^p \quad \textit{and} \tag{3.2}$$

$$\mathrm{OCE}^{\mu}(F + g\cdot G) = \inf_{|m|\le CM^p} \int_{\mathcal{X}} l(F(x) + g\cdot G(x) - m) + m\,\mu(dx) \tag{3.3}$$

for every $g \in \mathcal{G}$. The same holds true if the pair (μ, M) is replaced by (μ_N, M_N) (with the constant C in (3.2) not depending on N).

Proof. Assume without loss of generality that $M < \infty$; otherwise, there is nothing to show.

As l is increasing and of polynomial growth with degree p and $\mathcal{G}$ is bounded, we have that

$$\sup_{g\in\mathcal{G}} l(F + g\cdot G) \le \begin{cases} CJ^p & \text{if } p < \infty, \\ C & \text{if } p = \infty. \end{cases} \tag{3.4}$$

In particular, the choice $m = 0$ (in the definition of OCE) and the fact that $l \ge -1$ yield

$$\mathrm{OCE}^{\mu}(F + g\cdot G) \le \int_{\mathcal{X}} l(F(x) + g\cdot G(x))\,\mu(dx) \le CM^p$$

for all $g \in \mathcal{G}$, showing the upper bound in (3.2). Further, as $l \ge -1$ and $M \ge 1$, this also implies that the infimum over m in the definition of $\mathrm{OCE}^{\mu}(F + g\cdot G)$ can be restricted to $m \le CM^p$ for all $g \in \mathcal{G}$.

On the other hand, by convexity of l and the assumption that $\liminf_{x\to\infty} l(x)/x > 1$, there exist $a > 1$ and $b \in \mathbb{R}$ such that $l(x) \ge ax - b$ for every $x \in \mathbb{R}$. This implies

$$\begin{aligned} &\int_{\mathcal{X}} l(F(x) + g\cdot G(x) - m) + m\,\mu(dx) \\ &\ge \int_{\mathcal{X}} a(-CJ(x) - m) - b + m\,\mu(dx) \\ &\ge m(1-a) - CM^p, \end{aligned} \tag{3.5}$$

where we use that $\int_{\mathcal{X}} J\,d\mu \le M \le M^p$, which follows from Hölder's inequality and as $M \ge 1$. By the previous part, we already know that $\mathrm{OCE}^{\mu}(F + g\cdot G) \le CM^p$ for all $g \in \mathcal{G}$. Together with (3.5), this implies that the infimum over m in $\mathrm{OCE}^{\mu}(F + g\cdot G)$ can be restricted to $m \ge -CM^p$ for all $g \in \mathcal{G}$. In turn, using once more that $l \ge -1$, this also implies that $\mathrm{OCE}^{\mu}(F + g\cdot G) \ge -CM^p$ for all $g \in \mathcal{G}$ and, thus, completes the proof for (μ, M).

Observe that (3.2) and (3.3) with (μ, M) replaced by (μ_N, M_N) is obtained using exactly the same argument as before with (μ, M) replaced by (μ_N, M_N). In fact, by (3.4), we have $\mathrm{OCE}^{\mu_N}(F + g\cdot G) \le CM_N^p$, which implies that the infimum in the definition of $\mathrm{OCE}^{\mu_N}(F + g\cdot G)$ can be restricted to $m \le CM_N^p$ P-a.s. for all $g \in \mathcal{G}$. On the other hand, as in (3.5), we have

$$\int_{\mathcal{X}} l(F(x) + g\cdot G(x) - m) + m\,\mu_N(dx) \ge m(1-a) - CM_N^p,$$

from which we infer that the infimum in the definition of $\mathrm{OCE}^{\mu_N}(F + g\cdot G)$ can be restricted to $m \ge -CM_N^p$ P-a.s. for all $g \in \mathcal{G}$. This, thus, shows $\mathrm{OCE}^{\mu_N}(F + g\cdot G) \ge -CM_N^p$. □

Lemma 3.2. *Assume that l' has polynomial growth of degree $p-1$, let $m_0 \in \mathbb{R}$, and define*

$$\mathcal{H} := \{l(F + g\cdot G - m) + m : g \in \mathcal{G} \text{ and } m \in [-m_0, m_0]\}.$$

Then, for every $\varepsilon > 0$, we have that

$$\mathcal{N}(\mathcal{H}, \|\cdot\|_{L^2(\mu_N)}, \varepsilon) \le \left(\frac{C\|J\|^p_{L^{2p}(\mu_N)}}{\varepsilon}\right)^{e+1} \vee 1$$

if $p < \infty$, and $\mathcal{N}(\mathcal{H}, \|\cdot\|_{L^2(\mu_N)}, \varepsilon) \le (C/\varepsilon)^{e+1} \vee 1$ if $p = \infty$.

Proof. Without loss of generality, we work only on the set in which $\|J\|_{L^{2p}(\mu_N)} < \infty$ (otherwise, there is nothing to show). We proceed in two steps.

a. Pick two elements $H, \tilde{H} \in \mathcal{H}$ represented as

$$\begin{aligned} H &= l(F + g \cdot G - m) + m \quad \text{and} \\ \tilde{H} &= l(F + \tilde{g} \cdot G - \tilde{m}) + \tilde{m} \end{aligned}$$

and define the family of functions $(\varphi_t)_{t\in[0,1]}$ from $\mathcal{X}$ to $\mathbb{R}$ by

$$\varphi_t := F + g \cdot G - m + t((\tilde{g} - g) \cdot G + m - \tilde{m})$$

for every $t \in [0,1]$. Then, $H = l(\varphi_0) + m$, and $\tilde{H} = l(\varphi_1) + \tilde{m}$. As $\mathcal{G}$ is bounded, $|\varphi_t| \le CJ$ for all $t \in [0,1]$. By convexity of l, its right derivative l' is increasing. By the fundamental theorem of calculus, we have

$$\begin{aligned} \|H - \tilde{H}\|_{L^2(\mu_N)} &\le \left\| \int_0^1 l'(\varphi_t)\partial_t\varphi_t \, dt \right\|_{L^2(\mu_N)} + |m - \tilde{m}| \\ &\le \|l'(CJ)((\tilde{g} - g) \cdot G + m - \tilde{m})\|_{L^2(\mu_N)} + |m - \tilde{m}|. \end{aligned}$$

Now, note that

$$\|l'(CJ)J\|_{L^2(\mu_N)} \le \begin{cases} C\|J\|^p_{L^{2p}(\mu_N)} & \text{if } p < \infty, \\ C & \text{if } p = \infty. \end{cases}$$

Indeed, for $p < \infty$, this follows from the assumption that $l'(x) \le C(1 + |x|^{p-1})$ for all $x \in \mathbb{R}$ and the fact that $J \ge 1$. For $p = \infty$, one has, by assumption, that J is μ-almost surely bounded. Hence, P-almost surely, J is also μ_N-almost surely bounded (by the same constant). As l is bounded on bounded sets (by convexity), this implies that $l'(J)$ is μ_N-almost surely bounded.

To conclude, we use once more that $\mathcal{G}$ is bounded, and hence, $|(\tilde{g} - g) \cdot G| \le |\tilde{g} - g|J$. Therefore,

$$\|H - \tilde{H}\|_{L^2(\mu_N)} \le \begin{cases} C\|J\|^p_{L^{2p}(\mu_N)}(|g - \tilde{g}| + |m - \tilde{m}|) & \text{if } p < \infty, \\ C(|g - \tilde{g}| + |m - \tilde{m}|) & \text{if } p = \infty. \end{cases} \tag{3.6}$$

In the following, we restrict to $p < \infty$ and leave the obvious changes needed when $p = \infty$ to the reader.

b. Fix $\varepsilon > 0$ and let $A \subset [-m_0, m_0]$ be such that

$$\text{for all } m \in [-m_0, m_0] \quad \text{there is} \quad \tilde{m} \in A \text{ with} \quad |m - \tilde{m}| \le \frac{\varepsilon}{2C\|J\|^p_{L^{2p}(\mu_N)}}$$

and $B \subset \mathcal{G}$ such that

$$\text{for all} \quad g \in \mathcal{G} \quad \text{there is} \quad \tilde{g} \in B \quad \text{with} \quad |g - \tilde{g}| \le \frac{\varepsilon}{2C\|J\|^p_{L^{2p}(\mu_N)}}.$$

Then, if we define $\tilde{\mathcal{H}}$ exactly as $\mathcal{H}$ only with $[-m_0, m_0]$ replaced by A and $\mathcal{G}$ replaced by B, then, by (3.6), for every $H \in \mathcal{H}$, there is $\tilde{H} \in \tilde{\mathcal{H}}$ with $\|H - \tilde{H}\|_{L^2(\mu_N)} \le \varepsilon$.

This implies that

$$\begin{aligned} \mathcal{N}(\mathcal{H}, \|\cdot\|_{L^2(\mu_N)}, \varepsilon) &\le \operatorname{card}(\tilde{\mathcal{H}}) \\ &\le \operatorname{card}(A \times B) = \operatorname{card}(A)\operatorname{card}(B), \end{aligned}$$

where card means cardinality.

The set A can be constructed simply by an equidistant partition of $[-m_0, m_0]$ at cardinality $\operatorname{card}(A) \le (C\|J\|^p_{L^{2p}(\mu_N)}/\varepsilon) \vee 1$. In a similar manner, B can be constructed with $\operatorname{card}(B) \le (C\|J\|^p_{L^{2p}(\mu_N)}/\varepsilon)^e \vee 1$.

Combining both steps yields the proof. □

In order to apply results from theory of empirical processes, we need the following observation.

Lemma 3.3. *The set $\mathcal{H}$ defined in Lemma 3.2 satisfies Assumption A.1.*

Proof. Let $A \subset [-m_0, m_0]$ be countable and dense and define

$$\mathcal{H}' := \{l(F + g\cdot G - m) : g \in \mathcal{G} \text{ and } m \in A\}.$$

The set $\mathcal{H}'$ is clearly countable. Let $H = l(F + g\cdot G - m) \in \mathcal{H}$ and let $(m^n)_n \subset A$ be a sequence that converges to m. Then, $H^n := l(F + g\cdot G - m^n) \in \mathcal{H}'$ converges pointwise to H and in $L^2(\nu)$ for every measure ν such that $l(|F| + |G|) \in L^2(\nu)$ (by dominated convergence). □

Inspecting the proof actually yields the following result, which we state for later reference.

Corollary 3.1. *Let $m_0 \in \mathbb{R}$, let $f : \mathbb{R} \to \mathbb{R}$ be locally Lipschitz continuous, and assume that J is bounded. Then, it holds that*[7]

$$\mathcal{N}(\{f(F + g\cdot G - m) : g \in \mathcal{G} \text{ and } \quad m \in [-m_0, m_0]\}, \|\cdot\|_\infty, \varepsilon) \le \left(\frac{C}{\varepsilon}\right)^{e+1} \vee 1$$

for every $\varepsilon > 0$.

We are now ready for the following.

Proof of Theorem 3.1. For shorthand notation, set

$$\Delta_N := \sup_{g\in\mathcal{G}} |\mathrm{OCE}^{\mu}(F + g\cdot G) - \mathrm{OCE}^{\mu_N}(F + g\cdot G)|$$

for every $N \ge 1$. With M and M_N defined in (3.1), we write

$$E[\Delta_N] = E[\Delta_N 1_{M_N \le M+1}] + E[\Delta_N 1_{M_N > M+1}]$$

and investigate both terms separately.

a. We start with the first term. Lemma 3.1 guarantees that

$$\Delta_N 1_{M_N \le M+1} \le \sup_{H\in\mathcal{H}} \left| \int_{\mathcal{X}} H(x)\,(\mu - \mu_N)(dx) \right|$$

for every $N \ge 1$, where

$$\mathcal{H} := \{l(F + g\cdot G - m) + m : g \in \mathcal{G} \text{ and } |m| \le C(M+1)^p\}.$$

By Lemma 3.3, the set $\mathcal{H}$ satisfies Assumption A.1. Therefore, the "empirical process version" of Dudley's entropy-integral theorem (i.e., Theorem A.1) implies that

$$\begin{aligned} &E\left[\sup_{H\in\mathcal{H}} \left| \int_{\mathcal{X}} H(x)\,(\mu - \mu_N)(dx) \right|\right] \\ &\le \frac{C}{\sqrt{N}} \left(E[H^*(S)^2]^{\frac{1}{2}} + E\left[\int_0^\infty \sqrt{\log \mathcal{N}(\mathcal{H}, \|\cdot\|_{L^2(\mu_N)}, \varepsilon)}\, d\varepsilon \right] \right) \end{aligned}$$

for all $N \ge 1$, where $H^* := l(F + g^*\cdot G) \in \mathcal{H}$ for some $g^* \in \mathcal{G}$. By definition of J, we have that $E[H^*(S)^2]^{\frac{1}{2}} \le C\|J\|^p_{L^{2p}(\mu)}$. It remains to gain control over the entropy integral term.

Assume first that $p < \infty$. Then, estimating the covering numbers of $\mathcal{H}$ by means of Lemma 3.2 implies that

$$\begin{aligned} &E\left[\int_0^\infty \sqrt{\log \mathcal{N}(\mathcal{H}, \|\cdot\|_{L^2(\mu_N)}, \varepsilon)}\, d\varepsilon \right] \\ &\le CE\left[\int_0^\infty \sqrt{\log\left(\frac{C\|J\|^p_{L^{2p}(\mu_N)}}{\varepsilon} \vee 1 \right)}\, d\varepsilon \right] \\ &\le CE\left[\|J\|^p_{L^{2p}(\mu_N)} \int_0^\infty \sqrt{\log\left(\frac{1}{\tilde{\varepsilon}} \vee 1 \right)}\, d\tilde{\varepsilon} \right], \end{aligned}$$

where the last inequality follows from substituting ε by $\tilde{\varepsilon} := \varepsilon / C\|J\|^p_{L^{2p}(\mu_N)}$. In a final step, notice that

$$\int_0^\infty \sqrt{\log\left(\frac{1}{\varepsilon} \vee 1 \right)}\, d\varepsilon < \infty \quad \text{and} \quad E\left[\|J\|^p_{L^{2p}(\mu_N)} \right] \le C\|J\|^p_{L^{2p}(\mu)}.$$

The second statement follows from Jensen's inequality. Therefore,

$$E[\Delta_N 1_{M_N \le M+1}] \le \frac{C}{\sqrt{N}}$$

for all $N \ge 1$, showing that the first term behaves as required. If $p = \infty$, the same arguments apply (with $\|J\|^p_{L^{2p}(\mu_N)}$ replaced by a constant and Corollary 3.1 applied instead of Lemma 3.2), and we again obtain $E[\Delta_N 1_{M_N \le M+1}] \le C/\sqrt{N}$.

b. As for the second term, applying Hölder's inequality yields

$$E[\Delta_N 1_{M_N > M+1}] \le E[\Delta_N^2]^{1/2} P[M_N > M+1]^{1/2}. \tag{3.7}$$

We start by estimating $P[M_N > M+1]^{1/2}$. For $p = \infty$, one has $P[M_N > M+1] = 0$ for all N. For $p < \infty$, using first that $M, M_N \ge 1$ and then Chebycheff's inequality, we estimate

$$\begin{aligned} P[M_N - M > 1] &\le P[M_N^p - M^p > 1] \\ &\le E[(M_N^p - M^p)^2]. \end{aligned}$$

Further, making use of the fact that the $(S_1, \ldots, S_N)$ are independent with $M^p = E[J(S_n)^p]$ for all n, one has

$$\begin{aligned} E[(M_N^p - M^p)^2] &= E\left[\left(\frac{1}{N}\sum_{n=1}^N (J(S_n)^p - E[J(S_n)^p])\right)^2\right] \\ &= \frac{1}{N} E[(J(S_1)^p - E[J(S_1)^p])^2] \\ &\le \frac{2\|J\|^{2p}_{L^{2p}(\mu)}}{N}. \end{aligned}$$

This shows that $P[M_N > M+1]^{1/2} \le C/\sqrt{N}$.

Regarding $E[\Delta_N^2]$, use Lemma 3.1 to estimate

$$E[\Delta_N^2] \le C(M^{2p} + E[M_N^{2p}]).$$

The same arguments as before show that $E[M_N^{2p}] \le \|J\|^{2p}_{L^{2p}(\mu)}$. Plugging both estimates into (3.7) shows that

$$E[\Delta_N 1_{M_N > M+1}] \le \frac{C}{\sqrt{N}}$$

for all $N \ge 1$.

Putting both estimates together, we obtain $E[\Delta_N] \le C/\sqrt{N}$ for all $N \ge 1$. This completes the proof. □

4. General Law-Invariant Risk Measures

This section deals with general risk measures, which we start by briefly describing. First, in order to allow for unbounded F and G, one needs to define risk measures for unbounded functions. A function $\rho : L^p \to \mathbb{R}$ with $p \in [1, \infty]$ is again called a (convex) law-invariant risk measure if (a)–(d) of Definition 1.1 hold with L^∞ replaced by L^p. Further, recall that ρ is called "coherent" if, in addition, $\rho(\lambda X) = \lambda \rho(X)$ for all $X \in L^p$ and $\lambda \ge 0$.

As already mentioned, by Jouini et al. [31], every law-invariant risk measure automatically satisfies the Fatou property as well as the spectral representation[8]

$$\rho(X) = \sup_{\gamma \in \mathcal{M}} \left(\int_{[0,1)} \mathrm{AVaR}_u(X)\, \gamma(du) - \beta(\gamma) \right) \quad \text{for } X \in L^p. \tag{4.1}$$

See Gao et al. [22] for the case of unbounded random variables. Here, $\mathcal{M}$ is a subset of probability measures on $[0,1)$ armed with its Borel σ-field, $\beta : \mathcal{M} \to [0, \infty)$ is a convex function, and AVaR is the average value at risk defined in (1.1). Note that AVaR is evidently a coherent law-invariant risk measure. Recall the definition of $J := 1 + |F| + |G|$ already given in (3.1).

Before we are ready to state the generalization of part (i) of Theorem 2.1, the treatment of unbounded F, G requires one last definition: for every parameter $p \in [1, \infty]$ and $x \geq 0$ set

$$w_p(x) := \sup\{\rho(X) : \|X\|_{L^p} \leq x\}.$$

Note that w_p is convex, nonnegative, and w_p grows at least linearly. Moreover, in the important case of a coherent risk measure or if $p = \infty$, the function w_p is linear.

Theorem 4.1. *Let $1 < q \leq p \leq \infty$ and let $\rho : L^p \to \mathbb{R}$ be a law-invariant risk measure that is q-regular. Assume that $\mathcal{G}$ is bounded and that*

a. $\|w_p(tJ^p)\|_{L^2(\mu)} < \infty$ *for every $t \geq 0$ in the case that $p < \infty$.*

b. *J is bounded in the case that $p = \infty$.*

Then,

$$E\left[\sup_{g \in \mathcal{G}} |\rho^{\mu}(F + g \cdot G) - \rho^{\mu_N}(F + g \cdot G)|\right] \leq \frac{C}{N^{\frac{1/q - 1/2p}{2 - 1/p}}}$$

for all $N \geq 1$.

Note that, if w_p is linear, then $\|w_p(tJ^p)\|_{L^2(\mu)} < \infty$ simply means that $\|J\|_{L^{2p}(\mu)}$ is finite. In general, $\|w_p(tJ^p)\|_{L^2(\mu)} < \infty$ always implies that $\|J\|_{L^{2p}(\mu)} < \infty$ (by convexity of w_p). For convenience, we compute in Table 1 some values of the convergence rates obtained in Theorem 4.1.

Further observe that the rate $R_{p,q}$ is increasing in p and decreasing in q. The idea of the proof of Theorem 4.1 is the following: by Section 3, we understand the behavior of the mean error for the average value at risk (being a special case of the optimized certainty equivalents). By the spectral representation (4.1), AVaR forms the building block of every law-invariant risk measure, and we conclude via a (multiscale) approximation, keeping track of the risk-aversion parameter u of the average value at risk (which makes all constants explode when approaching $u \approx 1$) and the growth of measures $\gamma(du)$ in the spectral representation (4.1) (which only puts little mass on $u \approx 1$).

The preparatory work needed is done in the next few lemmas.

Lemma 4.1. *Let the assumptions of Theorem 4.1 be satisfied. Let X^* be Pareto-distributed with scale parameter one and shape parameter q. Then, we have that*

$$\mathrm{AVaR}_u(X^*) = \frac{q}{q-1}\frac{1}{(1-u)^{1/q}}$$

for every $u \in [0, 1)$.

Proof. The proof follows from an elementary calculation, for example, by involving the quantile representation

$$\mathrm{AVaR}_u(X^*) = \frac{1}{1-u}\int_u^1 q_{X^*}(t)\,dt$$

of the average value at risk (Föllmer and Schied [19, proposition 4.51]), where $q_{X^*}(t)$ denotes the t-quantile of X^*. □

Lemma 4.2. *Let the assumptions of Theorem 4.1 be satisfied. For every $p \in (1, \infty]$ and $X \in L^p$, we have that*

$$|\mathrm{AVaR}_u(X)| \leq \frac{\|X\|_{L^p}}{(1-u)^{1/p}}$$

for every $u \in [0, 1)$.

Proof. For $p = \infty$, the claim is trivial. For $p < \infty$, we again involve the quantile representation of the average value at risk and apply Hölder's inequality

$$|\mathrm{AVaR}_u(X)| = \frac{1}{1-u}\left|\int_{[0,1]} 1_{[u,1]}(t) q_X(t)\,dt\right| \leq \frac{(1-u)^{\frac{p-1}{p}}}{1-u}\left(\int_{[0,1]} q_X(t)^p\,dt\right)^{\frac{1}{p}}.$$

As the last integral equals $\|X\|_{L^p}$, this completes the proof. □

Table 1. Convergence rates for different values of p and q.

	$q \approx 1$	$q = 2$	$q = p$	$p = \infty$
$R_{p,q} := \frac{1/q - 1/2p}{2 - 1/p}$	$\approx \frac{1}{2}$	$\frac{p-1}{4p-2}$	$\frac{1}{2(2p-1)}$	$\frac{1}{2q}$

Lemma 4.3. *Let the assumptions of Theorem 4.1 be satisfied. For every fixed $a > 0$, there exists a constant $b > 0$ such that*

$$\rho(X) = \sup_{\gamma\in\mathcal{M}:\ \text{s.t.}\ \beta(\gamma)\le b} \left(\int_{[0,1)} \mathrm{AVaR}_u(X)\,\gamma(du) - \beta(\gamma) \right)$$

for all $X \in L^p$ with $\|X\|_{L^p} \le a$.

Proof. Let X^* be the random variable of Lemma 4.1.

a. In a first step, we show that $|\rho(X)| \le C$ for all $X \in L^p$ with $\|X\|_{L^p} \le a$. For such X, by Lemmas 4.1 and 4.2, one has that

$$\mathrm{AVaR}_u(|X|) \le \frac{a}{(1-u)^{1/p}} \le \frac{a}{(1-u)^{1/q}} = \mathrm{AVaR}_u(CX^*) \tag{4.2}$$

for every $u \in [0,1)$. Here, we use that $q \le p$; hence, $(1-u)^{1/p} \ge (1-u)^{1/q}$. Therefore,

$$\begin{aligned}
\rho(|X|) &\le \sup_{\gamma\in\mathcal{M}} \left(\int_{[0,1)} \mathrm{AVaR}_u(|X|)\,\gamma(du) - \beta(\gamma) \right)\\
&\le \sup_{\gamma\in\mathcal{M}} \left(\int_{[0,1)} \mathrm{AVaR}_u(CX^*)\,\gamma(du) - \beta(\gamma) \right)\\
&\le \sup_{n\in\mathbb{N}} \sup_{\gamma\in\mathcal{M}} \left(\int_{[0,1)} \mathrm{AVaR}_u(CX^* \wedge n)\,\gamma(du) - \beta(\gamma) \right) = \sup_{n\in\mathbb{N}} \rho(CX^* \wedge n)
\end{aligned}$$

for every X with $\|X\|_{L^p} \le a$, where the latter inequality follows by monotone convergence. Note that CX^* again follows a Pareto distribution with shape parameter q, and hence, $\sup_{n\in\mathbb{N}} \rho(CX^* \wedge n)$ is finite by definition of q-regularity.

It further follows by convexity and monotonicity of ρ together with $\rho(0) = 0$ that $|\rho(X)| \le \rho(|X|)$ for all $X \in L^p$. This implies that, indeed, $|\rho(X)| \le C$ for all $X \in L^p$ with $\|X\|_{L^p} \le a$.

b. We proceed to prove the claim. Define

$$\varphi : \mathbb{R}_+ \to [0,\infty] \quad \text{by} \quad \varphi(y) := \sup_{x\in\mathbb{R}_+} \left(xy - \sup_{n\in\mathbb{N}} \rho(xX^* \wedge n) \right).$$

Then, φ is convex, increasing, and as $\sup_{n\in\mathbb{N}} \rho(xX^* \wedge n) < \infty$ for all $x \in \mathbb{R}_+$, one can verify that $\varphi(y)/y \to \infty$ as $y \to \infty$. Now note that the (spectral) representation of ρ in (4.1) implies that

$$\sup_{n\in\mathbb{N}} \rho(xX^* \wedge n) \ge \int_{[0,1)} \mathrm{AVaR}_u(xX^*)\,\gamma(du) - \beta(\gamma)$$

for all $x \ge 0$ and $\gamma \in \mathcal{M}$. Therefore, one has

$$\begin{aligned}
\beta(\gamma) &\ge \sup_{x\ge 0} \left(\int_{[0,1)} \mathrm{AVaR}_u(xX^*)\,\gamma(du) - \rho(xX^*) \right)\\
&= \varphi\left(\int_{[0,1)} \mathrm{AVaR}_u(X^*)\,\gamma(du) \right)
\end{aligned}$$

for every $\gamma \in \mathcal{M}$. For every X with $\|X\|_{L^p} \le a$, by (4.2), one has

$$\begin{aligned}
\int_{[0,1)} \mathrm{AVaR}_u(X)\,\gamma(du) - \beta(\gamma) &\le C \int_{[0,1)} \mathrm{AVaR}_u(X^*)\,\gamma(du) - \beta(\gamma)\\
&\le C\varphi^{-1}(\beta(\gamma)) - \beta(\gamma),
\end{aligned} \tag{4.3}$$

where φ^{-1} denotes the (right)-inverse of φ.

As $\varphi(y)/y \to \infty$ when $y \to \infty$, one has that $\varphi^{-1}(x)/x \to 0$ when $x \to \infty$, which implies that

$$C\varphi^{-1}(\beta(\gamma)) - \beta(\gamma) \to -\infty \quad \text{when } \beta(\gamma) \to \infty. \tag{4.4}$$

Now recall that $\rho(X)$ equals the supremum over $\gamma \in \mathcal{M}$ of the left-hand side of (4.3) and that $|\rho(X)| \le C$ for all X with $\|X\|_{L^q} \le a$ by the first part of this proof. Therefore, (4.4) implies that there is some constant b such that only $\gamma \in \mathcal{M}$ for which $\beta(\gamma) \le b$ need to be considered in the computation of $\rho(X)$. □

Lemma 4.4. *Let the assumptions of Theorem 4.1 be satisfied. For every fixed $b \in \mathbb{R}_+$, we have*

$$\Gamma_b([r,1)) := \sup_{\gamma\in\mathcal{M}\ \text{s.t.}\ \beta(\gamma)\le b} \gamma([r,1)) \le C(1-r)^{1/q}$$

for every $r \in [0,1)$.

Proof. Let X^* be the random variable of Lemma 4.1. Then, it follows from interchanging two suprema in the spectral representation (4.1) (one over n and one over γ), monotone convergence (applied under each γ), and Lemma 4.1 that

$$\begin{aligned}\sup_n \rho(X^*\wedge n) &= \sup_{\gamma\in\mathcal{M}}\sup_n\left(\int_{[0,1)}\mathrm{AVaR}_u(X^*\wedge n)\,\gamma(du)-\beta(\gamma)\right)\\ &\geq \sup_{\gamma\in\mathcal{M}\text{ s.t. }\beta(\gamma)\leq b}\left(\int_{[0,1)}\frac{q}{q-1}\frac{1}{(1-u)^{1/q}}\,\gamma(du)-\beta(\gamma)\right).\end{aligned}\tag{4.5}$$

By assumption $\sup_n\rho(X^*\wedge n)\in\mathbb{R}$, which implies that

$$\sup_{\gamma\in\mathcal{M}\text{ s.t. }\beta(\gamma)\leq b}\int_{[0,1)}\frac{1}{(1-u)^{1/q}}\,\gamma(du)\leq\frac{q-1}{q}\left(\sup_n\rho(X^*\wedge n)+b+1\right)=C.$$

In particular, this implies that

$$\Gamma_b([r,1))\leq\sup_{\gamma\in\mathcal{M}\text{ s.t. }\beta(\gamma)\leq b}\int_{[r,1)}\left(\frac{1-r}{1-u}\right)^{1/q}\gamma(du)\leq C(1-r)^{1/q},$$

which proves the claim. □

Lemma 4.5. *Let the assumptions of Theorem 4.1 be satisfied. Let $0\leq b<a<1$. Then, it holds that*

$$\sum_{n\geq1}2^{-an}\cdot((x2^n)\wedge 2^{bn})\leq C(x^{\frac{a-b}{1-b}}\vee x)$$

for every $x\in[0,\infty)$ (where C does not depend on x).

Proof. For $x=0$, there is nothing to prove. We now consider the case $x\in(0,1]$, denote by s_n the summand, and set

$$n_N:=\frac{\log(1/x)}{(1-b)\log 2}.$$

Then, a quick computation reveals

$$s_n=\begin{cases}x2^{n(1-a)} & \text{if } n<n_N,\\ 2^{n(b-a)} & \text{if } n\geq n_N.\end{cases}$$

By properties of the geometric series, one has

$$\sum_{n<n_N}s_n=Cx\sum_{n<n_N}2^{n(1-a)}\leq Cx\frac{1-2^{n_N(1-a)}}{1-2^{1-a}}\leq Cx2^{n_N(1-a)}.$$

Moreover, as $s^{\log t}=t^{\log s}$ for $s,t>0$, the definition of n_N implies that

$$\begin{aligned}2^{n_N(1-a)}&=\left(2^{\frac{1-a}{(1-b)\log 2}}\right)^{\log(1/x)}\\ &=\left(\frac{1}{x}\right)^{\log\left(2^{\frac{1-a}{(1-b)\log 2}}\right)}=x^{\frac{a-1}{1-b}}.\end{aligned}\tag{4.6}$$

Putting everything together, this implies

$$\sum_{n<n_N}s_n\leq Cx\cdot x^{\frac{a-1}{1-b}}=Cx^{\frac{a-b}{1-b}}.$$

For the tail of the sum, the same computation as in (4.6) shows that $2^{n_N(b-a)}=x^{\frac{b-a}{1-b}}$. Therefore, another application of the geometric series properties implies that

$$\begin{aligned}\sum_{n\geq n_N}s_n&=\sum_{n\geq n_N}2^{n(b-a)}\\ &\leq\frac{2^{n_N(b-a)}}{1-2^{b-a}}\leq C2^{n_N(b-a)}=Cx^{\frac{a-b}{1-b}}.\end{aligned}$$

Hence, adding the sums over $n < n_N$ and $n \geq n_N$ and noting that $(a-b)/(1-b) \in (0,1)$ and, hence, $x \leq x^{(a-b)/(1-b)}$ for $x \in [0,1]$ yields the claim for $x \in (0,1]$.

For $x \geq 1$, we have $x \geq x^{(a-b)/(1-b)}$ and

$$\sum_{n\geq 1} 2^{-an} \cdot ((x2^n) \wedge 2^{bn}) \leq x \sum_{n\geq 1} 2^{-an} \cdot (2^n \wedge 2^{bn}) \leq Cx,$$

where the last inequality follows from convergence of the geometric series/the previous step. □

For every $N \geq 1$ and $u \in [0,1)$, define

$$\delta_u^N := \sup_{g\in\mathcal{G}} |\mathrm{AVaR}_u^{\mu}(F + g\cdot G) - \mathrm{AVaR}_u^{\mu_N}(F+g\cdot G)|. \tag{4.7}$$

The following lemma controls uniformly the behavior of δ. Observe that measurability of $\mathrm{AVaR}_u^{\mu_N}(F+g\cdot G)$ is addressed in Remark 2.2. In particular, δ_u^N is measurable for every $u \in [0,1)$, and a quick argument shows that $\sup_{u\in[0,v]}\delta_u^N$ remains measurable for every $v \in [0,1)$.

Lemma 4.6. *Let the assumptions of Theorem 4.1 be satisfied. We have that*

$$E\left[\sup_{u\in[0,v]} \delta_u^N\right] \leq \frac{C}{(1-v)\sqrt{N}} \wedge \frac{C}{(1-v)^{1/2p}}$$

for every $v \in (0,1)$.

Proof. We start with the easier estimate, namely, that

$$E\left[\sup_{u\in[0,v]} \delta_u^N\right] \leq \frac{C}{(1-v)^{1/2p}}. \tag{4.8}$$

As $|F + g\cdot G| \leq CJ$ for every $g \in \mathcal{G}$, monotonicity of AVaR_u implies $\mathrm{AVaR}_u^{\mu}(F+g\cdot G) \leq \mathrm{AVaR}_u^{\mu}(CJ)$ for every $g \in \mathcal{G}$ and similarly with μ replaced by μ_N. Now, Lemma 4.2 implies

$$\sup_{u\in[0,v]} \delta_u^N \leq \frac{\|CJ\|_{L^{2p}(\mu)} + \|CJ\|_{L^{2p}(\mu_N)}}{(1-v)^{1/2p}}.$$

Further, Jensen's inequality implies $E[\|CJ\|_{L^{2p}(\mu_N)}] \leq \|CJ\|_{L^{2p}(\mu)}$, and thus, we get (4.8).

To conclude the proof, we are left to prove that

$$E\left[\sup_{u\in[0,v]} \delta_u^N\right] \leq \frac{C}{(1-v)\sqrt{N}}, \tag{4.9}$$

which we do in several steps.

a. Define

$$\mathcal{H} := \{\varphi(F+g\cdot G) : \varphi : \mathbb{R} \to \mathbb{R} \text{ is 1-Lipschitz}, \varphi(0) = 0 \text{ and } g \in \mathcal{G}\}.$$

Then, it holds that

$$\sup_{u\in[0,v]} \delta_u^N \leq \frac{1}{1-v} \sup_{H\in\mathcal{H}} \left| \int_{\mathcal{X}} H\,(\mu - \mu_N)(dx) \right|. \tag{4.10}$$

Indeed, every function appearing in the definition of AVaR_u is of the form $\varphi(F+g\cdot G)/(1-u)$ for a 1-Lipschitz function; see (1.1). Subtracting $\varphi(F(0)+g\cdot G(0))/(1-u)$ does not change the value of the difference of two integrals, which yields the claim.

b. We proceed to compute the covering numbers of $\mathcal{H}$. First, observe that because $\mathcal{G}$ is bounded, there is a constant C_0 such that $|F+g\cdot G| \leq C_0 J$ and $|(g-\tilde{g})\cdot G| \leq C_0 J$ for all $g, \tilde{g} \in \mathcal{G} \cup \{0\}$. The value of C_0 is kept fixed throughout this proof. Let $\varepsilon > 0$ and set

$$a_\varepsilon := \begin{cases} \dfrac{(6C_0)^{1/p}\|J\|_{L^{2p}(\mu_N)}}{\varepsilon^{1/p}}, & \text{if } p < \infty, \\ \|J\|_{L^\infty(\mu_N)}, & \text{if } p = \infty. \end{cases} \tag{4.11}$$

First, let $\tilde{L}_\varepsilon$ be a set of 1-Lipschitz functions from $\mathbb{R}$ to $\mathbb{R}$ that vanish at zero such that, for every 1-Lipschitz function φ, there is $\tilde{\varphi} \in \tilde{L}_\varepsilon$ satisfying $\sup_{t\in[-C_0a_\varepsilon, C_0a_\varepsilon]}|\varphi(t) - \tilde{\varphi}(t)| \le \varepsilon/3$. Such a set $\tilde{L}_\varepsilon$ can be constructed with

$$\mathrm{card}(\tilde{L}_\varepsilon) \le \exp\left(\frac{C}{(\varepsilon/a_\varepsilon)\wedge 1}\right); \tag{4.12}$$

we detail this in step (c). Moreover, let $\tilde{\mathcal{G}}_\varepsilon \subset \mathcal{G}$ be such that, for every $g \in \mathcal{G}$, there is $\tilde{g} \in \tilde{\mathcal{G}}_\varepsilon$ satisfying $|g - \tilde{g}| \le \varepsilon/(3C_0a_\varepsilon)$. Such a set $\tilde{\mathcal{G}}_\varepsilon$ can be constructed with

$$\mathrm{card}(\tilde{\mathcal{G}}_\varepsilon) \le \left(\frac{C}{(\varepsilon/a_\varepsilon)\wedge 1}\right)^e, \tag{4.13}$$

using an equidistant grid of the bounded set $\mathcal{G} \subset \mathbb{R}^e$.

Now, set

$$\tilde{\mathcal{H}}_\varepsilon := \{\tilde{\varphi}(F + g\cdot G) : \tilde{\varphi} \in \tilde{L}_\varepsilon, \tilde{g} \in \tilde{\mathcal{G}}_\varepsilon\}.$$

We claim that, for every $H = \varphi(F + g\cdot G) \in \mathcal{H}$, there is $\tilde{H} = \tilde{\varphi}(F + \tilde{g}\cdot G) \in \tilde{\mathcal{H}}$ such that $\|H - \tilde{H}\|_{L^2(\mu_N)} \le \varepsilon$. If this is true, then

$$\begin{aligned}\mathcal{N}(\mathcal{H}, \|\cdot\|_{L^2(\mu_N)}, \varepsilon) &\le \mathrm{card}(\tilde{\mathcal{H}}_\varepsilon) \le \mathrm{card}(\tilde{L}_\varepsilon)\mathrm{card}(\tilde{\mathcal{G}}_\varepsilon)\\ &\le \exp\left(\frac{C}{(\varepsilon^{(p+1)/p}/\|J\|_{L^{2p}(\mu_N)})\wedge 1}\right)\cdot\left(\frac{C}{(\varepsilon^{(p+1)/p}/\|J\|_{L^{2p}(\mu_N)})\wedge 1}\right)^e\end{aligned} \tag{4.14}$$

for every $\varepsilon > 0$, where the last inequality holds by (4.12), (4.13), and the choice of a_ε in (4.11).

To prove this claim, let $\tilde{\varphi} \in \tilde{L}_\varepsilon$ be such that $\sup_{t\in[-C_0a_\varepsilon, C_0a_\varepsilon]}|\varphi(t) - \tilde{\varphi}(t)| \le \varepsilon/3$, $g \in \tilde{\mathcal{G}}_\varepsilon$ such that $|g - \tilde{g}| \le \varepsilon/(3C_0a_\varepsilon)$ and write

$$\|H - \tilde{H}\|_{L^2(\mu_N)} \le \|1_{J\le a_\varepsilon}(H - \tilde{H})\|_{L^2(\mu_N)} + \|1_{J> a_\varepsilon}(H - \tilde{H})\|_{L^2(\mu_N)}. \tag{4.15}$$

To estimate the first term in the right-hand side of (4.15), recall that $\tilde{\varphi}$ is 1-Lipschitz, and hence,

$$\begin{aligned}\|1_{J\le a_\varepsilon}(H - \tilde{H})\|_{L^2(\mu_N)} &\le \|1_{J\le a_\varepsilon}(\varphi(F + g\cdot G) - \tilde{\varphi}(F + g\cdot G))\|_{L^2(\mu_N)}\\ &\quad + \|1_{J\le a_\varepsilon}(\tilde{\varphi}(F + g\cdot G) - \tilde{\varphi}(F + \tilde{g}\cdot G))\|_{L^2(\mu_N)}\\ &\le \sup_{t\in[-C_0a_\varepsilon, C_0a_\varepsilon]}|\varphi(t) - \tilde{\varphi}(t)| + C_0a_\varepsilon|g - \tilde{g}| \le \frac{2\varepsilon}{3}\end{aligned}$$

by choice of $\tilde{\varphi}$ and $\tilde{g}$. As for the second term in the right-hand side of (4.15), first note that it is zero in case $p = \infty$ because $1_{J>a_\varepsilon} = 0$ μ_N-almost surely by the choice of a_ε in (4.11). Otherwise, if $p < \infty$, recalling that $|H|, |\tilde{H}| \le C_0J$, Markov's inequality and the choice of a_ε imply that

$$\begin{aligned}\|1_{J>a_\varepsilon}(H - \tilde{H})\|_{L^2(\mu_N)} &\le 2C_0\|1_{J>a_\varepsilon}J\|_{L^2(\mu_N)}\\ &\le \frac{2C_0\|J^p\|_{L^2(\mu_N)}}{a_\varepsilon^p} \le \frac{\varepsilon}{3}.\end{aligned}$$

This proves our claim that $\|H - \tilde{H}\|_{L^2(\mu_N)} \le \varepsilon$.

c. It remains to argue that the set $\tilde{L}_\varepsilon$ in step (b) exists. To that end, denote by L the set of all 1-Lipschitz functions $\varphi : \mathbb{R} \to \mathbb{R}$ satisfying $\varphi(0) = 0$. Further, for $\varphi \in L$, denote by

$$\mathcal{R}(\varphi) : [-1, 1] \to \mathbb{R}, \quad t \longmapsto \frac{\varphi(C_0a_\varepsilon t)}{C_0a_\varepsilon}$$

its rescaled restriction. Then, $\mathcal{R}(L)$ consists of 1-Lipschitz functions that are bounded by one, and van der Vaart and Wellner [50, theorem 2.7.1] imply that there exists a set R'_ε with cardinality at most $\exp\left(\frac{C}{(\varepsilon/a_\varepsilon)\wedge 1}\right)$ such that, for every $\varphi \in \mathcal{R}(L)$, there is $\varphi' \in R'_\varepsilon$ with $\sup_{t\in[-1,1]}|\varphi(t) - \varphi'(t)| \le \varepsilon/(6C_0a_\varepsilon)$. By the triangle inequality, there is a set $\tilde{R}_\varepsilon \subset \mathcal{R}(L)$ of the same cardinality as R'_ε such that, for every $\varphi \in \mathcal{R}(L)$, there is $\tilde{\varphi} \in \tilde{R}_\varepsilon$ satisfying $\sup_{t\in[-1,1]}|\varphi(t) - \tilde{\varphi}(t)| \le \varepsilon/(3C_0a_\varepsilon)$. Now, extend every $\varphi \in \tilde{R}_\varepsilon$ to a function with domain $\mathbb{R}$ via

$$\mathcal{E}(\varphi) : \mathbb{R} \to \mathbb{R}, \quad t \longmapsto C_0a_\varepsilon\varphi\left((-1)\vee\left(\frac{t}{C_0a_\varepsilon}\wedge 1\right)\right)$$

and note that

$$\sup_{t\in[-C_0a_\varepsilon,C_0a_\varepsilon]} |\varphi(t)-\mathcal{E}(\tilde\varphi(t))| = C_0a_\varepsilon \sup_{t\in[-1,1]}|\mathcal{R}(\varphi)(t)-\tilde\varphi(t)|.$$

Hence, $\tilde L_\varepsilon := \mathcal{E}(\tilde R_\varepsilon)$ is the desired set.

d. The set $\mathcal{H}$ satisfies Assumption A.1. Indeed, first observe that the set of continuous functions from $\mathbb{R}$ to $\mathbb{R}$ endowed with the topology of uniform convergence on compacts[9] is separable; hence, the subset of 1-Lipschitz functions is separable as well w.r.t. this topology. The rest of the argument follows from similar arguments as presented in Lemma 3.3.

e. We use the empirical process version of Dudley's entropy integral theorem, that is, Theorem A.1. Note that $H^* := 0 \in \mathcal{H}$, and therefore, Theorem A.1 implies

$$\begin{aligned} E\left[\sup_{u\in[0,v]}\delta_u^N\right] &\le \frac{C}{\sqrt{N}}E\left[\int_0^\infty \sqrt{\log\mathcal{N}(\mathcal{H},\|\cdot\|_{L^2(\mu_N)},\varepsilon)}\,d\varepsilon\right]\\ &= \frac{C}{\sqrt{N}}E\left[\|J\|_{L^{2p}(\mu_N)}\int_0^\infty \sqrt{\log\left(\exp\left(\frac{C}{\tilde\varepsilon^{(p+1)/p}\wedge 1}\right)\left(\frac{C}{\tilde\varepsilon^{(p+1)/p}\wedge 1}\right)^e\right)}\,d\tilde\varepsilon\right], \end{aligned}$$

where the last line follows from using (4.14) and substituting ε by $\tilde\varepsilon = \varepsilon/\|CJ^p\|_{L^2(\mu_N)}^{1/p}$. It remains to notice that the (now deterministic) integral over $d\tilde\varepsilon$ is finite. Moreover, Jensen's inequality implies $E[\|J\|_{L^{2p}(\mu_N)}] \le \|J\|_{L^{2p}(\mu)}$, and the latter term is finite by assumption.

In conclusion, we show (4.9), and the proof is complete. □

Proof of Theorem 4.1. Recall the definition of $M := \|J\|_{L^p(\mu)}$ and $M_N := \|J\|_{L^p(\mu_N)}$ given in (3.1). As in the proof of Theorem 3.1, we set

$$\Delta_N := \sup_{g\in\mathcal{G}}|\rho^\mu(F+g\cdot G)-\rho^{\mu_N}(F+g\cdot G)|$$

and consider both terms in

$$E^*[\Delta_N] = E^*[\Delta_N 1_{M_N\le M+1}] + E^*[\Delta_N 1_{M_N>M+1}]$$

separately (note that linearity of the outer expectation holds here because $\{M_N \le M+1\}$ is a measurable set).

a. As $\mathcal{G}$ is bounded, we have $\|F+g\cdot G\|_{L^p(\mu)} \le CM$. Therefore, by Lemma 4.4, there exists some b such that

$$\rho^\mu(F+g\cdot G) = \sup_{\gamma\in\mathcal{M}\text{ s.t. }\beta(\gamma)\le b}\left(\int_{[0,1)}\mathrm{AVaR}_u^\mu(F+g\cdot G)\,\gamma(du)-\beta(\gamma)\right)$$

for all $g\in\mathcal{G}$. Possibly making b larger, the same reasoning implies that, on the set $M_N\le M+1$, the same representation holds true if μ is replaced by μ_N. Recalling the definition of δ^N in (4.7) and the definition of Γ_b given in Lemma 4.4, we can write

$$\begin{aligned} \Delta_N 1_{M_N\le M+1} &\le \sup_{\gamma\in\mathcal{M}\text{ s.t. }\beta(\gamma)\le b}\int_{[0,1)}\delta_u^N\,\gamma(du)\\ &\le \sum_{n\ge1}\Gamma_b(I_n)\sup_{u\in I_n}\delta_u^N, \end{aligned}$$

where $I_n := [1-2^{-n+1}, 1-2^{-n})$ for every n, that is, $I_1 = [0,1/2)$, $I_2 = [1/2,3/4)$, and so forth.

Now, estimate $\Gamma_b(I_n) \le C2^{-n/q}$ by means of Lemma 4.4 and $E[\sup_{u\in I_n}\delta_u^N] \le C(2^n\sqrt{N}^{-1})\wedge 2^{n/2p}$ by means of Lemma 4.6. Then, an application of Lemma 4.5 implies that

$$\begin{aligned} E^*[\Delta_N 1_{M_N\le M+1}] &\le C\sum_{n\ge1}2^{-n/q}\left(\frac{2^n}{\sqrt{N}}\wedge 2^{n/2p}\right)\\ &\le \frac{C}{\sqrt{N}^{\frac{1/q-1/2p}{1-1/2p}}}\vee\frac{C}{\sqrt{N}} \le \frac{C}{\sqrt{N}^{\frac{1/q-1/2p}{1-1/2p}}}, \end{aligned}$$

where the last inequality holds as $\frac{1/q-1/2p}{1-1/2p}\in(0,1)$.

b. The second term is controlled in a similar way as in the proof of Theorem 3.1; namely, we first estimate

$$E^*[\Delta_N 1_{M_N>M+1}] \leq E^*[\Delta_N^2]^{1/2} P[M_N > M+1]^{1/2}$$
$$\leq \frac{CE^*[\Delta_N^2]^{1/2}}{\sqrt{N}}.$$

It, therefore, remains to check that $E^*[\Delta_N^2] \leq C$. In fact, if $p = \infty$, then $M_N \leq M$ almost surely, and there is nothing left to prove. So assume that $p < \infty$. Using monotonicity of ρ and the fact that $\mathcal{G}$ is bounded, this boils down to checking that $E[\rho^{\mu_N}(CJ)^2] \leq C$. To that end, by definition of w_p and as $J \geq 1$, one has that

$$\rho^{\mu_N}(CJ) \leq w_p\big(C\|J\|_{L^p(\mu_N)}\big) \leq w_p\left(C\frac{1}{N}\sum_{n\leq N} J(S_n)^p\right).$$

By convexity of $x \longmapsto w_p(x)^2$, we may further estimate

$$E^*[\rho^{\mu_N}(CJ)^2] \leq \frac{1}{N}\sum_{n\leq N} E[w_p(CJ(S_n)^p)^2]$$
$$= \int_{\mathcal{X}} w_p(CJ(x)^p)^2\,\mu(dx),$$

and the last term is finite by assumption.

Combining both steps completes the proof. □

5. Deviation Inequalities

In the following, we prove (a generalization of) part (ii) of Theorem 2.1 and part (ii) of Theorem 2.2.

Theorem 5.1. *Assume that F and G are bounded functions and that the set $\mathcal{G}$ is bounded. Moreover, let $q \in (1, \infty)$ and assume that ρ is q-regular. Then, there are constants $c, C > 0$ such that*

$$P^*\left[\sup_{g\in\mathcal{G}} |\rho^{\mu}(F + g\cdot\mathcal{G}) - \rho^{\mu_N}(F + g\cdot\mathcal{G})| \geq \varepsilon\right] \leq C\exp(-cN\varepsilon^{2q})$$

for all $\varepsilon > 0$ and $N \geq 1$.

Proof.

a. In a first step, recall that F, G, and $\mathcal{G}$ are bounded; hence, there is a constant a such that $|F + g\cdot G| \leq a$ for all $g \in \mathcal{G}$. As the optimal m in the definition of the average value at risk is given by a respective quantile, it follows that

$$\mathrm{AVaR}^{\mu}_u(F + g\cdot G) = \inf_{|m|\leq a} \frac{1}{1-u}\int_{\mathcal{X}} (F + g\cdot G - m)_+ + (1-u)m\,\mu(dx) \tag{5.1}$$

for every $u \in [0, 1)$ and $g \in \mathcal{G}$, and (5.1) remains true if μ is replaced by μ_N. Further, as $\int_{\mathcal{X}}(1-u)m(\mu - \mu_N)(dx) = 0$ for all $m \in \mathbb{R}$ and $u \in [0, 1)$, this implies that

$$|\mathrm{AVaR}^{\mu}_u(F + g\cdot G) - \mathrm{AVaR}^{\mu_N}_u(F + g\cdot G)| \leq \frac{\delta_0^N}{1-u}, \tag{5.2}$$

where we set

$$\delta_0^N := \left|\sup_{H\in\mathcal{H}} \int_{\mathcal{X}} H(x)(\mu - \mu_N)(dx)\right| \quad \text{and}$$
$$\mathcal{H} := \{(F + g\cdot G - m)_+ : |m| \leq a \text{ and } g \in \mathcal{G}\}.$$

Note that, by Lemma 3.3, the set $\mathcal{H}$ satisfies Assumption A.1.

b. In a second step, notice that the same arguments (again, actually simpler as J is bounded) as in the proof of Theorem 2.1 imply that there is some $b > 0$ such that the supremum over $\gamma \in \mathcal{M}$ in the spectral representation (4.1) of ρ can be restricted to those γ for which $\beta(\gamma) \leq b$. This implies

$$|\rho^{\mu}(F + g\cdot F) - \rho^{\mu_N}(F + g\cdot G)|$$
$$\leq \sup_{\gamma\in\mathcal{M} \text{ s.t. } \beta(\gamma)\leq b} \int_{[0,1)} |\mathrm{AVaR}^{\mu}_u(F + g\cdot G) - \mathrm{AVaR}^{\mu_N}_u(F + g\cdot G)|\,\gamma(du)$$
$$\leq \sum_{n\geq 1} \Gamma_b(I_n)\left(\sup_{u\in I_n} \frac{\delta_0^N}{1-u} \wedge C_1\right),$$

where $I_n := [1 - 2^{-n+1}, 1 - 2^{-n})$ for every n and the constant C_1 appears because F, G, and $\mathcal{G}$ are bounded. Without loss of generality, we may assume that $C_1 \geq 1$. Then, estimating $\Gamma_b(I_n) \leq C_2 2^{-n/q}$ by Lemma 4.4, we obtain

$$\sup_{g \in \mathcal{G}} |\rho^{\mu}(F + g \cdot F) - \rho^{\mu_N}(F + g \cdot G)| \leq C_1 C_2 \sum_{n \geq 1} 2^{-n/q}((2^n \delta_0^N) \wedge 1) \leq CC_1C_2((\delta_0^N)^{1/q} \vee \delta_0^N),$$

for all $N \geq 1$ almost surely, and the last inequality follows from Lemma 4.5. Finally, as $\delta_0^N \leq C_3$ almost surely, we conclude that

$$\sup_{g \in \mathcal{G}} |\rho^{\mu}(F + g \cdot F) - \rho^{\mu_N}(F + g \cdot G)| \leq C(\delta_0^N)^{1/q}. \tag{5.3}$$

c. In a last step, it remains to estimate δ_0^N. By Corollary 3.1, one has that

$$\mathcal{N}(\mathcal{H}, \|\cdot\|_\infty, \varepsilon) \leq \left(\frac{C}{\varepsilon}\right)^{e+1} \vee 1$$

for all $\varepsilon > 0$. Hence, because $\mathcal{H}$ satisfies Assumption A.1, Theorem A.2 implies that

$$P[\delta_0^N \geq \varepsilon] \leq C \exp\left(-\frac{N\varepsilon^2}{C}\right)$$

for all $\varepsilon > 0$ and $N \geq 1$. The proof is completed by plugging the last estimate into Equation (5.3). □

Theorem 5.2. *Assume that F and G are bounded functions, that the set $\mathcal{G}$ is bounded, and let $\rho = \mathrm{OCE}$ be the optimized certainty equivalent risk measure. Then, there are constants $c, C > 0$ such that*

$$P\left[\sup_{g \in \mathcal{G}} |\rho^{\mu}(F + g \cdot \mathcal{G}) - \rho^{\mu_N}(F + g \cdot \mathcal{G})| \geq \varepsilon\right] \leq C \exp(-cN\varepsilon^2)$$

for all $\varepsilon > 0$ and $N \geq 1$.

Proof. The proof is similar to the one given for Theorem 5.1, and we keep it short. By Lemma 3.1, one has

$$|\rho^{\mu}(F + g \cdot G) - \rho^{\mu_N}(F + g \cdot G)| \leq \sup_{H \in \mathcal{H}} \left| \int_{\mathcal{X}} H(x)\,(\mu - \mu_N)(dx) \right| =: \delta_0^N$$

almost surely for the set

$$\mathcal{H} := \{l(F + g \cdot G - m) + m : g \in \mathcal{G} \text{ and } |m| \leq a\}$$

with a such that $|F + g \cdot G| \leq a$ for all $g \in \mathcal{G}$. By Lemma 3.3, the set $\mathcal{H}$ satisfies Assumption A.1. Thus, an application of Theorem A.2 again implies that $P[\delta_0^N \geq \varepsilon] \leq C \exp(-cN\varepsilon^2)$ for some constants $c, C > 0$. This concludes the proof. □

6. Sharpness of Rates

Whenever investigating average errors involving a (linear) dependence on i.i.d. phenomena, the central limit theorem assures that the $1/\sqrt{N}$ rate cannot be improved. Indeed, take, for instance, $\rho(X) := E[X] = \mathrm{AVaR}_0(X)$. Then, if μ is a probability on $[0,1]$ and F is a (bounded) function that is equal to the identity on $[0,1]$, one has that

$$\rho^{\mu_N}(F) = \frac{1}{N}\sum_{n \leq N} F(S_n) \text{ approximately has the distribution } \mathcal{N}\left(\rho^{\mu}(F), \frac{\mathrm{Var}(F(S))}{N}\right)$$

for large N by the central limit theorem, in which $\mathcal{N}$ denotes the normal distribution and $\mathrm{Var}(F(S))$ is the variance of $F(S)$. In particular, $E[|\rho^{\mu}(F) - \rho^{\mu_N}(F)|]$ asymptotically behaves like $\sqrt{\mathrm{Var}(F(S))/N}$, and $P[|\rho^{\mu}(F) - \rho^{\mu_N}(F)| \geq \varepsilon]$ asymptotically behaves like $2\Phi(-\varepsilon^2 N/\mathrm{Var}(F(S)))$, where Φ is the cumulative distribution function of the standard normal distribution. We refer to Belomestny and Krätschmer [6], Beutner and Zähle [10], and Chen [11] for central limit theorems for risk measures.

In comparison with the $1/\sqrt{N}$ rate, the rates obtained for general risk measures, for example, in Theorem 2.1 are worse. As the proofs are presented, they depend on the notion of regularity of the risk measure given in Definition 2.1, and this section is devoted to showing the necessity of regularity; we prove Proposition 2.1. To that end, to ease the notation, for probabilities μ on $\mathbb{R}$ with bounded support, we write

$$\rho(\mu) := \rho(X) \quad \text{where} \quad X \sim \mu.$$

Remark 6.1. Without the assumption that ρ is q-regular, the proof of Proposition 2.1 becomes rather trivial: take $\rho(X) := \text{ess.sup} X$ and let μ be some probability with support $[0,1]$. As $\rho(\mu_N) = \max_{n\le N} X_n$ (where (X_n) is an i.i.d. sample of μ), one has

$$P[|\rho(\mu) - \rho(\mu_N)| \ge \varepsilon] = P\left[\max_{n\le N} X_n \le 1-\varepsilon\right] = \mu([0,1-\varepsilon])^N.$$

For suitable choices of μ, the latter term can converge arbitrarily slowly to zero. Therefore, $E[|\rho(\mu) - \rho(\mu_N)|] = \int_0^1 \mu([0,1-\varepsilon])^N \, d\varepsilon$ converges arbitrarily slowly as well.

The proof of Proposition 2.1 mimics the idea of Remark 6.1, simultaneously enforcing regularity of ρ. To ease notation, denote by

$$\text{Ber}(p) := (1-p)\delta_0 + p\delta_1 \tag{6.1}$$

the Bernoulli distribution with parameter of success $p \in [0,1]$. Then, for $\mu = \text{Ber}(p)$, the empirical measure μ_N of μ satisfies

$$\mu_N \equiv \text{Ber}(p)_N = \text{Ber}(\hat{p}_N) \quad \text{where} \quad \hat{p}_N := \frac{1}{N}\sum_{n\le N} X_n \tag{6.2}$$

(almost surely), where (X_n) are i.i.d. $\text{Ber}(p)$ distributed. This simple formula is actually the reason why we stick to the Bernoulli distribution as computations become a lot easier.

We start with two simple lemmas and leave their simple proofs to the reader.

Lemma 6.1. *Let $p \in (0,1)$. Then,*

$$\text{AVaR}_u(\text{Ber}(p)) = \frac{p}{1-u} \wedge 1$$

for every $u \in [0,1)$.

Lemma 6.2. *It holds that*

$$\sup_{x\ge 1}((1-x^{-\delta})a + x^{-\delta}((ax)\wedge 1)) = (1-a^{\delta})a + a^{\delta}$$

for every $a \in [0,1]$ and $\delta > 0$.

Proof of Proposition 2.1. For shorthand notation, set $\delta := 1/q$. Define $\rho : L^\infty \to \mathbb{R}$ by

$$\rho(X) := \sup_{x\ge 1}((1-x^{\delta})\text{AVaR}_0(X) + x^{-\delta}\text{AVaR}_{1-1/x}(X)). \tag{6.3}$$

As AVaR is a law-invariant coherent risk measure, ρ inherits all those properties.

To check that ρ is $(q+\varepsilon)$-regular for every $\varepsilon > 0$, fix such ε and denote by X^* a random variable with the Pareto distribution with scale parameter one and shape parameter $q+\varepsilon$. Then, X^* has finite qth moment, and the definition of ρ together with Lemma 4.2 imply that

$$\rho(X^* \wedge n) \le \|X^*\|_{L^q} \sup_{x\ge 1}((1-x^{-\delta})1 + x^{-\delta}x^{1/q}) < \infty$$

for all $n \in \mathbb{N}$. As the right-hand side does not depend on n, this shows that ρ is $(q+\varepsilon)$-regular.

Now, let $p_N := 1/N$ and let (X_n^N) be an i.i.d. sample of $\text{Ber}(p_N)$, that is, $P[X_n^N = 1] = p_N = 1/N$ for all n and N. Further, recall that the empirical measure of $\text{Ber}(p_N)$ is $\text{Ber}(\hat{p}_N)$, where $\hat{p}_N := \frac{1}{N}\sum_{n\le N} X_n^N$. We show that

$$\rho(\text{Ber}(p_N)) - E[\rho(\text{Ber}(\hat{p}_N))] \ge \frac{p_N^{\delta}}{C}$$

for all N. Using the triangle inequality, this clearly implies the statement of the proposition.

By Lemmas 6.1 and 6.2, we compute

$$\begin{aligned}\rho(\text{Ber}(p_N)) &= \sup_{x\ge 1}((1-x^{-\delta})p_N + x^{-\delta}((xp_N)\wedge 1)) \\ &= (1-p_N^{\delta})p_N + p_N^{\delta},\end{aligned}$$

and similarly,

$$\rho(\text{Ber}(\hat{p}_N)) = (1-\hat{p}_N^{\delta})\hat{p}_N + \hat{p}_N^{\delta}.$$

Now, recall that $E[\hat{p}_N] = p_N$, and by Jensen's inequality, $E[\hat{p}_N^\delta \hat{p}_N] \geq p_N^\delta p_N$; hence,

$$\rho(\text{Ber}(p_N)) - E[\rho(\text{Ber}(\hat{p}_N))] \geq p_N^\delta - E[\hat{p}_N^\delta].$$

For the set

$$A_N := \{\hat{p}_N = 0\} = \{X_n^N = 0 \quad \text{for all} \quad n \leq N\},$$

one computes

$$P[A_N] = (1 - p_N)^N = \exp\left(N \log\left(1 - \frac{1}{N}\right)\right) \geq \exp(-2)$$

for $N \geq 2$. Moreover $E[\hat{p}_N^\delta] = E[\hat{p}_N^\delta 1_{A_N^c}]$ and an application of Hölder's inequality (with exponents $p = 1/\delta$ and $q = 1/(p-1) = 1/(1-\delta)$) gives

$$\begin{aligned} E[\hat{p}_N^\delta] &\leq E[\hat{p}_N]^\delta P[A_N^c]^{1-\delta} \\ &\leq p_N^\delta (1 - \exp(-2))^{1-\delta} =: p_N^\delta c \end{aligned}$$

for all $N \geq 2$. Here, we also use that $E[\hat{p}_N] = p_N$ and the previous computation for (the limit of) $P[A_N]$.

In particular,

$$\rho(\text{Ber}(p_N)) - E[\rho(\text{Ber}(\hat{p}_N))] \geq p_N^\delta (1 - c)$$

for all $N \geq 2$. As $c \in (0, 1)$, this completes the proof (considering the case $n = 1$ separately). □

Remark 6.2. In the theory of risk measures, two continuity properties are often considered: the Fatou property and the stronger Lebesgue property. We refer the unfamiliar reader to Föllmer and Schied [19, section 4.2]. A result of Jouini et al. [31] ensures that every law-invariant risk measure automatically satisfies the Fatou property, and it is easy to see that a q-regular law-invariant risk measure satisfies the Lebesgue property.

Small modifications in the proof of Proposition 2.1 actually give the existence of a law-invariant risk measure that satisfies the Lebesgue property but for which no polynomial convergence rate holds true.

7. Additional Proofs

7.1. Remaining Proofs for Theorem 2.1

We finally provide the proof of Theorem 2.1 for the case that ρ is the shortfall risk measure.

a. Define the function $J : \mathbb{R} \to \mathbb{R}$ by

$$J(m) := \inf_{g \in \mathcal{G}} \int l(F + g \cdot G - m)\, \mu(dx),$$

and in the same way, define the (random) function J_N with μ replaced by μ_N. Further, let $a \geq 0$ such that $|F + g \cdot G| \leq a$ for every $g \in \mathbb{R}$. Then, $|\pi^\mu(F)| \leq a$, or in other words,

$$\pi^\mu(F) = \inf\{m \in [a, a] : J(m) \leq 0\}.$$

The same is true if μ is replaced by μ_N and J by J_N (almost surely).

b. We claim that there is $c > 0$ such that $J(\tilde{m}) \leq J(m) - c(\tilde{m} - m)$ for all $m, \tilde{m} \in [-a, a]$ with $m \leq \tilde{m}$. Indeed, as l is convex and strictly increasing, its (right) derivative l' is strictly positive. Now, let $g \in \mathcal{G}$ be optimal for $J(m)$ (for notational simplicity, otherwise, use some ε-optimal g), that is, $J(m) = \int l(F + g \cdot G - m)\, d\mu$. The fundamental theorem of calculus then implies

$$\begin{aligned} J(\tilde{m}) &\leq \int l(F + g \cdot G - \tilde{m})\, d\mu \\ &= \int l(F + g \cdot G - m) - (\tilde{m} - m) \int_0^1 l'(F + g \cdot G - m + t(\tilde{m} - m))\, dt\, d\mu. \end{aligned}$$

The term inside the second integral is larger than $c := \inf_{|t| \leq 2a} l'(t) > 0$. So $J(\tilde{m}) \leq J(m) - c(\tilde{m} - m)$, which is what we claim.

c. We claim that J and J_N are continuous. Indeed, this is an easy consequence of the continuity of $(m, g) \mapsto \int l(F + g \cdot G - m)\, d\mu$ together with the fact that $\mathcal{G}$ is relativity compact (similarly for J_N); we spare the details.

d. Step (b), in particular, implies that J is strictly increasing. Combining this with the continuity of J yields that $\pi^\mu(F)$ is the unique number satisfying $J(\pi^\mu(F)) = 0$. Similarly, $\pi^{\mu_N}(F)$ is the unique number satisfying $J_N(\pi^{\mu_N}(F)) = 0$, and therefore,

$$\begin{aligned}|J(\pi^{\mu_N}(F)) - J(\pi^\mu(F))| &= |J(\pi^{\mu_N}(F)) - J_N(\pi^{\mu_N}(F))| \\ &\le \sup_{|m|\le a} |J(m) - J_N(m)| =: \Delta_N.\end{aligned}$$

Making use of step (a), this implies $|\pi^{\mu_N}(F) - \pi^\mu(F)| \le c\Delta_N$, and so it remains to gain control over Δ_N. As

$$\Delta_N \le \sup_{H\in\mathcal{H}} \left| \int H\, d(\mu - \mu_N) \right| \quad \text{for}$$
$$\mathcal{H} := \{l(F + g\cdot G - m) : |m| \le a \text{ and } g \in \mathcal{G}\},$$

we can use Lemma 3.3, Corollary 3.1, and Dudley's theorem as in the proof of Theorem 2.2 to obtain $E[\Delta_N] \le C/\sqrt{N}$ for all $N \ge 1$. Similarly, Corollary 3.1 and the arguments given for the proof of Theorem 5.1 imply that $P[\Delta_N \ge \varepsilon] \le C\exp(-cN\varepsilon^2)$ for all $\varepsilon > 0, N \ge 1$, where $c > 0$ is some (new) small constant. This completes the proof.

7.2. The Proof of Proposition 2.2

We only sketch the proof of Proposition 2.2 as it is very similar to that of Theorem 2.1 on the optimized certainty equivalents. The only difference is the absence of the component m (in the definition of OCE), which actually makes the proof even simpler. In particular, we have

$$\mathcal{N}(\{U(F + g\cdot G) : g \in \mathcal{G}\}, \|\cdot\|_\infty, \varepsilon) \le \left(\frac{C}{\varepsilon}\right)^e \vee 1$$

for all $\varepsilon > 0$ by Corollary 3.1. To conclude the proof, copy the arguments given for the proofs of Theorems 3.1 and 5.2.

Acknowledgments

The authors thank Patrick Cheridito as well as the associate editor and two referees for extraordinarily helpful comments.

Appendix. Supplementary Results

In this appendix, we provide an additional result pertaining to the boundedness assumption on $\mathcal{G}$. Recall that $\mathcal{G}$ is said to be bounded as a subset of $\mathbb{R}^e$ equipped with the Euclidean norm. Further, we state two results from empirical process theory that we use in this article and comment on the measurability of the plug-in estimator.

A.1. The Set $\mathcal{G}$ Needs to be Bounded

Our setup also includes the case of risk-based hedging, in which case one rather writes

$$\pi^\mu(F) = \inf\{m \in \mathbb{R} : \text{ there is some } g \in \mathcal{G} \text{ such that } \rho^\mu(F - m + g\cdot G) \le 0\}.$$

(This expression follows from additivity on the constants of ρ^μ).

In prose, $\pi^\mu(F)$ is the minimal capital m needed such that, possibly after trading, the loss F reduced by m becomes acceptable. In this setting, one typically does not restrict to bounded strategies, that is, one takes $\mathcal{G} = \mathbb{R}^e$.

The goal of this section is to prove the next proposition, which states that requiring $\mathcal{G}$ to be bounded is not just a technical simplification we made but, in fact, necessary.

One precaution needs to be made though: assume, for instance, that $G_i = 0$ for all i; then, clearly $g \longmapsto \rho^\mu(F + g\cdot G)$ does not depend on g, and the size of $\mathcal{G}$ does not matter. To exclude such cases (without too much effort), we assume that (μ, G) is nondegenerate in the sense that, for every $g \in \mathbb{R}^e \setminus \{0\}$, one has $\mu(g\cdot G < 0) > 0$.

Proposition A.1. *Let $\rho : L^\infty \to \mathbb{R}$ be any law-invariant risk measure, let F and each G_i be bounded, and let (μ, G) be nondegenerate in the preceding sense. Assume that $\pi^\mu(F) \in \mathbb{R}$ and*

$$E[|\pi^\mu(F) - \pi^{\mu_N}(F)|] \to 0$$

as $N \to \infty$. Then, the set $\mathcal{G}$ needs to be bounded.

Proof. We show the negation, namely, that, if $\mathcal{G}$ is unbounded, convergence cannot be true. To that end, let (g^n) be a sequence in $\mathcal{G}$ witnessing that $\mathcal{G}$ is unbounded. After passing to a subsequence, there exists $g^* \in \mathbb{R}^e$ with $|g^*| = 1$ such that $g^n/|g^n| \to g^*$. By assumption, $\mu(g^*\cdot G < 0) > 0$, and hence, there is $\varepsilon > 0$ such that

$$\mu(U) > 0 \quad \text{where} \quad U := \{x \in \mathcal{X} : g^*\cdot G(x) < -\varepsilon\}.$$

By definition of π, one has

$$\pi^{\mu_N}(F) \le \rho^{\mu_N}(F + g^n \cdot G)$$

for every $n \in \mathbb{N}$. Moreover, it holds that

$$F + g^n \cdot G \le \sup_U F + \sup_U g^n \cdot G =: a_n \quad \mu_N \text{ a.s. on } \{\mu_N(U) = 1\}$$

for every $n \in \mathbb{N}$. By assumption, the first term in the definition of a_n is bounded. Further, as $g^n/|g^n|$ converges to g^*, one has that

$$\begin{aligned} g^n \cdot G &= |g^n|\left(g^* \cdot G + \left(\frac{g^n}{|g^n|} - g^*\right) \cdot G\right) \\ &\le |g^n|\left(-\varepsilon + C\left|\frac{g^n}{|g^n|} - g^*\right|\right) < -\frac{|g^n|\varepsilon}{2} \end{aligned}$$

on U for all large n. By monotonicity of ρ^{μ_N}, this implies

$$\rho^{\mu_N}(F + g^n \cdot G) \le \rho^{\mu_N}(a_n) = a_n \to -\infty \quad \text{on } \{\mu_N(U) = 1\}$$

as $n \to \infty$. Finally, as

$$P[\mu_N(U) = 1] = 1 - (1 - \mu(U))^N > 0$$

for every $N \ge 1$, we conclude that $\pi^{\mu_N}(F) = -\infty$ with positive probability. In particular, $E[|\pi^{\mu}(F) - \pi^{\mu_N}(F)|] = \infty$ for every $N \ge 1$, which proves the claim. □

Remark A.1. Let us argue that our standing assumption that $\mathcal{G}$ is a countable set (which is there to circumvent issues regarding measurability) can be made without loss of generality. If $\mathcal{G}$ is not necessarily countable, we take a subset $\mathcal{G}' \subset \mathcal{G}$ that is countable and dense. If ρ^{ν} is a risk measure that is finite on $L^p(\nu)$ for $p \in [1, \infty]$, then it is automatically continuous w.r.t. $\|\cdot\|_{L^p(\nu)}$. In particular, if F and $|G|$ are in $L^p(\nu)$, then for every $g \in \mathcal{G}$ and $(g^n)_n \subset \mathcal{G}'$ that converges to g, we have that $F + g^n \cdot G \to F + g \cdot G$ in $L^p(\nu)$. As a consequence, under these assumptions,

$$\pi^{\nu}(F) = \inf_{g \in \mathcal{G}} \rho^{\nu}(F + g \cdot G) = \inf_{g \in \mathcal{G}'} \rho^{\nu}(F + g \cdot G).$$

This shows that, in all considerations made in this paper, the set $\mathcal{G}$ can be replaced by the set $\mathcal{G}'$.

A.2. Two Inequalities from Empirical Process Theory

Recall that $S, (S_n)_{n \ge 1}$ are i.i.d. random variables taking their values in a Polish space $\mathcal{X}$ distributed according to μ and that $\mu_N := \frac{1}{N}\sum_{n \le N} \delta_{S_n}$ is the associated empirical measure. Moreover, let $\mathcal{F}$ be a set of measurable, μ-square integrable functions from $\mathcal{X}$ to $\mathbb{R}$, and recall that $\mathcal{N}(\mathcal{F}, \|\cdot\|_{L^p(\mu)}, \varepsilon)$ is the covering number of $\mathcal{F}$ w.r.t. the L^p-norm at scale ε. Finally, we present results only under the following assumption regarding measurability.

Assumption A.1. *There is a countable set $\mathcal{F}' \subset \mathcal{F}$ such that, for every $f \in \mathcal{F}$, there is a sequence $(f_n)_n$ in $\mathcal{F}'$ such that $f_n \to f$ pointwise and in $L^2(\mu)$.*

Note that Assumption A.1 implies that

$$\sup_{f \in \mathcal{F}} \left|\frac{1}{N}\sum_{n \le N} f(S_n) - E[f(S)]\right| = \sup_{f \in \mathcal{F}'} \left|\frac{1}{N}\sum_{n \le N} f(S_n) - E[f(S)]\right|;$$

in particular, the term on the right-hand side is measurable. Moreover, a set $\mathcal{F}$ that satisfies Assumption A.1 is pointwise measurable in the sense of van der Vaart and Wellner [50, section 2.3.3].

We first state Dudley's entropy integral theorem in the form needed here.

Theorem A.1. *Suppose that Assumption A.1 is satisfied and let $f^* \in \mathcal{F}$ be an arbitrary but fixed function in $\mathcal{F}$. Then, we have that*

$$\begin{aligned} &E\left[\sup_{f \in \mathcal{F}} \left|\frac{1}{N}\sum_{n \le N} f(S_n) - E[f(S)]\right|\right] \\ &\le \frac{C}{\sqrt{N}}\left(E[f^*(S)^2]^{\frac{1}{2}} + E\left[\int_0^{\infty} \sqrt{\log \mathcal{N}(\mathcal{F}, \|\cdot\|_{L^2(\mu_N)}, \varepsilon)}\, d\varepsilon\right]\right. \end{aligned}$$

for all $N \ge 1$, where C is an absolute constant.

Proof. For completeness, we provide the proof of this standard fact. Note that, in all of the following arguments, the set $\mathcal{F}$ can be replaced without loss of generality by $\mathcal{F}'$, ensuring that no measurability issues occur. By the symmetrization lemma (see van der Vaart and Wellner [50, lemma 2.3.1]), we have that

$$E\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{N}\sum_{n\le N}f(S_n)-E[f(S)]\right|\right]\le 2E\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{N}\sum_{n\le N}\varepsilon_n f(S_n)\right|\right],$$

where the $(\varepsilon_n)_n$ are i.i.d. random signs (i.e., $P[\varepsilon_n=\pm 1]=1/2$), which are stochastically independent of $(S_n)_n$. Now, Hoeffding's inequality (van der Vaart and Wellner [50, lemma 2.2.7]) implies that, conditionally on $(S_n)_n$, the process $(X_f)_{f\in\mathcal{F}}$, where $X_f:=\frac{1}{\sqrt{N}}\sum_{n=1}^N\varepsilon_n f(S_n)$ is sub-Gaussian w.r.t. to the $L^2(\mu_N)$ norm. Hence, Dudley's entropy integral theorem (van der Vaart and Wellner [50, corollary 2.2.8]) for sub-Gaussian processes applied conditionally on $(S_n)_n$ gives

$$E\left[\sup_{f\in\mathcal{F}}|X_f|\,\Big|(S_n)_n\right]\le E\left[|X_{f^*}|\,\Big|(S_n)_n\right]+C\int_0^\infty\sqrt{\log\mathcal{N}(\mathcal{F},\|\cdot\|_{L^2(\mu_N)},\varepsilon)}\,d\varepsilon. \tag{A.1}$$

Finally, as the $\varepsilon_n f^*(S_n)$ s are i.i.d. zero mean random variables, Hölder's inequality assures that

$$E[|X_{f^*}|]\le E\left[\left(\frac{1}{\sqrt{N}}\sum_{n\le N}\varepsilon_n f^*(S_n)\right)^2\right]^{\frac12}=E[f^*(S)^2]^{\frac12}.$$

Thus, the statement of the theorem follows by integrating (A.1). □

Theorem A.2. *Suppose that Assumption A.1 is satisfied, that there is M such that* $|f|\le M$ *for all* $f\in\mathcal{F}$*, and that*

$$\mathcal{N}(\mathcal{F},\|\cdot\|_\infty,\varepsilon)\le\left(\frac{a}{\varepsilon}\right)^b\vee 1 \tag{A.2}$$

for all $\varepsilon>0$*, where* $a,b>0$ *are two constants. Then, there is a constant C depending only on a, b, M such that*

$$P\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{N}\sum_{n\le N}f(S_n)-E[f(S)]\right|\ge\varepsilon\right]\le C\exp\left(-\frac{N\varepsilon^2}{C}\right)$$

for all $\varepsilon>0$ *and* $N\ge 1$.

Proof. The goal is to apply van der Vaart and Wellner [50, theorem 2.14.10]. To that end, we may assume without loss of generality that $0\le f\le 1$ for every $f\in\mathcal{F}$. In a first step, note that (A.2) implies that

$$\sup_Q\log\mathcal{N}(\mathcal{F},\|\cdot\|_{L^2(Q)},\varepsilon)\le\frac{K}{\varepsilon}$$

for all $\varepsilon>0$, where K is a constant depending only on a and b, and the supremum is taken over all probability measures. This is exactly van der Vaart and Wellner [50, equation (2.14.8)], which is the assumption needed to apply van der Vaart and Wellner [50, theorem 2.14.10]. Hence, with the notation of that theorem, we have $U=5/3<2$ so that, for $\delta=1/6$, there are constants α and β depending only on K such that

$$P\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{N}\sum_{n\le N}f(S_n)-E[f(S)]\right|\ge\varepsilon\right]\le\alpha\exp(\beta(\sqrt{N}\varepsilon)^{U+\delta})\exp(-2(\sqrt{N}\varepsilon)^2)$$

for every $\varepsilon>0$ and $N\ge 1$. Recalling that $U+\delta=11/12<2$, a quick computation shows that

$$\alpha\exp(\beta(\sqrt{N}\varepsilon)^{U+\delta})\exp(-2(\sqrt{N}\varepsilon)^2)\le C\exp\left(-\frac{N\varepsilon^2}{C}\right),$$

for some constant C depending on α and β. This completes the proof. □

Endnotes

[1] See, for instance, McNeil et. al. [42, section 2.2] for ample discussion.

[2] Observe that, in contrast to the original definition (Artzner et al. [2], Föllmer and Schied [19]), we take risk measures to be increasing. This means that X models the (discounted) loss, and $\rho(X)$ is the capital to be added to a position with loss X to make it acceptable; see, for example, McNeil et. al. [42, chapter 6] for a similar framework and more details. This is done for notational convenience and does not affect generality. In fact, putting $\tilde\rho(X):=\rho(-X)$, the functional $\tilde\rho$ is a risk measure in the sense of Artzner et al. [2] and Föllmer and Schied [19].

[3] Actually "law-determined" might be a more sensible term to describe this property, but in accordance with the literature, we use the term "law-invariant" because it is predominantly used.

[4] The issue of numerical simulation of the estimator $\rho^{\mu_N}(F)$ is considered in Chu and Tangpi [12].

[5] That is, there is a constant $C_{\mathcal{G}}>0$ such that $\sup_{x\in\mathcal{G}}|x|\le C_{\mathcal{G}}$, where $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^e$.

[6] For instance, this is always true for the law-invariant comonotonic risk measure; see Acciaio and Svindland [1, corollary 10]

[7] Observe that $\|f\|_\infty:=\sup_{x\in\mathbb{R}}|f(x)|$ represents the supremum norm of f and not the essential supremum norm that we denote $\|\cdot\|_{L^\infty}$.

[8] It also goes under the name "Kusuoka representation" as the L^∞-version was discovered by Kusuoka [40].

[9] That is, w.r.t. to the topology induced by the metric $d(\varphi,\overline{\varphi}) := \sum_{k\geq 1}(1 \wedge \sup_{t\in[-k,k]}|\varphi(t) - \overline{\varphi}(t)|)\cdot 2^{-k}$.

References

[1] Acciaio B, Svindland G (2013) Are law-invariant risk functions concave on distributions? *Dependence Model.* 1(3):54–64.
[2] Artzner P, Delbaen F, Eber JM, Heath D (1999) Coherent measures of risk. *Math. Finance* 9(3):203–228.
[3] Bartl D, Mendelson S (2022) On Monte-Carlo methods in convex stochastic optimization. *Ann. Appl. Probab.* 32(4):3146–3198.
[4] Bartl D, Drapeau S, Tangpi L (2020) Computational aspects of robust optimized certainty equivalents and option pricing. *Math. Finance* 30(1):287–309.
[5] Basel Committee on Banking Supervision (2013) Fundamental review of the trading book: A revised market risk framework. Technical report, Bank for International Settlements, Basel, Switzerland.
[6] Belomestny D, Krätschmer V (2012) Central limit theorems for law-invariant coherent risk measures. *J. Appl. Probab.* 49(1):1–21.
[7] Ben-Tal A, Taboulle M (1986) Expected utility, penalty functions and duality in stochastic nonlinear programming. *Management Sci.* 32(11):1445–1466.
[8] Ben-Tal A, Taboulle M (2007) An old-new concept of convex risk measures: The optimized certainty equivalent. *Math. Finance* 17(3):449–476.
[9] Bertsimas D, Gupta V, Kallus N (2017) Robust sample average approximation. *Math. Programming* 171:217–282.
[10] Beutner E, Zähle H (2010) A modified functional delta method and its application to the estimation of risk functionals. *J. Multivariate Anal.* 101(10):2452–2463.
[11] Chen SX (2008) Nonparametric estimation of expected shortfall. *J. Financial Econom.* 6(1):87–107.
[12] Chu J, Tangpi L (2021) Non-asymptotic estimation of risk measures using stochastic gradient Langevin dynamics. Preprint, submitted November 24, https://doi.org/10.48550/arXiv.2111.12248.
[13] Claus M, Krätschmer V, Schultz R (2017) Weak continuity of risk functionals with applications to stochastic programming. *SIAM J. Optim.* 27(1):91–109.
[14] Cont R, Deguest R, Scandolo G (2010) Robustness and sensitivity analysis of risk measurement procedures. *Quant. Finance* 10(6):593–606.
[15] Eckstein S, Kupper M, Pohl M (2020) Robust risk aggregation with neural networks. *Math. Finance* 30(4):1229–1272.
[16] Embrechts P, Hofert M (2014) Statistics and quantitative risk management for banking and insurance. *Annual Rev. Statist. Appl.* 1:493–514.
[17] Esfahani PM, Kuhn D (2018) Data-driven distributionally robust optimization using the Wassertein metric: Performance guarantees and tractable reformulations. *Math. Programming* 171:115–166.
[18] Föllmer H, Schied A (2002) Convex measures of risk and trading constraint. *Finance Stochastics* 6(4)429–447.
[19] Föllmer H, Schied A (2004) *Stochastic Finance: An Introduction in Discrete Time*, 2nd ed., de Gruyter Studies in Mathematics (Walter de Gruyter, Berlin).
[20] Fournier N, Guillin A (2015) On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* 162:707–738.
[21] Furman E, Zitikis R (2008) Weighted risk capital allocations. *Insurance Math. Econom.* 43(2):263–269.
[22] Gao N, Leung D, Munari C, Xanthos F (2018) Fatou property, representation, and extensions of law-invariant risk measures on general Orlicz spaces. *Finance Stochastics* 22:395–415.
[23] Glasserman P (2004) *Monte Carlo Methods in Financial Engineering* (Springer Science and Business Media, New York).
[24] Glasserman P, Heidelberger P, Shahabuddin P (2000) Variance reduction techniques for estimating value-at-risk. *Management Sci.* 46(10):1349–1364.
[25] Guigues V, Krätschmer V, Shapiro A (2018) A central limit theorem and hypothesis testing for risk-averse stochastic programs. *SIAM J. Optim.* 28(2):1337–1366.
[26] Holzmann H, Zwingmann T (2020) Weak convergence of quantile and expectile processes under general assumptions. *Bernoulli* 26(1):323–351.
[27] Hong LJ (2009) Estimating quantile sensitivities. *Oper. Res.* 57(1):118–130.
[28] Hong LJ, Hu Z, Liu G (2014) Monte Carlo methods for value-at-risk and conditional value-at-risk: A review. *ACM Trans. Model. Comput. Simulation* 24(4):1–37.
[29] Jin X, Fu MC, Xiong X (2003) Probabilistic error bounds for simulation quantile estimators. *Management Sci.* 14(2):230–246.
[30] Jorion P (2006) *Value at Risk*, 2nd ed. (McGraw-Hill, New York).
[31] Jouini E, Schachermayer W, Touzi N (2006) Law invariant risk measures have the Fatou property. Kusuoka S, Yamazaki A, eds. *Advances in Mathematical Economics*, vol. 9 (Springer, Japan), 49–71.
[32] Kim S, Pasupathy R, Henderson SG (2015) A guide to sample average approximation. Fu M, ed. *Handbook of Simulation Optimization*, vol. 216 (Springer, New York) 207–243.
[33] Kleywegt AJ, Shapiro A, and Homem de Mello T (2001) The sample average approximation method for stochastic discrete optimization. *SIAM J. Optim.* 12(2):479–502.
[34] Krätschmer V (2021) First order asymptotics of the sample average approximation method to solve risk averse stochastic programs. Preprint, submitted July 29, https://doi.org/10.48550/arXiv.2107.13863.
[35] Krätschmer V, Zähle H (2017) Statistical inference for expectile-based risk measures. *Scandinavian J. Statist.* 44(2):425–454.
[36] Krätschmer V, Schied A, Zähle H (2012) Qualitative and infinitesimal robustness of tail-dependent statistical functionals. *J. Multivariate Anal.* 103(1):35–47.
[37] Krätschmer V, Schied A, Zähle H (2014) Comparative and quantitative robustness for law-invariant risk measures. *Finance Stochastics* 18:271–295.
[38] Krätschmer V, Schied A, Zähle H (2015) Quasi-Hadamard differentiability of general riskk functionals and its application. *Statist. Risk Model.* 32(1):25–47.

[39] Krätschmer V, Schied A, Zähle H (2017) Domains of weak continuity of statistical functions with a view toward robust statistics. *J. Multivariate Anal.* 158(4):1–19.
[40] Kusuoka S (2001) On law invariant coherent risk measures. *Advances in Mathematical Economics* (Springer), 83–95.
[41] Markowitz HM (1952) Portfolio selection. *J. Finance* 7(1):77–91.
[42] McNeil AJ, Frey R, Embrechts P (2005) *Quantitative Risk Management* (Princeton University Press).
[43] Nonnenmacher DJF, Zagst R (1995) A new form of Jensen's inequality and its application to statistical experiments. *ANZIAM J.* 36(4):389–398.
[44] Obłój J, Wiesel J (2021) Statistical estimation of superhedging prices. *Ann. Statist.* 49(1):508–530.
[45] Pal S (2006) On capital requirements and optimal strategies to achieve acceptability. Unpublished PhD thesis, Columbia University, New York.
[46] Pal S (2007) Computing strategies for achieving acceptability: A Monte Carlo approach. *Stochastic Processes Appl.* 117(11):1587–1605.
[47] Pitera M, Schmidt T (2018) Unbiased estimation of risk. *J. Banking Finance* 91:133–145.
[48] Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. *J. Risk* 2(3):21–42.
[49] Shapiro A, Dentcheva D, Ruszczyński A (2014) *Lectures on Stochastic Programming: Modeling and Theory* (Society for Industrial and Applied Mathematics).
[50] van der Vaart AW, Wellner JA (1996) Weak convergence. *Weak Convergence and Empirical Processes with Applications to Statistics* (Springer, New York).
[51] Weber S (2007) Distribution-invariant risk measures, entropy, and large deviations. *J. Appl. Probab.* 44(1):16–40.
[52] Young VR (2004) Premium principles. Teugels JL, Sundt B, eds. *Encyclopedia of Acturial Science* (Wiley, Hoboken, NJ).