To appear in *Quantitative Finance*, Vol. 00, No. 00, Month 20XX, 1–30

Quantitative Statistical Robustness for Tail-Dependent Law Invariant Risk Measures

WEI WANG⋆† , HUIFU XU∗‡ and TIEJUN MA§

(*Received 00 Month 20XX; in final form 00 Month 20XX*)

When estimating the risk of a financial position with empirical data or Monte Carlo simulations via a tail-dependent law invariant risk measure such as the Conditional Value-at-Risk (CVaR), it is important to ensure the robustness of the plug-in estimator particularly when the data contain noise. Krätschmer *et al.* (2014) propose a new framework to examine the qualitative robustness of such estimators for the tail-dependent law invariant risk measures on Orlicz spaces, which is a step further from an earlier work by Cont *et al.* (2010) for studying the robustness of risk measurement procedures. In this paper, we follow the stream of the research to propose a quantitative approach for verifying the statistical robustness of tail-dependent law invariant risk measures. A distinct feature of our approach is that we use the Fortet-Mourier metric to quantify variation of the true underlying probability measure in the analysis of the discrepancy between the law of the plug-in estimator of the risk measure based on true data and the one based on perturbed data, this approach enables us to derive an explicit error bound for the discrepancy when the risk functional is Lipschitz continuous over a class of admissible sets. Moreover, the newly introduced notion of Lipschitz continuity allows us to examine the degree of robustness for tail-dependent risk measures. Finally, we apply our quantitative approach to some well-known risk measures to illustrate our results and give an illustrative example about the tightness of the proposed error bound.

Keywords: Quantitative robustness, tail-dependent law invariant risk measures, Fortet-Mourier metric, admissible sets, index of quantitative robustness.

JEL Classification: D81

1. Introduction

One of the main purposes of quantitative modeling and measuring in finance is to quantify the loss of a financial position. Over the past two decades, various risk measures have been proposed for measuring such loss, see Föllmer and Weber (2015) for an overview. A risk measure is represented as a mapping assigning an extended real number (a measure of risk) to a random loss under an implicit assumption that the true loss probability distribution is known. In practice, however, such distribution is often unknown or it is prohibitively expensive to calculate the risk with the true distribution. Thus, in applications, evaluating the risk of a financial position often involves two steps: estimating the probability distribution from available empirical data or Monte Carlo

⋆Email:ww1e17@soton.ac.uk. The work of this author is partly carried out during his visit to the Department of Systems Engineering & Engineering Management, The Chinese University of Hong Kong supported by CUHK Direct Grant. His work is also supported by ESRC SCDTP with grant No. ES/R501025/1.
∗Corresponding author. Email: hfxu@se.cuhk.edu.hk. The work of this author is supported by a GAR grant No. 14500620.
§Email:tiejun.ma@soton.ac.uk.

sampling of the random loss and then plugging the estimated distribution into a risk measure (more precisely an associated risk functional) to quantify the financial loss. This kind of approach is based on the fact that risk measures are mostly law invariant, that is, they are determined only by the probability distributions of the underlying random variables. For the loss of a financial position, a measure of risk computed with the estimated distribution is known as a *plug-in estimate* for the risk measure (Zähle 2011).

Let X denote the random loss of a financial position on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and ρ be a law invariant risk measure. The plug-in estimate of $\rho(X)$ is given by $\varrho(\widehat{P})$, where $\widehat{P}$ is the empirical distribution based on available observations of X and ϱ is a *risk functional* defined by

$$\varrho(P) = \rho(X), \quad \text{if } X \text{ has law } P, \tag{1}$$

see, e.g. Beutner and Zähle (2010), Belomestny and Krätschmer (2012). Cont *et al.* (2010) first study the quality of the plug-in estimators of law invariant risk measures using Hampel's classical concept of qualitative robustness (Hampel 1971), that is, the plug-in estimator of a risk functional is said to be *qualitatively robust* if it is insensitive to the variation of sampling data. The research is important because perceived data (particularly empirical data) may contain noise and without such insensitivity, financial activities based on the risk measures may cause damages. For instance, when $\rho(X)$ is applied to allocate the risk capital for an insurance company, altering the capital allocation may be costly. According to Hampel's theorem, Cont *et al.* (2010) demonstrate that the qualitative robustness of a plug-in estimator is equivalent to the weak continuity of the risk functional and that value at risk (VaR) is qualitatively robust whereas conditional value at risk (CVaR) is not.

Krätschmer *et al.* (2012) argue that the use of Hampel's classical concept of qualitative robustness may be problematic because it requires the risk measure essentially to be insensitive with respect to the tail behaviour of the random variable and the recent financial crisis shows that a faulty estimate of tail behaviour can lead to a drastic underestimation of the risk. Consequently, they propose a refined notion of qualitative robustness that applies also to tail-dependent statistical functionals and that allows one to compare statistical functionals in regards to their degree of robustness. The new concept captures the trade-off between robustness and sensitivity and can be quantified by an index of qualitative robustness. Furthermore, under the new concept, Krätschmer *et al.* (2014) analyze the qualitative robustness of the law invariant convex risk measure on Orlicz spaces and show that CVaR and spectral risk measures are all qualitatively robust when the perturbation of the true probability distribution is restricted to a finer topological space. Alternative generalizations of Hampel's theorem can be found for strong mixing data (see, e.g., Zähle (2014) and Zähle (2015)) and for stochastic processes in various ways (see, e.g., Boente *et al.* (1987) and Strohriegl and Hable (2016)). For comprehensive study of statistical robustness, we refer readers to Huber and Ronchetti (2011), Zähle (2016), Hampel *et al.* (2011), Maronna *et al.* (2019), and references therein.

In this paper, we take a step further by deriving an error bound for the plug-in estimator of law invariant risk measures in terms of the variation of data. This is achieved by adopting different metrics to measure the discrepancy of the estimators and the variation of data. Specifically, we propose to use the Kantorovich metric for measuring the former and the Fortet-Mourier metric for the latter, as opposed to Lévy distance in Cont *et al.* (2010) or the Prokohorov and weighted Kolmogorov metrics in Krätschmer *et al.* (2012). We call this type of analysis quantitative because there is no explicit error bound in the existing qualitative robust analysis. Moreover, we introduce a new notion of the so-called *admissible sets*, which effectively restrict the scope of data variation but still large enough for our analysis. The new metrics enable us to establish an explicit relationship between the discrepancy of the laws of the plug-in estimators of law invariant risk measures based on the true data and perturbed data respectively and the discrepancy of the associated probability distributions of the data. The research is inspired by the recent work of Guo and Xu (2020) for preference robust optimization models under the Kantorovich metric and Wang and Xu (2020) for the preference robust spectral risk optimization models under the Fortet-Mourier metric.

The main contributions of the paper can be summarized as follows:

First, we introduce a notion of admissible set induced by a probability metric, which is a class of probability distributions whose discrepancy with the law of the Dirac measure at 0 is finite. The admissibility effectively restricts the range of data perturbation. Using the notion, we compare the admissibility under ϕ-topology and that under the Fortet-Mourier metric, and find that they coincide for some specific choice of ϕ.

Second, we introduce the concept of quantitative statistical robustness for a plug-in estimator. By using the Kantorovich metric to measure the variation of a plug-in estimator and the Fortet-Mourier metric to measure the variation of the true probability distributions, we prove that the plug-in estimator of a law invariant risk measure is quantitatively statistically robust over a class of admissible sets where the associated risk functional is Lipschitz continuous. Moreover, we find that the risk functionals associated with the general moment-type convex risk measures and the law invariant coherent risk measures are all Lipschitz continuous on some specified admissible sets, and consequently these risk measures are all quantitatively statistically robust.

Third, based on the concept of Lipschitz continuity for a risk functional on a class of admissible sets induced by the Fortet-Morier metric, we find that for the Lipschitz continuous risk functional, the parameter of the Fortet-Mourier metric allows us to compare the tail-dependent risk measures with regard to their degree of robustness, i.e., the index of quantitative statistical robustness.

Fourth, we apply the proposed approach to examine the quantitative statistical robustness of a range of well-known law invariant risk measures, including CVaR, optimized certainty equivalent, shortfall risk measure and conclude that under mild conditions, they are all quantitative robust. An illustrative example about the tightness of the proposed error bound is given.

The rest of the paper is organized as follows. Section 2 sets up the background of the problem for research. Section 3 introduces the concepts of Fortet-Mourier metric and admissible sets. Section 4 establishes the quantitative statistical robustness results. Section 5 showcases quantitative statistical robustness of some well-known risk measures. Section 6 gives an illustrative example about the tightness of the error bound. Finally Section 7 concludes. Some technical details are given in the appendices.

2. Problem statement

In this section, we discuss the background of statistical robustness in the context of law invariant risk measures. We begin by a brief review of law invariant convex risk measures and its estimation, and then move to explain the issues when the data may contain noise.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an atomless probability space, where Ω is a sample space with sigma algebra $\mathcal{F}$ and $\mathbb{P}$ is a probability measure. Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ be a financial loss and $F_X(x) := \mathbb{P}(X \leq x)$ the cumulative distribution function (CDF for short) of X. Let $P := \mathbb{P} \circ X^{-1}$ be the push-forward probability measure on $\mathbb{R}$ induced by X. Since $\mathbb{P}(X \leq x)$ coincides with $P((-\infty, x])$ ($P(x)$ for short), we also call P the distribution or the law of X interchangeably throughout the paper. For $t \in (0, 1)$, let $F_P^{-1}(t) := \inf\{x \in \mathbb{R} : F_P(x) \geq t\} = \{x \in \mathbb{R} : F_X(x) \geq t\}$ be the quantile function of X. For $p \geq 1$, let $L^p(\Omega, \mathcal{F}, \mathbb{P})$ (L^p for short) denote the space of random variables mapping from $(\Omega, \mathcal{F}, \mathbb{P})$ to $\mathbb{R}$ with finite p-th order moments, i.e., $\|X\|_p = (\int_\Omega |X(\omega)|^p \mathbb{P}(d\omega))^{1/p} < +\infty$. Let $\mathscr{P}(\mathbb{R})$ denote the set of all probability measures on $\mathbb{R}$ and $\mathscr{P}_p(\mathbb{R}) := \{P = \mathbb{P} \circ X^{-1} : X \in L^p\}$ for $p \geq 1$. We say that a map $\rho : L^1 \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ is a *convex risk measure*[1] (Föllmer and Schied 2002) if it satisfies the following properties:

(i) *Monotonicity*: $\rho(X) \leq \rho(Y)$ for $X, Y \in L^1$ with $X \leq Y$ $\mathbb{P}$-almost surely;
(ii) *Translation invariance*: $\rho(X + c) = \rho(X) + c$ for $X \in L^1$ and $c \in \mathbb{R}$;
(iii) *Convexity*: $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y)$ for $X, Y \in L^1$ and $\lambda \in [0, 1]$.

[1]We note that the canonical model space for law invariant convex risk measure is L^1 (Filipović and Svindland 2012).

Moreover, if ρ satisfies *positive homogeneity*, i.e., for any $\alpha \geq 0$, $\rho(\alpha X) = \alpha\rho(X)$, then ρ is a *coherent risk measure*, see Artzner *et al.* (1999), Föllmer and Schied (2002) for the original definitions of these concepts. Furthermore, a risk measure ρ is said to be *law invariant* if $\rho(X) = \rho(Y)$ for X and Y having the same law.

As discussed in Cont *et al.* (2010), Krätschmer *et al.* (2012), it is a widely-accepted procedure to estimate the risk of a financial loss by means of a Monte Carlo method or from a set of available observations. Such a procedure is particularly sensible when ρ is law invariant. The following proposition states that the law invariance of a risk measure ρ is equivalent to the existence of a *risk functional* ϱ in (1).

PROPOSITION 2.1 *If $\rho : L^1 \to \overline{\mathbb{R}}$ is a law invariant risk measure, then there exists a unique* risk functional $\varrho : \mathscr{P}_1(\mathbb{R}) \to \overline{\mathbb{R}}$ *associated with ρ such that for any $X \in L^1$,*

$$\varrho(P) = \rho(F_P^{-1}(U)) = \rho(X), \tag{2}$$

where $P = \mathbb{P} \circ X^{-1}$, F_P^{-1} is the quantile function of X and U is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ that is uniformly distributed over $(0, 1)$.

The result is well-known, see, e.g., Krätschmer *et al.* (2014, 2017), Claus *et al.* (2017), Delage *et al.* (2019). The risk functional $\varrho(P)$ with the law $P = \mathbb{P} \circ X^{-1}$ can be used in a natural way to construct an estimator for the risk $\rho(X)$ of $X \in L^1$. All one needs to do is to take an estimate P_N of P based on the available observations of X and then plug this estimator into the risk functional ϱ to obtain the desired estimator of $\rho(X)$, i.e.,

$$\widehat{\varrho}_N(\xi^1, \xi^2, \ldots, \xi^N) := \varrho(P_N), \tag{3}$$

where in this paper, P_N can be seen as the empirical distribution of an independent and identically distributed (i.i.d., for short) sequence $\xi^1, \xi^2, \ldots, \xi^N$ of historical observations or Monte Carlo simulations, i.e.,

$$P_N(x) := \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\{\xi^i \leq x\}}, \quad x \in \mathbb{R}. \tag{4}$$

Here and later on $\mathbf{1}_A$ denotes the indicator function of event A and by slightly abusing the notation, we use ξ^k to denote both a realization and a random variable depending on the context. In practice, P_N can be a fairly general estimate, for instance, P_N can be a smoothed empirical distribution based on uncensored data or empirical distribution based on censored data (see, e.g., Zähle (2011)) or empirical distribution based on identically distributed dependent data (see, e.g., Zähle (2015)). In this paper, our focus is the i.i.d. case.

We can see that $\widehat{\varrho}_N$ is a mapping from $\mathbb{R}^{\otimes N}$ to $\mathbb{R}$. Figure 1 illustrates the relationship between the risk functionals, their estimators and the related spaces.

In practice, the samples obtained from empirical data may contain noise. In that case, we might regard the samples as generated by a perturbed random variable Y with law Q, that is, $Q = \mathbb{P} \circ Y^{-1}$. Let $\tilde{\xi}^1, \cdots, \tilde{\xi}^N$ be i.i.d samples from Y. Then the practical empirical distribution function for estimating the law of X is

$$Q_N(x) := \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\{\tilde{\xi}^i \leq x\}}, \quad x \in \mathbb{R}, \tag{5}$$

and the practical estimator is $\widetilde{\varrho}_N = \widehat{\varrho}_N(\tilde{\xi}^1, \cdots, \tilde{\xi}^N) := \varrho(Q_N)$ whereas $\widehat{\varrho}_N$ is a statistical estimator with noise being detached. Since we are unable to obtain the latter, we tend to use the former as

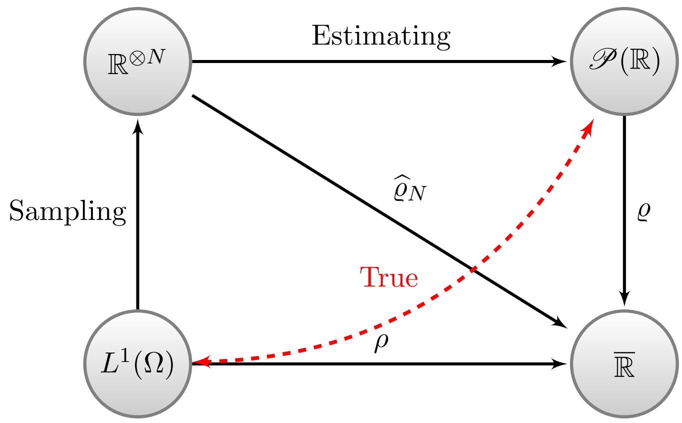


Figure 1. The diagram for risk functionals, their estimators and related spaces

a statistical estimator of $\rho(X)$ and this works only if the two estimators are sufficiently close.

To quantify the closeness, we may look into the discrepancy between the laws of the two estimators under some metric dl, i.e.,

$$\mathsf{dl}(\text{law}\{\varrho(P_N)\}, \text{law}\{\varrho(Q_N)\}) = \mathsf{dl}\left(P^{\otimes N} \circ \widehat{\varrho}_N^{\,-1}, Q^{\otimes N} \circ \widehat{\varrho}_N^{\,-1}\right), \tag{6}$$

where $P^{\otimes N}$ and $Q^{\otimes N}$ denote the probability measures on measurable space $\left(\mathbb{R}^{\otimes N}, \mathcal{B}(\mathbb{R})^{\otimes N}\right)$ with marginals P and Q on each $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ respectively, $\mathcal{B}(\mathbb{R})$ denotes the corresponding Borel sigma algebra on $\mathbb{R}$. Since neither P nor Q is known, we want the discrepancy to be uniformly small for all P and Q over a subset of admissible sets on $\mathscr{P}(\mathbb{R})$ so long as Q is sufficiently close to P under some metric dl'. The uniformity may be interpreted as robustness. *Qualitative robustness* refers to the case that the relationship between $\mathsf{dl}\left(P^{\otimes N} \circ \widehat{\varrho}_N^{\,-1}, Q^{\otimes N} \circ \widehat{\varrho}_N^{\,-1}\right)$ and $\mathsf{dl}'(P, Q)$ is implicit whereas *quantitative robustness* refers to the case that the relationship is explicit, i.e., a function of the latter can be used to bound the former, and this is what we aim to achieve in this paper because qualitative robustness have been well investigated, for instance, in Cont *et al.* (2010), Krätschmer *et al.* (2014, 2012).

3. ζ-metrics and admissible sets

There are two essential elements in investigating both the qualitative and quantitative statistical robustness of a risk functional: One is the specific choice of probability metrics but not just the topologies generated by them (see, e.g., Huber and Ronchetti (2011), Cont *et al.* (2010), Krätschmer *et al.* (2012)), to quantify the change of the law P and to estimate the discrepancy between the laws of two estimators, i.e., (6); the other is the determination of the subset $\mathcal{M}$ of admissible sets in $\mathscr{P}(\mathbb{R})$ (see, e.g., Krätschmer *et al.* (2012), Zähle (2015), Krätschmer *et al.* (2017)), containing all empirical distributions: $\mathcal{M}_{1,\text{emp}} \subset \mathcal{M}$, to restrict the perturbation of the law P. For instance, the subset $\mathcal{M}$ may be specified via some generalized moment conditions, which are interesting in econometric or financial applications (Cont *et al.* 2010).

To introduce these two essential elements thoroughly, some preliminary notions and results in probability theory and statistics such as ϕ-weak topology are required. We first give a sketch of them to prepare our discussions in the follow-up sections. Let $\phi : \mathbb{R} \to [0, \infty)$ be a continuous function and

$$\mathcal{M}_1^{\phi} := \left\{ P' \in \mathscr{P}(\mathbb{R}) : \int_{\mathbb{R}} \phi(t) P'(dt) < \infty \right\},$$

which defines a subset of probability measures in $\mathscr{P}(\mathbb{R})$ having finite generalized moment condition of ϕ and we may call it the admissible set of ϕ-weak topology for the consistency of the paper. In the particular case when $\phi(\cdot) := |\cdot|^p$ and p is a positive number, we write $\mathcal{M}_1^p$ for $\mathcal{M}_1^{|\cdot|^p}$. Note that $\mathcal{M}_1^{p_2} \subset \mathcal{M}_1^{p_1}$ for any positive numbers p_1, p_2 with $p_1 < p_2$ due to Hölder inequality.

Definition 1 (ϕ-weak topology) Let $\phi : \mathbb{R} \to [0, \infty)$ be a *gauge function*, that is, ϕ is continuous and $\phi \geq 1$ holds outside a compact set. Define $\mathcal{C}_1^\phi$ the linear space of all continuous functions $h : \mathbb{R} \to \mathbb{R}$ for which there exists a positive constant c such that $|h(t)| \leq c(\phi(t) + 1)$, $\forall t \in \mathbb{R}$. The *ϕ-weak topology*, denoted by τ_ϕ, is the coarsest topology on $\mathcal{M}_1^\phi$ for which the mapping $g_h : \mathcal{M}_1^\phi \to \mathbb{R}$ defined by

$$g_h(P') := \int_{\mathbb{R}} h(t)P'(dt), \ \forall h \in \mathcal{C}_1^\phi,$$

is continuous. A sequence $\{P_l\} \subset \mathcal{M}_1^\phi$ is said to *converge ϕ-weakly* to $P \in \mathcal{M}_1^\phi$ written $P_l \xrightarrow{\phi} P$ if it converges w.r.t. τ_ϕ.

Clearly, ϕ-weak topology is finer than the weak topology, and the two topologies coincide if and only if ϕ is bounded. It is well-known (see, e.g., Krätschmer *et al.* 2012, Lemma 3.4) that ϕ-weak convergence is equivalent to weak convergence, denoted by $P_l \xrightarrow{w} P$, together with $\int_{\mathbb{R}} \phi(t)P_l(dt) \to \int_{\mathbb{R}} \phi(t)P(dt)$. Moreover, it follows by Krätschmer *et al.* (2012, 2014) that the ϕ-weak topology on $\mathcal{M}_1^\phi$ is generated by the metric $\mathsf{dl}_\phi : \mathcal{M}_1^\phi \times \mathcal{M}_1^\phi \to \mathbb{R}$ defined by

$$\mathsf{dl}_\phi(P, Q) := \mathsf{dl}_{\mathrm{Prok}}(P, Q) + \left| \int_{\mathbb{R}} \phi(t)P(dt) - \int_{\mathbb{R}} \phi(t)Q(dt) \right|, \tag{7}$$

for $P, Q \in \mathcal{M}_1^\phi$, where $\mathsf{dl}_{\mathrm{Prok}} : \mathscr{P}(\mathbb{R}) \times \mathscr{P}(\mathbb{R}) \to \mathbb{R}_+$ is the *Prokhorov metric* defined by

$$\mathsf{dl}_{\mathrm{Prok}}(P, Q) := \inf\{\epsilon > 0 : P(A) \leq Q(A^\epsilon) + \epsilon, \forall A \in \mathcal{B}(\mathbb{R})\}, \tag{8}$$

where $A^\epsilon := A + B_\epsilon(0)$ denotes the Minkowski sum of A and the open ball centred at 0 on $\mathbb{R}$ and $\mathcal{B}(\mathbb{R})$ is the corresponding Borel sigma algebra on $\mathbb{R}$. Note that the Prokhorov metric metrizes the weak topology on $\mathbb{R}$, see, e.g., Gibbs and Su (2002).

3.1. *ζ-metrics*

Instead of exploiting the widely-used probability metrics such as the Lévy metric, the Prokhorov metric and the weighted Kolmogorov metric in the literature of qualitative robustness (Cont *et al.* 2010, Krätschmer *et al.* 2012), we will switch to the so-called ζ-metrics to establish the quantitative statistical robustness framework for a risk functional. Specifically, we will use Kantorovich metric and Fortet-Mourier metrics. We begin with a formal definition of ζ-metrics and then clarify the relationships between ζ-metrics and those used in Mizera (2010), Cont *et al.* (2010), Krätschmer *et al.* (2012).

Definition 2 Let $P, Q \in \mathscr{P}(\mathbb{R}^k)$ be the probability measures on $\mathbb{R}^k$ and $\mathcal{F}$ be a class of measurable functions from $\mathbb{R}^k$ to $\mathbb{R}$. The ζ-metrics between P and Q is defined by

$$\mathsf{dl}_{\mathcal{F}}(P, Q) := \sup_{\psi \in \mathcal{F}} \left| \int_{\mathbb{R}^k} \psi(\boldsymbol{\xi})P(d\boldsymbol{\xi}) - \int_{\mathbb{R}^k} \psi(\boldsymbol{\xi})Q(d\boldsymbol{\xi}) \right|. \tag{9}$$

From the definition, we can see that $\mathsf{dl}_{\mathcal{F}}(P,Q)$ is the maximum difference of the expected values of the class of measurable functions $\mathcal{F}$ with respect to P and Q. ζ-metrics are widely used in the stability analysis of stochastic programming, see Römisch (2003) for an overview. The specific ζ-metrics that we consider in this paper are the Kantorovich metric and the Fortet-Mourier metric. The next definition gives a precise description of the two notions.

Definition 3 (Fortet-Mourier metric) Let

$$\mathcal{F}_p(\mathbb{R}^k) := \left\{\psi : \mathbb{R}^k \to \mathbb{R} : |\psi(\boldsymbol{\xi}) - \psi(\tilde{\boldsymbol{\xi}})| \le c_p(\boldsymbol{\xi}, \tilde{\boldsymbol{\xi}})\|\boldsymbol{\xi} - \tilde{\boldsymbol{\xi}}\|, \forall \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}} \in \mathbb{R}^k\right\}, \tag{10}$$

where $\|\cdot\|$ denotes some norm on $\mathbb{R}^k$ and $c_p(\boldsymbol{\xi}, \tilde{\boldsymbol{\xi}}) := \max\{1, \|\boldsymbol{\xi}\|, \|\tilde{\boldsymbol{\xi}}\|\}^{p-1}$ for all $\boldsymbol{\xi}, \tilde{\boldsymbol{\xi}} \in \mathbb{R}^k$ and $p \ge 1$ describes the growth of the local Lipschitz constants. The p-th order *Fortet-Mourier metric* for $P, Q \in \mathscr{P}(\mathbb{R}^k)$ is defined by

$$\mathsf{dl}_{FM,p}(P,Q) := \sup_{\psi \in \mathcal{F}_p(\mathbb{R}^k)} \left| \int_{\mathbb{R}^k} \psi(\boldsymbol{\xi}) P(d\boldsymbol{\xi}) - \int_{\mathbb{R}^k} \psi(\boldsymbol{\xi}) Q(d\boldsymbol{\xi}) \right|. \tag{11}$$

In the case when $p = 1$, it is known as the *Kantorovich metric* denoted by $\mathsf{dl}_K(P,Q)$.

From the definition, we can see that for any positive numbers $p \ge p' \ge 1$,

$$\mathsf{dl}_{FM,p}(P,Q) \ge \mathsf{dl}_{FM,p'}(P,Q) \ge \mathsf{dl}_K(P,Q), \tag{12}$$

which means $\mathsf{dl}_{FM,p}(P,Q)$ becomes tighter as p increases and they are all tighter than $\mathsf{dl}_K(P,Q)$. Moreover, the Fortet–Mourier metric metricizes weak convergence on sets of probability measures possessing uniformly a p-th moment (Pflug and Pichler 2011, p. 350). On $\mathbb{R}$, the Fortet–Mourier metric may be equivalently written as

$$\mathsf{dl}_{FM,p}(P,Q) = \int_{\mathbb{R}} \max\{1, |x|^{p-1}\}|P(x) - Q(x)|dx, \text{ for } P, Q \in \mathscr{P}(\mathbb{R}), \tag{13}$$

see, e.g., (Rachev 1991, p. 93). In this paper, we will mainly focus on the case with $k = 1$ because most risk measures are defined on the set of one-dimensional random variables.

In the next example, we illustrate the relationship between the existing probability metrics used in qualitative statistical robustness and the ζ-metrics.

Example 3.1 *We list some well-known probability metrics that are used in statistical robustness.*

(i) The Kantorovich (or Wasserstein) metric. *Let $\mathcal{F}_1$ be the set of all Lipschitz continuous functions with modulus being bounded by* 1. *The Kantorovich metric is defined as*

$$\mathsf{dl}_K(P,Q) := \int_{-\infty}^{+\infty} |P(x) - Q(x)|dx = \mathsf{dl}_{\mathcal{F}_1}(P,Q). \tag{14}$$

Moreover, $\mathsf{dl}_{\mathrm{Prok}}(P,Q)^2 \le \mathsf{dl}_K(P,Q)$*, see, e.g., Gibbs and Su (2002, Theorem 2).*

(ii) The weighted Kolmogorov metric *(Krätschmer et al. 2012). Let ϕ be a u*-shaped function, *i.e., a continuous function $\phi : \mathbb{R} \to [1, +\infty)$ which is non-increasing on $(-\infty, 0)$ and non-decreasing on $(0, +\infty)$. The weighted Kolmogorov metric is defined as*

$$\mathsf{dl}_{(\phi)}(P,Q) := \sup_{x \in \mathbb{R}} |P(x) - Q(x)|\phi(x) \le \mathsf{dl}_{\mathcal{F}}(P,Q),$$

where $\mathcal{F}$ is the set of all functions bounded by ϕ. In the case that $\mathcal{F}$ is the set of all indicator functions $\mathbf{1}_B$, where $B := \{(-\infty, \xi], \xi \in \mathbb{R}\}$, then $\mathsf{dl}_{\mathcal{F}}(P,Q) = \mathsf{dl}_{(1)}(P,Q)$, $\mathsf{dl}_{(\phi)}$ reduces to the

Kolmogorov metric. *If $\mathcal{F}$ is the set of all weighted indicator functions with weighting ϕ, then we obtain* $\mathsf{dl}_{\mathcal{F}}(P,Q) = \mathsf{dl}_{(\phi)}(P,Q)$.

(iii) The Lévy distance *(Rachev 1991). Let $\mathcal{F}$ be the set of all measurable functions bounded by 1. The Lévy distance is defined as*

$$\mathsf{dl}_{\text{Lévy}}(P,Q) := \inf\{\epsilon > 0 : Q(x-\epsilon) - \epsilon \le P(x) \le Q(x+\epsilon) + \epsilon,\ \forall\, x \in \mathbb{R}\} \le \mathsf{dl}_{\mathcal{F}}(P,Q).$$

Moreover, $\mathsf{dl}_{\text{Lévy}}(P,Q) \le \mathsf{dl}_{\text{Prok}}(P,Q)$ *and* $\mathsf{dl}_{\text{Lévy}}(P,Q) \le \mathsf{dl}_{(\phi)}(P,Q)$ *for any $\phi \ge 1$, see, e.g., Gibbs and Su (2002).*

(iv) The Prokhorov metric *(Krätschmer et al. 2012). Let $\mathcal{F}$ be the set of all measurable functions bounded by 1. Then from Gibbs and Su (2002), we have*

$$\mathsf{dl}_{\text{Prok}}(P,Q) := \inf\{\epsilon > 0 : P(A) \le Q(A^\epsilon) + \epsilon, \forall A \in \mathcal{B}(\mathbb{R})\} \le \frac{1}{2}\mathsf{dl}_{\mathcal{F}}(P,Q),$$

where $A^\epsilon := \{x \in \mathbb{R} : \inf_{y\in A} |x-y| \le \epsilon\}$. *Moreover,* $\mathsf{dl}_{\text{Prok}}(P,Q)^2 \le \mathsf{dl}_K(P,Q)$.

3.2. *Admissible sets*

We now turn to discuss another important component in statistical robust analysis, that is, the subset $\mathcal{M}$ of admissible sets in $\mathscr{P}(\mathbb{R})$ which describes the scope of the perturbation of the law P by a metric. This can be motivated by ensuring the finiteness of $\mathsf{dl}(P,Q)$. To this effect, we formally introduce the concept of admissible sets induced by probability metrics.

Definition 4 (Admissible sets induced by probability metrics) Let dl be a probability metric on $\mathscr{P}(\mathbb{R})$. The admissible sets induced by dl are defined as

$$\mathcal{P}_{\mathsf{dl}}(\mathbb{R}) := \{P \in \mathscr{P}(\mathbb{R}) : \mathsf{dl}(P,\delta_0) < +\infty\}, \tag{15}$$

where δ_0 denotes the Dirac measure at 0.

Let $\mathcal{P}_p(\mathbb{R})$ denote the admissible sets induced by the Fortet-Mourier metrics with parameter p on $\mathscr{P}(\mathbb{R})$. By Definition 4, we have

$$\begin{aligned}\mathcal{P}_p(\mathbb{R}) &:= \{P \in \mathscr{P}(\mathbb{R}) : \mathsf{dl}_{FM,p}(P,\delta_0) < +\infty\} \\ &= \left\{P \in \mathscr{P}(\mathbb{R}) : \sup_{\psi\in\mathcal{F}_p(\mathbb{R})} \left|\int_{\mathbb{R}} \psi(\xi)P(d\xi) - \int_{\mathbb{R}} \psi(\xi)\delta_0(d\xi)\right| < +\infty\right\} \\ &= \left\{P \in \mathscr{P}(\mathbb{R}) : \sup_{\psi\in\mathcal{F}_p(\mathbb{R})} \left|\int_{\mathbb{R}} \psi(\xi)P(d\xi) - \psi(0)\right| < +\infty\right\}. \end{aligned} \tag{16}$$

By triangle inequality, this ensures $\mathsf{dl}_{FM,p}(P,Q) < +\infty$ for any $P, Q \in \mathcal{P}_p(\mathbb{R})$.

In the following example, we compare the admissible sets induced by different probability metrics.

Example 3.2 (Admissible sets induced by probability metrics) *We reconsider the admissible sets induced by probability metrics defined in Example 3.1.*

(i) The admissible sets induced by the Kantorovich (or Wasserstein) metric *are defined as*

$$\mathcal{P}_K(\mathbb{R}) := \{P \in \mathscr{P}(\mathbb{R}) : \mathsf{dl}_K(P,\delta_0) < +\infty\} = \mathcal{M}_1^1,$$

see details in Appendix A.

(ii) The admissible sets induced by the weighted Kolmogorov metric *are defined as*

$$\begin{aligned}\mathcal{P}_{(\phi)}(\mathbb{R}) &:= \{P \in \mathscr{P}(\mathbb{R}) : \mathsf{dl}_{(\phi)}(P, \delta_0) < +\infty\} \\ &= \left\{P \in \mathscr{P}(\mathbb{R}) : \sup_{x \le 0} |P(x)\phi(x)| + \sup_{x>0} |(1 - P(x))\phi(x)| < +\infty\right\},\end{aligned}$$

which coincides with the set $\mathcal{M}_1^{(\phi)}(\mathbb{R})$ defined in Krätschmer et al. (2012, subsection 3.2). If ϕ is bounded on $\mathbb{R}$, then it is straight that $\mathcal{P}_{(\phi)}(\mathbb{R}) = \mathscr{P}(\mathbb{R})$. In the case when ϕ is unbounded on $\mathbb{R}$, then

$$\mathcal{M}_1^{\phi} \subset \mathcal{P}_{(\phi)}(\mathbb{R}) \subset \bigcap_{\epsilon>0} \mathcal{M}_1^{\phi^{1-\epsilon}}. \tag{17}$$

The inclusions in (17) are strict, see details in Appendix A.

(iii) The admissible sets induced by the Lévy distance *are defined as*

$$\mathcal{P}_{\text{Lévy}}(\mathbb{R}) := \{P \in \mathscr{P}(\mathbb{R}) : \mathsf{dl}_{\text{Lévy}}(P, \delta_0) < +\infty\} = \mathscr{P}(\mathbb{R}).$$

Since $\mathsf{dl}_{\text{Lévy}} \le 1$, then the admissible sets coincide with $\mathscr{P}(\mathbb{R})$.

(iv) The admissible sets induced by the Prokhorov metric *are defined as*

$$\mathcal{P}_{\text{Prok}}(\mathbb{R}) := \{P \in \mathscr{P}(\mathbb{R}) : \mathsf{dl}_{\text{Prok}}(P, \delta_0) < +\infty\} = \mathscr{P}(\mathbb{R}).$$

Since $\mathsf{dl}_{\text{Prok}} \le 1$, then the admissible sets coincides with $\mathscr{P}(\mathbb{R})$.

3.3. *Relationship with ϕ-weak topology*

Since ϕ-weak topology has been widely used for qualitative robust analysis in the literature whereas we use the topology induced by the Fortet-Mourier metrics for quantitative robust analysis, it would be helpful to look into potential connections of the two apparently completely different notions. In the next proposition, we look into such connection.

PROPOSITION 3.3 *Let $p \ge 1$ be fixed and*

$$\phi_p(t) := \begin{cases} |t|, & \text{for } |t| \le 1, \\ |t|^p, & \text{otherwise.} \end{cases}$$

The following assertions hold.

- *(i)* $\mathcal{P}_p(\mathbb{R}) = \mathcal{M}_1^{\phi_p} (= \mathcal{M}_1^p)$.
- *(ii)* $\mathsf{dl}_{\phi_p}(P, Q) \le \sqrt{\mathsf{dl}_{FM,p}(P, Q)} + p\mathsf{dl}_{FM,p}(P, Q),\ \forall P, Q \in \mathcal{P}_p(\mathbb{R})$.
- *(iii) (Rachev 1991, Theorem 6.3.1) The Fortet-Mourier metric $\mathsf{dl}_{FM,p}$ generates the $\|\cdot\|^p$-weak topology.*

Part (i) of the proposition says that the admissible set $\mathcal{P}_p(\mathbb{R})$ coincides with the set of laws on $\mathbb{R}$ having the finite generalized moment condition of ϕ_p. Part (ii) indicates that $\mathsf{dl}_{FM,p}$ is tighter than dl_{ϕ_p} defined in (7). Part (iii) indicates that despite Fortet-Mourier metric $\mathsf{dl}_{FM,p}$ and dl_{ϕ_p} are different metrics, they generate the same topology. For completeness, we include a proof in the Appendix C.

To conclude this section, we remark that the subset $\mathcal{M}$ to be used in the definition of qualitative robust analysis will be confined to the subset of admissible sets when we adopt the Fortet-Mourier metric for quantitative robust analysis in the next section.

4. Statistical robustness

We are now ready to return our discussions to the robustness of statistical estimators of law invariant risk measures that are outlined in Section 2.

4.1. *Qualitative statistical robustness*

To position our research properly, we begin by a brief overview of the existing results about the qualitative statistical robustness.

Definition 5 (Qualitative $\mathcal{C}$-Robustness (Cont *et al.* 2010, Krätschmer *et al.* 2014)) Let $\mathcal{C}$ be a subset of $\mathscr{P}(\mathbb{R})$ and $P \in \mathcal{C}$. A sequence of statistical estimators $\{\widehat{\varrho}_N\}_{N\in\mathbb{N}}$ is said to be *qualitatively $\mathcal{C}$-robust* at P w.r.t. $(\mathsf{dl}, \mathsf{dl}')$ if for every $\epsilon > 0$ there exist $\delta > 0$ and $N_0 \in \mathbb{N}$ such that for all $Q \in \mathcal{C}$ and $N \geq N_0$

$$\mathsf{dl}(P, Q) \leq \delta \implies \mathsf{dl}'(P^{\otimes\mathbb{N}} \circ \widehat{\varrho}_N^{-1}, Q^{\otimes\mathbb{N}} \circ \widehat{\varrho}_N^{-1}) \leq \epsilon. \tag{18}$$

If, in addition, $\{\widehat{\varrho}_N\}_{N\in\mathbb{N}}$ arises as in (3) from a risk functional ϱ, then ϱ is called *qualitatively $\mathcal{C}$-robust* at P w.r.t. $(\mathsf{dl}, \mathsf{dl}')$.

The definition above captures two versions of qualitative statistical robustness proposed by Cont *et al.* (2010) for i.i.d. observations on $\mathbb{R}$ with dl and dl' being Lévy distance and Krätschmer *et al.* (2012) for i.i.d. observations on $\mathbb{R}$ with $\mathsf{dl} = \mathsf{dl}_{(\phi)}$ and $\mathsf{dl}' = \mathsf{dl}_{\text{Prok}}$ respectively. Zähle (2016, Definition 1.1) extends the notion to "qualitative robustness" without any restriction on N, "asymptotic qualitative robustness" for all $N \geq N_0$ and "finite-sample robust" for $N \leq N_0$. Since $\mathsf{dl}_{\text{Prok}}$ is tighter than $\mathsf{dl}_{\text{Lévy}}$ (see Example 3.1(iii)), it means Krätschmer *et al.* (2012) examines the discrepancy of the laws with a tighter metric. On the other hand, from the definition of $\mathsf{dl}_{(\phi)}$, we can see that it is also tighter than $\mathsf{dl}_{\text{Lévy}}$ (see Example 3.1(iii)) and allows one to capture the difference of distributions at the tail, it means the robust analysis in Krätschmer *et al.* (2012) is restricted to a smaller class of probability distributions when Q is perturbed from P. This explains why CVaR is robust under the criterion of the latter but not the former.

A key result that Krätschmer *et al.* (2012) establish is the Hampel's theorem which states the equivalence between qualitative statistical robustness and stability/continuity of a risk functional (with respect to perturbation of the probability distribution) under the Uniform Glivenko-Cantelli (UGC) property of empirical distributions over a specified set.

Definition 6 ($\mathcal{C}$-Continuity (Krätschmer *et al.* 2012)) Let $P \in \mathscr{P}(\mathbb{R})$ and $\mathcal{C}$ be a subset of $\mathscr{P}(\mathbb{R})$. Then ϱ is called *$\mathcal{C}$-continuous* at P w.r.t. $(\mathsf{dl}, |\cdot|)$ if for every $\epsilon > 0$, there exists $\delta > 0$ such that for all $Q \in \mathcal{C}$

$$\mathsf{dl}(P, Q) \leq \delta \implies |\varrho(P) - \varrho(Q)| \leq \epsilon.$$

Definition 7 (UGC Property (Krätschmer *et al.* 2012)) Let $\mathcal{C}$ be a subset of $\mathscr{P}(\mathbb{R})$. Then we say that the metric space $(\mathcal{C}, \mathsf{dl})$ has *the UGC property* if for every $\epsilon > 0$ and $\delta > 0$, there exists $N_0 \in \mathbb{N}$ such that for all $P \in \mathcal{C}$ and $N \geq N_0$

$$P^{\otimes N}\left[(\xi^1, \ldots, \xi^N) \in \mathbb{R}^{\otimes N} : \mathsf{dl}(P, P_N) \geq \delta\}\right] \leq \epsilon.$$

The UGC property means that convergence in probability of the empirical probability measure to the true marginal distribution uniformly in $\mathcal{C}$ on $\mathscr{P}(\mathbb{R})$. Examples for metrics spaces $(\mathcal{C}, \mathsf{dl})$ having the UGC property can be found in (Krätschmer *et al.* 2012, Section 3). In particular, it is shown that there exists a subset of the admissible sets induced by the weigthed Kolmogorov metric enjoys the UGC property, see (Krätschmer *et al.* 2012, Theorem 3.1).

THEOREM 4.1 (Hampel's Theorem (Krätschmer *et al.* 2012)) *Let $\mathcal{C}$ be a subset of $\mathscr{P}(\mathbb{R})$ and $P \in \mathcal{C}$. Assume that $(\mathcal{C}, \mathsf{dl})$ has the UGC property and $\mathcal{M}_{1,\text{emp}} \subset \mathcal{C}$. If the mapping ϱ is $\mathcal{M}_{1,\text{emp}}$-continuous at P w.r.t. $(\mathsf{dl}, |\cdot|)$, then the sequence $\{\widehat{\varrho}_N\}_{N\in\mathbb{N}}$ is qualitatively $\mathcal{C}$-robust at P w.r.t. $(\mathsf{dl}, \mathsf{dl}_{\text{Prok}})$.*

Krätschmer *et al.* (2017) take a step further to investigate how the local robustness set $\mathcal{C}$ can be effectively identified particularly for maximum likelihood estimators. To this effect, they introduce a new concept "w-set" as follows.

DEFINITION 8 (w-set (Krätschmer *et al.* 2017)) Let $\{\psi_k\}_{k\in\mathbb{N}}$ be any sequence of gauge functions and $M \subseteq \mathcal{M}_1^{\{\psi_k\}}$. Then M is said to be a *w-set* in $\mathcal{M}_1^{\{\psi_k\}}$ if $\mathcal{O}_{\{\psi_k\}} \cap M = \mathcal{O}_w \cap M$, where $\mathcal{O}_w$ is the weak topology on $\mathscr{P}(\mathbb{R})$ and $\mathcal{O}_{\{\psi_k\}}$ is the $\{\psi_k\}$-weak topology on $\mathcal{M}_1^{\{\psi_k\}}$.

The thrust of the concept is that on the w-sets one can work with metrics generating the classical weak topology. Using the concept of w-set, Krätschmer *et al.* (2017) extend Theorem 4.1 to the following.

THEOREM 4.2 (Krätschmer *et al.* (2017)) *If for any w-set $M \subseteq \mathcal{M} \subseteq \mathcal{M}_1^{\{\psi_k\}}$ the mapping ϱ is continuous at every $P \in M$ for the relative $\{\psi_k\}$-weak topology $\mathcal{O}_{\{\psi_k\}} \cap \mathcal{M}$, then the sequence $\{\widehat{\varrho}_N\}_{N\in\mathbb{N}}$ is qualitatively robust (in the sense of Zähle) on M.*

4.2. *Quantitative statistical robustness*

We now move on to discuss our central topic, quantitative statistical robustness for the plug-in estimators of law invariant risk measures. Intuitively speaking, quantitative statistical robustness of a risk functional ϱ means that for any two admissible laws P and Q on a specified admissible set on $\mathcal{P}(\mathbb{R})$, the distance between the laws of their plug-in estimators $\varrho(P_N)$ and $\varrho(Q_N)$ is bounded by the distance between P and Q.

DEFINITION 9 (Quantitative statistical h-robustness) Let $\mathsf{dl}, \mathsf{dl}'$ be probability metrics on $\mathscr{P}(\mathbb{R})$ and $\mathcal{M} \subset \mathcal{P}_{\mathsf{dl}'}(\mathbb{R})$ denote a subset of admissible laws on $\mathbb{R}$. A sequence of statistical estimators $\{\widehat{\varrho}_N\}_{N\in\mathbb{N}}$ is said to be *quantitatively statistically h-robust* on $\mathcal{M}$ w.r.t. $(\mathsf{dl}, \mathsf{dl}')$ if there exists a non-decreasing continuous function $h : \mathbb{R}_+ \to \mathbb{R}_+$ with $h(0) = 0$ such that for all $P, Q \in \mathcal{M}$ and $N \in \mathbb{N}$

$$\mathsf{dl}(P^{\otimes N} \circ \widehat{\varrho}_N^{-1}, Q^{\otimes N} \circ \widehat{\varrho}_N^{-1}) \le h(\mathsf{dl}'(P, Q)) < +\infty. \tag{19}$$

If, in addition, $\{\widehat{\varrho}_N\}_{N\in\mathbb{N}}$ arises as in (3) from a risk functional ϱ, then ϱ is called *quantitatively statistically robust* on $\mathcal{M}$ at P w.r.t. $(\mathsf{dl}, \mathsf{dl}')$. In a particular case when $\mathsf{dl} = \mathsf{dl}_K$, $h(t) = Lt$ and $\mathsf{dl}' = \mathsf{dl}_{FM,p}$, inequality (19) reduces to

$$\mathsf{dl}_K\left(P^{\otimes N} \circ \widehat{\varrho}_N^{-1}, Q^{\otimes N} \circ \widehat{\varrho}_N^{-1}\right) \le L\mathsf{dl}_{FM,p}(P, Q) < +\infty. \tag{20}$$

In comparison with the qualitative statistical robustness introduced by Krätschmer *et al.* (2014) or Cont *et al.* (2010), the definition (20) here has several advantages. First, we use Kantorovich metric instead of Prokhorov metric to quantify the discrepancy between $P^{\otimes N} \circ \widehat{\varrho}_N^{-1}$ and $Q^{\otimes N} \circ \widehat{\varrho}_N^{-1}$. This enables us to capture the tail behaviour of the two laws and facilitate us to derive an explicit bound for the difference. Second, we use the Fortet-Mourier metric to quantify the perturbation of P because it facilitates us to derive inequality (20) under condition (34), this can be seen clearly from the proof of Theorem 4.5. Third, inequality (20) gives an error bound for the discrepancy of the two laws and the bound is valid for all Q in $\mathcal{M}$ instead of those in a neighborhood of P.

Next, we introduce a definition on the Lipschitz continuity of a general statistical mapping from $\mathscr{P}(\mathbb{R})$ to $\mathbb{R}$, which strengthens the earlier definition of $\mathcal{C}$-continuity for a general statistical functional.

DEFINITION 10 (Lipschitz continuity) Let $\varrho : \mathscr{P}(\mathbb{R}) \to \mathbb{R}$ be a general statistical functional and $\mathcal{M}$ be a subset of $\mathscr{P}(\mathbb{R})$. ϱ is said to be *Lipschitz continuous* on $\mathcal{M}$ w.r.t. dl if there exists a positive constant L such that

$$|\varrho(P) - \varrho(Q)| \le L\mathsf{dl}(P, Q) < +\infty, \quad \forall\, P, Q \in \mathcal{M}. \tag{21}$$

There are a few points to note about the above definition of Lipschitz continuity:

1. The Lipschitz continuity is global instead of local over $\mathcal{M}$. The condition is strong but we will find that many risk functionals are global Lipschitz continuous on some $\mathcal{M}$ indeed.
2. The magnitude of the continuity depends on the metric dl which measures the distance between P and Q. In a specific case when $\mathsf{dl} = \mathsf{dl}_{FM,p}$, (21) reduces to

$$|\varrho(P) - \varrho(Q)| \le L \int_{\mathbb{R}} |P(x) - Q(x)| c_p(x) dx < +\infty, \quad \forall\, P, Q \in \mathcal{M}, \tag{22}$$

where $c_p(x) = \max\{1, |x|^{p-1}\}$. The exponent p plays an important role in (22) because it interacts with the tails of $P(\cdot)$ and $Q(\cdot)$. Moreover, if $\mathcal{M} \subset \mathcal{P}_p(\mathbb{R})$, then consequently the finiteness holds. We will come back to this later.
3. Let P_N and Q_N be empirical distributions on $\mathbb{R}$. By plugging P_N and Q_N into (22), we obtain

$$\begin{aligned}
|\varrho(P_N) - \varrho(Q_N)| &\le L \int_{\mathbb{R}} |P_N(x) - Q_N(x)| c_p(x) dx \\
&= L \int_{\mathbb{R}} \left| \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{\xi^i \le x\}} - \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{\tilde{\xi}^i \le x\}} \right| c_p(x) dx \\
&\le \frac{L}{N} \sum_{i=1}^N \int_{\mathbb{R}} \left| \mathbf{1}_{\{\xi^i \le x\}} - \mathbf{1}_{\{\tilde{\xi}^i \le x\}} \right| c_p(x) dx = \frac{L}{N} \sum_{i=1}^N \int_{\min\{\xi^i, \tilde{\xi}^i\}}^{\max\{\xi^i, \tilde{\xi}^i\}} c_p(x) dx \\
&= \frac{L}{N} \sum_{i=1}^N \left| \int_{\xi^i}^{\tilde{\xi}^i} c_p(x) dx \right| \le \frac{L}{N} \sum_{i=1}^N |\xi^i - \tilde{\xi}^i| \max\{c_p(\xi^i), c_p(\tilde{\xi}^i)\} \\
&= \frac{L}{N} \sum_{i=1}^N |\xi^i - \tilde{\xi}^i| c_p(\xi^i, \tilde{\xi}^i), \quad \forall \xi^i, \tilde{\xi}^i \in \mathbb{R},
\end{aligned} \tag{23}$$

where $c_p(\xi, \tilde{\xi}) = \max\{1, |\xi|, |\tilde{\xi}|\}^{p-1}$ for all $\xi, \tilde{\xi} \in \mathbb{R}$.
4. In the case when ϱ is continuous on $\mathcal{M}$ with respect to the topology of weak convergence, the Lipschitz continuity (22) on $\mathcal{M}$ is equivalent to the Lipschitz continuity (23) on the set of all empirical distributions $\mathcal{M}_{1,\text{emp}}$ (see the first inequality of equation (23)) because $\mathcal{M}_{1,\text{emp}}$ is dense in $\mathscr{P}(\mathbb{R})$ with respect to the topology of weak convergence, see e.g., Claus (2016, Theorem 2.49).

EXAMPLE 4.3 (p-th moment functional) *For $p \ge 1$, we consider the* p-th moment functional $T^{(p)}$ *on $\mathcal{M}_1^p = \mathcal{P}_p(\mathbb{R})$ as defined by:*

$$T^{(p)}(P) := \int_{-\infty}^{+\infty} x^p dP(x) < +\infty, \quad \forall P \in \mathcal{M}_1^p.$$

Analogous to Example B.1, we have

$$T^{(p)}(P) = -\int_{-\infty}^{0} P(x)px^{p-1}dx + \int_{0}^{+\infty}(1-P(x))px^{p-1}dx. \tag{24}$$

Thus, for any $P, Q \in \mathcal{M}_1^p$,

$$\begin{aligned}|T^{(p)}(P) - T^{(p)}(Q)| &= \left|\int_{-\infty}^{+\infty}(P(x)-Q(x))px^{p-1}dx\right| \leq p\int_{-\infty}^{+\infty}|P(x)-Q(x)||x|^{p-1}dx \\ &\leq p\int_{-\infty}^{+\infty}|P(x)-Q(x)|c_p(x)dx < +\infty,\end{aligned} \tag{25}$$

where $c_p(x) = \max\{1, |x|^{p-1}\}$. From (22), we can see that the p-th moment functional $T^{(p)}$ is Lipschitz continuous w.r.t. $\mathsf{dl}_{FM,p}$ on $\mathcal{M}_1^p$. Furthermore, let $f(x)$ be a polynomial of degree $p \geq 1$. Then based on the above analysis, the risk functional

$$T^f(P) := \int_{-\infty}^{+\infty} f(x)dP(x) < +\infty, \quad \forall P \in \mathcal{M}_1^p,$$

is also Lipschitz continuous w.r.t. $\mathsf{dl}_{FM,p}$ on $\mathcal{M}_1^p$. We call the associated risk measure defined by $\rho(X) = \mathbb{E}[f(X)]$ as the general moment-type risk measure *throughout the paper.*

Lemma 4.4 *Let $\boldsymbol{\xi} := (\xi^1, \cdots, \xi^N) \in \mathbb{R}^{\otimes N}$ and $\Phi_{N,p}$ be a set of functions from $\mathbb{R}^{\otimes N}$ to $\mathbb{R}$, i.e.,*

$$\Phi_{N,p} := \left\{\varphi : \mathbb{R}^{\otimes N} \to \mathbb{R} : |\varphi(\boldsymbol{\xi}) - \varphi(\tilde{\boldsymbol{\xi}})| \leq \frac{1}{N}\sum_{k=1}^{N} c_p(\xi^k, \tilde{\xi}^k)|\xi^k - \tilde{\xi}^k|, \ \forall \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}} \in \mathbb{R}^{\otimes N}\right\}, \tag{26}$$

where $c_p(\xi, \tilde{\xi}) := \max\{1, |\xi|, |\tilde{\xi}|\}^{p-1}$ for all $\xi, \tilde{\xi} \in \mathbb{R}$ and $p \geq 1$. Then

$$\mathsf{dl}_{\Phi_{N,p}}(P^{\otimes N}, Q^{\otimes N}) \leq \mathsf{dl}_{FM,p}(P, Q) < +\infty, \quad \forall P, Q \in \mathcal{P}_p(\mathbb{R}), \tag{27}$$

where $\mathsf{dl}_{\Phi_{N,p}}$ is defined by (9).

Before presenting a proof, it might be helpful for us to explain why we consider a specific set of functions $\Phi_{N,p}$. For fixed $N \in \mathbb{N}$, let $\mathcal{M}_{1,\text{emp}}^N$ denote the set of all empirical laws P_N over $\mathbb{R}$, then $\mathcal{M}_{1,\text{emp}} = \bigcup_{N\in\mathbb{N}} \mathcal{M}_{1,\text{emp}}^N$. Thus $\Phi_{N,p}$ may be regarded as a set of functions derived from a class of Lipschitz continuous functional on $\mathcal{M}_{1,\text{emp}}^N$ with $L = 1$ and $\mathsf{dl} = \mathsf{dl}_{FM,p}$ (by writing $T(P_N)$ as a function of samples). Lemma 4.4 says that for any $N \in \mathbb{N}$, the discrepancy between $P^{\otimes N}$ and $Q^{\otimes N}$ under the metric $\mathsf{dl}_{\Phi_{N,p}}$ can be bounded by $\mathsf{dl}_{FM,p}(P, Q)$. Note that the conclusion also holds for the case when P and Q are defined on $\mathcal{P}(\mathbb{R}^k)$. We restrict our discussion to $\mathcal{P}(\mathbb{R})$ because in that case we can readily use (13) to derive (23) from (21). It is unclear however whether (13) holds on $\mathcal{P}(\mathbb{R}^k)$.

Proof. Let $\xi^{-j} := \{\xi^1, \cdots, \xi^{j-1}, \xi^{j+1}, \cdots, \xi^N\}$. For any $\mu_1, \cdots, \mu_N \in \mathcal{P}_p(\mathbb{R})$ and $j \in \{1, \cdots, N\}$, denote

$$\mu_{-j}(d\xi^{-j}) := \mu_1(d\xi^1)\cdots\mu_{j-1}(d\xi^{j-1})\mu_{j+1}(d\xi^{j+1})\cdots\mu_N(d\xi^N)$$

and for any $\varphi \in \Phi_{N,p}$, let $h^{\varphi}_{\xi^{-j}}(\xi^j) := \int_{\mathbb{R}^{\otimes(N-1)}} \varphi(\xi^{-j}, \xi^j)\mu_{-j}(d\xi^{-j})$. We want to show

$$\int_{\mathbb{R}^N} \varphi(\boldsymbol{\xi})\mu(d\boldsymbol{\xi}) = \int_{\mathbb{R}} h^{\varphi}_{\xi^{-j}}(\xi^j)\mu_j(d\xi^j) < +\infty. \tag{28}$$

To this end, we take four steps. Observe first that for any fixed $\bar{\xi}$ and $\mu \in \mathcal{P}_p(\mathbb{R})$,

$$\int_{\mathbb{R}} c_p(\xi, \bar{\xi})|\xi - \bar{\xi}|\mu(d\xi) < +\infty. \tag{29}$$

Second, for fixed $\xi_0 \in \mathbb{R}^{\otimes N}$ and any $\varphi \in \Phi_{N,p}$,

$$|\varphi(\xi^{-j}, \xi^j)| \leq |\varphi(\xi_0^{-j}, \xi^j)| + \frac{1}{N}\sum_{k=1,k\neq j}^{N} c_p(\xi^k, \xi_0^k)|\xi^k - \xi_0^k|.$$

By taking expectation on both sides with respect to μ_{-j} and applying (29) for $\mu = \mu_k$ and $\bar{\xi} = \xi_0^k$ ($k \neq j$), we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^{\otimes(N-1)}} |\varphi(\xi^{-j}, \xi^j)|\mu_{-j}(d\xi^{-j}) \\
&\leq \int_{\mathbb{R}^{\otimes(N-1)}} |\varphi(\xi_0^{-j}, \xi^j)|\mu_{-j}(d\xi^{-j}) + \frac{1}{N}\sum_{k=1,k\neq j}^{N} \int_{\mathbb{R}^{\otimes(N-1)}} c_p(\xi^k, \xi_0^k)|\xi^k - \xi_0^k|\mu_{-j}(d\xi^{-j}) \\
&= |\varphi(\xi_0^{-j}, \xi^j)| + \frac{1}{N}\sum_{k=1,k\neq j}^{N} \int_{\mathbb{R}} c_p(\xi^k, \xi_0^k)|\xi^k - \xi_0^k|\mu_k(d\xi^k) < +\infty,
\end{aligned} \tag{30}$$

where the last equality is due to the fact that the integrand in the first integral is a constant. This shows $h^{\varphi}_{\xi^{-j}}(\xi^j)$ is well-defined. Third, for any $\varphi \in \Phi_{N,p}$,

$$\begin{aligned}
|h^{\varphi}_{\xi^{-j}}(\xi^j) - h^{\varphi}_{\xi^{-j}}(\tilde{\xi}^j)| &\leq \int_{\mathbb{R}^{\otimes(N-1)}} \left|\varphi(\xi^{-j}, \xi^j) - \varphi(\xi^{-j}, \tilde{\xi}^j)\right| \mu_{-j}(d\xi^{-j}) \\
&\leq \int_{\mathbb{R}^{\otimes(N-1)}} \frac{1}{N} c_p(\xi^j, \tilde{\xi}^j)|\xi^j - \tilde{\xi}^j|\mu_{-j}(d\xi^{-j}) \\
&\leq \frac{1}{N} c_p(\xi^j, \tilde{\xi}^j)|\xi^j - \tilde{\xi}^j|,
\end{aligned}$$

which means $h^{\varphi}_{\xi^{-j}}$ is measurable on $\mathbb{R}$ and $Nh^{\varphi}_{\xi^{-j}} \in \mathcal{F}_p(\mathbb{R})$ by the definition of $\mathcal{F}_p(\mathbb{R})$. Fourth, By

(30) and the property of φ,

$$
\begin{aligned}
&\int_{\mathbb{R}}\int_{\mathbb{R}^{\otimes(N-1)}} |\varphi(\xi^{-j},\xi^j)|\mu_{-j}(d\xi^{-j})\mu_j(d\xi^j) \\
&\le \int_{\mathbb{R}} |\varphi(\xi_0^{-j},\xi^j)|\mu_j(d\xi^j) + \frac{1}{N}\sum_{k=1,k\neq j}^{N}\int_{\mathbb{R}} c_p(\xi^k,\xi_0^k)|\xi^k-\xi_0^k|\mu_k(d\xi^k) \\
&\le \int_{\mathbb{R}} \left(|\varphi(\xi_0^{-j},0)| + \frac{1}{N}c_p(0,\xi^j)|\xi^j|\right)\mu_j(d\xi^j) + \frac{1}{N}\sum_{k=1,k\neq j}^{N}\int_{\mathbb{R}} c_p(\xi^k,\xi_0^k)|\xi^k-\xi_0^k|\mu_k(d\xi^k) \\
&= |\varphi(\xi_0^{-j},0)| + \frac{1}{N}\int_{\mathbb{R}} c_p(0,\xi^j)|\xi^j|\mu_j(d\xi^j) + \frac{1}{N}\sum_{k=1,k\neq j}^{N}\int_{\mathbb{R}} c_p(\xi^k,\xi_0^k)|\xi^k-\xi_0^k|\mu_k(d\xi^k) \\
&< +\infty,
\end{aligned}
$$

where the first inequality is due to the fact that the second summand in the last line of (30) is a constant, and the last inequality is from the fact that the third summand is finite due to the assumption $\mu_k \in \mathcal{P}_p(\mathbb{R})$. Consequently we can use Tonelli's theorem to obtain

$$
\int_{\mathbb{R}^N} |\varphi(\boldsymbol{\xi})|\mu(d\boldsymbol{\xi}) = \int_{\mathbb{R}}\int_{\mathbb{R}^{\otimes(N-1)}} |\varphi(\xi^{-j},\xi^j)|\mu_{-j}(d\xi^{-j})\mu_j(d\xi^j) < +\infty, \tag{31}
$$

which gives rise to (28) as desired.

Let $\mathcal{H}$ denote the set of functions $h^{\varphi}_{\xi^{-j}}$ generated by $\varphi \in \Phi_{N,p}$. Then for any $\mu_1,\cdots,\mu_N \in \mathcal{P}_p(\mathbb{R})$, by the definition of $\mathsf{dl}_{\Phi_{N,p}}$ and (28),

$$
\begin{aligned}
\mathsf{dl}_{\Phi_{N,p}}(\mu_{-j}\otimes\mu_j,\mu_{-j}\otimes\widetilde{\mu}_j) &= \sup_{\varphi\in\Phi_{N,p}}\bigg|\int_{\mathbb{R}}\int_{\mathbb{R}^{\otimes(N-1)}} \varphi(\xi^{-j},\xi^j)\mu_{-j}(d\xi^{-j})\mu_j(d\xi^j) \\
&\quad - \int_{\mathbb{R}}\int_{\mathbb{R}^{\otimes(N-1)}} \varphi(\xi^{-j},\xi^j)\mu_{-j}(d\xi^{-j})\tilde{\mu}_j(d\xi^j)\bigg| \\
&= \sup_{h^{\varphi}_{\xi^{-j}}\in\mathcal{H}}\left|\int_{\mathbb{R}} h^{\varphi}_{\xi^{-j}}(\xi^j)\mu_j(d\xi^j) - \int_{\mathbb{R}} h^{\varphi}_{\xi^{-j}}(\xi^j)\tilde{\mu}_j(d\xi^j)\right| \\
&\le \frac{1}{N}\mathsf{dl}_{FM,p}(\mu_j,\tilde{\mu}_j),
\end{aligned} \tag{32}
$$

where the inequality is due to $Nh_{\xi^{-j}} \in \mathcal{F}_p(\mathbb{R})$ and the definition of $\mathsf{dl}_{FM,p}(P,Q)$. It is important to note that the first equality in (32) does not depend on the order of how the measures are combined via product due to Tonelli's theorem. Finally, by the triangle inequality of the pseudo-metric, we have

$$
\begin{aligned}
\mathsf{dl}_{\Phi_{N,p}}\left(P^{\otimes N},Q^{\otimes N}\right) &\le \mathsf{dl}_{\Phi_{N,p}}\left(P^{\otimes N},P^{\otimes(N-1)}\otimes Q\right) + \mathsf{dl}_{\Phi_{N,p}}\left(P^{\otimes(N-1)}\otimes Q,P^{\otimes(N-2)}\otimes Q^{\otimes 2}\right) \\
&\quad + \cdots + \mathsf{dl}_{\Phi_{N,p}}\left(P\otimes Q^{\otimes(N-1)},Q^{\otimes N}\right) \\
&\le \frac{1}{N}\mathsf{dl}_{FM,p}(P,Q)\times N = \mathsf{dl}_{FM,p}(P,Q).
\end{aligned}
$$

The proof is complete. $\square$

Remark 1 It might be interesting to ask whether Lemma 4.4 can be established with other ζ-metrics[1]. The answer is yes but maybe under some undesirable conditions on $\Phi_{N,p}$. Consider for example the total variation metric and define the set

$$\Phi_N := \left\{ \varphi_N : \mathbb{R}^{\otimes N} \to \mathbb{R} : |\varphi_N(\boldsymbol{\xi})| \le \frac{L}{N}, \forall \boldsymbol{\xi} \in \mathbb{R}^{\otimes N} \right\}. \tag{33}$$

Following a similar analysis to the proof of the lemma, we can show that $\mathsf{dl}_{\Phi_N}(P^{\otimes N}, Q^{\otimes N}) \le 2L\mathsf{dl}_{TV}(P,Q)$. However, we feel the condition $|\varphi_N(\boldsymbol{\xi})| \le \frac{L}{N}$ is too strong as the bound goes to zero when $N \to \infty$.

With the intermediate technical result, we are now ready to present our main result of quantitative statistical robustness for the plug-in estimator of a general risk functional.

Theorem 4.5 *Let $\varrho : \mathscr{P}(\mathbb{R}) \to \mathbb{R}$ be a general statistical functional and $\mathcal{M}$ be a subset of $\mathcal{P}_p(\mathbb{R})$ with $p \ge 1$. Assume, for any $N \in \mathbb{N}$, there exists a positive constant L such that*

$$|\varrho(P_N) - \varrho(Q_N)| \le \frac{L}{N} \sum_{k=1}^{N} c_p(\xi^k, \tilde{\xi}^k)|\xi^k - \tilde{\xi}^k|, \ \forall \xi^k, \tilde{\xi}^k \in \mathbb{R}, \tag{34}$$

where $c_p(\xi^k, \tilde{\xi}^k)$ was introduced subsequent to (10), P_N and Q_N are given by (4) and (5) based on the samples from P and Q in $\mathcal{M}$ respectively. Let $\widehat{\varrho}_N$ be defined as in (3). Then the sequence of the plug-in estimators $\{\widehat{\varrho}_N\}_{N\in\mathbb{N}}$ is quantitatively robust on $\mathcal{M}$ w.r.t. $(\mathsf{dl}_K, \mathsf{dl}_{FM,p})$.

Before presenting a proof, we note that the conclusion in Theorem 4.5 also holds for the case when P and Q are defined on $\mathcal{P}_p(\mathbb{R}^k)$.

Proof. For any $N \in \mathbb{N}$ and $P, Q \in \mathcal{M}$, by definition

$$\begin{aligned} & \mathsf{dl}_K\left(P^{\otimes N} \circ \widehat{\varrho}_N^{-1}, Q^{\otimes N} \circ \widehat{\varrho}_N^{-1}\right) \\ = & \sup_{\psi \in \mathcal{F}_1(\mathbb{R})} \left| \int_{\mathbb{R}} \psi(t) P^{\otimes N} \circ \widehat{\varrho}_N^{-1}(dt) - \int_{\mathbb{R}} \psi(t) Q^{\otimes N} \circ \widehat{\varrho}_N^{-1}(dt) \right| \\ = & \sup_{\psi \in \mathcal{F}_1(\mathbb{R})} \left| \int_{\mathbb{R}^{\otimes N}} \psi(\varrho(\boldsymbol{\xi}^N)) P^{\otimes N}(d\boldsymbol{\xi}^N) - \int_{\mathbb{R}^{\otimes N}} \psi(\varrho(\boldsymbol{\xi}^N)) Q^{\otimes N}(d\boldsymbol{\xi}^N) \right|, \end{aligned} \tag{35}$$

where we write $\boldsymbol{\xi}^N$ for $(\xi^1, \cdots, \xi^N)$ and $\varrho(\boldsymbol{\xi}^N)$ for $\widehat{\varrho}_N$ to indicate its dependence on $\xi^1, \cdots, \xi^N$.

To see well-definiteness of the pseudo-metric, it suffices to show that for any $\psi \in \mathcal{F}_1(\mathbb{R})$, both $\int_{\mathbb{R}^{\otimes N}} \psi(\varrho(\boldsymbol{\xi}^N)) P^{\otimes N}(d\boldsymbol{\xi}^N)$ and $\int_{\mathbb{R}^{\otimes N}} \psi(\varrho(\boldsymbol{\xi}^N)) Q^{\otimes N}(d\boldsymbol{\xi}^N)$ are well-defined. We take two steps. First, for each $\psi \in \mathcal{F}_1(\mathbb{R})$, by the definition of $\mathcal{F}_1(\mathbb{R})$ and (34)

$$|\psi(\varrho(\boldsymbol{\xi}^N)) - \psi(\varrho(\boldsymbol{\xi}_0^N))| \le |\varrho(\boldsymbol{\xi}^N) - \varrho(\boldsymbol{\xi}_0^N)| \le \frac{L}{N} \sum_{k=1}^{N} c_p(\xi^k, \xi_0^k)|\xi^k - \xi_0^k|, \tag{36}$$

where $\boldsymbol{\xi}_0^N$ is fixed. Since $\boldsymbol{\xi}_0^N$ can be anywhere, the inequality above implies that $\psi(\varrho(\cdot))$ is pointwise continuous and hence it is measurable. Second, for any $P \in \mathcal{M}$, by using inequality (36) and

[1] ζ-metric is needed because when we use the Kantorovich metric to measure the distance between two statistical estimators, the dual formulation of the metric enables us to reformulate the distance as the difference between $P^{\otimes N}$ and $Q^{\otimes N}$ under certain ζ-metric, see (35) in the proof of Theorem 4.4.

applying Tonelli's theorem to the integral $\int_{\mathbb{R}^{\otimes N}} |\psi(\varrho(\boldsymbol{\xi}^N))| P^{\otimes N}(d\boldsymbol{\xi}^N)$ by switching the order of integration, we obtain

$$\begin{aligned}\int_{\mathbb{R}^{\otimes N}} |\psi(\varrho(\boldsymbol{\xi}^N))| P^{\otimes N}(d\boldsymbol{\xi}^N) &\leq \left|\psi(\varrho(\boldsymbol{\xi}_0^N))\right| + \frac{L}{N}\int_{\mathbb{R}^{\otimes N}} \sum_{k=1}^N c_p(\xi^k, \xi_0^k)|\xi^k - \xi_0^k| P^{\otimes N}(d\boldsymbol{\xi}^N) \\ &= \left|\psi(\varrho(\boldsymbol{\xi}_0^N))\right| + \frac{L}{N}\sum_{k=1}^N \int_{\mathbb{R}} c_p(\xi, \xi_0^k)|\xi - \xi_0^k| P(d\xi) < +\infty, \quad (37)\end{aligned}$$

where the equality is due to the fact that the integrand depends on ξ^k only and the last inequality holds because

$$\int_{\mathbb{R}} c_p(\xi_k, \xi_0^k)|\xi^k - \xi_0^k| P(d\xi^k) < +\infty, \text{ for } P \in \mathcal{M} \subset \mathcal{P}_p(\mathbb{R}), k = 1, \ldots, N.$$

The boundedness and measurablility ensure the well-definedness of $\int_{\mathbb{R}^{\otimes N}} \psi(\varrho(\boldsymbol{\xi}^N)) P^{\otimes N}(d\boldsymbol{\xi}^N)$ as desired. Similar conclusion can be drawn when P is replaced by Q.

To complete the proof, we note that (36) implies that $\frac{1}{L}\psi(\varrho(\cdot))$ belongs to $\Phi_{N,p}$ and by the definition of Fortet-Mourier metric, (35) implies

$$\mathsf{dl}_K\left(P^{\otimes N} \circ \widehat{\varrho}_N^{-1}, Q^{\otimes N} \circ \widehat{\varrho}_N^{-1}\right) \leq L \mathsf{dl}_{\Phi_{N,p}}(P^{\otimes N}, Q^{\otimes N}).$$

The rest follows from Lemma 4.4.

□

Note that Theorem 4.5 may be established under strengthened condition (22)) (from condition (23). The next corollary states this.

Corollary 4.6 *Let $\varrho : \mathscr{P}_1(\mathbb{R}) \to \mathbb{R}$ be a general statistical functional and $\mathcal{M}$ be a subset of $\mathcal{P}_p(\mathbb{R})$ with $p \geq 1$. If ϱ satisfies condition (22), then ϱ is quantitatively statistically robust on $\mathcal{M}$ w.r.t.* $(\mathsf{dl}_K, \mathsf{dl}_{FM,p})$.

From Example 3.1, we have $\mathsf{dl}_{\text{Prok}}(P, Q) \leq \sqrt{\mathsf{dl}_K(P, Q)}$ for all $P, Q \in \mathscr{P}(\mathbb{R})$, then we have the following corollary.

Corollary 4.7 *Let $\varrho : \mathscr{P}_1(\mathbb{R}) \to \mathbb{R}$ be a general statistical functional. Assume that ϱ is Lipschitz continuous w.r.t. $\mathsf{dl}_{FM,p}$ on $\mathcal{M} \subset \mathcal{P}_p(\mathbb{R})$ for the constant L, where $p \geq 1$. Then ϱ is quantitatively robust on $\mathcal{M}$ w.r.t.* $(\mathsf{dl}_{\text{Prok}}, \mathsf{dl}_{FM,p})$, *i.e.,*

$$\mathsf{dl}_{\text{Prok}}\left(P^{\otimes N} \circ \widehat{\varrho}_N^{-1}, Q^{\otimes N} \circ \widehat{\varrho}_N^{-1}\right) \leq \sqrt{L \mathsf{dl}_{FM,p}(P, Q)} < +\infty, \ \forall\, P, Q \in \mathcal{M}$$

for all $N \in \mathbb{N}$.

Corollary 4.8 *Let $\rho : L^1 \to \mathbb{R}$ be a law invariant coherent risk measure and ϱ its associated risk functional as defined in (2). Then ϱ is quantitatively robust on $\mathcal{P}_1(\mathbb{R})$ w.r.t.* $(\mathsf{dl}_K, \mathsf{dl}_{FM,1})$.

Proof. It follows from Ruszczyński and Shapiro (2006, Corollary 3.1) that ρ is continuous with respect to $\|\cdot\|_1$-norm. Moreover, due to the norm continuity, Inoue (2003, Lemma 2.1) ensures that ρ is Lipschitz-continuous with respect to $\|\cdot\|_1$-norm. We now show that the associated risk functional ϱ is Lipschitz continuous with respect to $\mathsf{dl}_{FM,1}$. By Proposition 2.1, $\rho(X) = \rho(F_P^{-1}(U)) = \varrho(P)$, where P is the law of X and U is a random variable uniformly distributed over $[0, 1]$. Moreover it is known (see e.g. Panaretos and Zemel (2019)) that

$$\|F_P^{-1}(U) - F_Q^{-1}(U)\|_1 = \mathsf{dl}_{W,1}(P,Q), \tag{38}$$

where $\mathsf{dl}_{W,p}$ is the p-th order Wasserstein metric:

$$\mathsf{dl}_{W,p}(P,Q) = \left(\inf_{\pi\in\Pi(P,Q)} \int_{\mathbb{R}} |x-y|^p d\pi(x,y)\right)^{\frac{1}{p}},$$

where $p \geq 1$, Q is the law of Y, $\Pi(P,Q)$ is the set of probability measures on $\mathbb{R}\times\mathbb{R}$ with marginals P and Q, i.e., for all Borel subsets $A, B \in \mathbb{R}$, $P(A) = \pi(A\times\mathbb{R})$ and $Q(B) = \pi(\mathbb{R}\times B)$. Let L be the Lipschitz modulus of $\rho(\cdot)$ (w.r.t. $\|\cdot\|_1$-norm). By (38), we obtain

$$\begin{aligned}|\varrho(P) - \varrho(Q)| &= |\rho(X) - \rho(Y)| = |\rho(F_P^{-1}(U)) - \rho(F_Q^{-1}(U))| \\ &\leq L\|F_P^{-1}(U) - F_Q^{-1}(U)\|_1 = L\mathsf{dl}_{W,1}(P,Q) < +\infty. \end{aligned}\tag{39}$$

Moreover, it follows by Panaretos and Zemel (2019) that

$$\mathsf{dl}_{W,1}(P,Q) = \int_{\mathbb{R}} |P(t) - Q(t)| dt \leq \mathsf{dl}_{FM,1}(P,Q). \tag{40}$$

Combining (39) and (40), we arrive at

$$|\varrho(P) - \varrho(Q)| \leq L\mathsf{dl}_{FM,1}(P,Q), \tag{41}$$

which implies

$$|\varrho(P_N) - \varrho(Q_N)| \leq \frac{L}{N}\sum_{k=1}^{N} |\xi^k - \tilde{\xi}^k|, \ \forall \xi^k, \tilde{\xi}^k \in \mathbb{R}. \tag{42}$$

Inequality (42) is a special case of condition (34) with $p = 1$. By Theorem 4.5, the sequence of $\{\widehat{\varrho}_N\}_{N\in\mathbb{N}}$ is quantitatively robust on $\mathcal{P}_1(\mathbb{R})$ w.r.t. $(\mathsf{dl}_K, \mathsf{dl}_{FM,1})$. The conclusion follows from the Definition 9. $\square$

Next, we take a step further to consider the index of quantitative robustness for a general statistical functional. Since the Lipschitz continuity of the statistical functional ϱ on $\mathcal{P}_p(\mathbb{R})$ under $\mathsf{dl}_{FM,p}$ provides a key sufficient condition (34) in Theorem 4.5, we use p associated with the Lipschitz continuity to define the index of quantitative robustness as below.

Definition 11 (Index of quantitative robustness) Let $\varrho : \mathscr{P}(\mathbb{R}) \to \mathbb{R}$ be a general statistical functional. If ϱ is Lipschitz continuous w.r.t. $\mathsf{dl}_{FM,p}$ on $\mathcal{P}_p(\mathbb{R})$ for the constant L for some $p \geq 1$, then we can define an *index of quantitative robustness of a statistical functional* ϱ as

$$\mathrm{iqr}(\varrho) := \left(\inf\{p \in [1,+\infty) : \varrho \text{ is Lipschitz continuous w.r.t. } \mathsf{dl}_{FM,p} \text{ on } \mathcal{P}_p(\mathbb{R})\}\right)^{-1}. \tag{43}$$

This index is a quantitative measurement for the degree of robustness of a statistical functional. A larger index reflects a higher degree of robustness. For a general statistical functional ϱ, (22) may hold for uncountable many p, see e.g., the 2-th moment functional $T^{(2)}$ satisfying (22) for any $p \geq 2$ on $\mathcal{P}_p(\mathbb{R}) = \mathcal{M}_1^p$. From Definition 11, we conclude that the p-th moment functional $T^{(p)}$ has the index $\mathrm{iqr}(T^{(p)}) = \frac{1}{p}$. Definition 11 coincides with the index of qualitative robustness proposed by Krätschmer *et al.* (2012) when ϱ is Lipschitz continuous w.r.t. $\mathsf{dl}_{FM,p}$ on $\mathcal{P}_p(\mathbb{R})$. The main advantage of Definition 11 is that it is easy to calculate and we will illustrate this in the next section.

5. Application to law invariant risk measures

As we discussed in Proposition 2, a law invariant risk measure of a random variable can be represented as a composition of a risk functional and law of the random variable. In practice, risk of a random variable is often calculated with empirical data, this is because either the true probability distribution is unknown or it might be prohibitively expensive to calculate the risk of a random variable with the true probability distribution. This raises a question as to whether the estimated risk measure based on empirical data is reliable or not when the data are potentially contaminated. In this section, we apply the quantitative robustness results established in Theorem 4.5 to some well-known risk measures. The next proposition synthesizes Proposition 2 and Theorem 4.5.

PROPOSITION 5.1 *Let ρ be a tail-dependent law invariant convex risk measure ϱ its associated risk functional as defined in (2). Assume that there exists a positive number $p \geq 1$ such that (34) holds true (for any $N \in \mathbb{N}$), where $c_p(\xi^k, \tilde{\xi}^k)$ as introduced subsequent to (10), P_N and Q_N are given by (4) and (5) based on the samples from P and Q in $\mathcal{P}_p(\mathbb{R})$ respectively. Then ϱ is quantitatively robust on $\mathcal{P}_p(\mathbb{R})$ w.r.t. $(\mathsf{dl}_K, \mathsf{dl}_{FM,p})$.*

Since the law invariant risk measure and its associated risk functional is one-to-one from Proposition 2.1, in this section, we will also say ρ is quantitatively robust if its associated risk functional is. In what follows, we verify condition (34) for some well-known risk measures and hence show that they satisfy the proposed quantitative statistical robustness (20).

EXAMPLE 5.2 *The* expectation *of $G \in \mathscr{P}_1(\mathbb{R})$ defined by $\mathbb{E}(G) := \int_{\mathbb{R}} \xi dG(\xi)$ satisfies*

$$|\mathbb{E}(P_N) - \mathbb{E}(Q_N)| = \left| \int_{\mathbb{R}} \xi d(P_N - Q_N)(\xi) \right| \leq \frac{1}{N} \sum_{i=1}^{N} |\xi^i - \tilde{\xi}^i|.$$

Then the expectation *is quantitatively robust on $\mathcal{P}_1(\mathbb{R})$ w.r.t. $(\mathsf{dl}_K, \mathsf{dl}_K)$ with $L = 1$ and the index of quantitative robustness is* $\mathrm{iqr}(\mathbb{E}) = 1$.

EXAMPLE 5.3 *The* conditional value-at-risk *of $G \in \mathscr{P}_1(\mathbb{R})$ at level $\tau \in (0,1)$ defined by*

$$\mathrm{CVaR}_\tau(G) := \inf \left\{ r + \frac{1}{1-\tau} \int_{\mathbb{R}} \max\{0, \xi - r\} dG(\xi), \forall r \in \mathbb{R} \right\}$$

satisfies

$$\begin{aligned}
|\mathrm{CVaR}_\tau(P_N) - \mathrm{CVaR}_\tau(Q_N)| &\leq \frac{1}{1-\tau} \sup_{r \in \mathbb{R}} \left| \int_{\mathbb{R}} \max\{0, \xi - r\} d(P_N - Q_N)(\xi) \right| \\
&= \frac{1}{1-\tau} \sup_{r \in \mathbb{R}} \frac{1}{N} \left| \sum_{i=1}^{N} \max\{0, \xi^i - r\} - \max\{0, \tilde{\xi}^i - r\} \right| \\
&\leq \frac{1}{1-\tau} \times \frac{1}{N} \sum_{i=1}^{N} |\xi^i - \tilde{\xi}^i|,
\end{aligned}$$

where the last inequality is due to the fact that $|\max\{0, x\} - \max\{0, y\}| \leq |x - y|$ for all $x, y \in \mathbb{R}$. Thus, the CVaR_τ is quantitatively robust on $\mathcal{P}_1(\mathbb{R})$ w.r.t. $(\mathsf{dl}_K, \mathsf{dl}_K)$ with $L = \frac{1}{1-\tau}$ and the index *of quantitative robustness is* $\mathrm{iqr}(\mathrm{CVaR}_\tau) = 1$ *for* $\tau \in (0,1)$*. Similar conclusions may be drawn for spectral risk measures, see Wang and Xu (2020).*

EXAMPLE 5.4 *The* upper semi-deviation $sd_+(G)$ *of* $G \in \mathscr{P}_1(\mathbb{R})$ *defined by*

$$sd_+(G) := \int_{\mathbb{R}} \max\left\{0, \xi - \int_{\mathbb{R}} u dG(u)\right\} dG(\xi),$$

satisfies

$$\begin{aligned}|sd_+(P_N) - sd_+(Q_N)| &= \left| \frac{1}{N}\sum_{j=1}^N \max\left\{0, \xi^j - \frac{1}{N}\sum_{i=1}^N \xi^i\right\} - \frac{1}{N}\sum_{j=1}^N \max\left\{0, \tilde{\xi}^j - \frac{1}{N}\sum_{i=1}^N \tilde{\xi}^i\right\}\right| \\ &\le \frac{1}{2}\sum_{j=1}^N \left(\left|\xi^j - \tilde{\xi}^j\right| + \frac{1}{N}\sum_{i=1}^N |\xi^i - \tilde{\xi}^i|\right) = \frac{2}{N}\sum_{i=1}^N |\xi^i - \tilde{\xi}^i|.\end{aligned}$$

Then the upper semi-derivation *is quantitatively robust on* $\mathcal{P}_1(\mathbb{R})$ *w.r.t.* $(\mathsf{dl}_K, \mathsf{dl}_K)$ *with* $L = 2$ *and the index of quantitative robustness is* $\mathrm{iqr}(\mathrm{sd}_+) = 1$.

EXAMPLE 5.5 *The* Optimized Certainty Equivalent (OCE) *(Ben-Tal and Teboulle 2007) of* $G \in \mathscr{P}_1(\mathbb{R})$ *is given by*

$$S_u(G) := \sup_{\eta\in\mathbb{R}} \left\{\eta + \int_{\mathbb{R}} u(\xi - \eta) dG(\xi)\right\},$$

where $u : \mathbb{R} \to [-\infty, \infty)$ *is a proper concave and non-decreasing utility function satisfying the normalized property:* $u(0) = 0$ *and* $1 \in \partial u(0)$*, where* $\partial u(\cdot)$ *denotes the subdifferential map of* u*. By the essential of (Ben-Tal and Teboulle 2007, Proposition 2.1), we have*

$$S_u(P_N) = \sup_{\eta\in\mathbb{R}} \left\{\eta + \frac{1}{N}\sum_{i=1}^N u(\xi^i - \eta)\right\} = \sup_{\eta\in[\xi_{\min},\xi_{\max}]} \left\{\eta + \frac{1}{N}\sum_{i=1}^N u(\xi^i - \eta)\right\},$$

where $\xi_{\min} = \min\{\xi^1, \ldots, \xi^N; \tilde{\xi}^1, \ldots, \tilde{\xi}^N\}$ *and* $\xi_{\max} = \max\{\xi^1, \ldots, \xi^N; \tilde{\xi}^1, \ldots, \tilde{\xi}^N\}$*. Let* $\varrho(G) := -S_u(G)$*. Then it satisfies*

$$\begin{aligned}|\varrho(P_N) - \varrho(Q_N)| &\le \sup_{\eta\in[\xi_{\min},\xi_{\max}]} \left|\left(\eta + \int_{\mathbb{R}} u(\xi - \eta) dP_N(\xi)\right) - \left(\eta + \int_{\mathbb{R}} u(\xi - \eta) dQ_N(\xi)\right)\right| \\ &= \sup_{\eta\in[\xi_{\min},\xi_{\max}]} \left|\frac{1}{N}\sum_{i=1}^N u(\xi^i - \eta) - \frac{1}{N}\sum_{i=1}^N u(\tilde{\xi}^i - \eta)\right| \\ &\le \frac{1}{N}\sum_{i=1}^N u'_-(\xi_{\min})|\xi^i - \tilde{\xi}^i|,\end{aligned}$$

where $u'_-(t)$ *denotes the left derivative of* u *at* t *and the last inequality is due to the fact that* u *is non-decreasing and concave, subsequently,* $u'_-(t)$ *is non-increasing.*

In the following, we consider two interesting cases. One is that $\sup_{\eta\in\mathbb{R}} u'_-(\eta) < +\infty$*, in which case* ϱ *is quantitatively robust on* $\mathcal{P}_1(\mathbb{R})$ *w.r.t.* $(\mathsf{dl}_K, \mathsf{dl}_K)$ *with* $L = \sup_{\eta\in\mathbb{R}} u'_-(\eta)$ *and the index of quantitative robustness for this case is* 1*. The other is that there exists some smallest positive number* $p > 1$ *and positive constant* $\tilde{L}$ *such that* $u'_-(\xi_{\min}) \le \tilde{L} c_p(\xi^i, \tilde{\xi}^i)$*, where* $c_p(\xi^i, \tilde{\xi}^i) = \max\{1, |\xi^i|, |\tilde{\xi}^i|\}^{p-1}$*. In* *this case, ϱ is quantitatively robust on $\mathcal{P}_p(\mathbb{R})$ w.r.t. $(\mathsf{dl}_K, \mathsf{dl}_{FM,p})$ with $L = \tilde{L}$ and the index of quantitative robustness for this case is $\frac{1}{p}$.*

To see how the above two interesting cases could possibly be satisfied, we consider two specific utility functions: piecewise linear utility function and quadratic utility function, both of which are extracted from (Ben-Tal and Teboulle 2007).

(a) Piecewise linear utility function with $u(t) := \gamma_1[t]_+ + \gamma_2[-t]_+$, where $0 \le \gamma_1 < 1 < \gamma_2$ and $[z]_+ = \max\{0, z\}$. A simple calculation yields

$$|\varrho(P_N) - \varrho(Q_N)| \le \frac{\gamma_2}{N}\sum_{i=1}^{N} |\xi^i - \tilde{\xi}^i|.$$

Thus in this case, ϱ is quantitatively robust on $\mathcal{P}_1(\mathbb{R})$ w.r.t. $(\mathsf{dl}_K, \mathsf{dl}_K)$ with $L = \gamma_2$ and the index of quantitative robustness is $\mathrm{iqr}(-S_u) = 1$.

(b) Quadratic utility with $u(t) := (t - \frac{1}{2}t^2)\mathbf{1}_{(-\infty,1)}(t) + \frac{1}{2}\mathbf{1}_{[1,+\infty)}(t)$. It is easy to observe that the function is locally Lipschitz continuous over $[\xi_{\min}, \xi_{\max}]$ with modulus being bounded by $|1 - \xi_{\min}|$. Thus

$$|\varrho(P_N) - \varrho(Q_N)| \le \sup_{\eta \in [\xi_{\min}, \xi_{\max}]} \frac{1}{N}\sum_{i=1}^{N} \left|u(\xi^i - \eta) - u(\tilde{\xi}^i - \eta)\right| \le \frac{1}{N}\sum_{i=1}^{N} |1 - \xi_{\min}||\xi^i - \tilde{\xi}^i|.$$

Moreover, if $\xi_{\min} \le -1$, then $|1 - \xi_{\min}| \le 2|\xi_{\min}|$. Subsequently,

$$|\varrho(P_N) - \varrho(Q_N)| \le \frac{2}{N}\sum_{i=1}^{N} c_2(\xi^i, \tilde{\xi}^i)|\xi^i - \tilde{\xi}^i|,$$

where $c_2(\xi^i, \tilde{\xi}^i) = \max\{1, |\xi^i|, |\tilde{\xi}^i|\}$. Thus in this case, ϱ is quantitatively robust on $\mathcal{P}_2(\mathbb{R})$ w.r.t. $(\mathsf{dl}_K, \mathsf{dl}_{FM,2})$ with $L = 2$ provided that $\xi_{\min} < -1$ and the index of quantitative robustness is $\mathrm{iqr}(-S_u) = \frac{1}{2}$.

Example 5.6 *Suppose that $l : \mathbb{R} \to \mathbb{R}$ is an increasing convex loss function which is not identically constant. Let x_0 be an interior point in the range of l. The* Shortfall Risk Measure *(Föllmer and Schied 2002) of $G \in \mathscr{P}(\mathbb{R})$ is defined by*

$$\varrho_l(G) := \inf\left\{m \in \mathbb{R} : \int_{\mathbb{R}} l(\xi - m)dG(\xi) \le x_0\right\}. \tag{44}$$

Following a similar analysis to Guo et al. (2017), we can recast the formulation above as

$$\varrho_l(G) = \inf_{m\in\mathbb{R}} \sup_{\lambda\ge 0}\left\{m + \lambda\left(\int_{\mathbb{R}} l(\xi - m)dG(\xi) - x_0\right)\right\}. \tag{45}$$

Swapping the inf and sup operations, we can obtain the Lagrange dual of the problem. Moreover, if we assume that the inequality constraint in (44) satisfies the well-known Slater condition, i.e., there exists m_0 such that $\int_{\mathbb{R}} l(\xi - m_0)dG(\xi) - x_0 < 0$, then the Lagrange multipliers of (44) is bounded and the strong duality holds. Consequently, we can rewrite (45) as

$$\varrho_l(G) = \inf_{m\in\mathbb{R}} \sup_{\lambda\in[a,b]}\left\{m + \lambda\left(\int_{\mathbb{R}} l(\xi - m)dG(\xi) - x_0\right)\right\}, \tag{46}$$

where a, b are some positive numbers. By the essential of (Ben-Tal and Teboulle 2007, Proposition

2.1), we have

$$\varrho_l(P_N) = \sup_{\lambda\in[a,b]} \inf_{m\in\mathbb{R}} \left\{ m + \lambda \left(\frac{1}{N}\sum_{i=1}^N l(\xi^i - \eta) - x_0 \right) \right\}$$
$$= \sup_{\lambda\in[a,b]} \inf_{m\in[\xi_{\min},\xi_{\max}]} \left\{ m + \lambda \left(\frac{1}{N}\sum_{i=1}^N l(\xi^i - m) - x_0 \right) \right\},$$

where $\xi_{\min} = \min\{\xi^1,\dots,\xi^N;\tilde{\xi}^1,\dots,\tilde{\xi}^N\}$ *and* $\xi_{\max} = \max\{\xi^1,\dots,\xi^N;\tilde{\xi}^1,\dots,\tilde{\xi}^N\}$*. Subsequently,*

$$|\varrho_l(P_N) - \varrho_l(Q_N)| \le b \sup_{m\in[\xi_{\min},\xi_{\max}]} \left| \frac{1}{N}\sum_{i=1}^N l(\xi^i - m) - \frac{1}{N}\sum_{i=1}^N l(\tilde{\xi}^i - m) \right|$$
$$\le \frac{b}{N}\sum_{i=1}^N \sup_{m\in\mathbb{R}} l'_+(m)|\xi^i - \tilde{\xi}^i|,$$

where $l'_+(t)$ *denote the right derivative of* l *at* t *and the last inequality is due to the fact that* l *is non-decreasing convex, subsequently,* $l'_+(t)$ *is non-decreasing.*

In the following, we consider two interesting cases. One is that $\sup_{m\in\mathbb{R}} l'_+(m) < +\infty$*, then* ϱ_l *is quantitatively robust on* $\mathcal{P}_1(\mathbb{R})$ *w.r.t.* $(\mathsf{dl}_K, \mathsf{dl}_K)$ *with* $L = \sup_{m\in\mathbb{R}} l'_+(m)$ *and the index of quantitative robustness for this case is 1. The other is that there exists some smallest positive number* $p > 1$ *and positive constant* $\tilde{L}$ *such that* $l'_+(\xi^i - \xi_{\min}) \vee l'_+(\tilde{\xi}^i - \xi_{\min}) \le \tilde{L}c_p(\xi^i, \tilde{\xi}^i)$*, where* $c_p(\xi^i, \tilde{\xi}^i) = \max\{1, |\xi^i|, |\tilde{\xi}^i|\}^{p-1}$*, then* ϱ_l *is quantitatively robust on* $\mathcal{P}_1(\mathbb{R})$ *w.r.t.* $(\mathsf{dl}_K, \mathsf{dl}_{FM,p})$ *with* $L = \tilde{L}$ *and the index of quantitative robustness for this case is* $\frac{1}{p}$*.*

To see how the above two interesting cases could possibly be satisfied, we consider two specific loss functions: deposit insurance loss function (Chen et al. 2013) and p-th power loss function (Föllmer and Schied 2002).

(a) Deposit insurance loss function, $l(x) = [x]_+$*, where* $[x]_+ = \max\{x, 0\}$*. Then* $\sup_{m\in\mathbb{R}} l'_+(m) = 1 < +\infty$*. Thus, in this case, the* shortfall risk measure *is quantitatively robust on* $\mathcal{P}_1(\mathbb{R})$ *w.r.t.* $(\mathsf{dl}_K, \mathsf{dl}_K)$ *with* $L = 1$ *and the index of quantitative robustness is 1.*

(b) For $x_0 > 0$*, we consider the p-th power loss function,*

$$l(x) = \begin{cases} \frac{1}{p}x^p, & \text{if } x \ge 0, \\ 0, & \text{otherwise,} \end{cases}$$

where $p > 1$*. We have* $l'_+(x) = x^{p-1}$ *for* $x \ge 0$ *and* $l'_+(x) = 0$ *for* $x < 0$*. Then, if* $\xi_{\min} \ge 0$*, then* $0 \le \xi^i - \xi_{\min} \le |\xi^i|$ *and subsequently* $l'_+(\xi^i - \xi_{\min}) \vee l'_+(\tilde{\xi}^i - \xi_{\min}) \le c_p(\xi^i, \tilde{\xi}^i)$*. Thus, in this case, the* shortfall risk measure *is quantitatively robust on* $\mathcal{P}_p(\mathbb{R})$ *w.r.t.* $(\mathsf{dl}_k, \mathsf{dl}_{FM,p})$ *with* $L = 1$ *provided that* $\xi_{\min} \ge 0$ *and the index of quantitative robustness is* $\frac{1}{p}$*.*

6. An illustrative example

It might be interesting to ask whether the error bound established in Theorem 4.5 is tight. To address this question, we have carried out some numerical analysis on CVaR in Example 5.2. We consider a random loss X which follows a normal distribution, i.e., $X \sim \mathcal{N}(\mu, \sigma)$, with law

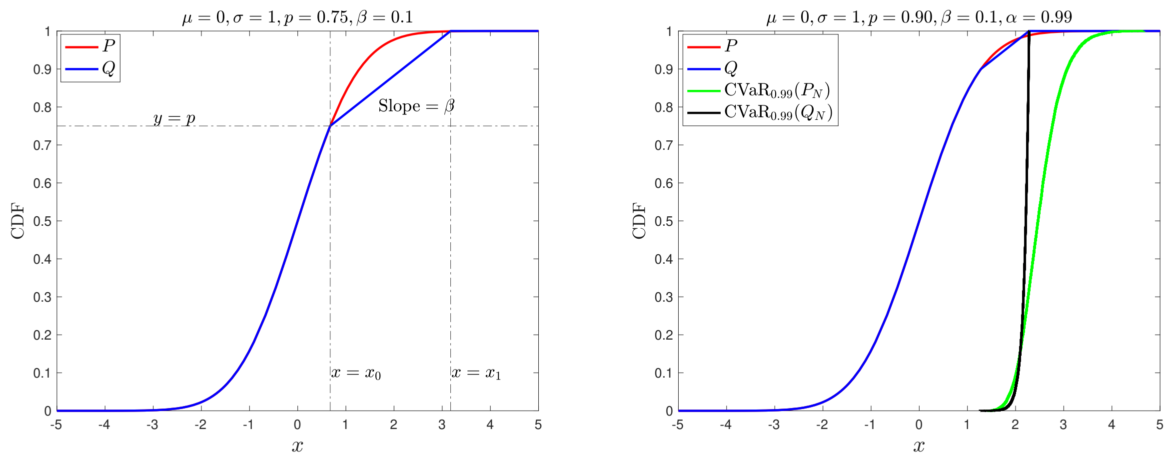
$$P(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt, \quad \forall x \in \mathbb{R}. \tag{47}$$

Since CVaR is a law-invariant tail-dependent risk measure, to illustrate our quantitative statistical robustness, we construct a perturbation $Q(\cdot)$ from $P(\cdot)$ as

$$Q(x) = \begin{cases} P(x), & \text{for } x \le x_0, \\ p + \beta(x - x_0), & \text{for } x_0 \le x \le x_1, \\ 1, & \text{for } x > x_1, \end{cases} \tag{48}$$

where $x_0 = F_P^{-1}(p)$, $x_1 = x_0 + \frac{1}{\beta}(1-p)$ and $p \in (0,1)$. Figure 2(a) depicts the CDFs of P and Q with $\mu = 0, \sigma = 1, p = 0.75$ and $\beta = 0.1$. From Figure 2(a) we can see that P and Q differs only on the tail beyond the pth-quantile.

Next, we use P and Q to randomly generate samples $\xi^1, \ldots, \xi^N$ and $\tilde{\xi}^1, \ldots, \tilde{\xi}^N$ respectively with fixed sample size $N = 100$ and then construct P_N and Q_N via (4) and (5). We compute CVaR with the confidence level $\alpha = 0.99$ based on the generated empirical distributions P_N and Q_N and construct the respective approximated CDFs after running 10000 sampling, see Figure 2(b). From Figure 2(b), we can see that the discrepancy between $\mathrm{CVaR}_{0.99}(P_N)$ and $\mathrm{CVaR}_{0.99}(Q_N)$ is significantly larger than that of P and Q and $\mathrm{CVaR}_{0.99}(Q_N)$ is upper-bounded because $\mathrm{CVaR}_\alpha(Q) \le F_Q^{-1}(1)$ for any $\alpha \in (0,1)$.



(a) **(b)**

Figure 2. **(a)** The CDFs of P and Q. **(b)** The CDFs of $\mathrm{CVaR}_{0.99}(P_N)$ and $\mathrm{CVaR}_{0.99}(Q_N)$.

We now look into the accuracy or tightness of the error bound established in Example 5.2 for CVaR with P and Q defined in (47) and (48). To this effect, we let

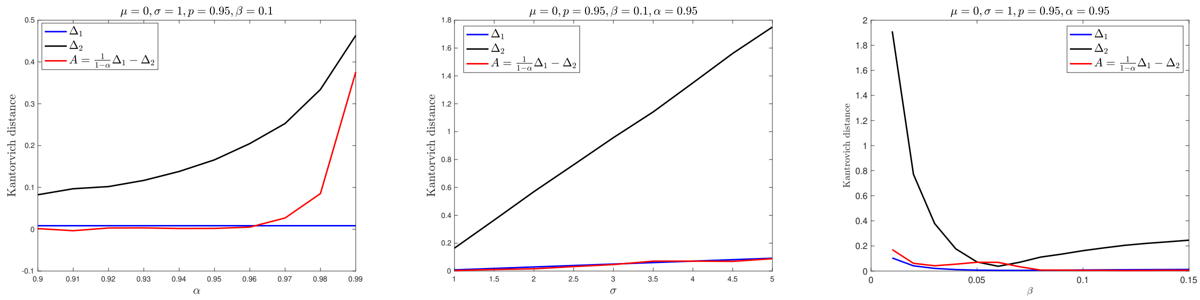
$$\Delta_1 := \mathsf{dl}_K(P^{\otimes N} \circ \mathrm{CVaR}_\alpha(P_N)^{-1}, Q^{\otimes N} \circ \mathrm{CVaR}_\alpha(Q_N)) \quad \text{and} \quad \Delta_2 := \mathsf{dl}_K(P, Q),$$

and quantify the accuracy/tightness of the error bound by

$$A := \Delta_1 - \frac{1}{1-\alpha}\Delta_2, \quad \text{for } \alpha \in (0,1).$$

From Example 5.2, we know that $A \ge 0$ and a smaller A means a tighter bound. We have carried out some static analysis on quantities Δ_1, Δ_2 and A with respect to the variation of (a) confidence level α, (b) variance σ and (c) perturbation parameter β, and depict the results in Figure 3(a)-(c) respectively.

Figure 3(a) shows that for fixed P and Q (determined by μ, σ, p and β), A is small when α is in the range $(0.9, 0.96)$, whereas A is relatively large when $\alpha \in (0.96, 0.99)$. This is partly because $L = \frac{1}{1-\alpha}$ is much larger when α is in the upper band of the confidence and subsequently more conservative



(a) **(b)** **(c)**

Figure 3. **(a)** Change of accuracy w.r.t. variation of α. **(b)** Change of accuracy w.r.t. variation of σ. **(c)** Change of accuracy w.r.t. variation of β.

weight is added to Δ_1 to ensure the upper bound valid. Figure 3(b) shows that A is relatively small for different choice of $\sigma \in (1, 5)$ and Figure 3(c) demonstrates that for $\beta \in (0.075, 0.15)$, A is also small. Based on these analysis, we may conclude that the proposed error bound established in Theorem 4.5 is useful to estimate the discrepancy between laws of risk estimators although more comprehensive analysis may be needed before it can be effectively applied in practice.

7. Concluding remarks

In this paper we take a step further from the existing research on qualitative statistical robustness of law invariant risk measures to quantitative statistical robustness by deriving explicitly an error bound for the laws of statistical estimators and this is achieved by adopting the Kantorovich metric and the Fortet-Mourier metric. Both metrics are easy to compute in practice. During the revision of the paper, we are inspired by one of the referees to show that statistical estimator of any law invariant coherent risk measure on L^1 is quantitatively statistically robust. The error bound gives a guidance on impact of perturbation of the true distribution on the statistical estimators. In particular, if we know the true data are generated by some known distribution but the perceived data might be contaminated, then our result provides an interval for the statistical estimator based on true data if we know the range of the data perturbation (maximum distance between Q and P). We illustrate how this works through CVaR.

There are a number of issues remain to explore. For instance, it might be interesting to investigate whether Theorem 4.5 holds for vector-valued statistical estimators and whether the quantitative robustness results can be extended to more general metrics. We leave all these for future research.
Acknowledgements. The authors would like to thank the two anonymous referees for insightful comments and constructive suggestions which have significantly help us strengthened the paper. They are also thankful to the associate editor for organizing an efficient review.

References

Artzner, P., Delbaen, F., Eber, J.M. and Heath, D., Coherent measures of risk. *Mathematical finance*, 1999, **9**, 203–228.

Belomestny, D. and Krätschmer, V., Central limit theorems for law-invariant coherent risk measures. *Journal of Applied Probability*, 2012, **49**, 1–21.

Ben-Tal, A. and Teboulle, M., An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 2007, **17**, 449–476.

Beutner, E. and Zähle, H., A modified functional delta method and its application to the estimation of risk functionals. *Journal of Multivariate Analysis*, 2010, **101**, 2452–2463.

Boente, G., Fraiman, R., Yohai, V.J. *et al.*, Qualitative robustness for stochastic processes. *The Annals of Statistics*, 1987, **15**, 1293–1312.

Chen, C., Iyengar, G. and Moallemi, C.C., An axiomatic approach to systemic risk. *Management Science*, 2013, **59**, 1373–1388.

Claus, M., Advancing stability analysis of mean-risk stochastic programs: Bilevel and two-stage models. PhD thesis, PhD thesis, University of Duisburg-Essen, 2016.

Claus, M., Krätschmer, V. and Schultz, R., Weak continuity of risk functionals with applications to stochastic programming. *SIAM Journal on Optimization*, 2017, **27**, 91–109.

Cont, R., Deguest, R. and Scandolo, G., Robustness and sensitivity analysis of risk measurement procedures. *Quantitative finance*, 2010, **10**, 593–606.

Delage, E., Kuhn, D. and Wiesemann, W., "Dice"-sion–Making Under Uncertainty: When Can a Random Decision Reduce Risk?. *Management Science*, 2019, **65**, 3282–3301.

Filipović, D. and Svindland, G., The Canonical Model Space for Law-Invariant Convex Risk Measures Is L1. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 2012, **22**, 585–589.

Föllmer, H. and Schied, A., Convex measures of risk and trading constraints. *Finance and stochastics*, 2002, **6**, 429–447.

Föllmer, H. and Weber, S., The axiomatic approach to risk measures for capital determination. *Annual Review of Financial Economics*, 2015, **7**, 301–337.

Gibbs, A.L. and Su, F.E., On choosing and bounding probability metrics. *International statistical review*, 2002, **70**, 419–435.

Guo, S. and Xu, H., Statistical robustness in utility preference robust optimization models. *Submitted to Mathematical Programming*, 2020.

Guo, S., Xu, H. and Zhang, L., Convergence analysis for mathematical programs with distributionally robust chance constraint. *SIAM Journal on optimization*, 2017, **27**, 784–816.

Hampel, F.R., A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, 1971, pp. 1887–1896.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A., *Robust statistics: the approach based on influence functions*, Vol. 196, , 2011, John Wiley & Sons.

Huber, P.J. and Ronchetti, E.M., *Robust statistics*, 2011, Springer.

Inoue, A., On the worst conditional expectation. *Journal of Mathematical Analysis and Applications*, 2003, **286**, 237–247.

Krätschmer, V., Schied, A. and Zähle, H., Qualitative and infinitesimal robustness of tail-dependent statistical functionals. *Journal of Multivariate Analysis*, 2012, **103**, 35–47.

Krätschmer, V., Schied, A. and Zähle, H., Comparative and qualitative robustness for law-invariant risk measures. *Finance and Stochastics*, 2014, **18**, 271–295.

Krätschmer, V., Schied, A. and Zähle, H., Domains of weak continuity of statistical functionals with a view toward robust statistics. *Journal of Multivariate Analysis*, 2017, **158**, 1–19.

Maronna, R.A., Martin, R.D., Yohai, V.J. and Salibián-Barrera, M., *Robust statistics: theory and methods (with R)*, 2019, John Wiley & Sons.

Mattila, P., *Geometry of sets and measures in Euclidean spaces: fractals and rectifiability*, 44, 1999, Cambridge university press.

Mizera, I., Qualitative robustness and weak continuity: the extreme unction. *Nonparametrics and robustness in modern statistical inference and time series analysis: a Festschrift in honor of Professor Jana Jurecková*, 2010, **1**, 169.

Panaretos, V.M. and Zemel, Y., Statistical aspects of Wasserstein distances. *Annual review of statistics and its application*, 2019, **6**, 405–431.

Pflug, G.C. and Pichler, A., Approximations for probability distributions and stochastic optimization problems. In *Stochastic optimization methods in finance and energy*, pp. 343–387, 2011, Springer.

Rachev, S.T., *Probability metrics and the stability of stochastic models*, Vol. 269, , 1991, John Wiley & Son Ltd.

Römisch, W., Stability of stochastic programming problems. *Handbooks in operations research and management science*, 2003, **10**, 483–554.

Ruszczyński, A. and Shapiro, A., Optimization of convex risk functions. *Mathematics of operations research*, 2006, **31**, 433–452.

Strohriegl, K. and Hable, R., Qualitative robustness of estimators on stochastic processes. *Metrika*, 2016, **79**, 895–917.

Wang, W. and Xu, H., Robust spectral risk optimization when information on risk spectrum is incomplete.

SIAM Journal on Optimization, 2020, **30**, 3198–3229.
Zähle, H., Rates of almost sure convergence of plug-in estimates for distortion risk measures. *Metrika*, 2011, **74**, 267–285.
Zähle, H., Qualitative robustness of von Mises statistics based on strongly mixing data. *Statistical Papers*, 2014, **55**, 157–167.
Zähle, H., Qualitative robustness of statistical functionals under strong mixing. *Bernoulli*, 2015, **21**, 1412–1434.
Zähle, H., A definition of qualitative robustness for general point estimators, and examples. *Journal of Multivariate Analysis*, 2016, **143**, 12–31.

Appendix A: Proof of Example 3.1

(i) *The admissible sets induced by the Kantorovich (or Wasserstein) metric* are defined as

$$\begin{aligned}\mathcal{P}_K(\mathbb{R}) &:= \{P \in \mathscr{P}(\mathbb{R}) : \mathsf{dl}_K(P,\delta_0) < +\infty\}\\ &= \left\{P \in \mathscr{P}(\mathbb{R}) : \int_{-\infty}^{0} P(x)dx + \int_{0}^{+\infty} (1-P(x))dx < +\infty\right\} (= \mathscr{P}_1(\mathbb{R}))\\ &= \left\{P \in \mathscr{P}(\mathbb{R}) : \int_{\mathbb{R}} |x| dP(x) < +\infty\right\}\\ &= \mathcal{M}_1^1,\end{aligned}$$

where the second equality follows from the definition of the Kantorovich metric (see, (14)). To see how the third equality holds, we note that for any $t < 0$, we have

$$\begin{aligned}+\infty > \int_{-\infty}^{0} P(x)dx &= \int_{t}^{0} P(x)dx + \int_{2t}^{t} P(x)dx + \int_{-\infty}^{2t} P(x)dx\\ &\geq \int_{t}^{0} P(x)dx + \frac{1}{2}P(2t)|2t|.\end{aligned}$$

Since $-2tP(2t) \geq 0$, then let $t \to -\infty$, then we have $\lim_{t\to-\infty} tP(t) = 0$. Similarly, we have $\lim_{t\to+\infty} t(1-P(t)) = 0$. By using integration-by-parts formula (more precisely (Mattila 1999, Theorem 1.15)), we obtain the right hand side of the third equality. The last equality follows from the definition of ϕ-topology in which case $\phi = |\cdot|$.

(ii) *The admissible sets induced by the weighted Kolmogorov metric* are defined as

$$\begin{aligned}\mathcal{P}_{(\phi)}(\mathbb{R}) &:= \{P \in \mathscr{P}(\mathbb{R}) : \mathsf{dl}_{(\phi)}(P,\delta_0) < +\infty\}\\ &= \left\{P \in \mathscr{P}(\mathbb{R}) : \sup_{x\leq 0} |P(x)\phi(x)| + \sup_{x>0} |(1-P(x))\phi(x)| < +\infty\right\},\end{aligned}$$

which coincides with the set $\mathcal{M}_1^{(\phi)}$ defined in Krätschmer *et al.* (2012, subsection 3.2).

If ϕ is bounded on $\mathbb{R}$, then it is straight that $\mathcal{P}_{(\phi)}(\mathbb{R}) = \mathcal{P}(\mathbb{R})$. In the case when ϕ is unbounded on $\mathbb{R}$, then

$$\mathcal{M}_1^{\phi} \subset \mathcal{P}_{(\phi)}(\mathbb{R}) \subset \bigcap_{\epsilon>0} \mathcal{M}_1^{\phi^{1-\epsilon}}. \tag{A1}$$

The inclusions in (A1) are strict because we can find a counterexample showing equality may fail, see Example B.1 in the appendix. In what follows, we give a proof for (17). Let $P \in \mathcal{M}_1^{\phi}$, since ϕ

is a u-shaped function, then for any $M > 0$ and $N \leq 0$, we have

$$\begin{aligned} +\infty > \int_{\mathbb{R}} \phi(x)dP(x) &= \int_0^{+\infty} \phi(x)dP(x) + \int_{-\infty}^0 \phi(x)dP(x) \\ &\geq \int_0^M \phi(x)dP(x) + \phi(M)(1-P(M)) + \int_N^0 \phi(x)dP(x) + \phi(N)P(N) \\ &\geq \phi(M)(1-P(M)) + \phi(N)P(N), \end{aligned}$$

and consequently $\int \phi(x)dP(x) \geq \sup_{x\leq 0}|P(x)\phi(x)| + \sup_{x>0}|(1-P(x))\phi(x)|$. Thus, $\mathcal{M}_1^\phi \subset \mathcal{P}_{(\phi)}(\mathbb{R})$.

On the other hand, for any $\epsilon > 0$, if we let $\phi_\epsilon(x) := \phi(x)^{1-\epsilon}$ for $x \in \mathbb{R}$, then ϕ_ϵ is a gague function. Moreover, for any $P \in \mathcal{P}_{(\phi)}(\mathbb{R})$, there exists a $k < +\infty$ such that $k = \sup_{x\leq 0}|P(x)\phi(x)| + \sup_{x>0}|(1-P(x))\phi(x)|$. To ease the exposition, we can assume that the law $P(x) > 0$ for any $x \in \mathbb{R}$. Then

$$\phi(x) \leq \frac{k}{P(x)} \text{ for } x \leq 0 \text{ and } \phi(x) \leq \frac{k}{1-P(x)} \text{ for } x > 0.$$

Thus

$$\begin{aligned} \int_{\mathbb{R}} \phi_\epsilon(x)dP(x) &= \int_{-\infty}^0 \phi_\epsilon(x)dP(x) + \int_0^{+\infty} \phi_\epsilon(x)dP(x) \\ &\leq k^{1-\epsilon}\int_{-\infty}^0 \frac{1}{P(x)^{1-\epsilon}}dP(x) + k^{1-\epsilon}\int_0^{+\infty} \frac{1}{(1-P(x))^{1-\epsilon}}dP(x) \\ &= k^{1-\epsilon}\left[\frac{1}{\epsilon}P(x)^\epsilon\right]_{-\infty}^0 - k^{1-\epsilon}\left[\frac{1}{\epsilon}(1-P(x))^\epsilon\right] \\ &= \frac{1}{\epsilon}k^{1-\epsilon}[P(0)^\epsilon - (1-P(0))^\epsilon] < +\infty \end{aligned}$$

which implies $P \in \mathcal{M}_1^{\phi_\epsilon}$. Summarizing the discussions above, we obtain (17).

Appendix B: A counterexample

Example B.1 *In this example, we show that both inclusions in (17) are strict. We first show that* $\mathcal{M}_1^\phi \neq \mathcal{P}_{(\phi)}(\mathbb{R})$, *i.e., there exists a* $P \in \mathcal{P}_{(\phi)}(\mathbb{R})$ *such that* $P \notin \mathcal{M}_1^\phi$. *Let* ϕ *be a unbounded u-shaped function. Then by the continuity of* ϕ, *there exist* $a < 0$ *and* $b > 0$ *with* $\phi(a) = 2 = \phi(b)$. *Let*

$$P(x) = \begin{cases} \frac{1}{\phi(x)}, & \text{for } x \leq a, \\ \frac{1}{2}, & \text{for } a \leq x \leq b\ , \\ 1 - \frac{1}{\phi(x)}, & \text{for } x \geq b. \end{cases}$$

Since $\phi(x) \geq 1$ *for all* x *outside* $[a,b]$, *then* $P(x)$ *is well-defined on* $\mathbb{R}$. *By the monotonicity and unboundedness of* ϕ, *we have* $P \in \mathscr{P}(\mathbb{R})$. *Moreover, since*

$$\sup_{x\leq 0}|P(x)\phi(x)| + \sup_{x>0}|(1-P(x))\phi(x)| = 2,$$

then $P \in \mathcal{P}_{(\phi)}(\mathbb{R})$. *However, by change of variables in integration, we have*

$$\begin{aligned}\int_{\mathbb{R}} \phi(x) dP(x) &= \int_{-\infty}^{a} \phi(x) d\left(\frac{1}{\phi(x)}\right) + \int_{b}^{+\infty} \phi(x) d\left(1 - \frac{1}{\phi(x)}\right) \\ &= \int_{0}^{\frac{1}{2}} \frac{1}{t} dt + \int_{\frac{1}{2}}^{1} \frac{1}{1-t} dt = 2\int_{0}^{\frac{1}{2}} \frac{1}{t} dt = +\infty,\end{aligned}$$

which means $P \notin \mathcal{M}_1^{\phi}$.

Now we show that $\mathcal{P}_{(\phi)}(\mathbb{R}) \neq \bigcap_{\epsilon>0} \mathcal{M}_1^{\phi^{1-\epsilon}}$, *i.e., there exists a* $P \in \bigcap_{\epsilon>0} \mathcal{M}_1^{\phi^{1-\epsilon}}$ *such that* $P \notin \mathcal{P}_{(\phi)}(\mathbb{R})$. *Let* ϕ *be an unbounded u-shaped function. Then there exists an unbounded u-shaped function* ψ *such that* $\lim_{|x|\to+\infty} \psi(x)/\phi(x) = 0$. *More precisely, for any* $\epsilon \in (0,1)$, *there exists an unbounded u-shaped function* ψ *such that*

$$\lim_{|x|\to+\infty} \psi(x)/\phi(x)^{1-\epsilon} = 0. \tag{B1}$$

We construct such ψ *as follows: since* ϕ *is an unbounded u-shape function, then there exist* $a < 0$ *and* $b > 0$ *with* $\phi(a) = e^2 = \phi(b)$. *Let*

$$\psi(x) = \begin{cases} \ln(\phi(x)), & \text{for } x \leq a, \\ 2, & \text{for } a \leq x \leq b, \\ \ln(\phi(x)), & \text{for } x \geq b. \end{cases}$$

Then ψ *is an unbounded u-shaped function and satisfies (B1). Let*

$$P(x) = \begin{cases} \frac{1}{\psi(x)}, & \text{for } x \leq a, \\ \frac{1}{2}, & \text{for } a \leq x \leq b \ , \\ 1 - \frac{1}{\psi(x)}, & \text{for } x \geq b. \end{cases}$$

Since $\psi(x) \geq 1$ *for all* x, *then* $P(x)$ *is well-defined on* $\mathbb{R}$. *By the monotonicity and unboundedness of* ψ, *we have* $P \in \mathscr{P}(\mathbb{R})$.

For fixed $\epsilon \in (0,1)$, *by change of variables in integration, we have*

$$\begin{aligned}\int_{\mathbb{R}} \phi(x)^{1-\epsilon} dP(x) &= \int_{-\infty}^{a} \phi(x)^{1-\epsilon} d\left(\frac{1}{\psi(x)}\right) + \int_{b}^{+\infty} \phi(x)^{1-\epsilon} d\left(1 - \frac{1}{\psi(x)}\right) \\ &= \int_{-\infty}^{a} \phi(x)^{1-\epsilon} d\left(\frac{1}{\ln\phi(x)}\right) + \int_{b}^{+\infty} \phi(x)^{1-\epsilon} d\left(1 - \frac{1}{\ln\phi(x)}\right) \\ &= \int_{0}^{\frac{1}{\ln 2}} e^{\frac{1-\epsilon}{t}} dt + \int_{\frac{1}{\ln 2}}^{1} e^{\frac{1-\epsilon}{1-t}} dt \\ &< +\infty.\end{aligned}$$

Since for $\epsilon \geq 1$, $\phi^{1-\epsilon}$ *is bounded on* $\mathbb{R}$, *then* $\mathcal{M}_1^{\phi^{1-\epsilon}} = \mathscr{P}(\mathbb{R})$. *Thus,* $P \in \bigcap_{\epsilon>0} \mathcal{M}_1^{\phi^{1-\epsilon}}$. *However,*

$$\sup_{x\leq 0} |P(x)\phi(x)| + \sup_{x>0} |(1-P(x))\phi(x)| \geq \sup_{x\leq a} \left|\frac{\phi(x)}{\ln(\phi(x))}\right| + \sup_{x\geq b} \left|\frac{\phi(x)}{\ln(\phi(x))}\right| = +\infty,$$

which means $P \notin \mathcal{P}_{(\phi)}(\mathbb{R})$.

Appendix C: Proof of Proposition 3.3

Proof. Part (i). Since for any $p \geq 1$, $\frac{1}{p}\phi_p \in \mathcal{F}_p(\mathbb{R})$, then by the definition of $\mathcal{P}_p(\mathbb{R})$, we have that $P \in \mathcal{P}_p(\mathbb{R})$ implies $P \in \mathcal{M}_1^{\phi_p}$ and subsequently, $\mathcal{P}_p(\mathbb{R}) \subset \mathcal{M}_1^{\phi_p}$.

On the other hand, let $P \in \mathcal{M}_1^{\phi_p}$, then $\int_{\mathbb{R}} \phi_p(\xi)P(d\xi) < \infty$. For any $\psi \in \mathcal{F}_p(\mathbb{R})$, we have

$$|\psi(\xi) - \psi(0)| \leq \max\{1, |\xi|^{p-1}\}|\xi| \leq \max\{|\xi|, |\xi|^p\}, \text{ for all } \xi \in \mathbb{R},$$

and consequently,

$$\begin{aligned}\left|\int_{\mathbb{R}} \psi(\xi)P(d\xi) - \psi(0)\right| &= \left|\int_{\mathbb{R}} (\psi(\xi) - \psi(0))P(d\xi)\right| \leq \int_{\mathbb{R}} |\psi(\xi) - \psi(0)|P(d\xi) \\ &= \int_{\mathbb{R}} \max\{|\xi|, |\xi|^p\}P(d\xi) \leq \int_{\mathbb{R}} \phi_p(\xi)P(d\xi).\end{aligned}$$

Therefore, we have

$$\sup_{\psi \in \mathcal{F}_p(\mathbb{R})} \left|\int_{\mathbb{R}} \psi(\xi)P(d\xi) - \psi(0)\right| \leq \int_{\mathbb{R}} \phi_p(\xi)P(d\xi) < \infty,$$

and consequently, $\mathcal{M}_1^{\phi_p} \subset \mathcal{P}_p(\mathbb{R})$.

Part (ii). Since $\frac{1}{p}\phi_p \in \mathcal{F}_p(\mathbb{R})$, then for any $P, Q \in \mathcal{P}_p(\mathbb{R})$,

$$\begin{aligned}\left|\int_{\mathbb{R}} \phi_p(\xi)P(d\xi) - \int_{\mathbb{R}} \phi_p(\xi)Q(d\xi)\right| &\leq p\left|\int_{\mathbb{R}} \frac{1}{p}\phi_p(\xi)P(d\xi) - \int_{\mathbb{R}} \frac{1}{p}\phi_p(\xi)Q(d\xi)\right| \\ &\leq p\mathsf{dl}_{FM,p}(P, Q) < +\infty.\end{aligned}$$

From Example 3.1(i) and (12), we have $\mathsf{dl}_{\text{Prok}}(P, Q) \leq \sqrt{\mathsf{dl}_{FM,p}(P, Q)}$. Finally, by the definition of dl_{ϕ_p}, i.e., (7), we obtain the conclusion.

Part (iii) follows straightforwardly from Part (ii). □