

## Article

# The Wasserstein Metric and Robustness in Risk Management

Rüdiger Kiesel <sup>1,\*</sup>, Robin Rühlicke <sup>1</sup>, Gerhard Stahl <sup>2</sup> and Jinsong Zheng <sup>2</sup>

<sup>1</sup> Chair for Energy Trading and Finance, University of Duisburg-Essen, Campus Essen, Universitätsstraße 12, Essen 45141, Germany; robin.ruehlicke@uni-due.de

<sup>2</sup> Group Risk Management, Talanx AG, Riethorst 2, Hannover 30659, Germany; gerhard.stahl@talanx.com (G.S.); jinsong.zheng@talanx.com (J.Z.)

\* Correspondence: ruediger.kiesel@uni-due.de; Tel.: +49-201-183-4963

Academic Editor: Mogens Steffensen

Received: 13 May 2016; Accepted: 15 August 2016; Published: 31 August 2016

**Abstract:** In the aftermath of the financial crisis, it was realized that the mathematical models used for the valuation of financial instruments and the quantification of risk inherent in portfolios consisting of these financial instruments exhibit a substantial model risk. Consequently, regulators and other stakeholders have started to require that the internal models used by financial institutions are robust. We present an approach to consistently incorporate the robustness requirements into the quantitative risk management process of a financial institution, with a special focus on insurance. We advocate the Wasserstein metric as the canonical metric for approximations in robust risk management and present supporting arguments. Representing risk measures as statistical functionals, we relate risk measures with the concept of robustness and hence continuity with respect to the Wasserstein metric. This allows us to use results from robust statistics concerning continuity and differentiability of functionals. Finally, we illustrate our approach via practical applications.

**Keywords:** risk management; Wasserstein metric; robustness; risk measures

**MSC:** 91B30; 62G35; 60B10; 60F05

## 1. Introduction

In the aftermath of the financial crisis, it was realized that the mathematical models for financial markets carry a substantial amount of model risk. Ref. [1], with a discussion from a banking point of view, and [2], who provide an overview with a focus on systemic risk, investigate various aspects of model risk, while [3,4] give some early examples. Accounting for model risk is particularly important for the forecast of risk inherent to portfolios consisting of financial instruments as this forecast is used to calculate economic and regulatory capital within financial institutions. Hence, it is no surprise that if a financial institution opts (e.g., under Solvency II or Basel III) for an application of an internal model, its stakeholders, such as regulators and rating agencies, require that such internal models are *reliable*. For that reason, the EU directive for Solvency II imposes explicitly that internal models have to meet *statistical quality*, *calibration*, and *validation standards* and have to pass the *use test* (see Articles 120, 121, 122 and 124, respectively, of the Solvency II directive [5]). By means of these quantitative and qualitative requirements, it is intended to achieve a high level of comfort with respect to the reliability of internal models. Furthermore, regulators require explicitly the *robustness* of a model as an overarching *modelling principle* (see [6]). This is especially important because both internal and external model users, who include the results of the models in their decision-making process, are often non-experts in stochastic models. In this respect, robustness of models is a necessary condition to introduce a sustainable *risk culture*.

Motivated by the above, we pose the problem of a consistent robust modeling approach by considering an axiomatic setting for monetary risk measures that ensures robustness on the one hand and emphasizes the *approximative nature* of internal models on the other hand by making the need for formulating a suitable *distance* explicit. Thus, we focus on the statistical robustness of the risk measurement procedure and in doing so complement recent approaches to evaluate model risk as in [7] or to use statistical concepts to evaluate the performance of risk measures as suggested in e.g., [8] or [9].

Within our approach, we consider risk functionals, which treat risk measures on the level of distributions, and allow so to exploit established statistical theory. Risk functionals are typically not continuous with respect to the weak topology as this topology ignores the tails of the distribution. Ref. [10] pointed out this fact and suggested to use a stronger topology. We will thus use robustness in the sense of continuity with respect to an appropriate metric, namely the Wasserstein metric, which was introduced in a general setting in [11] and in the financial context in [12] (p. 79), as the uniform metric between inverse distribution functions. Our approach is in line with [13,14], who provide extensions and introduce an index of qualitative robustness, and the approaches suggested by [15], who prefer robust methods for estimating parameters (Value-at-Risk, expectations, etc.) of forecast distributions. Our framework allows for utilisation of several results that have been established recently in [16].

In addition, we provide various real-life examples to support and to emphasise the applicability of our results. We thus extend and clarify results in [17], who provide real-life examples that highlight the need for robust models for both valuation and calculation of risk in order to control model risk. In addition, in an insurance context [18] recently, applied robust methods to the estimation of loss triangles.

## 2. A Suitable Distance for Risk Management

We consider monetary risk measures that are related to capital requirements. These risk measures,  $\rho$ , are typically introduced for (random) financial positions  $X$  defined on an appropriate probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , by

$$\rho : L^\infty(\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}, \quad (1)$$

where  $L^\infty$  denotes the space of real valued bounded functions.

We refer the reader for definitions and the various variants of monetary risk measures to [19,20]. Important for us is the contribution by [21], who introduced *natural risk statistics*. This concept bridges the gap between risk measures and statistics by defining natural risk statistics on a *sample space* in contrast to the definition of risk measures on a probability space (1).

In situations when we are dealing with possibly unbounded financial positions  $X$ , the space  $L^\infty$  is not suitable. Following the observation of [22] that Orlicz spaces or Orlicz hearts (see Proposition 2 below) are appropriate in this case, [16] show that for  $p \in [0, \infty)$ ,  $L^p$  is an Orlicz heart and thus can serve as a space of financial positions. In fact, based on [23], Theorem 2.10 of [16] shows that law-invariant convex risk measures on  $L^\infty$  has a unique extension to  $L^1$  (convex, monotone, and lower semicontinuous with respect to the  $L^1$ -norm), which inherits continuity properties.

In order to discuss the robustness (estimation) of risk measures, we need the following concept:

**Definition 1.** *Statistical functional.* Let  $\mathbb{F}$  denote the set of all cumulative distribution functions on  $\mathbb{R}$  and  $\bar{\mathbb{F}} \subset \mathbb{F}$  an appropriate subset. We consider maps

$$T : \bar{\mathbb{F}} \subseteq \mathbb{F} \longrightarrow \mathbb{R}. \quad (2)$$

Now, let  $X_1, \dots, X_n$  be a sample from a population with distribution function  $F \in \bar{\mathbb{F}}$  and empirical distribution  $F_n$ . If a statistic  $T_n = T_n(X_1, \dots, X_n)$  can be written as a functional  $T$  of  $F_n$ , say  $T_n = T(F_n)$ , then we will call  $T$  a statistical functional.

This is particularly important as we assess the risk  $\rho(X)$  of a financial position  $X$  with distribution function  $F$ . Typically, we will use a Monte Carlo simulation or a sequence of historical data. For law-invariant  $\rho$ , a natural estimate of  $\rho(X)$  is  $T_\rho(F_n)$ , where  $F_n$  is the empirical distribution function and

$$T_\rho(F) = \rho(X), \quad \text{for } X \text{ with distribution function } F. \quad (3)$$

Now, robustness of the risk measure  $\rho$  can be related to continuity of  $T_\rho$ . Here, robustness refers to a certain insensitivity of  $\rho$  to variations in the approximating distribution for  $F$ . For this, the concept of continuity is crucial.

Note that any statistical functional  $T(F)$  can be approximated with  $T(F_n)$ , provided that  $F_n$  is in a neighborhood of  $F$  and the functional  $T$  is sufficiently well behaved. This leads us to consider  $F$  as a “point” in  $\mathbb{F}$  and to investigate continuity, differentiability and other regularity properties for functionals defined on (subsets of)  $\mathbb{F}$ .

In order to define these concepts properly, we need to introduce:

**Definition 2.** *Simple probability metric.* Let  $\bar{\mathbb{F}} \subset \mathbb{F}$  be a convex class of distribution functions on  $\mathbb{R}$  containing all degenerate distributions. A mapping

$$d : \bar{\mathbb{F}} \times \bar{\mathbb{F}} \rightarrow [0, \infty)$$

is called a simple probability metric (or distance) if

1.  $F(x) = G(x)$  for all  $x \in \mathbb{R}$  iff  $d(F, G) = 0$ ,
2.  $d(F, G) = d(G, F)$  for all  $F, G \in \bar{\mathbb{F}}$ ,
3.  $d(F, G) \leq d(F, H) + d(G, H)$  for all  $F, G, H \in \bar{\mathbb{F}}$ .

Unfortunately, there are plenty of possible metrics to choose. In order to select a specific metric as a preferable one from a practitioner’s point of view, two questions have to be answered:

1. Does the chosen metric fit well in the context of application, i.e., has the chosen metric a natural interpretation? Is it perhaps the canonical metric for a particular field of application?
2. Is the metric not too strong? This means—do we have a sufficiently rich set of continuous objects?

The following example lists very common risk measures that should fit into the risk measurement framework from a practitioner’s point of view.

**Example 1.** *Examples of statistical functionals and risk measures:* The central piece of many (internal) risk management models is risk measures such as Value-at-Risk (VaR) and Tail-VaR (TVaR). We give representations of these important risk measures as statistical functionals. Let  $F_n$  be the empirical distribution function related to a given data set. In addition, recall that, for a cumulative distribution function  $F$ , the (upper) quantile of order  $\alpha \in (0, 1)$  is defined as  $F^{-1}(\alpha) = \inf\{x \in \mathbb{R} : F(x) > \alpha\}$ .

- **Value-at-Risk.** For a fixed  $0 < \alpha < 1$  (usually  $\alpha \leq 0.05$ ), we define the Value-at-Risk at level  $\alpha$  ( $\text{VaR}_\alpha$ ) as

$$\text{VaR}_\alpha = -F^{-1}(\alpha). \quad (4)$$

The statistical functional for estimating Value-at-Risk is then

$$T(F_n) = -F_n^{-1}(\alpha) = - \int_0^1 F_n^{-1}(\tau) d\delta_\alpha(\tau),$$

where  $\delta_\alpha$  puts mass one to  $\alpha \in (0, 1)$ .

- **Tail-VaR (TVaR).**<sup>1</sup> The statistical functional for estimating TVaR is

$$T(F_n) = -(1 - \alpha)^{-1} \int_{\alpha}^1 F_n^{-1}(x) dx. \quad (5)$$

If we substitute  $F$  for  $F_n$  in the examples above, we obtain the associated *risk measures*. In particular, any law-invariant<sup>2</sup> risk measure can be written as a statistical functional (see [19] for further discussion). For the VaR, this yields  $T(F) = -F^{-1}(\alpha)$ . From (4) and (5), we conclude the important role of the empirical distribution for determining risk measures. Hence, a metric for risk management should be based on the whole cumulative distribution function.

We now turn to the question of robustness of risk measures in more detail. Given that we want to use the empirical distribution in praxis, we need to specify how to quantify the effect of a small deviation from the true distribution on the risk measure. We need the following definition:

**Definition 3.** *Continuity of statistical functionals.* Let  $\bar{\mathbb{F}} \subset \mathbb{F}$  be a convex class of distribution functions on  $\mathbb{R}$  containing all degenerate distributions. A functional  $T$  defined on  $\bar{\mathbb{F}}$  is continuous at  $F \in \bar{\mathbb{F}}$  if

$$T(G) - T(F) = o(1) \text{ as } d(G, F) = o(1), \quad F, G \in \bar{\mathbb{F}},$$

where  $d(G, F)$  denotes a distance of two distribution  $G$  and  $F$ .

In the axiomatic approach of [24], the continuity of a risk measure is not required. Consequently, the class of coherent risk measures includes discontinuous risk measures such as the TVaR. Given the approximative character of stochastic models, we think that continuity, and, hence, robustness, are, from a practitioner's point of view, even more important than subadditivity.

**Example 2.** *TVaR is not continuous w.r.t. the weak topology:* Consider the sequence

$$G_n = \frac{n-1}{n} U[0, 1] + \frac{1}{n} \mathbf{1}_{n^2},$$

where  $U[0, 1]$  denotes the uniform distribution on  $[0, 1]$  and  $\mathbf{1}_{n^2}$  puts its mass one to  $n^2$ ,  $n \in \mathbb{N}$ . Obviously,  $G_n$  converges weakly to  $U[0, 1]$ :

$$G_n \xrightarrow{w} U[0, 1].$$

However, the associated TVaRs do not. For all  $\alpha$ , we have:

$$TVaR_{\alpha}(G_n) = \int_{-VaR_{\alpha}}^{\infty} x dG_n > n,$$

whereas

$$TVaR_{\alpha}(U[0, 1]) = \int_{-VaR_{\alpha}}^1 x dx \leq 1$$

is finite.

The above example clearly highlights the fact that risk measures, which put emphasis on the tail of a distribution cannot be expected to be continuous with respect to topologies that ignore the tails of distributions. In contrast, it can be shown that the VaR at level  $\alpha$  is continuous w.r.t the weak topology (see Theorem 3.7 in [25]) at  $F \in \mathbb{F}$  if  $F^{-1}$  is continuous at  $\alpha$ . Given that VaR and TVaR are the risk

<sup>1</sup> TVaR is also referred to as Expected Shortfall (ES)

<sup>2</sup> [15] calls them distribution-free

measures that are of overwhelming importance in practice, we suggest using a metric that implies the continuity of the associated functionals.

We will now discuss properties of a metric that are particularly useful for the risk management process. In terms of practical applications, it is important to strike a balance between properties in the centre of a distribution and its tail behaviour. Thus, as outlined in Example 1, the following requirements are natural in order that the statistical functionals allow for estimating all of the risk measures given in Example 1 consistently:

$$F_n \xrightarrow{w} F \quad \text{and} \quad \int_{\mathbb{R}} x^p dF_n(x) \longrightarrow \int_{\mathbb{R}} x^p dF(x) \quad (6)$$

for some  $p \geq 0$  and  $F_n \in \bar{\mathbb{F}} \subset \mathbb{F}$ . Note that the linear functional  $\int_{\mathbb{R}} x^p dF(x)$  is now continuous in  $F$  w.r.t. a given metric. For applications in risk management, the cases  $p = 1$  and  $p = 2$  are of particular importance. Thus, we are looking for a metric that combines these two properties.

We will show that with the Wasserstein metric (especially, the case  $p = 2$ ), both types of risk measures are considered: one-tail measures (such as VaR) and two-tail measures (such as volatility). Hence, this approach comprises both concepts.

We will only need the Wasserstein metric for two distributions  $F, G$  on the real line.

**Definition 4.** *Wasserstein Metric.* Let  $X \sim F, Y \sim G$  random variables with their corresponding distribution functions. The Wasserstein distance,  $W_p(X, Y) = W_p(F, G)$  for two distributions is given by:

$$W_p(X, Y) = W_p(F, G) = \int_0^1 |F^{-1}(u) - G^{-1}(u)|^p du, \quad p \geq 1. \quad (7)$$

For a general definition of the Wasserstein metric in Banach spaces, the reader will find appropriate definitions in [11]. Refs. [26,27] show the equivalence of the above definition with the general case. An example for the use of the Wasserstein metric in the context of investments is [28].

**Proposition 1.** *Properties of Wasserstein Metric.* The Wasserstein metric has the following properties:

1. For  $X, Y$  random variables we have the following scaling properties of  $W_p$ :

$$W_p(aX, aY) = |a|W_p(X, Y), \quad \text{for any scale } a. \quad (8)$$

The scaling property (8) is important for risk management because it allows the change of numéraire, e.g., change of home currency.

2. For  $F_1, F_2, G_1, G_2$  in  $\Gamma_2$  and  $\epsilon \in (0, 1)$

$$W_2(\epsilon F_1 + (1 - \epsilon)F_2, \epsilon G_1 + (1 - \epsilon)G_2) \leq \epsilon W_2(F_1, G_1) + (1 - \epsilon)W_2(F_2, G_2). \quad (9)$$

If  $\epsilon = 0.5$  and  $G_1 = F_1$ , then we have the regularity property:

$$W_2(F_1 + F_2, F_1 + G_2) \leq W_2(F_2, G_2). \quad (10)$$

3. Let  $U_j, V_i, i, j = 1, \dots, m$  be independent and assume that the laws are in  $\Gamma_2$ . Then,

$$W_2\left(\sum_{i=1}^m U_j, \sum_{j=1}^m V_j\right) \leq \sum_{j=1}^m W_2(U_j, V_j). \quad (11)$$

This property is key to proving the consistency for bootstrapping mean statistics (see [11]).

4. The Wasserstein metric is a one-ideal metric (see the Appendix A for a definition of one-ideal).

**Example 3.** *Wasserstein and  $L^p$  distance:* For  $F, G$  distribution functions, the  $L^p$ -distance,  $\Theta_p(F, G)$  is defines as:

$$\Theta_p(F, G) = \left( \int_{-\infty}^{\infty} |F(u) - G(u)|^p du \right)^{1/p}, \quad p \geq 1. \quad (12)$$

We compare this distance with the Wasserstein metric for a standard normal distribution  $\Phi$ , which we perturbate such that there is a difference in the right tail. Thus, let

$$G(x) = (1 - \varepsilon)\Phi(x) + \varepsilon \text{Exp}(x; \lambda),$$

where  $\text{Exp}(x; \lambda)$  is an exponential distribution. Table 1 shows that the Wasserstein distance increases in  $p$  while the  $L^p$  distance decreases. Thus, the weighting of differences in the tail increase for the Wasserstein metric in  $p$ , while it decreases for  $L^p$ .

**Table 1.** Wasserstein and  $L^p$  distance,  $\varepsilon = 0.1, \lambda = 0.5$ .

<b>p</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
$W_p$	0.054	0.067	0.075	0.085
$\Theta_p$	0.054	0.039	0.038	0.038

The convexity properties are closely related to the independence axiom of the Von-Neumann-Morgenstern approach to utility theory. Therefore, the Wasserstein metric fits smoothly into this framework. As the results of the risk measuring exercise need to have an impact on decisions within the company (the use test in Solvency II) and as utility theory is widely used in this context, this is a very desirable property.

Recently, [16] have investigated the robustness properties of convex risk measures, in particular regarding the Wasserstein metric.

**Proposition 2.** [16] (Theorems 2.5 & 2.7). *Let  $X_1, X_2, \dots \in L^p(\Omega, \mathcal{F}, \mathbb{P})$  be a stationary and ergodic sequence of random variables with the same law as  $X \in H^p, p \in [1, \infty)^3$  with*

$$H^p(\Omega, \mathcal{F}, \mathbb{P}) = \left\{ X \in L^0 : E[|cX|^p / p] < \infty \quad \forall c > 0 \right\}.$$

*Then, for a law-invariant convex risk measure  $\rho$  with corresponding statistical functional  $T_\rho$ , the estimator*

$$\hat{\rho}_n = T_\rho \left( \frac{1}{n} \sum_{i=1}^n \delta_{X_i} \right)$$

*is consistent, that is,  $\hat{\rho}_n \rightarrow \rho(X)$  as  $n \rightarrow \infty$   $\mathbb{P}$ -a.s., and continuous with respect to the  $p$ -Wasserstein metric<sup>4</sup>.*

The proposition implies, for instance, the continuity of the TVaR w.r.t. the Wasserstein metric.

Many statistical functionals are so-called  $L$  - functionals (see [21] for a discussion of the role of  $L$ -statistics in terms of natural risk measures). We can relate them to the Wasserstein-metric with the following:

<sup>3</sup> The space  $H^p$  is called Orlicz heart, see [22]. Actually, as the so-called  $\Delta_2$  condition is satisfied, the Orlicz heart  $H^p$  coincides with  $L^p$  in the current setting.

<sup>4</sup> This is a consequence of the fact that the  $\Delta_2$  condition is satisfied, see [16].

**Theorem 1.** *Continuity of L-statistical functional w.r.t the Wasserstein metric. Let  $T$  be an L-statistical functional of the form*

$$T(G) = \int_0^1 F^{-1}(t)J(t)dt = \int_{-\infty}^{\infty} x dK(G(x)), \quad G \in \mathbb{F} \quad (13)$$

for  $\int_0^1 F^{-1}(t)J(t)dt < \infty$ ,  $K(t) = \int_0^t J(u)du$ , and  $J$  a weights generating function bounded on  $[0,1]$ . Then,  $T$  is continuous w.r.t to Wasserstein metric  $W_p$ ,  $p \in [1, \infty)$ .

**Proof.** The claim follows as a special case from [16], Proposition 2.22. We provide an independent proof in the Appendix B.  $\square$

Theorem 1 implies the continuity of mean, trimmed mean, and TVaR w.r.t. the Wasserstein metric.  $\square$

As we discussed above, we extended the classical concept of qualitative robustness, which is based on the weak topology, by using a more suitable topology, and therefore we made it useful for tail sensitive functionals such as most risk measures. Ref. [16] propose the alternative notion of  $\psi$ -robustness. Here,  $\psi : \mathbb{P}_+ \rightarrow [0, \infty]$  is a left-continuous, nondecreasing convex (Young) function, which we call weight function in the following. We denote by  $\mathcal{M}_1$  the set of all probability measures<sup>5</sup> on  $\mathbb{R}$  and by  $\mathcal{M}_1^\psi \subset \mathcal{M}_1$  the set of all probability measures  $\mu$  on  $\mathbb{R}$  with  $\int_{\mathbb{R}} \psi d\mu < \infty$ . Additionally, we have to impose a product structure on the underlying probability space, setting  $\Omega = \mathbb{R}^N$ ,  $X_i(\omega) = \omega(i)$  for  $\omega \in \Omega$ ,  $i \in \mathbb{N}$ ,  $\mathcal{F} = \sigma\{X_1, X_2, \dots\}$ , and, for  $\mu \in \mathcal{M}_1$ ,  $\mathbb{P}_\mu = \mu^{\otimes N}$ .

**Definition 5.** *Qualitative robustness and  $\psi$ -robustness. Let  $\mathcal{N} \subset \mathcal{M}_1$ , and let  $d_A$  and  $d_B$  be metrics on  $\mathcal{N}$  and  $\mathcal{M}_1$ , respectively. The functional  $T_\rho$  is called (qualitatively)  $\mathcal{N}$ -robust with respect to  $d_A$  and  $d_B$ , if for all  $\mu \in \mathcal{M}_1$  and  $\epsilon > 0$ , there exist  $\delta > 0$  and  $n_0 \in \mathbb{N}$  such that*

$$\nu \in \mathcal{N}, d_A(\mu, \nu) \leq \delta \Rightarrow d_B(\mathbb{P}_\mu \circ \hat{\rho}_n^{-1}, \mathbb{P}_\nu \circ \hat{\rho}_n^{-1}) \leq \epsilon \text{ for } n \geq n_0.$$

Let  $\psi$  be a weight function. A risk functional  $\mathcal{R}_\rho$  is called  $\psi$ -robust on  $\mathcal{M} \subset \mathcal{M}_1$  if  $\mathcal{R}_\rho$  is robust with respect to  $d_A$  and  $d_B$  on all subsets  $\mathcal{N} \subset \mathcal{M}$ , which are uniformly integrating, that is

$$\lim_{M \rightarrow \infty} \sup_{\nu \in \mathcal{N}} \int_{\psi \geq M} \psi d\nu = 0.$$

The key inside (already pointed out in [10]) is that, when insisting on distances that generate the usual weak topology of measures, one can have distributions that are rather close in the weak topology but will have completely different tail behaviour. Thus, the notion of robustness should allow an insensitivity with respect to tail behaviour. As [16] point out, the Prohorov distance

$$d_{\text{Proh}}(\mu, \nu) := \inf \{ \epsilon > 0 : \mu(A) \leq \nu(A^\epsilon) + \epsilon \text{ for all } A \in \mathcal{B}(\mathbb{R}) \} \quad (14)$$

and the Prohorov  $\psi$ -metric

$$d_\psi(\mu, \nu) := d_{\text{Proh}}(\mu, \nu) + \left| \int \psi d\mu - \int \psi d\nu \right|, \quad \mu, \nu \in \mathcal{M}_1 \quad (15)$$

are appropriate choices. Ref. [16] (Theorem. 2.14 rso Theorem 2.25) characterize robust risk measures resp. show that it is sufficient to study law-invariant convex risk measures on  $\mathcal{M}(L^\infty)$ .

<sup>5</sup> We restrict ourselves to measures on the real line.



A special case for  $\psi_p = |x|^p/p$ , the function that is also implicitly used in Proposition 2, allows for stating the result for the  $p$ -Wasserstein metric.

**Proposition 3.** Every law-invariant convex risk measure  $\rho : L^p \rightarrow \mathbb{R}$ ,  $p \in [1, \infty)$  is  $\psi_p$ -robust with respect to the  $p$ -Wasserstein metric.

In particular, the proposition shows that the TVaR is  $\psi_p$ -robust for any  $p \geq 1$ . The value at risk is qualitatively robust even in the classical sense, that is, with  $d_A$  and  $d_B$  equal to the Prohorov metric and  $\mathcal{N} = \mathcal{M}_1$  (see ([14], Example 4.3)).

**Example 4.** Wasserstein metric for  $p=1$ : for heavy tailed distributions as they appear in finance, the mean absolute deviation is preferable, due to its robustness properties. Hence,

$$\int |x - F^{-1}(0.5)| dF,$$

where  $F^{-1}(0.5)$  denotes the median, is a suitable substitute for  $E(X^2)$ , and thus for volatility-sensitive risk measures, in (6) which is much more in line with our overall robustness approach. In requiring

$$F_n \xrightarrow{w} F \quad \text{and} \quad \int |x - F_n^{-1}(0.5)| dF_n \longrightarrow \int |x - F^{-1}(0.5)| dF. \quad (16)$$

We are in the  $L^1$  setting of the Wasserstein framework. Note that for the  $L^1$  case, we have:

$$d_{L^1}(F_n, F) = \int_{-\infty}^{\infty} |F_n - F| dx = \int_0^1 |F_n^{-1} - F^{-1}| dx = d_{\ell_1}(F_n, F). \quad (17)$$

This may be used with advantage, when we consider Fréchet differentiable functions. Following [29], we consider the following norms:

$$\begin{aligned} \|F - G\|_{\infty+p} &= \|F - G\|_{\infty} + \|F - G\|_p \\ &= \sup_x |F(x) - G(x)| + \left( \int |F - G|^p dx \right)^{1/p}. \end{aligned} \quad (18)$$

For  $p = 1$ , this means that if a statistical functional is strongly differentiable w.r.t supremum norm, then it is automatically differentiable with respect to the sum of the Wasserstein and supremum norm.

This norm implies, for statistical functionals which satisfy a Fréchet differentiability condition, that jackknifing and bootstraps yield consistent estimators.

### 3. Application of the Wasserstein Metric in Risk Management

This section is devoted to exemplifying the importance of the Wasserstein Metric by analyzing application from real-life risk management.

**Example 5.** Consistency of summary statistics and estimated capital requirement: nearly any statistical analysis starts with the computation of a table of summary statistics comprising means such as the expectation, standard deviation, skewness and kurtosis, and quantiles such as the interquantile range and other characteristics.

Of course, the minimum requirements for interpreting these statistics would be that the statistics are consistent estimators for the associated parameter. Obviously, for means and quantile types of statistics, a convergence with respect to the Wasserstein distance would imply consistency. Hence, also for statistical analysis, the Wasserstein metric is quite natural for bridging the gap between descriptive statistics and inferential type approaches. This is very much in line with the approach advocated by [30].



To see the importance of consistency in regulation, we consider an important risk measure under Solvency II: the Solvency Capital Required, denoted by SCR. Note that SCR is defined by a difference<sup>6</sup> of a quantile and an expectation, which we can write as a difference of statistical functionals i.e.,

$$\text{SCR} := E(X) + \text{VaR}_\alpha(X) \quad (19)$$

$$= \int_0^1 F_n^{-1}(x) dx - \int_0^1 F_n^{-1}(\tau) d\delta_\alpha(\tau) \quad (20)$$

$$= T_1(F_n) - T_2(F_n) = S(F_n). \quad (21)$$

In the light of (19), the Wasserstein Metric for  $p = 1$  is sufficient to ensure the continuity of the functional  $S$  at  $F$ :

$$S(F_n) \rightarrow S(F), \quad (22)$$

if

$$W_1(F_n, F) \rightarrow 0. \quad (23)$$

**Example 6.** *Robustness and materiality:* the implementation of internal models calls for a concept of materiality that allows for working with various proxies. In light of Example 2, the sequence of distributions  $(G_n)$  seems to provide a good approximation for the uniform distribution  $U$ , due to the fact that

$$d(G_n, U) \leq \frac{1}{n^2}, \quad (24)$$

w.r.t. the weak topology; roughly speaking: the error is arbitrarily small, hence immaterial. However, neither the TVaR nor the expectation is a continuous functional at  $U$  with respect to the weak convergence. If the SCR is based on (19)—or if we substitute  $\text{VaR}_\alpha$  by  $\text{TVaR}_\alpha$ , as it is proposed under Basel III—the SCR is difficult to be used as a yardstick to gauge the materiality of proxies, if the considered risk functional is not continuous or respectively robust. However, the choice of the Wasserstein metric for  $p = 1$  allows for judging the materiality of the error. Note that from a mathematical point of view, materiality is closely related to robustness and continuity.

**Example 7.** *Test statistic consistently chosen to the natural distance:* in light of the previous examples, it is natural to use the Wasserstein Metric for  $p = 1$  in the context of risk management. In the process of the application of an internal model, a number of statistical analyses are necessary, especially in the validation process (see Article 124 of the Solvency Directive [5]).

Due to simplicity in practical risk management applications, tools in software packages available are often applied for statistical tests. A point in case is the application of the Kolmogoroff–Smirnow test statistic, which is based on the Kolmogoroff–Smirnow distance as defined in (25):

$$d_K(F_n, F) = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|. \quad (25)$$

However, from a principal point of view, it might be questioned whether this is a suitable test statistic for a given problem at hand.

Again, we advocate the application of the Wasserstein distance ( $p \geq 1$ ) as a test statistic. The distribution function of this test may be determined by the application of the bootstrap. For a sample of size  $n = 10000$ , Table 2 contains the distances

$$W_1(F_n(x), F_{\hat{\theta}}) \quad (26)$$

<sup>6</sup> of course this is due to the fact that we consider the loss in the VaR definition.

between the empirical distribution function,  $F_n(x)$ , and some estimated parametric models  $F_{\hat{\theta}}$ , namely the Normal, Gamma and Generalized Extreme Value distribution.

Real life examples have shown that results based on the application of the Kolmogoroff–Smirnow test yield very different results comparable to those based on the Wasserstein Metric. Due to the lengthy exposition, however, we omitted these examples in this exposition.

Instead, we consider in the following situation only the output of a Monte-Carlo simulation from an internal model in terms of the Wasserstein metric. Though Monte-Carlo simulations have a number of advantages, it is often very cumbersome to figure out relationships. For that reason, we approximate the output by some parametric densities. Table 2 shows for various companies (denoted by  $E_i$ ) the Wasserstein distance between the empirical distribution function and fitted parametric models as outlined above. From this table, we may conclude which parametric model might be chosen per company.

This table may be understood in analogy to those cases where the well-known Kolmogoroff distance is applied and used as a test statistic. From Table 2, we might conclude that the Normal distribution as a simple fitting does not perform badly from an overall perspective. Of course, the Gamma distribution seems to be preferable; however, it has more parameters than the Normal distribution. The performance of the Generalized Extreme Value (GEV) distribution is very volatile; it performs well in some cases, very badly in others. Of course, it would be possible to support these judgements by some formal tests by bootstrapping the Wasserstein distance. However, given the case that the simulation is based on 10,000 scenarios, the result is quite clear.

**Table 2.** Wasserstein metric, multiplied by 100.

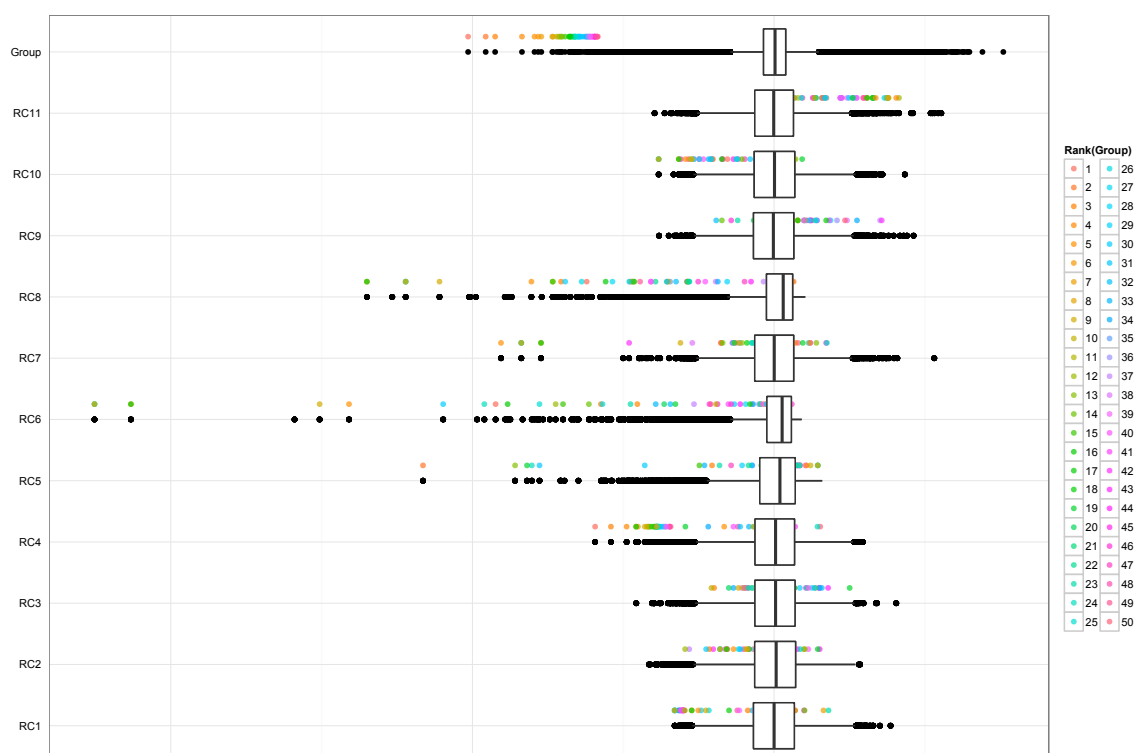
	Normal	Gamma	GEV
$E_1$	5.14	4.08	14.36
$E_2$	4.74	1.55	1.18
$E_3$	5.17	1.78	1.27
$E_4$	4.95	1.59	1.60
$E_5$	6.41	5.45	9.22
$E_6$	9.90	5.92	1.15
$E_7$	7.55	4.53	1.39

**Example 8.** *Is the focus of risk management just on the tails? Typically, risk management is focused very much on the tails of distributions, where very adverse events are modeled. The risk measures  $\text{VaR}_\alpha$ ,  $\text{TVaR}$  and all the progress made in the application of Extreme Value Theory over the last decade confirms this attitude. We think this is completely right; however, the following real-life example highlights that tails alone do not tell the whole story.*

This example is taken from the aggregation of risks. Aggregation in itself is of highest importance due to the measurement of the amount of diversification induced by the considered method of aggregation.

The aggregation considered in Figure 1 is based on a so-called factor approach. In this framework,  $n$  scenarios of a common set of risk factors related to a number of risk categories are simulated. The result is a  $n \times d$  matrix. For each column, the portfolio is evaluated and yields  $n$  potential changes of the considered portfolio.

The first box plot of Figure 1 is related to an insurance company, whereas the other eleven ones belong to risk categories such as market risk, natural catastrophes and others. With coloured points, the fifty most adverse scenarios or events are depicted. It is important to note that, for the risk categories, these fifty scenarios are related to realizations that are not necessarily located in the tail. It is interesting to see that these fifty scenarios are spread across the whole body of the distributions involved. In order to build an adequate risk model, it is of crucial importance to ensure that the whole distribution function is correctly specified and not only their tails. The Wasserstein distance fits very well to these needs because it averages the quantiles of a distribution as given in (4).



**Figure 1.** The sequence of box plots and related outliers represent distributions from eleven risk categories that are aggregated and displayed in the first box plot from above. The coloured points above the box plot for the aggregate depict the fifty worst case scenarios. These worst case scenarios are traced back to the risk categories by coloured dots. As the graphs show, these fifty worst cases for the aggregate are not necessarily located in the tail for the contributing risk categories. Hence, the risk management cannot only focus on the tails but on the whole distribution function. The Wasserstein distance fits to this need in a very particular way.

**Example 9.** *Ranking opportunities: in financial applications, one is often requested to choose one investment out of several ones. This might be achieved by the means a stochastic ordering of the random variables associated with the investment possibilities.*

An important approach for ordering is stochastic dominance. Let  $X$  and  $Y$  be random variables with distribution functions  $F_X$  and  $F_Y$ . Then,  $X$  is said to be stochastically dominated by  $Y$ , denoted by  $X \leq_{st} Y$ , if

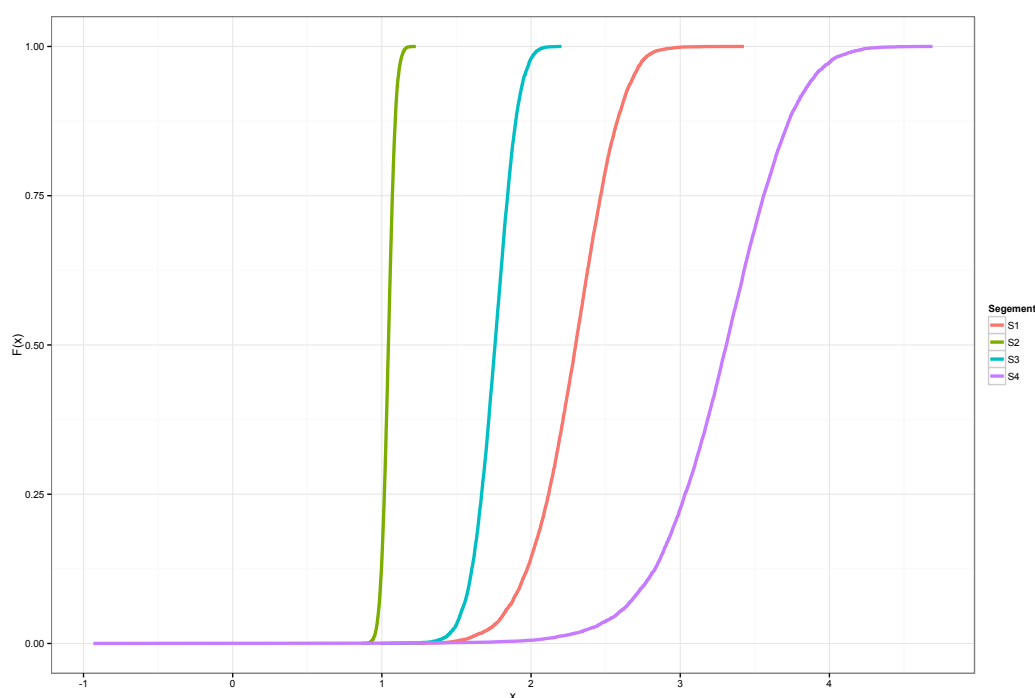
$$F_X(t) \geq F_Y(t), \quad (27)$$

for all  $t \in \mathbb{R}$ . In many practical circumstances, especially in capital allocation, the random variables involved are stochastically ordered by  $\leq_{st}$ .

Figure 2 contains such a real-life situation, where the random variables represent divisions of an insurance group. In this situation, the Wasserstein Metric can be easily calculated, if stochastic dominance,  $X \leq_{st} Y$ , can be assumed, then:

$$W_1(X, Y) = |E(X) - E(Y)|. \quad (28)$$

Again, in this situation, the Wasserstein Metric is easily interpreted and calculated.



**Figure 2.** The graph depicts distribution functions that follow a stochastic dominance ordering. The distribution functions may be interpreted as those functions that describe the risk of particular portfolios or segments.

#### 4. Conclusions

In this paper, we provide a consistent framework for a robust, quantitative risk management process. We argue that various approximations take place in this process and that we can take this into account by using a suitable metric throughout. The Wasserstein metric has convenient properties for applications in risk management, especially its focus on the tails of the distributions. It also features in recent work on the robustness of statistical functionals such as [14,16]. This strengthens our case for the Wasserstein metric as the canonical distance for risk management. The robustness results allow us—together with the concept of natural risk statistics—to recognize the approximative nature of the risk management process in the risk estimation step. We combine these results into the risk management space, adding an ordering to facilitate decision making based on the risk measurements. Several examples show how the process can be applied in practice and provide evidence for the use of the Wasserstein metric in a number of applications within a financial institution.

**Acknowledgments:** Robin Rühlicke is grateful for partial financial support by Talanx AG.

**Author Contributions:** G.S., J.Z. and R.R. conceived and designed the experiments; J.Z. and R.R. performed the experiments; G.S., R.R. and R.K. worked on the theoretical contributions, G.S. and R.K. structured the results, G.S. and R.K. wrote the final version paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### Appendix A Ideal Metric

We start our introduction of such a metric by recalling the definition of a probability metric.

Ideal metrics are used in the context of central limit theorems, but their properties are also desirable in financial applications. For example, when comparing portfolios, the regularity property states that adding an independent common subportfolio to two portfolios does not increase their distance (see Chapter 9.3 in [12] for this and other examples).

**Definition A1** (s-ideal metric). Let  $d$  be a metric on  $\bar{\mathbb{F}} \subset \mathbb{F}$ . For any random variables  $X, Y$  such that their distribution functions  $F_X, F_Y$ , respectively, are in  $\bar{\mathbb{F}}$ ,  $d$  induces a metric

$$d(X, Y) = d(F_X, F_Y).$$

Such a metric is called s-ideal if

1. for any rvs  $X, Y, Z$  with  $F_X, F_Y, F_Z \in \bar{\mathbb{F}}$ ,  $X$  and  $Y$  independent of  $Z$ ,

$$d(X + Z, Y + Z) \leq d(X, Y), \text{ and}$$

2. for any rvs  $X, Y$  with  $F_X, F_Y \in \bar{\mathbb{F}}$  and  $c \in [0, \infty)$ , it holds that

$$d(cX, cY) = c^s d(X, Y).$$

## Appendix B Proof of Theorem 1

Let  $T$  be defined in (13) with a function  $J$  bounded on  $[0, 1]$ :

$$\begin{aligned} T(G) - T(F) &= \int_{-\infty}^{\infty} x dK(G(x)) - \int_{-\infty}^{\infty} x dK(F(x)) \\ &= - \int_{-\infty}^{\infty} [K(G(x)) - K(F(x))] dx \quad (\text{integrated by parts}) \\ &= - \int_{-\infty}^{\infty} \int_{F(x)}^{G(x)} J(u) du dx, \end{aligned} \quad (\text{A1})$$

then

$$\begin{aligned} |T(G) - T(F)| &= \left| \int_{-\infty}^{\infty} \int_{F(x)}^{G(x)} J(u) du dx \right| \\ &\leq \int_{-\infty}^{\infty} \left| \int_{F(x)}^{G(x)} J(u) du \right| dx \quad (\text{Jensen's inequality}) \\ &\leq \|J\|_{\infty} \int_{-\infty}^{\infty} |G(x) - F(x)| dx \\ &= \|J\|_{\infty} \int_0^1 |G^{-1}(x) - F^{-1}(x)| dx, \end{aligned} \quad (\text{A2})$$

where  $\|h\|_{\infty} = \sup_x |h(x)|$ . According to Hölder's inequality,

$$\int_S |f(x)g(x)| dx \leq \left( \int_S |f(x)|^p dx \right)^{1/p} \left( \int_S |g(x)|^q dx \right)^{1/q}. \quad (\text{A3})$$

Let  $S$  be the  $[0, 1]$ ,  $g(x) = 1$ , and  $f(x) = G^{-1}(x) - F^{-1}(x)$ , then we have

$$\left\{ \int_0^1 |G^{-1}(t) - F^{-1}(t)| dt \right\} \leq \left\{ \int_0^1 |G^{-1}(t) - F^{-1}(t)|^p dt \right\}^{1/p}. \quad (\text{A4})$$

In summary, we can have  $|T(G) - T(F)| \leq \|J\|_{\infty} W_p(G, F)$ . Therefore, we can have  $|T(G) - T(F)| = o_p(1)$  as  $W_p(G, F) \rightarrow 0$ .

## References

1. Morini, M. *Understanding and Managing Model Risk: A Practical Guide for Quants, Traders and Validators*; Wiley: New York, NY, USA, 2011.
2. Fouque, J.; Langsam, J. *Handbook on Systemic Risk*; Cambridge University Press: Cambridge, UK, 2013.

3. Jorion, P. Risk Management Lessons from the Credit Crisis. *Eur. Financ. Manag.* **2009**, *15*, 923–933.
4. Hull, J.C. The Credit Crunch of 2007: What Went Wrong? Why? What Lessons Can Be Learned? *J. Credit Risk* **2009**, *5*, 3–18.
5. European Parliament and Council of the European Union. Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance. *Off. J. Eur. Union* **2009**, *20*, 7–25.
6. Committee of European Insurance and Occupational Pensions Supervisors (CEIOPS). *CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: Articles 120 to 126 Tests and Standards for Internal Model Approval*; CEIOPS: Frankfurt am Main, Germany, 2009.
7. Barrieu, P.; Scandolo, G. Assessing financial model risk. *Eur. J. Oper. Res.* **2014**, *242*, 546–556.
8. Gneiting, T. Making and evaluating point forecasts. *J. Am. Stat. Assoc.* **2011**, *106*, 746–762.
9. Ziegel, J. Coherence and elicibility. *Math. Financ.* **2013**, doi:10.1111/mafi.12080.
10. Weber, S. Distribution-invariant risk measures, information and dynamic consistency. *Math. Financ.* **2006**, *16*, 419–441.
11. Bickel, P.; Freedman, D. Some asymptotic theory for the bootstrap. *Ann. Stat.* **1981**, *9*, 1196–1217.
12. Rachev, S.; Stoyanov, S.; Fabozzi, F. *Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization*; John Wiley & Sons: New York, NY, USA, 2008.
13. Krätschmer, V.; Schied, A.; Zähle, H. Sensitivity of risk measures with respect to the normal approximation of total claim distributions. *Insur. Math. Econ.* **2011**, *49*, 335–344.
14. Krätschmer, V.; Schied, A.; Zähle, H. Qualitative and infinitesimal robustness of tail-dependent statistical functionals. *J. Multivar. Anal.* **2012**, *103*, 35–47.
15. Cont, R.; Deguest, R.; Scandolo, G. Robustness and sensitivity analysis of risk measurement procedures. *Quant. Financ.* **2010**, *10*, 593–606.
16. Krätschmer, V.; Schied, A.; Zähle, H. Comparative and qualitative robustness for law-invariant risk measures. *Financ. Stoch.* **2014**, *18*, 271–295.
17. Sibbertsen, P.; Stahl, G.; Luedtke, C. Measuring model risk. *J. Risk Model Valid.* **2008**, *2*, 65–81.
18. Busse, M.; Müller, U.; Dacorogna, M. Robust Estimation of Reserve Risk. *ASTIN Bull.* **2010**, *40*, 453–489.
19. Föllmer, H.; Knipsel, T. Convex Risk Measures: Basic Facts, Law-invariance and beyond, Asymptotics for Large Portfolios. In *Handbook of the Fundamentals of Financial Decision Making*; MacLean, L., Ziemba, W., Eds.; World Scientific: Singapore, 2013; Volume II, pp. 507–554.
20. Cont, R.; Deguest, R.; He, X. Loss-based risk measures. *Stat. Decis.* **2013**, *30*, 133–167.
21. Heyde, C.; Kou, S.; Peng, X. *What Is a Good External Risk Measure: Bridging the Gaps between Robustness, Subadditivity, and Insurance Risk Measures*; Columbia University: New York, NY, USA, 2007.
22. Cheridito, P.; Li, T. Risk measures on Orlicz hearts. *Math. Financ.* **2009**, *19*, 189–214.
23. Filipovic, D.; Svindland, G. The canonical model space for law-invariant risk measures is  $L^1$ . *Math. Financ.* **2012**, *22*, 585–589.
24. Artzner, P.; Delbaen, F.; Eber, J.; Heath, D. Coherent Measures of risk. *Math. Financ.* **1999**, *9*, 203–228.
25. Huber, P.J.; Ronchetti, E.M. *Robust Statistics*, 2nd ed.; Wiley: New York, NY, USA, 2009.
26. Dall'Aglio, G. Sugli estremi dei momenti delle funzioni di ripartizione doppia. *Ann. Sc. Norm. Super. Pisa* **1956**, *3*, 35–74.
27. Mallows, C.L. A Note on Asymptotic Joint Normality. *Ann. Math. Stat.* **1972**, *43*, 508–515.
28. Pflug, G.; Pichler, A.; Wozabal, D. The  $1/N$  investment strategy is optimal under high model ambiguity. *J. Bank. Financ.* **2012**, *36*, 410–417.
29. Clarke, B.R. A Review of Differentiability in Relation to Robustness with an Application to Seismic Data Analysis. *Proc. Indian Natl. Sic. Acad.* **2000**, *66*, 467–482.
30. Davies, P. Approximating Data. *J. Korean Stat. Soc.* **2008**, *37*, 191–211.

**Sample Availability:** Samples of the compounds are available from the authors.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).