NON-ASYMPTOTIC RATES FOR THE ESTIMATION OF RISK MEASURES

DANIEL BARTL AND LUDOVIC TANGPI

Abstract. Consider the problem of computing the riskiness \( \rho(F(S)) \) of a financial position \( F \) written on the underlying \( S \) with respect to a general law invariant risk measure \( \rho \); for instance, \( \rho \) can be the average value at risk. In practice the true distribution of \( S \) is typically unknown and one needs to resort to historical data for the computation. In this article we investigate rates of convergence of \( \rho(F(S_N)) \) to \( \rho(F(S)) \), where \( S_N \) is distributed as the empirical measure of \( S \) with \( N \) observations. We provide (sharp) non-asymptotic rates for both the deviation probability and the expectation of the estimation error. Our framework further allows for hedging, and the convergence rates we obtain depend neither on the dimension of the underlying stocks nor on the number of options available for trading.

1. Introduction and Main results

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. To this aim, a rigorous axiomatic approach to risk management was initiated by Artzner et. al. \cite{Artzner1999} and matured into an impressive theory of risk measures. We refer the unfamiliar reader to Definition \ref{def:RiskMeasure}. Prime examples of risk measures are the average value at risk of Rockafellar and Uryasev \cite{Rockafellar2000} the optimized certainty equivalents of Ben-Tal and Teboulle \cite{Ben-Tal1997, Ben-Tal1998}, or the shortfall risk of Föllmer and Schied \cite{Foellmer2002}. In this paper we discuss the estimation of risk measures.

Let us first consider the case of plain risk measures, without trading or optimization issues. Denote the underlying by \( S \) and by \( \mu \) its distribution, that is, \( \mu \) is a probability measure on some measurable space \( X \) and \( S \) is a random variable distributed according to \( \mu \). Given a financial position \( F: X \rightarrow \mathbb{R} \) written on \( S \), the task is to compute \( \rho^{\mu}(F) := \rho(F(S)) \) where \( \rho \) is a law invariant convex risk measure. In practice however, the true distribution \( \mu \) is unavailable, and one often resort to historical data. This means that instead of \( \rho^{\mu}(F) \) one computes the (plug-in) estimator \( \rho^{\mu_N}(F) \), where \( \mu_N \) is the empirical measure built from \( N \) i.i.d. historical observations of the underlying \( S \). As we will soon observe, while this estimator is consistent, it typically underestimates the true risk \( \rho^{\mu}(F) \). Thus an essential question for risk managers is:

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

\textit{Date}: March 25, 2020.

\textit{2010 Mathematics Subject Classification}. 91B82, 91B30, 91B16.

\textit{Key words and phrases}. Risk measure, estimation, non-asymptotic rates, portfolio optimization, empirical processes.
To make this question rigorous, one of course needs to specify what ‘far’ means 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 above question by providing (sharp)
non-asymptotic rates on the expected estimation error and on the probability that
the estimation error exceeds some prescribed threshold.

Before presenting our main results, let us generalize the discussion to the more
practically relevant situation where hedging is also possible. In fact, we can ad-
ditionally consider options \(G_1, \ldots, G_e: \mathcal{X} \to \mathbb{R}\) available for
trading at prices \(p_1, \ldots, p_e \in \mathbb{R}\), respectively (where \(e \in \mathbb{N}\)). Trading according to a strategy \(g \in \mathbb{R}^e\) then yields the outcome
\(F + \sum_{i=1}^e g_i(G_i - p_i)\) so that the risk manager’s task is to
estimate the minimal risk incurred when trading in the option market, that is, to
compute

\[
\pi^\mu(F) := \inf_{g \in \mathcal{G}} \rho^\mu\left( F + \sum_{i=1}^e g_i(G_i - p_i) \right),
\]

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, if \(G = \{ g \in [0, 1]^e : g_1 + \cdots + g_e = 1 \}\) corresponds to portfolio optimization; see \([42]\) for some
background. Notice that if 0 is the only admissible trading strategy, i.e. \(\mathcal{G} = \{0\}\),
then we have \(\pi^\mu = \rho^\mu\) and hence all results obtained for \(\pi\) translate to \(\rho\).

1.1. Results for AVaR, OCE, and SF risk measures. While the mathematical
challenges of the present article lie in the treatment of general risk measures, we
start with an easy-to-state result for two specific and widely used risk measures.

For any \(F: \mathcal{X} \to \mathbb{R}\), the shortfall risk measure \([22]\) is defined as

\[
\text{SF}^\mu(F) := \inf \left\{ m \in \mathbb{R} : \mathbb{E}^\mu[l(F(S) - m)] \leq 1 \right\}.
\]

Here \(\mathbb{E}^\mu[\cdot]\) denotes the expectation under which \(S \sim \mu\) and \(l: \mathbb{R} \to \mathbb{R}_+\) is a loss
function, meaning that \(l\) is increasing and convex such that \(1 \in \partial l(0)\) (the subdiffer-
tential at point 0). In other words, \(\text{SF}^\mu(F)\) is the smallest capital \(m\) needed to reduce
the loss \(F\) to make it acceptable, meaning that the expected loss \(\mathbb{E}^\mu[l(F(S) - m)]\)
is below the threshold 1.

In a similar spirit, the optimized certainty equivalent (OCE) is defined via

\[
\text{OCE}^\mu(F) := \inf_{m \in \mathbb{R}} \left( \mathbb{E}^\mu[l(F(S) - m)] + m \right),\]

see \([7, 6]\). 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 or the entropic risk
measure, see \([21]\) below for the OCE representation of average value at risk.

For the rest of this introduction we assume that \(F, G_1, \ldots, G_e: \mathcal{X} \to \mathbb{R}\) are
bounded measurable functions of the underlying and that \(\mathcal{G} \subset \mathbb{R}^e\) is a bounded set.

**Theorem 1.1 (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

\[
\mathbb{E} \left[ |\pi^\mu(F) - \pi^{\mu N}(F)| \right] \leq \frac{C}{\sqrt{N}}
\]

for all \(N \geq 1\).
(ii) We have the matching deviation inequality

\[ P \left[ \left| \pi^\mu(F) - \pi^\mu_N(F) \right| \geq \varepsilon \right] \leq C \exp \left( -cN\varepsilon^2 \right) \]

for all \( N \geq 1 \) 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 diameter of \( G \). Three remarks are in order.

Remark 1.2.

(a) The rates obtained in both parts of Theorem 1.1 are the usual rates dictated by the central limit theorem and in particular optimal, see Section 5.

(b) While boundedness of \( F \) and \( G \) can be relaxed to some extend (see Theorem 3.1), the boundedness requirement on \( G \) is necessary. Indeed, in Proposition 6.1 we will show convergence of \( \pi^\mu_N(F) \) to \( \pi^\mu(F) \) (at any rate) already implies that \( G \) is bounded.

(c) An important observation is that throughout this paper the rates will never depend on the number \( e \) of options, nor on the ‘dimension’ of the underlying space \( X \). In addition \( F \) and \( G_1, \ldots, G_e \) are not subject to any continuity condition.

One could wonder whether, at least if \( X = \mathbb{R}^d \) and \( F, G \) are Lipschitz continuous, the statements of Theorem 1.1 would 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 to the true one. While this technique certainly works for dimension \( d = 1 \), in the present general, multidimensional setting this approach would force the convergence rates to be significantly worse: In dimension \( d \geq 3 \), the Wasserstein distance converges with rate \( N^{-1/d} \) instead of \( N^{-1/2} \) obtained above, see [23].

Before discussing the generalization of Theorem 1.1 beyond OCE and SF risk measures, let us present a few statistical properties of the estimator. First of all, it follows as a direct consequence of the Borel-Cantelli lemma that part (ii) of Theorem 1.1 implies the following strong consistency property:

**Corollary 1.3** (Consistency). In the setting of Theorem 1.1 we have that

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

\( P \)-almost surely.

It is clear that if \( F \) and \( G \) are bounded continuous functions, the claim of Corollary 1.3 is a consequence of weak convergence of the empirical measure to the true one. Recall however that here we merely assumed \( F \) and \( G \) to be measurable.

Despite the above strong consistency property, the estimator is typically biased as alluded to above. In fact \( \rho^\mu_N(F) \) often underestimates \( \rho(F) \). This is seen the easiest in the case of the optimized certainty equivalent: taking the infimum over \( m \) in its definition outside the expectation \( E[\cdot] \) shows that

\[ E\left[ OCE^\mu_N(F) \right] \leq \inf_{m \in \mathbb{R}} E \left[ E^\mu_N \left[ l(F(S) - m) + m \right] \right] = OCE^\mu(F). \]

The same applies in presence of trading, namely \( E[\pi^\mu_N(F)] \leq \pi^\mu(F) \). More generally, a quick inspection of OCE and SF reveals that both are concave considered as mappings of \( \mu \). Now, more general as above, for every concave mapping
$\mu \mapsto \rho^\mu(F)$, applying Jensen’s inequality formally as in the real-valued case yields $E[\rho^{\mu_N}(F)] \leq \rho^{E[\mu_N]}(F) = \rho^\mu(F)$ (where we used $E[\mu_N] = \mu$). As a matter of fact, while not all general law invariant risk measures are concave in $\mu$, this is often the case see Acciaio and Svindland [1]. In particular, the above discussion also applies to general risk measures presented in the next section, and we refer to Pitera and Schmidt [10] for further discussion on the issue of biasedness and some empirical evidence.

1.2. Results for general risk measures. It is natural to ask whether Theorem 1.1 extends to general risk measures. Let us recall the definition

**Definition 1.4 (Risk measure).** A functional $\rho: L^\infty \to \mathbb{R}$ over a standard probability space is a law invariant (convex) risk measure if

(a) $\rho(X + m) = \rho(X) + m$ and $\rho(0) = 0$ for all $X$ and $m \in \mathbb{R}$,
(b) $\rho(X) \leq \rho(Y)$ if $X \leq Y$ almost surely,
(c) $\rho(\lambda X + (1 - \lambda) Y) \leq \lambda \rho(X) + (1 - \lambda) \rho(Y)$ for $\lambda \in [0, 1]$,
(d) $\rho(X) = \rho(Y)$ if $X \sim Y$, that is, if $X$ and $Y$ have the same distribution.

As above, for a bounded function $F: \mathcal{X} \to \mathbb{R}$ and a probability $\mu$ on $\mathcal{X}$, we write $\rho^\mu(F) := \rho(F(S))$ where $S \sim \mu$.

In addition to the properties (a)-(d) stated above, it is customary to assume that risk measures satisfy some regularity condition.

**Definition 1.5.** Let $\rho: L^\infty \to \mathbb{R}$ be a risk measure and let $X, X_n \in L^\infty$ such that $\sup_n \|X_n\|_\infty < \infty$ and $X = \lim_n X_n$ almost surely. Then $\rho$ is said to have the

(a) Fatou property, if $\rho(X) \leq \liminf_n \rho(X_n)$;
(b) Lebesgue property, if $\rho(X) = \lim_n \rho(X_n)$.

Recall that by a result of Jouini, Schachermayer and Touzi [30], every law invariant risk measure automatically satisfies the Fatou property. Perhaps surprisingly, our first (negative) result states that Theorem 1.1 cannot be extended to general risk measures solely under the Lebesgue property; namely convergence can happen at an arbitrarily slow rate:

**Proposition 1.6 (No rates in general).** Assume that $\mathcal{X}$ is not a singleton. Then there exists a (sublinear) law invariant risk measure $\rho: L^\infty \to \mathbb{R}$ which satisfies the Lebesgue property and a bounded function $F: \mathcal{X} \to \mathbb{R}$ such that

there is no rate $\varepsilon > 0$ such that $E \left| \pi^\mu(F) - \pi^\mu_N(F) \right| \leq \frac{C}{N^\varepsilon}$

for all $N \geq 1$ and all probability measures $\mu$ (which are supported on two fixed distinct points in $\mathcal{X}$).

In view of the above negative result, the next step is to identify what causes the lack of convergence rates and to come up with a (hopefully) natural and easy-to-check (regularity) property for $\rho$ which guarantees convergence at some prescribed rate.
Interestingly, all it takes is for \( \rho \) to be finite for certain non-bounded random variables: roughly speaking, the more a risk measure behaves like \( \rho_{\text{max}}(X) := \text{ess.sup } X \), the worse the rates of convergence. This is due to the fact that changes of \( X \) on almost negligible sets can result in significant changes of \( \rho_{\text{max}}(X) \) and almost negligible events will not be exhibited properly by the sample. On the other extreme of the spectrum, the more a risk measure behaves like \( \rho_{\text{min}} := E[X] \), the better the rates as \( \rho_{\text{min}} \) as small changes of \( X \) will result in small changes of \( \rho_{\text{min}}(X) \).

For these two mentioned (extreme) examples, one clearly has that \( \rho_{\text{max}}(|X|) := \sup_n \rho_{\text{max}}(|X| \wedge n) \) is finite if and only if \( X \) is bounded, while \( \rho_{\text{min}}(|X|) \) is finite for all integrable \( X \). We have just explained that \( \rho_{\text{max}} \) does not allow for any convergence rates, while \( \rho_{\text{min}} \) allows for the usual \( 1/\sqrt{N} \) rates. Using convexity and monotonicity, we will be able to interpolate the above extreme cases in the sense that, roughly speaking, if an arbitrary risk measure \( \rho \) takes finite values for random variables with finite \( q \)-th moments, \( \rho \) is regular in the sense that the rates of convergence are of order \( N^{-1/2q} \).

To make these observations rigorous, we need one more definition, discussed after the proceeding theorem: a random variable \( X \) is said to have finite weak \( q \)-th moment if there is some \( C \geq 0 \) such that \( P[|X| \geq t] \leq Ct^{-q} \) for all \( t \geq 1 \).

**Theorem 1.7** (Rates for general risk measures). Let \( q \in (1, \infty) \) and let \( \rho: L^\infty \to \mathbb{R} \) be a law invariant risk measure taking finite values for random variables with finite weak \( q \)-th moment.\(^3\) Then there are constants \( c, C > 0 \) such that the following hold.

(i) We have the moment bound

\[
E \left[ \left| \pi^H(F) - \pi^{\mu_N}(F) \right| \right] \leq \frac{C}{N^{1/2q}}
\]

for all \( N \geq 1 \).

(ii) We have the matching deviation inequality

\[
P \left[ \left| \pi^H(F) - \pi^{\mu_N}(F) \right| \geq \varepsilon \right] \leq C \exp \left( -cN\varepsilon^{2q} \right)
\]

for all \( N \geq 1 \) and all \( \varepsilon > 0 \).

To make our assumption more tractable, recall that \( E[|X|^q] = q \int t^{q-1} P[|X| \geq t] \, dt \). Therefore if \( X \) has finite \( q \)-th moment, then it has finite weak \( q \)-th moment; and if \( X \) has finite weak \( q \)-th moment, then it has finite \( (q - \varepsilon) \)-th moment for all \( \varepsilon > 0 \). In particular, the assumption in the above theorem is satisfied whenever \( \rho(|X|) < \infty \) for all \( X \) with finite \( (q - \varepsilon) \)-th moment.

Note that \( F \) and \( G \) are (again) merely measurable and the rate depends solely on the assumption made on \( \rho \) and not on the dimension or the number of options traded. Moreover, the Borel-Cantelli lemma again implies that \( \pi^{\mu_N}(F) \) is a consistent estimator:

**Corollary 1.8.** In the setting of Theorem 1.7 we have that \( \lim_{N \to \infty} \pi^{\mu_N}(F) = \pi^H(F) \) \( P \)-almost surely.

---

\(^3\) We write \( \wedge \) for the minimum and \( \vee \) for the maximum.

\(^4\) By this we mean that \( \rho(|X|) := \sup_n \rho(|X| \wedge n) < \infty \) for all random variables \( X \) with finite weak \( q \)-th moment.
For values \( q \approx 1 \), the rates obtained in Theorem 1.7 almost coincide with the rates obtained in Theorem 1.1 which are the optimal (standard) rates when investigating i.i.d. phenomena. On the other hand, as \( q \) increases, the rates get worse and for \( q = \infty \), Proposition 1.6 tells us that no (polynomial) rates are available at all. The latter is in line with Theorem 1.7 and naturally triggers the question whether the results of Theorem 1.7 are optimal for all values of \( q \), which is part of the next result.

**Proposition 1.9 (Sharpness).** Assume that \( \mathcal{X} \) is not as singleton. Then, for every \( q \in [1, \infty) \), there exists a law invariant risk measure \( \rho : L^\infty \to \mathbb{R} \) taking finite values for random variables with finite \( q \)-th moment and a constant \( c > 0 \) such that: For all (large) \( N \geq 1 \) there is a probability \( \mu \) (supported on two distinct points of \( \mathcal{X} \)) satisfying

\[
E[|\pi^\mu(F) - \pi^{\mu_N}(F)|] \geq cN^{-1/q}.
\]

In other words, the rate(s) obtained in Theorem 1.7 are sharp up to a factor of two. Currently, the authors do not know whether the rates of Theorem 1.7 are actually sharp (without the factor two). One indication that this might be true is the following (explained in more details in Remark 5.3): For \( q \approx 1 \) the lower bound of Proposition 1.9 is approximately \( 1/N \) and we already know that the actual best possible rates are \( 1/\sqrt{N} \), that is, for \( q \approx 1 \) the lower bound is off exactly by the factor of two.

### 1.3. 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 which seem to fit to our framework and method include the estimation of risk premium principles in insurance, (see e.g. Young [46] or Furman and Zitikis [24] 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 \( u^\mu(F) := E[\mu(F(S))] \). Similar as before, allowing the agent to invest in a market, one obtains the utility maximization problem

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

In this case, we have the following:

**Proposition 1.10 (Utility maximization).** There are constants \( c, C > 0 \) such that

\[
E \left[ |u^{\mu}_{\text{max}}(F) - u^{\mu_N}_{\text{max}}(F)| \right] \leq \frac{C}{\sqrt{N}},
\]

\[
P \left[ |u^{\mu}_{\text{max}}(F) - u^{\mu_N}_{\text{max}}(F)| \geq \varepsilon \right] \leq C \exp \left( -cN \varepsilon^2 \right)
\]

for all \( N \geq 1 \) 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 that \( u^{\mu}_{\text{max}}(F) \) is a strongly consistent estimator which typically overestimates its true value (as we deal with maximization instead of minimization this time).
1.4. 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, Embrechts and Frey [2] for an in-depth treatment. See also the book of Glasserman [27, 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. We refer to Embrechts and Hofert [18] for an excellent review of the main lines of research in this direction. Concerning statistical estimation of risk measures, one of the earliest work is that of Weber [45] who considered the problem of estimating \( \rho(F) \) (without trading) in an asymptotic fashion as \( N \to \infty \). By means of the theory of large deviations, he showed that if \( \rho \) is sufficiently regular, then \( \rho^N(F) \) satisfies a large deviation principle. Along the same lines, [9, 5, 12] obtained central limit theorems for \( \rho^N(F) \); see also [42, Chapter 6].

This is a good place to highlight the difference between asymptotic rates and non-asymptotic rates and estimations: while there are instances where the asymptotic rates suggest a much faster convergence, this is only true within the asymptotic regime. In other words, no matter how large the sample size \( N \), asymptotic rates give no indication about how close \( \rho^N(F) \) is to \( \rho(F) \). Non-asymptotic rates however hold for every \( N \), and give an order of magnitude of the sample size needed to achieve a desired estimation accuracy.

Aside from large deviation and central limit theorem results, some authors investigate estimation of specific risk measures and (super)hedging functionals. These include Pal [37, 38] who analyzes hedging under risk measures which can be written as the finite maxima of expectations. Let us further refer to [17, 10, 40] for other (asymptotic) estimation results, mostly for the average value at risk and under some assumptions on the distribution \( \mu \); see e.g. Hong, Hu and Liu [28] for a review. A deviation-type inequality for the value at risk is proposed by Jin, Fu and Xiong [29]. The problem of strict superhedging was recently considered by Oblój and Wiesel [36]; this problem depends solely on the (topological) support of the underlying measure and therefore no rates are available in general.

When the estimation of \( \rho(F) \) is performed repeatedly or periodically, it is important that the estimator \( \rho^N(F) \) be stable, i.e. insensitive to small changes of \( \mu_N \). Such insensitivity is often referred to as robustness of the risk measure and was first analyzed by Cont, Deguest and Scandolo [15] who investigated a concept of robustness essentially equivalent to continuity of \( \rho \) w.r.t. weak convergence of measures. Alternative approaches to robustness were later proposed and analyzed by Krätschmer, Schied and Zähle [34, 33] and Cambou and Filipović [11]. Along the same lines some authors have investigated risk measures (and other stochastic maximization problems) under model uncertainty to account for the effect of possible misspecification of the estimated model, see e.g. [4, 19, 17, 20, 26, 28] where it is often assumed that the true model belongs to a Wasserstein ball. At this point, we should also mention Pichler [39] who studies the continuity of risk measures (in Wasserstein distance). Related but with a different agenda, Cheridito and Li [13, 14] characterize conditions under which risk measures on Orlicz spaces take finite values measures [13, Theorem 6.9]. The latter, as we shall see through the later proofs, also reflects on robustness.

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. The bulk of the literature in this direction is concerned with convergence issues and questions related to computational complexity of the estimators, see e.g. [32] and the book chapter [31] for a recent overview.

Somewhat related to this article, the recent years brought up a number of articles investigating non-asymptotic convergence rates of empirical measure in Wasserstein distance, see e.g. [23] and reference therein. However, as already mentioned, this approach would be restricted to the case $X = \mathbb{R}^d$, would requires strong continuity conditions on $F$ and $G$, and most importantly yields suboptimal rates.

1.5. **Organization of the rest of the paper.** We start by defining our basic notation and by proving a generalization to unbounded $F, G$ of Theorem 1.7 part (i) on mean speed convergence in Section 2 for the case of the optimized certainty equivalent. Section 3 is the main part of this paper and deals with the proof (of the generalization to unbounded $F, G$ of) Theorem 1.7 part (i). The deviation inequalities in Theorem 1.4 and Theorem 1.7 (that is, parts (ii) thereof) are proven in Section 4. Finally, sharpness of the rates for general risk measures is discussed in Section 5 and all remaining proofs are presented in Section 6.

2. 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 kind 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,...) but not on $N$. Moreover, the value of $C$ is allowed increase from line to line, for instance $\sup_N (xy - y^2) = Cx^2 \leq Cx^2 / 2$ or $C \sqrt{e+1} \leq C \sqrt{e}$ for all $e \in \mathbb{N}$, but not $N \leq C$ or $\sqrt{e+1} \leq \sqrt{e}/C$.

For a metric space $(S, d)$ and $\varepsilon > 0$, denote by $\mathcal{N}(S, d, \varepsilon)$ the covering numbers at scale $\varepsilon$, that is, $\mathcal{N}(S, d, \varepsilon)$ is the smallest number for which there is a subset $\tilde{S}$ with that cardinality satisfying: For every $s \in S$ there is $\tilde{s} \in \tilde{S}$ with $d(s, \tilde{s}) \leq \varepsilon$. In other words, $\mathcal{N}(S, d, \varepsilon)$ the smallest number of balls of radius $\varepsilon$ which covers $S$. The latter suggest this to be some measurement of compactness, and in fact, it is an important tool in understanding the behavior of empirical processes, see [13].

Recall that $e \in \mathbb{N}$ is a fixed number and $F, G_1, \ldots, G_e: X \to \mathbb{R}$ are positions written on $S$. 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 bounded. The necessity of this assumption is shown in Proposition 6.1.

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 [22, Section 4.3] for an overview. As we shall treat the average value at risk as a special case of the optimized certainty equivalents, the following definition / representation

\begin{equation}
\text{AVaR}_u(X) := \inf_{m \in \mathbb{R}} E \left[ \frac{1}{1-u} (X - m)^+ + m \right]
\end{equation}
seems best suited. Here $u \in [0, 1)$ is called the risk aversion parameter. From (2.1) it is clear that the average value at risk is a special case of the optimized certainty equivalent, recalled for the convenience of the reader:

$$\text{OCE}(X) := \inf_{m \in \mathbb{R}} E[l(X - m) + m],$$

where $l: \mathbb{R} \to \mathbb{R}_+$ is a convex increasing function with $1 \in \partial l(0)$. 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 $x \geq 0$. This assumption is there because $F$ and $G$ are possibly not bounded (in contrast to the introduction), but not needed if this is the case. We shall 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 $l'(x) \leq C(1 + |x|^{p-1})$ for all $x \in \mathbb{R}$ (with the convention $|x|^{p-1} = \infty$ for $p = \infty$). Note that in case $p = \infty$ this 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 2.1, or rather the following generalization thereof.

**Theorem 2.1.** Let $p \in [1, \infty]$, assume that $l'$ has polynomial growth of degree $p - 1$, and that $\|F\|_{L^2p(\mu)}$ and $\|G\|_{L^2p(\mu)}$ are finite. Then

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

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

We now turn to the proof of Theorem 2.1. 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 will be useful to keep track of the following quantities

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

The first result in this spirit is

**Lemma 2.2.** Assume that $l'$ has polynomial growth of degree $p - 1$. Then we have that

(2.3) \quad $|\text{OCE}^\mu(F + g \cdot G)| \leq CM^p$ and

(2.4) \quad $\text{OCE}^\mu(F + g \cdot G) = \inf_{m \leq CM^p} \int_X l(F(x) + g \cdot G(x) - m) + m \mu(dx)$

for every $g \in \mathcal{G}$. The same holds true if the pair $\mu, M$ is replaced by $\mu_N, M_N$.

**Proof.** We only focus on $\mu, M$, the proof for $\mu_N, M_N$ works analogously. 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

$$\sup_{g \in \mathcal{G}} l(F + g \cdot G) \leq \begin{cases} C J^p & \text{if } p < \infty, \\ C & \text{if } p = \infty. \end{cases}$$
In particular, the choice \( m = 0 \) (in the definition of OCE) and the fact that \( l \geq 0 \) yield
\[
\text{OCE}^\mu(F + g \cdot G) \leq \int_X l(F(x) + g \cdot G(x)) \, \mu(dx) \leq CM^p
\]
for all \( g \in \mathcal{G} \), showing the upper bound in (2.3). Further, as \( l \geq 0 \), this also implies that the infimum over \( m \) in the definition of \( \text{OCE}^\mu(F + g \cdot G) \) can be restricted to \( m \leq CM^p \) 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 \( l(x) \geq ax - b \) for every \( x \in \mathbb{R} \). This implies
\[
\int_X l(F(x) + g \cdot G(x) - m) + m \, \mu(dx)
\]
(2.5)
\[
\geq \int_X a(-CM(x) - m) - b + m \, \mu(dx)
\]
\[
\geq m(1 - a) - CM^p,
\]
where we used that \( \int_X J \, d\mu \leq M \leq M^p \) which follows from Hölder’s inequality and as \( M \geq 1 \). By the previous part we already know that \( \text{OCE}^\mu(F + g \cdot G) \leq CM^p \) for all \( g \in \mathcal{G} \). Together with (2.3) this implies that the infimum over \( m \) in \( \text{OCE}^\mu(F + g \cdot G) \) can be restricted to \( m \geq -CM^p \) for all \( g \in \mathcal{G} \). In turn, using once more that \( l \geq 0 \), this also implies that \( \text{OCE}^\mu(F + g \cdot G) \geq -CM^p \) for all \( g \in \mathcal{G} \) and thus completes the proof.

\( \square \)

Lemma 2.3. Assume that \( l' \) has polynomial growth of degree \( p - 1 \), let \( m_0 \in \mathbb{R} \), and define
\[
\mathcal{H} := \left\{ l(F + g \cdot G - m) + m : g \in \mathcal{G} \text{ and } m \in [-m_0, m_0] \right\}.
\]

Then, for every \( \varepsilon > 0 \), we have that
\[
\mathcal{N}(\mathcal{H}, \| \cdot \|_{L^2(\mu_N)}, \varepsilon) \leq \left( \frac{C\|J\|^p_{L^p(\mu_N)}}{\varepsilon} \right)^{\varepsilon + 1} + 1
\]
if \( p < \infty \); and \( \mathcal{N}(\mathcal{H}, \| \cdot \|_{L^2(\mu_N)}, \varepsilon) \leq (C/\varepsilon)^{\varepsilon + 1} + 1 \) if \( p = \infty \).

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

(a) Pick two elements \( H, \tilde{H} \in \mathcal{H} \) represented as
\[
\begin{align*}
H &= l(F + g \cdot G - m) + m \\
\tilde{H} &= l(F + \tilde{g} \cdot G - \tilde{m}) + \tilde{m}
\end{align*}
\]
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, \( |\varphi_t| \leq CJ \) for all \( t \in [0,1] \). By convexity of \( l \), its right derivative \( l' \) is increasing. By the fundamental theorem of calculus, we have
\[
\begin{align*}
\|H - \tilde{H}\|_{L^2(\mu_N)} &\leq \left\| \int_0^1 l'(\varphi_t)\varphi_t' \, dt \right\|_{L^2(\mu_N)} + |m - \tilde{m}|
\leq \left\| l'(CJ)((\tilde{g} - g) \cdot G + m - \tilde{m}) \right\|_{L^2(\mu_N)} + |m - \tilde{m}|.
\end{align*}
\]
Now note that
\[
\|l'(CJ)J\|_{L^2(\mu_N)} \leq \begin{cases} 
C\|J\|_{L^{2p}(\mu_N)}^p & \text{if } p < \infty, \\
C & \text{if } p = \infty.
\end{cases}
\]
Indeed, for \( p < \infty \) this follows from the assumption that \( l'(x) \leq C(1 + |x|^{p-1}) \)
for all \( x \in \mathbb{R} \), and the fact that \( J \geq 1 \). For \( p = \infty \), one has by assumption that
\( J \) is \( \mu \)-almost surely bounded. Hence, \( P \)-almost surely, \( J \) is also \( \mu_N \)-almost surely bounded (by the same constant). As \( l \) is bounded on bounded sets (by convexity), this implies that \( l'(J) \) is \( \mu_N \)-almost surely bounded.

To conclude, we use once more that \( \mathcal{G} \) is bounded and hence \( |(\tilde{g} - g) \cdot G| \leq |\tilde{g} - g| J \). Therefore
\[
\|H - \tilde{H}\|_{L^2(\mu_N)} \leq \begin{cases} 
C\|J\|_{L^{2p}(\mu_N)}^p |g - \tilde{g}| + |m - \tilde{m}| & \text{if } p < \infty, \\
C(|g - \tilde{g}| + |m - \tilde{m}|) & \text{if } p = \infty.
\end{cases}
\]

In the following we restrict to \( p < \infty \) and leave the obvious change to the reader.

(b) Fix \( \varepsilon > 0 \) and let \( A \subset \mathbb{R} \) be such that

for all \( m \in [-m_0, m_0] \) there is \( \tilde{m} \in A \) with \( |m - \tilde{m}| \leq \varepsilon \)

and \( B \subset \mathbb{R}^c \) such that

for all \( g \in \mathcal{G} \) there is \( \tilde{g} \in B \) with \( |g - \tilde{g}| \leq \varepsilon \).

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 \( B \), by (4), for every \( H \in \mathcal{H} \) there is \( \tilde{H} \in \tilde{\mathcal{H}} \) with \( \|H - \tilde{H}\|_{L^2(\mu_N)} \leq \varepsilon \).

This implies that
\[
\mathcal{N}(\mathcal{H}, \| \cdot \|_{L^2(\mu_N)}, \varepsilon) \leq \text{card}(\tilde{\mathcal{H}}) \leq \text{card}(A \times B) = \text{card}(A) \text{card}(B)
\]

where \( \text{card} \) means cardinality.

The set \( A \) can be constructed simply by a equidistant partition of \([-m_0, m_0] \) at cardinality \( \text{card}(A) \leq (C\|J\|_{L^{2p}(\mu_N)}^p)/\varepsilon \) \( \lor \) \( 1 \). In a similar manner, \( B \) can be constructed with \( \text{card}(B) \leq (C\|J\|_{L^{2p}(\mu_N)}^p)/\varepsilon \) \( \lor \) \( 1 \).

Combining both steps yields the proof. \( \square \)

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

**Corollary 2.4.** Let \( m_0 \in \mathbb{R} \), let \( f : \mathbb{R} \rightarrow \mathbb{R} \) be locally Lipschitz continuous, and assume that \( J \) is bounded. Then it holds that
\[
\mathcal{N}\left( \{ f(F + g \cdot G - m) : g \in \mathcal{G} \text{ and } m \in [-m_0, m_0] \}, \| \cdot \|_\infty, \varepsilon \right) \leq \left( \frac{C}{\varepsilon} \right)^{\varepsilon+1} \lor 1
\]
for every \( \varepsilon > 0 \).

We are now ready for the
Proof of Theorem 2.1. For shorthand notation, set
\[ \Delta_N := \sup_{g \in \mathcal{G}} \left| \text{OCE}^\mu(F + g \cdot G) - \text{OCE}^{\mu_N}(F + g \cdot G) \right| \]
for every \( N \geq 1 \). With \( M \) and \( M_N \) defined in (2.2), we write
\[ E[\Delta_N] = E[\Delta_N 1_{M_N \leq M+1}] + E[\Delta_N 1_{M_N > M+1}] \]
and investigate both terms separately.

(a) We start with the first term. Lemma 2.2 guarantees that
\[ \Delta_N 1_{M_N \leq M+1} \leq \sup_{H \in \mathcal{H}} \left| \int_X H(x)(\mu - \mu_N)(dx) \right| \]
for every \( N \geq 1 \), where
\[ \mathcal{H} := \{l(F + g \cdot G - m) + m : g \in \mathcal{G} \text{ and } |m| \leq C(M + 1)^p \}. \]
By the ‘empirical process version’ of Dudley’s entropy-integral theorem (see for instance Corollary 2.2.8 and Lemma 2.3.1 in [43]) one has that
\[ E\left[ \sup_{H \in \mathcal{H}} \left| \int_X H(x)(\mu - \mu_N)(dx) \right| \right] \leq C \sqrt{N} E\left[ \int_0^\infty \sqrt{\log N(\mathcal{H}, \| \cdot \|_{L^2(\mu_N)}, \varepsilon)} \, d\varepsilon \right] \]
for all \( N \geq 1 \).
Assume first that \( p < \infty \). Then, estimating the covering numbers of \( \mathcal{H} \) by means of Lemma 2.3 implies that
\[ E\left[ \sup_{H \in \mathcal{H}} \left| \int_X H(x)(\mu - \mu_N)(dx) \right| \right] \leq \frac{C}{\sqrt{N}} E\left[ \int_0^\infty \sqrt{\log N(\mathcal{H}, \| \cdot \|_{L^2(\mu_N)}, \varepsilon)} \, d\varepsilon \right] \]
where the last inequality follows from substituting \( \varepsilon \) by \( \tilde{\varepsilon} := \varepsilon/C \| J \|_{L^2(p,\mu_N)} \).
In a final step, notice that
\[ \int_0^\infty \sqrt{\log \left( \frac{1}{2} \vee 1 \right)} \, d\varepsilon < \infty \quad \text{and} \quad E[\| J \|_{L^2(p,\mu_N)}^p] \leq C \| J \|_{L^2(p,\mu)}^p. \]
The second statement follows from Jensen’s inequality. Therefore
\[ E[\Delta_N 1_{M_N \leq M+1}] \leq \frac{C}{\sqrt{N}} \]
for all \( N \geq 1 \), showing that the first term behaves as required. If \( p = \infty \) the same arguments apply (with \( \| J \|_{L^2(p,\mu_N)}^p \) replaced by a constant) and we again obtain \( E[\Delta_N 1_{M_N \leq M+1}] \leq C/\sqrt{N} \).

(b) As for the second term, applying Hölder’s inequality yields
\[ E[\Delta_N 1_{M_N > M+1}] \leq E[\Delta_N^2]^{1/2} P[M_N > M + 1]^{1/2}. \]
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 $M, M_N \geq 1$ and then Chebycheff’s inequality, we estimate

$$P[M_N > M + 1] \leq P[M_N^p - M^p > 1] \leq E[(M_N^p - M^p)^2].$$

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

$$E[(M_N^p - M^p)^2] = E\left[\left(\frac{1}{N} \sum_{n=1}^{N} (J(X_n)^p - E[J(X_n)^p])\right)^2\right]$$

$$= \frac{1}{N} E[(J(X_1)^p - E[J(X_1)^p])^2]$$

$$\leq \frac{2\|J\|_{L^{2p}(\mu)}}{N}.$$

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

Regarding $E[\Delta_N^2]$, use Lemma 2.2 to estimate

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

The same arguments as above show that $E[\Delta_N^{2p}] \leq \|J\|_{L^{2p}(\mu)}^{2p}$. Plugging both estimates in (2.7) shows that $E[\Delta_N^{1_{M_N > M+1}}] \leq C/\sqrt{N}$ for all $N \geq 1$.

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

3. 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 (convex) law invariant risk measure if (a)-(d) of Definition 1.4 hold with $L^\infty$ replaced by $L^p$. Further recall that $\rho$ is called sublinear if in addition $\rho(\lambda X) = \lambda \rho(X)$ for all $X \in L^p$ and $\lambda \geq 0$.

As already mentioned, by [30], every law invariant risk measure automatically satisfies the Fatou-property, which in turn guarantees a general dual representation [16]. Moreover, every law-invariant risk measure further satisfies the following spectral representation [4]

$$(3.1) \quad \rho(X) = \sup_{\gamma \in \mathcal{M}} \left( \int_{[0,1]} \text{AVaR}_u(X) \gamma(du) - \beta(\gamma) \right) \quad \text{for } X \in L^p,$$

see [20]. Here $\mathcal{M}$ is a subset of probability measures on $[0,1)$ armed with its Borel $\sigma$-field, $\beta: \mathcal{M} \to [0, \infty)$ is a convex function, and AVaR is the average value at

\[\text{It also goes under the name Kusuoka representation as the } L^\infty\text{-version was discovered by Kusuoka [22].}\]
risk, defined in (2.1). Note that AVaR is evidently a sublinear law invariant risk measure.

Before we are ready to state the generalization of part (i) of Theorem 1.7, the treatment of unbounded $F, G$ requires one last definition: for every parameter $p \in [1, \infty]$ and $x \geq 0$

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

Note that $w_p$ is convex, nonnegative, and $w_p$ grows at least linearly.

**Theorem 3.1.** Let $1 < q < 2p \leq \infty$ and let $\rho : L^p \rightarrow \mathbb{R}$ be a law invariant risk measures such that $\rho(|X|) < \infty$ whenever $X$ has finite weak $q$-th moment. Assume that $G$ is bounded and that $\int w_p(t(|F| + |G|)^p) \, d\mu < \infty$ for every $t \geq 0$ when $p < \infty$ and that $F$ and $G$ are bounded when $p = \infty$. Then

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

for all $N \geq 1$.

Note that $w_p$ is linear if either $p = \infty$ or $\rho$ is sublinear. In this case the integrability condition on $F$ and $G$ imposed in the above theorem simply means that $\|F\|_{L^p(\mu)}$ and $\|G\|_{L^p(\mu)}$ should be finite. To interpret the rates, note that the higher the integrability of $F$ and $G$, the better the rates. Similarly, more regularity of $\rho$ (i.e. a lower value of $q$) will also improve the rates. For convenience, we computed some values

| $1/q - 1/2p$ | $q \approx 1$ | $q = 2$ | $q = p$ | $q \approx 2p$ | $p = \infty$ |
|---------------|--------------|----------|---------|--------------|-----------|
| $1/2$         | $\frac{1}{2}$ | $\frac{1}{2p-2}$ | $\frac{1}{2p-1}$ | $0$         | $\frac{1}{2q}$ |

**Table 1.** Convergence rate for different values of $p$ and $q$.

The idea for the proof of Theorem 3.1 is the following: By Section 2 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 (3.1) they form the building block of every law invariant risk measure and we conclude via a (multiscale) approximation, keeping track of the risk aversion parameter $\mu$ of the average value at risk (which will make all constants explode when approaching $u \approx 1$) and the growth of measures $\gamma(du)$ in the spectral representation (3.1) (which only put little mass on $u \approx 1$).

The preparatory work needed is done in the next few lemmas. Some of them are rather elementary but crucial, thus we decided to include all proofs.

**Lemma 3.2.** Let $q \in (1, \infty)$ and $f_X(x) = q1_{[1,\infty)}(x)x^{-(q+1)}$ for $x \in [0, \infty)$ be the density of the distribution of the random variable $X$. Then $X$ has finite weak $q$-th moment and

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

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

**Proof.** Clearly $P[X \geq t] = t^{-q}$ for $t \geq 1$ so that $X$ has finite weak $q$-th moment by definition. Moreover, as $m \mapsto (x-m)_+$ is Lebesgue almost surely differentiable, the optimal $m^*$ for $\text{AVaR}_u(X)$ (recall (2.1)) is characterized by the first order condition

$$\int_{\mathbb{R}} \frac{1}{1-u} 1_{[m^*, \infty)}(x) + 1 F_X(dx) = 0.$$
A quick computation shows that
\[ \int_a^\infty F_X(dx) = a^{-q} \quad \text{and} \quad \int_a^\infty x F_X(dx) = \frac{q a^{-q+1}}{q-1} \]
for \( a \geq 1 \). Thus the optimal \( m \) equals \( m^* = (1-u)^{-q} \). Therefore, the value of \( \text{AVaR}_u(X) \) equals
\[ \int_\mathbb{R} \frac{1}{1-u} (x - m^*)_+ + m^* F_X(dx) = \frac{1}{1-u} \int_{m^*}^\infty x - m^* F_X(dx) + m^* \]
\[ = \frac{1}{1-u} \left( \frac{q (m^*)^{-q+1}}{q-1} - m^* (m^*)^{-q} \right) + m^*. \]
Plugging in the value of the optimal \( m^* = (1-u)^{-q} \) and simplifying the terms yields the claim. \( \square \)

**Lemma 3.3.** Let \( p \in (1, \infty] \) and \( X \in L^p \). Then we have
\[ |\text{AVaR}_u(X)| \leq \frac{\|X\|_{L^p}}{(1-u)^{1/p}} \]
for every \( u \in [0, 1) \).

**Proof.** For notational simplicity assume first that \( X \) has a strictly increasing continuous distribution function \( F_X \) and define \( m^* := F^{-1}_X(1-u) \). Plugging this choice into the definition of the average value at risk yields
\[ \text{AVaR}_u(X) \leq \int_\mathbb{R} \frac{1}{1-u} (x - m^*)_+ + m^* F_X(dx) \]
\[ = \frac{1}{1-u} E[X 1_{X \geq m^*}] \]
\[ \leq \frac{1}{1-u} \|X\|_{L^p} P[X \geq m^*(p-1)/p] \]
where we made use of Hölder’s inequality in the last step. As \( P[X \geq m^*] = 1-u \) this yields \( \text{AVaR}_u(X) \leq \|X\|_{L^p} / (1-u)^{1/p} \).

Sublinearity now implies \( |\text{AVaR}_u(X)| \leq \text{AVAR}_u(|X|) \) which shows the claim in case that \( F \) is continuous and strictly increasing.

In general, approximate \( F \) by strictly increasing functions (for instance, add independent Gaussian random variables to \( X \) with vanishing variance). \( \square \)

**Lemma 3.4.** Let \( 1 < q < 2p \leq \infty \) and assume that \( \rho(|X|) < \infty \) for all \( X \) with finite weak \( q \)-th moment. Then, for every fixed \( a > 0 \), there exists a constant \( b > 0 \) such that
\[ \rho(X) = \sup_{\beta \in \mathcal{M} : \beta(\gamma) \leq b} \left( \int_{(0,1)} \text{AVaR}_u(X) \gamma(du) - \beta(\gamma) \right) \]
for all \( X \in L^p \) with \( \|X\|_{L^p} \leq a \).

**Proof.** Let \( X^* \) be the random variable of Lemma 3.2.

(a) In a first step we show that \( |\rho(X)| \leq C \) for all \( X \in L^p \) with \( \|X\|_{L^p} \leq a \). For such \( X \), by Lemma 3.2 and Lemma 3.3 one has that
\[ \text{AVaR}_u(|X|) \leq \frac{a}{(1-u)^{1/p}} \leq \text{AVaR}_u(CX^*) \]
for every $u \in [0, 1)$. Therefore
\[
\rho(|X|) \leq \sup_{\beta \in \mathcal{M}} \left( \int_{[0,1)} \text{AVaR}_u(|X|) \gamma(du) - \beta(\gamma) \right)
\]
\[
\leq \sup_{\beta \in \mathcal{M}} \left( \int_{[0,1)} \text{AVaR}_u(CX^*) \gamma(du) - \beta(\gamma) \right) = \rho(CX^*)
\]
for every $X$ with $\|X\|_{L^p} \leq a$.

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

(b) We proceed to prove the claim. Define
\[
\varphi : \mathbb{R}_+ \to [0, \infty] \text{ by } \varphi(y) := \sup_{x \in \mathbb{R}_+} (xy - \rho(xX^*)).
\]
Then $\varphi$ is convex, increasing, and as $\rho(xX^*) < \infty$ 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 $\rho$ in (3.1) implies that
\[
\rho(xX^*) \geq \int_{[0,1)} \text{AVaR}_u(xX^*) \gamma(du) - \beta(\gamma)
\]
for all $x \geq 0$ and $\gamma \in \mathcal{M}$. Therefore, one has
\[
\beta(\gamma) \geq \sup_{x \geq 0} \left( \int_{[0,1)} \text{AVaR}_u(xX^*) \gamma(du) - \rho(xX^*) \right)
\]
\[
= \varphi\left( \int_{[0,1)} \text{AVaR}_u(X^*) \gamma(du) \right)
\]
for every $\gamma \in \mathcal{M}$. For every $X$ with $\|X\|_{L^p} \leq a$, by (3.2) one has
\[
(3.3) \quad \int_{[0,1)} \text{AVaR}_u(X) \gamma(du) - \beta(\gamma) \leq C \int_{[0,1)} \text{AVaR}_u(X^*) \gamma(du) - \beta(\gamma)
\]
\[
\leq C\varphi^{-1}(\beta(\gamma)) - \beta(\gamma),
\]
where $\varphi^{-1}$ denotes the (right)-inverse of $\varphi$.

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
\[
(3.4) \quad C\varphi^{-1}(\beta(\gamma)) - \beta(\gamma) \to -\infty \text{ when } \beta(\gamma) \to \infty.
\]

Now recall that $\rho(X)$ equals the supremum over $\gamma \in \mathcal{M}$ of the left hand side of (3.3) and that $|\rho(X)| \leq C$ for all $X$ with $\|X\|_{L^p} \leq a$ by the first part of this proof. Therefore (3.4) implies that there is some constant $b$ such that only $\gamma \in \mathcal{M}$ for which $\beta(\gamma) \leq b$ need to be considered in the computation of $\rho(X)$.

\[\Box\]

**Lemma 3.5.** Let $q \in (1, \infty)$ and assume that $\rho(|X|) < \infty$ for all $X$ with finite weak $q$-th moment. Then, for every fixed $b \in \mathbb{R}_+$, we have
\[
\Gamma_b([r, 1)) := \sup_{\gamma \in \mathcal{M} \text{ s.t. } \beta(\gamma) \leq b} \gamma([r, 1)) \leq C(1 - r)^{1/q}
\]
for every $r \in [0, 1)$.
Proof. Let $X^*$ be the random variable of Lemma 3.2. Then it follows from interchanging two suprema in the spectral representation (3.1) (one over $n$ and one over $\gamma$), monotone convergence (applied under each $\gamma$), and Lemma 3.2 that

$$
\sup_n \rho(X^* \wedge n) = \sup_{\gamma \in \mathcal{M}} \sup_n \left( \int_{[0,1]} \text{AVaR}_u^\gamma(X^* \wedge n) \gamma(du) - \beta(\gamma) \right)
$$

(3.5)

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

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

By properties of the geometric series one has

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

(3.6)

Putting everything together, this implies

$$
\sum_{n < n_N} s_n \leq C x \cdot x^{a-b} = C x^{a-b}.
$$
For the tail of the sum, the same computation as in (3.6) shows that 
\[ 2^{nN(b-a)} = x^{\frac{b-a}{1-b}}. \] 
Therefore, another application of the geometric series implies that 
\[ \sum_{n\geq n_N} s_n = \sum_{n\geq n_N} 2^{n(b-a)} \leq \frac{2^{nN(b-a)}}{1 - 2^{b-a}} \leq C 2^{nN(b-a)} = C x^{\frac{b-a}{1-b}}. \]
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
\[ \sum_{n\geq 1} 2^{-an} \cdot \left( x^{2^n} \land 2^{bn} \right) \leq x \sum_{n\geq 1} 2^{-an} \cdot \left( 2^n \land 2^{bn} \right) \leq Cx, \]
where the last inequality follows from convergence of the geometric series / the previous step. \( \square \)

For every \( N \geq 1 \) and \( u \in [0,1] \) define
\[ \delta^N_u := \sup_{g \in G} |\text{AVaR}_u^\mu(F + g \cdot G) - \text{AVaR}_u^\mu(F + g \cdot G)|. \] 
The following lemma controls uniformly the behavior of \( \delta \).

**Lemma 3.7.** Let \( 1 < q < 2p \leq \infty \). Then it holds
\[ E \left[ \sup_{u \in [0,v]} \delta^N_u \right] \leq \frac{C}{(1-v)\sqrt{N}} \wedge \frac{C}{(1-v)^{1/2p}} \]
for every \( v \in (0,1) \).

**Proof.** We start with the easier estimate, namely that
\[ E \left[ \sup_{u \in [0,v]} \delta^N_u \right] \leq \frac{C}{(1-v)^{1/2p}}. \]
As \( |F + g \cdot G| \leq C J \) for every \( g \in G \), monotonicity of AVaR\( \mu \) implies \( \text{AVaR}_u^\mu(F + g \cdot G) \leq \text{AVaR}_u^\mu(CJ) \) for every \( g \in G \); similarly with \( \mu \) replaced by \( \mu_N \). Now Lemma 3.3 implies
\[ \sup_{u \in [0,v]} \delta^N_u \leq \frac{\|CJ\|_{L^{2p}(\mu)} + \|CJ\|_{L^{2p}(\mu_N)}}{(1-v)^{1/2p}}. \]
Further Jensen’s inequality implies \( E[\|CJ\|_{L^{2p}(\mu_N)}] \leq \|CJ\|_{L^{2p}(\mu)} \) and thus we get (3.8).

To conclude the proof, we are left to prove that
\[ E \left[ \sup_{u \in [0,v]} \delta^N_u \right] \leq \frac{C}{(1-v)\sqrt{N}}, \]
which we shall do in several steps.

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

Then it holds that
\[ \sup_{u \in [0,v]} \delta^N_u \leq \frac{1}{1 - v} \sup_{H \in \mathcal{H}} \left| \int_X H(\mu - \mu_N)(dx) \right|. \]
Indeed, every function appearing in the definition of AVaR\(a\) is of the form \(\varphi(F + g \cdot G)/(1 - u)\) for a 1-Lipschitz function, see \(\cite{21}\). 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 \(\mathcal{H}\). Fix \(a \geq 1\) to be chosen later. For two functions

\[
H = \varphi(F + g \cdot G) \quad \text{and} \quad \bar{H} = \tilde{\varphi}(F + \tilde{g} \cdot G) \quad \text{in} \quad \mathcal{H}
\]

we write

\[
\|H - \bar{H}\|_{L^2(\mu_N)} \leq \|1_{J\leq a}(H - \bar{H})\|_{L^2(\mu_N)} + \|1_{J > a}(H - \bar{H})\|_{L^2(\mu_N)}
\]

and estimate both terms separately. As \(\mathcal{G}\) is bounded, \(|F + g \cdot G| \leq CJ\) for every \(g \in \mathcal{G}\); thus we have for the first term

\[
\|1_{J\leq a}(H - \bar{H})\|_{L^2(\mu_N)} \leq \sup_{|t| \leq Ca} |\varphi(t) - \tilde{\varphi}(t)| + Ca|g - \tilde{g}|.
\]

On the other hand, the second term equals zero if \(p = \infty\) (that is, if \(J\) is bounded) and \(a\) is large enough. For \(p < \infty\) we have

\[
\|1_{J > a}(H - \bar{H})\|_{L^2(\mu_N)} \leq C\|1_{J > a}J\|_{L^2(\mu_N)}
\]

\[
\leq C\|J^p\|_{L^2(\mu_N)} \frac{a^p}{a^p}.
\]

In the following, we shall consider only the (slightly more difficult) case \(p < \infty\) and leave the minor changes for the case \(p = \infty\) to the reader. In conclusion we have shown

\[
\|H - \bar{H}\|_{L^2(\mu_N)} \leq \sup_{|t| \leq Ca} |\varphi(t) - \tilde{\varphi}(t)| + Ca|g - \tilde{g}| + \frac{C\|J^p\|_{L^2(\mu_N)}}{a^p}.
\]

With this preparatory work out of the way, we proceed to compute \(\mathcal{N}(\mathcal{H}, \|\cdot\|_{L^2(\mu_N)}, \varepsilon)\) by making all three terms in the right hand side of \(\text{(3.11)}\) smaller than \(\varepsilon/3\). Let \(\hat{L}\) be a set of (Lipschitz) functions from \(\mathbb{R} \rightarrow \mathbb{R}\) such that for every 1-Lipschitz function \(\varphi\) there is \(\tilde{\varphi} \in \hat{L}\) such that \(\sup_{|t| \leq Ca} |\varphi(t) - \tilde{\varphi}(t)| \leq \varepsilon/3\). Such set \(\hat{L}\) can be constructed with

\[
\text{card}(\hat{L}) \leq \exp \left( \frac{C}{(\varepsilon/a) \wedge 1} \right),
\]

see \(\cite{33}\) Theorem 2.7.1.\footnote{Indeed, while \(\cite{33}\) considers Lipschitz functions from \([0, 1] \rightarrow [0, 1]\), the mapping \(\varphi \rightarrow (t \mapsto \varphi(2\alpha(t - 1/2))/(2\alpha))\) forms a bijection from 1-Lipschitz function with domain \([0, 1]\) to the ones with domain \([-\alpha, \alpha]\). The latter can be extended to function with domain \(\mathbb{R}\) and this is exactly how our set \(\hat{L}\) is obtained.} Moreover, let \(\hat{G}\) such that for every \(g \in \mathcal{G}\) there is \(\tilde{g} \in \hat{G}\) satisfying \(|g - \tilde{g}| \leq \varepsilon/3Ca\). Such set can be constructed with

\[
\text{card}(\hat{G}) \leq \left( \frac{C}{(\varepsilon/a) \wedge 1} \right)^{\varepsilon},
\]

by an equidistant grid of the bounded set \(\mathcal{G} \subset \mathbb{R}^\varepsilon\). Finally, if

\[
a := \left( \frac{C\|J^p\|_{L^2(\mu_N)}}{\varepsilon} \right)^{1/p},
\]

...
then the last term in (3.11) is smaller than $\varepsilon/3$ as well. Therefore, writing more compactly $\|J\|_{L^{2p}(\mu_N)} = \|J^p\|_{L^2(\mu_N)}^{1/p}$, we conclude

$$N(\mathcal{H}, \|\cdot\|_{L^2(\mu_N)}, \varepsilon) \leq \text{card}(\bar{L})\text{card}(\bar{G})$$

(3.12)

$$\leq \exp \left( \frac{C}{(\varepsilon(p+1)/p/C\|J\|_{L^{2p}(\mu_N)}) \wedge 1} \right) \cdot \left( \frac{C}{(\varepsilon(p+1)/p/C\|J\|_{L^{2p}(\mu_N)}) \wedge 1} \right)^{\epsilon}$$

for every $\varepsilon > 0$.

(c) Using (3.10) and Dudley’s theorem as in the proof of Theorem 1.7, we obtain

$$E \left[ \sup_{u \in [0, \varepsilon]} \delta_u^N \right] \leq \frac{C}{\sqrt{N}} \int_0^\infty \sqrt{\log N(\mathcal{H}, \|\cdot\|_{L^2(\mu_N)}, \varepsilon)} \, d\varepsilon$$

$$= \frac{C}{\sqrt{N}} \int_0^\infty \log \left( \exp \left( \frac{C}{(\varepsilon(p+1)/p/C\|J\|_{L^{2p}(\mu_N)}) \wedge 1} \right) \cdot \left( \frac{C}{(\varepsilon(p+1)/p/C\|J\|_{L^{2p}(\mu_N)}) \wedge 1} \right)^{\epsilon} \right) \, d\varepsilon$$

where the last line followed from using (3.12) and substituting $\varepsilon$ by $\tilde{\varepsilon} = \varepsilon/\|CJ^p\|_{L^2(\mu_N)}^{1/p}$. It remains to notice that the (now deterministic) integral over $d\varepsilon$ is finite. Moreover, Jensen’s inequality implies $E[\|J\|_{L^{2p}(\mu_N)}] \leq \|J\|_{L^{2p}(\mu)}$ and the latter term is finite by assumption.

In conclusion, we have shown (3.9) and the proof is complete. \qed

Proof of Theorem 3.7. Recall the definition of $M := \|J\|_{L^p(\mu)}$ and $M_N := \|J\|_{L^p(\mu_N)}$ given in (2.2). As in the proof of Theorem 2.1 we set

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

and consider both terms in

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

separately.

(a) As $\mathcal{G}$ is bounded, we have $\|F + g \cdot G\|_{L^p(\mu)} \leq CM$. Therefore, by Lemma 3.5, there exists some $b$ such that

$$\rho^\mu(F + g \cdot G) = \sup_{\beta \in M \text{ s.t. } \beta(\gamma) \leq b} \left( \int_{[0,1]} \text{VaR}^\mu_{\beta}(F + g \cdot G) \gamma(du) - \beta(\gamma) \right)$$

for all $g \in \mathcal{G}$. Possibly making $b$ larger, the same reasoning implies that, on the set $M_N \leq M + 1$, the same representation holds true if $\mu$ is replaced by $\mu_N$.

Recalling the definition of $\delta^N$ in (3.7) and the definition of $\Gamma_b$ given in Lemma 3.4, we can write

$$\Delta_N 1_{M_N \leq M+1} \leq \sup_{\beta \in M \text{ s.t. } \beta(\gamma) \leq b} \int_{[0,1]} \delta^N_{\mu} \gamma(du)$$

$$\leq \sum_{n \geq 1} \Gamma_b(I_n) \sup_{u \in I_n} \delta^N_{\mu},$$

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 $\Gamma_b(I_n) \leq C2^{-n/q}$ by means of Lemma 3.5 and $E[\sup_{u \in I_n} \delta^N_{\mu}] \leq C(2^n \sqrt{N} - 1) \wedge 2^{n/2p}$ by means of and Lemma 5.7. Then, an application of
Lemma 3.6 implies that
\[
E[\Delta N_{M_N \leq M+1}] \leq C \sum_{n \geq 1} 2^{-n/q} \left( \frac{2^n}{\sqrt{N}} \wedge 2^{n/2p} \right)
\]
\[
\leq \frac{C}{\sqrt{N}^{1/2 - 1/2q}} \vee \frac{C}{\sqrt{N}^{1/2 - 1/2p}} \leq \frac{C}{\sqrt{N}^{1/2 - 1/2q}}
\]
where the last inequality as \((1/q - 1/2p)/(1 - 1/2p) \in (0, 1)\).

(b) The second term is controlled in a similar way as in the proof of Theorem 2.1, namely we first estimate
\[
E[\Delta N_{M_N > M+1}] \leq E[\Delta^2_{N}]^{1/2} P[M_N > M + 1]^{1/2}
\]
\[
\leq CE[\Delta^2_{N}]^{1/2}.
\]
It therefore remains to check that \(E[\Delta^2_{N}] \leq C\). 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 \(\rho\) and the fact that \(G\) 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
\[
\rho^{\mu_N}(CJ) \leq w_p(C\|J\|_{L^p(\mu_N)}) \leq w_p\left(C \frac{1}{N} \sum_{n \leq N} J(X_n)^p\right).
\]
By convexity of \(x \mapsto w_p(x)^2\) we may further estimate
\[
E[\rho^{\mu_N}(CJ)^2] \leq \frac{1}{N} \sum_{n \leq N} E\left[w_p\left(CJ(X_n)^p\right)^2\right]
\]
\[
= \int_X w_p(CJ(x)^p)^2 \mu(dx)
\]
and the last term is finite by assumption.

Combining both steps completes the proof. \(\square\)

4. Deviation inequalities

In the following, we prove (the following generalizations of) part (ii) of Theorem 1.7 and part (ii) of Theorem 1.1 stated in the introduction.

**Theorem 4.1.** Assume that \(F\) and \(G\) are bounded functions and that the set \(G\) is bounded. Moreover, assume that \(\rho(|X|) < \infty\) for all \(X\) with finite weak \(q\)-th moment, where \(q \in (1, \infty)\). Then there are constants \(c, C > 0\) such that
\[
P\left[\sup_{g \in G} |\rho^{\mu}(F + g \cdot G) - \rho^{\mu_N}(F + g \cdot G)| \geq \varepsilon\right] \leq C \exp\left(-c N \varepsilon^{2q}\right)
\]
for all \(\varepsilon > 0\) and \(N \geq 1\).

**Proof.** (a) In a first step, as \(F\), \(G\), and \(G\) are bounded, the same arguments as given for Lemma 2.2 (simpler, in fact, due to the boundedness of the function \(J\) therein) show that
\[
\text{AVaR}_u^\mu(F + g \cdot G) = \inf_{|m| \leq u} \frac{1}{1 - u} \int_X (F + g \cdot G - m)_+ + (1 - u)m \mu(dx)
\]
for every $u \in [0, 1)$ and $g \in \mathcal{G}$, where $a$ depends on $\|F\|_\infty$, $\|G\|_\infty$, and the size of $\mathcal{G}$. Moreover, (4.11) remains true if $\mu$ is replaced by $\mu_N$. 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
\[
(4.2) \quad |\text{AVaR}_u(F + g \cdot G) - \text{AVaR}_{\mu_N}(F + g \cdot G)| \leq \frac{\delta^N_0}{1-u},
\]

where we set
\[
\delta^N_0 := \sup_{H \in \mathcal{H}} \int_{\mathcal{X}} H(x) (\mu - \mu_N)(dx)
\]
and
\[
\mathcal{H} := \{(F + g \cdot G - m) : |m| \leq a \text{ and } g \in \mathcal{G}\}.
\]

(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 (3.1) of $\rho$ can be restricted to those $\gamma$ for which $\beta(\gamma) \leq b$. This implies
\[
|\rho^N(F + g \cdot F) - \rho^N_{\mu}(F + g \cdot G)|
\]
\[
\leq \sup_{\gamma \in \mathcal{M} \text{ s.t. } \beta(\gamma) \leq b} \int_{[0,1]} |\text{AVaR}_{\mu}(F + g \cdot G) - \text{AVaR}_{\mu_N}(F + g \cdot G)| \gamma(du)
\]
\[
\leq \sum_{n \geq 1} \Gamma_b(I_n) \sup_{u \in I_n} \frac{\delta^N_0}{1-u}
\]

where $I_n := [1 - 2^{-n+1}, 1 - 2^{-n})$ for every $n$. Estimating $\Gamma_b(I_n)C2^{-n/q}$ by Lemma 3.3 and using Lemma 3.6 one arrives at
\[
(4.3) \quad \sup_{g \in \mathcal{G}} |\rho^N(F + g \cdot F) - \rho^N_{\mu}(F + g \cdot G)| \leq C\left(\delta^N_0\right)^{1/q} \vee \delta^N_0
\]
\[
\leq C(\delta^N_0)^{1/q}
\]

for all $N \geq 1$ almost surely, where the last inequality holds as $\delta^N_0 \leq C$ almost surely (and hence $\delta^N_0 \leq C(\delta^N_0)^{1/q}$ almost surely).

(c) In a last step, it remains to estimate $\delta^N_0$. The goal is to apply [43, Theorem 2.14.10]. By Corollary 2.1 one has that
\[
\mathcal{N}(\mathcal{H}, \|\|_{\infty}, \varepsilon) \leq \left(\frac{C}{\varepsilon}\right)^{\varepsilon+1} \vee 1
\]
for all $\varepsilon > 0$. As $\mathcal{N}(\mathcal{H}, \|\cdot\|_{L^2(\nu)}, \varepsilon) \leq \mathcal{N}(\mathcal{H}, \|\cdot\|_{\infty}, \varepsilon)$ for every probability $\nu$ on $\mathcal{X}$, the requirement (2.14.8) in [43] for [43, Theorem 2.14.10] is satisfied.

Therefore an application of that theorem shows that
\[
P\left[\sqrt{N} \delta^N_0 \geq \varepsilon\right] \leq C \exp\left(\frac{\varepsilon^a}{C}\right) \exp(-2\varepsilon^2),
\]
for some $a \in (0, 2)$ (with the notation of that theorem: as $U < 2$, chose $\delta > 0$ small so that $U + \delta < 2$). Finally, note that
\[
\exp\left(\frac{\varepsilon^a}{C}\right) \exp(-2\varepsilon^2) \leq C \exp\left(-\frac{\varepsilon^2}{C}\right),
\]
which implies
\[
P[\delta^N_0 \geq \varepsilon] \leq C \exp\left(-\frac{N\varepsilon^2}{C}\right).
\]

\[8\] In fact, the cited theorem is stated for classes of functions taking values in $[0,1]$, however, the present situation does not affect the statement up to constants.
for all $\varepsilon > 0$ and $N \geq 1$. The proof is completed by plugging the last estimate in equation (4.3). □

**Theorem 4.2.** Assume that $F$ and $G$ are bounded functions, that the set $G$ is bounded, and let $\rho = OCE$ be the optimized certainty equivalent risk measure. Then there are constants $c, C > 0$ such that

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

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

**Proof.** The proof is similar to the one given for Theorem 4.1 and we shall keep it short. By Lemma 2.2 one has

$$
|\rho^{\mu}(F + g \cdot G) - \rho^{\mu,N}(F + g \cdot G)| \leq \sup_{H \in \mathcal{H}} \left| \int_X H(x)(\mu - \mu_N)(dx) \right| =: \delta^N_0
$$

almost surely, for the set

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

Again apply Corollary 2.4 and [43, Theorem 2.14.10] to deduce that $P[\delta^N_0 \geq \varepsilon] \leq C \exp(-cN\varepsilon^2)$ for some constants $c, C > 0$. This concludes the proof. □

5. 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 $\rho(X) := E[X] = \text{AVaR}_0(X)$. Then, if $\mu$ is a probability on $[0,1]$ and $F$ is a (bounded) function which is equal to the identity on $[0,1]$, one simply has

$$
\rho^{\mu,N}(F) = \frac{1}{N} \sum_{n \leq N} F(S_n) \approx \text{Normal}\left( \rho^{\mu}(F(S)), \frac{\text{Var}(F(S))}{N} \right)
$$

for large $N$ by the central limit theorem. In particular $E[|\rho^{\mu}(F) - \rho^{\mu,N}(F)|] \approx \sqrt{\text{Var}(F(S))}/\sqrt{N}$ for all large $N$ and $P[|\rho^{\mu}(F) - \rho^{\mu,N}(F)| \geq \varepsilon] \approx 2\Phi(-\varepsilon \sqrt{N/\text{Var}(F(S))})$ where $\Phi$ is the cumulative distribution function of the normal distribution.

In contrast to the above $1/\sqrt{N}$ rate, the rates obtained for general risk measures e.g. in Theorem 4.1 are worse. As the proofs are presented, they depend on the continuity (integrability) of the risk measure. This section is devoted to showing that the integrability conditions imposed are necessary to obtain any rates and that the rates are in fact sharp up to a factor of 2 (comments on the factor 2 are given in Remark 5.3 below).

To ease the notation, for probabilities $\mu$ on $\mathbb{R}$ with bounded support, we shall write

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

(this corresponds, of course, to $\rho^{\mu}(F)$ for $X = \mathbb{R}$ and $F$ the identity). With this notation, Proposition 1.6 reads as follows.
Proposition 5.1. Let \( \varepsilon > 0 \). Then there exists a sublinear law invariant risk measure \( \rho : L^\infty \to \mathbb{R} \) satisfying the Lebesgue property (see below) as well as a constant \( c > 0 \) such that
\[
\sup_{\mu \text{ probability on } \{0,1\}} E\left[ |\rho(\mu) - \rho(\mu_N)| \right] \geq \frac{c}{N\varepsilon} \]
for all (large) \( N \).

Remark 5.2. Without the assumption that \( \rho \) satisfies the Lebesgue property, the proof of Proposition 5.1 becomes rather trivial: take \( \rho(X) := \text{ess.sup} X \) and let \( \mu \) be some probability with support \([0,1]\). As \( \rho(\mu_N) = \max_{n \leq N} X_n \) (where \((X_n)\) is an i.i.d. sample of \( \mu \)) one has
\[
P\left[ |\rho(\mu) - \rho(\mu_N)| \geq \varepsilon \right] = P\left[ \max_{n \leq N} X_n \leq 1 - \varepsilon \right] = \mu([0,1-\varepsilon])^N.
\]
For suitable choices of \( \mu \), the latter term can converge arbitrarily slow to zero. Therefore \( E[|\rho(\mu) - \rho(\mu_N)|] = \int_0^1 \mu([0,1-\varepsilon])^N d\varepsilon \) converges arbitrarily slow as well.

The proof of Proposition 5.1 below mimics the idea of Remark 5.2 while simultaneously enforcing the Lebesgue property. Moreover, it actually also reveals the following.

Remark 5.3. Combining Theorem 1.7 and Proposition 1.9 gives the following: For every law invariant risk measure as in the theorem, there are two constants \( c \geq 0 \) and \( C > 0 \) such that
\[
\frac{c}{N^{1/q}} \leq \sup_{\mu \text{ probability on } \{x_0,x_1\}} E\left[ |\pi^\mu(F) - \pi^{\mu_N}(F)| \right] \leq \frac{C}{\sqrt{N}^{1/q}}
\]
for all \( N \geq 1 \). Further, for certain choices of \( \rho \), the constant \( c \) can be chosen strictly positive \( c > 0 \).

This means the following: there is a gap between the rates which we are able to prove (the right hand side of (5.1)) and the rates which we are able to prove to be sharp (the left hand side of (5.1)); they are off by a factor of two.

However, we have already seen at the beginning of this section that the rate \( 1/\sqrt{N} \) can never be beaten, hence the left hand side of (5.1) certainly can be improved (at least for \( q \in (1,2) \)). This suggests that the proof of Proposition 1.9 might be optimized for general \( q \) in order to obtain sharper lower bounds. At this stage, unfortunately, the authors are unaware how.

To ease notation, denote by
\[
\text{Ber}(p) := (1-p)\delta_0 + p\delta_1
\]
the Bernoulli distribution with parameter of success \( p \in [0,1] \). Then, for \( \mu = \text{Ber}(p) \), the empirical measure \( \mu_N \) of \( \mu \) satisfies
\[
\mu_N \equiv \text{Ber}(p)_N = \text{Ber}(\hat{p}_N) \quad \text{where} \quad \hat{p}_N := \frac{1}{N} \sum_{n \leq N} X_n
\]
(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. As they are important, we include their (short) proofs.
**Lemma 5.4.** Let $p \in (0, 1)$. Then
\[
\text{AVaR}_u(Ber(p)) = \frac{p}{1-u} \wedge 1
\]
for every $u \in [0, 1)$.

**Proof.** By definition, we have
\[
\text{AVaR}_u(Ber(p)) = \inf_{m \in \mathbb{R}} \left( \frac{1}{1-u} \left( p(1-m)^+ + (1-p)(-m)^+ \right) + m \right)
\]
\[
= \frac{1}{1-u} \inf_{m \in \mathbb{R}} \begin{cases} 
1-u & \text{if } m \geq 1 \\
(m(1-u-p) + p & \text{if } 0 < m < 1 \\
p - um & \text{if } m \leq 0
\end{cases}
\]
\[
= \frac{p}{1-u} \wedge 1
\]
which shows the claim. \(\square\)

**Lemma 5.5.** It holds that
\[
\sup_{x \geq 1} \left( (1-x^{-\varepsilon})a + x^{-\varepsilon}((ax) \wedge 1) \right) = (1-a^{\varepsilon})a + a^{\varepsilon}
\]
for every $a \in [0, 1]$ and $\varepsilon > 0$.

**Proof.** For $a = 0$ or $a = 1$ the claim is clear. If $a \in (0, 1)$, the supremum can be restricted over $x \in [1, 1/a]$. For those $x$ the value equals $a(1 + x^{1-\varepsilon} - x^{-\varepsilon})$ which is increasing as a function in $x$. Hence the optimal $x$ is $1/a$ which yields the claim. \(\square\)

**Proof of Proposition 5.1.** Let $\varepsilon > 0$ be arbitrary and define $\rho: L^\infty \to \mathbb{R}$ by
\begin{equation}
\rho(X) := \sup_{x \geq 1} \left( (1-x^{-\varepsilon})\text{AVaR}_0(X) + x^{-\varepsilon}\text{AVaR}_{1-1/x}(X) \right)
\end{equation}
As AVaR is a law invariant sublinear risk measure, $\rho$ inherits all those properties. Moreover, a quick computation shows that $\text{AVaR}_u(X)$ satisfies the Lebesgue property for every $u \in [0, 1)$. As $x^{-\varepsilon} \to 0$ when $x \to \infty$, this then implies that $\rho$ satisfies the Lebesgue property as well.

For every $N$, we shall choose $\mu := Ber(p_N)$ with $p_N := 1/N$ in the supremum over all probabilities on $[0, 1]$ appearing in the statement of the proposition. So let $(X_n^N)$ be an i.i.d. sample of $Ber(p_N)$, that is, $P[X_n^N = 1] = p_N = 1/N$ for all $n$ and $N$, and recall that $\mu_N = Ber(\hat{p}_N)$ where $\hat{p}_N := \frac{1}{N} \sum_{n \leq N} X_n^N$. We will show that
\[
\rho(\text{Ber}(p_N)) - E[\rho(\text{Ber}(\hat{p}_N))] \geq \frac{p_N^\varepsilon}{C}
\]
for all (large) $N$. Using the triangle inequality, this clearly implies the statement of the proposition.

By Lemma 5.4 and Lemma 5.5 we compute
\[
\rho(\text{Ber}(p_N)) = \sup_{x \geq 1} \left( (1-x^{-\varepsilon})p_N + x^{-\varepsilon}((xp_N) \wedge 1) \right)
\]
\[
= (1 - p_N^\varepsilon)p_N + p_N^\varepsilon
\]
and similarly
\[
\rho(\text{Ber}(\hat{p}_N)) = (1 - \hat{p}_N^\varepsilon)\hat{p}_N + \hat{p}_N^\varepsilon.
\]
Now recall that $E[\hat{p}_N] = p_N$ and, by Jensen’s inequality, $E[\hat{p}_N^p] \geq p_N^p$; hence
\[
\rho(\text{Ber}(p_N)) - E[\rho(\text{Ber}(\hat{p}_N))] \geq p_N^\rho - E[\hat{p}_N].
\]

For the set
\[ A_N := \{ \hat{p}_N = 0 \} = \{ X_n^N = 0 \text{ for all } n \leq N \}, \]
one computes
\[
P[A_N] = (1 - p_N)^N \to \exp(-1) \in (0, 1)
\]
as $N \to \infty$. Moreover $E[\hat{p}_N] = E[\hat{p}_N^1] 1_{A_N}$ and an application of Hölder’s inequality (with exponents $p = 1/\varepsilon$ and $q = 1/(p - 1) = 1/(1 - \varepsilon)$) gives
\[
E[\hat{p}_N] \leq E[\hat{p}_N]^\varepsilon P[A_N]^{1-\varepsilon}
\]
\[
\leq p_N \left(1 - \frac{\exp(-1)}{2}\right)^{1-\varepsilon} =: p_N c
\]
for all $N$ large enough. Here we also used that $E[\hat{p}_N] = p_N$ and the previous computation for (the limit of) $P[A_N]$.

In particular
\[
\rho(\text{Ber}(p_N)) - E[\rho(\text{Ber}(\hat{p}_N))] \geq p_N^\rho (1 - c)
\]
for all $N$ large enough. As $c \in (0, 1)$, this completes the proof. \qed

**Proof of Proposition 6.4.** We use the notation as in the proof of Proposition 5.1. Define $\rho: L^\infty \to \mathbb{R}$ as in (5.1) with $\varepsilon := 1/q$. We need to check that $\rho(|X|) < \infty$ for all $X$ with finite $q$-th moment. By definition of $\rho$ and Lemma 5.3 it follows that
\[
\rho(|X| \wedge n) \leq \|X\|_{L^q} \sup_{x \geq 1} (1 - x^{-\varepsilon}) 1 + x^{-\varepsilon} x^{1/p} < \infty
\]
for all $n \in \mathbb{N}$. This shows that $\rho(|X|) < \infty$.

At this point we may copy the rest of the proof of Proposition 5.1 and obtain that $\rho(\text{Ber}(p_N)) - E[\rho(\text{Ber}(\hat{p}_N))] \geq N^{-1/q}/C$ for all large $N \geq 1$, which implies the claim. \qed

6. **Additional results**

In this last section we provide an additional result pertaining to the boundedness assumption on $\mathcal{G}$, and the remaining proofs, notably of the estimation of shortfall risk measure and of utility maximization.

6.1. **The set $\mathcal{G}$ needs to be bounded.** Our set up also includes the case of risk based hedging, in which case one would rather write
\[
\pi^\mu(F) = \inf\left\{ m \in \mathbb{R} : \text{there is some } g \in \mathcal{G} \text{ such that } \rho^\mu(F - m + g \cdot G) \leq 0 \right\}.
\]
(This expression follows from additivity on the constants of $\rho^\mu$.)

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

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 $g \mapsto \rho^\mu(F + g \cdot G)$ 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 \( (\mu, G) \) non-degenerate in the sense that for every \( g \in \mathbb{R}^e \setminus \{0\} \) one has \( \mu(g \cdot G < 0) > 0 \).

**Proposition 6.1.** Let \( \rho: L^\infty \to \mathbb{R} \) be any law invariant risk measure, let \( F \) and each \( G_i \) be bounded, and let \( (\mu, G) \) be non-degenerate in the above 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 \( G \) needs to be bounded.

**Proof.** We show the negation, namely that if \( G \) is unbounded, convergence cannot be true. To that end, let \( (g^n) \) be a sequence in \( G \) witnessing that \( G \) 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, \( \mu(g^* \cdot G < 0) > 0 \), hence there is \( \varepsilon > 0 \) such that

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

By definition of \( \pi \) one has

\[
\pi \mu N(F) \leq \rho \mu N(F + g^n \cdot G)
\]

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

\[
F + g^n \cdot G \leq \sup_U F + \sup_U g^n \cdot G := a_n \quad \mu N\text{-a.s. on } \{ \mu N(U) = 1 \}
\]

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

\[
g^n \cdot G = |g^n| \left( g^* \cdot G + \left( \frac{g^n}{|g^n|} - g^* \right) \cdot G \right)
\]

\[
\leq |g^n| \left( -\varepsilon + C |\frac{g^n}{|g^n|} - g^*| \right) < -\frac{|g^n| \varepsilon}{2}
\]
on \( U \) for all large \( n \). By monotonicity of \( \rho \mu N \), this implies

\[
\rho \mu N(F + g^n \cdot G) \leq \rho \mu N(a_n) \leq a_n \to -\infty \quad \text{on } \{ \mu N(U) = 1 \}
\]
as \( n \to \infty \). Finally, as

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

for every \( N \geq 1 \), we conclude that \( \pi \mu N(F) = -\infty \) with positive probability. In particular \( E[|\pi \mu(F) - \pi \mu N(F)|] = \infty \) for every \( N \geq 1 \), which proves the claim. \( \square \)

### 6.2. The proof of Proposition 1.10

We only sketch the proof of Proposition 1.10 as it is very similar to that of Theorem 1.7 on the optimized certainty equivalents. The only difference is the absence of the component \( m \) (in the definition of OCE), which actually makes life even simpler. In particular, we have

\[
\mathcal{N} \left( \{ U(F + g \cdot G) : g \in G \}, \| \cdot \|_\infty, \varepsilon \right) \leq \left( \frac{C}{\varepsilon} \right)^e \wedge 1
\]

for all \( \varepsilon > 0 \) by Corollary 2.4. To conclude the proof, copy the arguments given for the proofs of Theorem 2.1 and Theorem 4.2.
6.3. Remaining proofs for Theorem [1.7] We finally provide the proof of Theorem 1.7 for the case that \( \rho \) 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 \( \mu \) replaced by \( \mu_N \).

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

\[
\pi^\mu(F) = \inf\{m \in [a, a] : J(m) \leq 1\}.
\]

The same is true if \( \mu \) is replaced by \( \mu_N \) 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 \( \varepsilon \)-optimal \( g \)), that is, \( J(m) = \int l(F + g \cdot G - m) d\mu \). The fundamental theorem of calculus then implies

\[
J(\tilde{m}) \leq \int l(F + g \cdot G - \tilde{m}) d\mu
= \int l(F + g \cdot G - m) - (\tilde{m} - m) \int l'(F + g \cdot G - m + t(\tilde{m} - m)) dt d\mu.
\]

The term inside the 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 claimed.

(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 \( \mathcal{G} \) it relative 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 \( J(\pi^\mu(F)) = 1 \). Similarly, \( \pi^\mu_N(F) \) is the unique number satisfying \( J(\pi^\mu_N(F)) = 1 \) and therefore

\[
|J(\pi^\mu_N(F)) - J(\pi^\mu(F))| = |J(\pi^\mu_N(F)) - J_N(\pi^\mu_N(F))|
\leq \sup_{|m| \leq a} |J(m) - J_N(m)| =: \Delta_N.
\]

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

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

we can use Corollary 2.4 and Dudley’s theorem as in the proof of Theorem 1.1 to obtain \( E[\Delta_N] \leq C/\sqrt{N} \) for all \( N \geq 1 \). Similarly, Corollary 2.4 and the arguments given for the proof of Theorem 1.1 imply that \( P[\Delta_N \geq \varepsilon] \leq C \exp(-cN\varepsilon^2) \) for all \( \varepsilon > 0 \), \( N \geq 1 \), where \( c > 0 \) is some (new) small constant. This completes the proof.
ACKNOWLEDGMENTS: The authors would like to thank Patrick Cheridito for helpful discussions.

The first authors is grateful for financial support through the Austrian Science Fund (FWF) under project P28661.

REFERENCES

[1] B. Acciaio and G. Svindland. Are law-invariant risk functions concave on distributions? Dependence Modeling, 1:54–64, 2013.
[2] P. E. Alexander J. McNeil and R. Frey. Quantitative Risk Management. Princeton University Press, 2015.
[3] P. Artzner, F. Delbaen, J. M. Eber, and D. Heath. Coherent measures of risk. Math. Finance, 9:203–228, 1999.
[4] D. Bartl, S. Drapeau, and L. Tangpi. Computational aspects of robust optimized certainty equivalents and option pricing. Math. Finance, 30:287–309, 2020.
[5] D. Belomestny and V. Krätschmer. Central limit theorems for law-invariant coherent risk measures. Journal of Applied Probability, 49(1):1–21, 2012.
[6] A. Ben-Tal and M. Taboure. An old-new concept of convex risk measures: The optimized certainty equivalent. Math. Finance, 17:449–476, 2007.
[7] A. Ben-Tal and M. Teboulle. Expected utility, penalty functions and duality in stochastic nonlinear programming. Management Science, 32:1445–1466, 1986.
[8] D. Bertsimas, V. Gupta, and N. Kallus. Robust sample average approximation. Math. Programming, 171:217–282, 2017.
[9] E. Beutner and H. Zähle. A modified functional delta method and its application to the estimation of risk functionals. Journal of Multivariate Analysis, 101:2452–2463, 2010.
[10] T. Bielecki, I. Cialenco, M. Pitera, and T. Schmidt. Fair estimation of capital risk allocations. Preprint, 2019.
[11] M. Cambou and D. Filipović. Model uncertainty and scenario aggregation. Math. Finance, 27(2):534–567, 2015.
[12] S. X. Chen. Nonparametric estimation of expected shortfall. Journal of Financial Econometrics, 6(1):87–107, 2008.
[13] P. Cheridito and T. Li. Dual characterization of properties of risk measures on orlicz hearts. Mathematics and Financial Economics, 2(1):29, 2008.
[14] P. Cheridito and T. Li. Risk measures on orlicz hearts. Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics, 19(2):189–214, 2009.
[15] R. Cont, R. Deguest, and G. Scandolo. Robustness and sensitivity analysis of risk measurement procedures. Quantitative finance, 10(6):593–606, 2010.
[16] F. Delbaen. Monetary Utility Functions. Osaka University Press, 2012.
[17] S. Eckstein, M. Kupper, and M. Pohl. Robust risk aggregation with neural networks. Preprint, 2018.
[18] P. Embrechts and M. Hofert. Statistics and quantitative risk management for banking and insurance. Annual Review of Statistics and Its Application, 1:493–514, 2014.
[19] P. M. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. Math. Programming, 171:115–166, 2018.
[20] Y. Feng, R. Rudd, C. Baker, Q. Mashalaba, M. Mavuso, and E. Schlögl. Quantifying the model risk inherent in the calibration and recalibration of option pricing models. Available at SSRN 3267775, 2018.
[21] H. Föllmer and A. Schied. Convex measures of risk and trading constraint. Finance and Stochastics, 6(4):429–447, 2002.
[22] H. Föllmer and A. Schied. Stochastic Finance. An Introduction in Discrete Time. de Gruyter Studies in Mathematics. Walter de Gruyter, Berlin, New York, 2 edition, 2004.
[23] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. Probab. Theory Relat. Fields, 162:707–738, 2015.
[24] E. Furman and R. Zitikis. Weighted risk capital allocations. Insurance: Mathematics and Economics, 43(269-269), 2008.
[25] N. Gao, D. Leung, C. Munari, and F. Xanthos. Fatou property, representation, and extensions of law-invariant risk measures on general Orlicz spaces. *Finance Stoch.*, 22:395–415, 2018.
[26] R. Gao and A. J. Kleywegt. Distributionally robust stochastic optimization with Wasserstein distance. *Preprint*, 2016.
[27] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer Science and Business Media, New York, 2004.
[28] L. J. Hong, Z. Hu, and G. Liu. Monte Carlo methods for value-at-risk and conditional value-at-risk: A review. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 24(4):22, 2014.
[29] X. Jin, M. C. Fu, and X. Xiong. Probabilistic error bounds for simulation quantile estimators. *Management Science*, 14(2):230–246, 2003.
[30] E. Jouini, W. Schachermayer, and N. Touzi. Law invariant risk measures have the fatou property. In S. Kusuoka and A. Yamazaki, editors, *Advances in Mathematical Economics*, volume 9 of *Advances in Mathematical Economics*, pages 49–71. Springer Japan, 2006.
[31] S. Kim, R. Pasupathy, and S. G. Henderson. A guide to sample average approximation. In M. Fu, editor, *Handbook of Simulation Optimization*, volume 216, pages 207–243. Springer, New York, 2015.
[32] A. J. Kleywegt, A. Shapiro, and T. H. de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM J. Optim.*, 12(2):479–502, 2001.
[33] V. Krätschmer, A. Schied, and H. Zähle. Qualitative and infinitesimal robustness of tail-dependent statistical functionals. *Journal of Multivariate Analysis*, 103:35–47, 2012.
[34] V. Krätschmer, A. Schied, and H. Zähle. Comparative and quantitative robustness for law-invariant risk measures. *Finance Stoch.*, 2014.
[35] S. Kusuoka. On law invariant coherent risk measures. In *Advances in mathematical economics*, pages 83–95. Springer, 2001.
[36] J. Obłój and J. Wiesel. Statistical estimation of superhedging prices. *arXiv preprint arXiv:1807.04211*, 2018.
[37] S. Pal. On capital requirements and optimal strategies to achieve acceptability. PhD thesis, Citeseer, 2006.
[38] S. Pal. Computing strategies for achieving acceptability: A monte carlo approach. *Stochastic processes and their applications*, 117(11):1587–1605, 2007.
[39] A. Pichler. Evaluations of risk measures for different probability measures. *SIAM Journal on Optimization*, 23(1):530–551, 2013.
[40] M. Pitera and T. Schmidt. Unbiased estimation of risk. *Journal of Banking and Finance*, 91:133–145, 2018.
[41] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.
[42] A. Shapiro, D. Dentcheva, and A. Ruszczyski. Lectures on stochastic programming: modeling and theory. *SIAM Journal on Optimization*, 2014.
[43] A. W. Van Der Vaart and J. A. Wellner. Weak convergence. In *Weak convergence and empirical processes*, pages 16–28. Springer, 1996.
[44] B. Verweij, S. Ahmed, A. J. Kleywegt, G. Nemhauser, and A. Shapiro. The sample average approximation method applied to stochastic routing problems: A computational study. *Computational Optimization and Applications*, 24(2):99–353, 2003.
[45] S. Weber. Distribution-invariant risk measures, Entropy, and large deviations. *J. Appl. Prob.*, 44:16–40, 2007.
[46] V. R. Young. Premium principles. In J. Teugels and B. Sundt, editors, *Encyclopedia of Actuarial Science*. Wiley, 2004.
[47] T. Zwinge and H. Holzmann. Asymptotics for the expected shortfall. *Preprint*, 2016.