ARBEITSBERICHT
PROZESS- UND PRODUKT-ENGINEERING: **METHODEN**

2010/23, ISSN 1814-1307 (Print), ISSN 1814-1315 (Online)

ROBUST RISK MAESUREMENT UNDER MODEL UNCERTAINTY

Thomas Breuer thomas.breuer@fhv.at
Imre Csiszár csiszar@renyi.hu

24. November 2010

Fachhochschule Vorarlberg
Forschungszentrum
Prozess- und Produkt-Engineering

Robust Risk Measurement under Model Uncertainty*

Thomas Breuer Imre Csiszár

24 November 2010

Abstract

We propose a scenario-based risk measure which is coherent, robust under model uncertainty, and law-invariant: maximum expected loss over all alternative risk factor distributions whose relative entropy does not exceed a certain threshold. We determine explicitly the maximum expected loss as well as the worst case distribution. Practical implementations of this method do not require any numerical optimisation.

Keywords: stress tests, multiple priors, model risk, ambiguity aversion, relative entropy, maximum entropy principle, exponential family

JEL classification: C18, C44, C60, G01, G32, M48
AMS classification: 62C20, 90B50, 91B30, 94A17

1 Introduction

Risk measures assign to a portfolio a number interpreted as risk capital. Artzner et al. [1999] and Föllmer and Schied [2004] formulated requirements for risk measures and coined the terms 'coherent' resp. 'convex' for risk measures fulfulling them. But is coherency or convexity enough? Traditional risk measures, like Value at Risk or Expected Shortfall, assign risk numbers on the basis of the profit loss distribution, which arises from the portfolio when a risk factor distribution is given. For a fixed portfolio, a different risk factor distribution gives rise to a different profit loss distribution, and therefore to a different risk capital requirement. Thus, traditional risk measures rely on a specific model, which perhaps uses inappropriate risk factors, or which works with a misspecified type of risk factor distribution, or with

---

[*] Thomas Breuer, PPE Research Centre, FH Vorarlberg, `thomas.breuer@fhv.at`. Imre Csiszár, Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, `csiszar@renyi.hu`. We gratefully acknowledge comments by Klaus Böcker, Freddy Delbaen, Georg Pflug, Chris Rogers, Walter Schachermayer, Martin Summer, as well as to seminar participants in London, Munich, and Vienna.

misestimated distribution parameters. How robust is the risk capital requirement to different risk factor distributions? This kind of robustness is an additional requirement on top of coherency or convexity.

Scenario-based risk measures are intended to be robust under—or even independent of—changes in the risk factor distribution. Fixing some set of scenarios such a risk measure assigns to a portfolio the maximal portfolio loss over the scenario set. Traditionally a scenario is considered as simultaneous realisation of the risk factors. Such scenarios could be called pure scenarios. They are represented by an element $\boldsymbol{r}$ of the sample space, which often but not always is some subset of a Euclidean space. The evaluation of a scenario-based risk measure amounts to performing stress tests. The consequences of a pure scenario $\boldsymbol{r}$ for the portfolio is described by a loss funtion $L(\boldsymbol{r})$. The loss function characterises the portfolio.[1] From now on we will use the terms 'stress tests' and 'scenario-based risk measure' interchangeably.

By construction such a scenario-based risk measure is independent of the risk factor distribution. This is probably the reason why regulators require stress tests to complement traditional risk measurement. Additionally, scenario-based risk measures are coherent (modulo some technical issues). Furthermore, knowledge of the dangerous scenarios suggests action reducing risk if desired. But scenario-based risk measures have an important disadvantage against traditional risk measures. Often they are not law-invariant: Portfolios might have the same profit loss distribution without having the same worst case loss.

We propose a risk measure which combines the advantages of distribution based risk measures and scenario based risk measures: maximum expected loss over a suitable set of plausible scenarios. This risk measure by construction is coherent and robust under model uncertainty, but it turns out to be also law-invariant. A key idea is to use *mixed scenarios*, which are distributions of pure scenarios. This terminology is in analogy with game theory, which uses mixed strategies along with pure strategies, or with physics, which uses mixed states along with pure states.[2] Mixed scenarios can be interpreted in different ways. In the context of risk measurement, a mixed scenario is naturally interpreted as an alternative risk factor distribution.

We measure the plausiblity of a scenario Q by its relative entropy with respect to some reference distribution ν, denoted by $D(Q||\nu)$. (Synonyms for relative entropy are Kullback-Leibler distance or I-divergence.) The reference distribution ν results from some estimation procedure and is the best guess of the risk factor distribution. But other distributions are not

---

[1] For stress testers it might seem a bold assumption to know explicitly the loss function $L(\boldsymbol{r})$. After all, stress testers often need days to evaluate a complex portfolio in a given scenario. On the other hand, all standard quantitative risk management frameworks do work with a loss function, see e.g. McNeil et al. [2005, Chpt 2.1].

[2] Sometimes the term generalised scenario is used for mixed scenarios, see Delbaen [2002].

excluded. Regarding as admissible those distributions for which the relative entropy does not exceed some threshold k we propose the following risk measure:

$$\mathrm{MaxLoss}(L,k) := \sup_{Q:D(Q||\nu)\leq k} \mathbb{E}_Q(L). \tag{1}$$

This risk measure specifies to what level the expected loss of the portfolio can increase when the true risk factor distribution is not ν but some other distribution in the ball $\{Q : D(Q||\nu) \leq k\}$. The larger the radius k the larger the set of distributions one considers as plausible enough to be taken into account. Large k represent high robustness requirements.

Observe that this problem is 'dual' to a problem of maximum entropy inference. If an unknown distribution Q had to be inferred when the available information specified only a feasible set of distributions, and a distribution ν were given as a prior guess of Q, the maximum entropy[3] principle would suggest to infer the feasible distribution Q which minimizes $D(Q||\nu)$. In particular, if the feasible distributions were those with $\mathbb{E}_Q(L) = b$, for a constant b, we would arrive at the problem

$$\sup_{Q:\mathbb{E}_Q(L)=b} D(Q||\nu). \tag{2}$$

Note that the objective function of the worst case problem (1) is the constraint in the maximum entropy problem (2), and vice versa (Fig. 1). It is therefore intuitively expected that (taking k and b suitably related) both problems are solved by the same distribution $\overline{Q}$,

$$\arg \sup_{Q:D(Q||\nu)\leq k} \mathbb{E}_Q(L) = \arg \sup_{Q:\mathbb{E}_Q(L)=b} D(Q||\nu) =: \overline{Q}. \tag{3}$$

The literature on the maximum entropy problem establishes that (under some regularity conditions) the solution $\overline{Q}$ is a member of the exponential family of distributions with statistic L, which have a ν-density of the form $\exp(\theta L(\boldsymbol{r}))$ times a normalisation factor. Call a typical distribution in the exponential family $Q(\theta)$.

For members of the exponential family the two relevant quantities $D(Q(\theta)||\nu)$ and $\mathbb{E}_{Q(\theta)}(L)$ can be expressed with the help of a function

$$\Lambda(\theta, L) := \log\left(\int e^{\theta L(\boldsymbol{r})} d\nu(\boldsymbol{r})\right), \tag{4}$$

where θ is a positive real number. If the loss function L is clear from the context, we will simply write $\Lambda(\theta)$. The expected loss can be written as

$$\mathbb{E}_{Q(\theta)}(L) = \int L(\boldsymbol{r}) \exp(\theta L(\boldsymbol{r}) - \Lambda(\theta)) d\nu(\boldsymbol{r}) = \Lambda'(\theta), \tag{5}$$

---

[3] This name refers to the special case when the prior guess ν is the uniform distribution; then minimising $D(Q||\nu)$ is equivalent to maximising the entropy of Q.

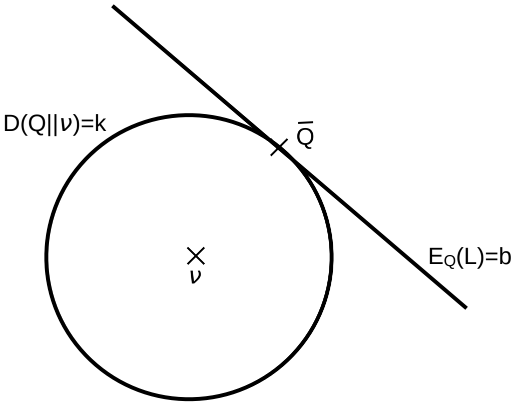

Figure 1: **'Duality' of Worst Case and Maximum Entropy** What is the objective function in the worst case problem (1) is the constraint in the maximum relative entropy problem (2), and vice versa.

(where $\Lambda'(\theta)$ is the derivative of $\Lambda(\theta, L)$ with respect to θ) and the relative entropy as

$$\begin{aligned} D(Q(\theta)||\nu) &= \int \log \frac{dQ(\theta)}{d\nu}(\boldsymbol{r})dQ(\theta)(\boldsymbol{r}) = \int (\theta L(\boldsymbol{r}) - \Lambda(\theta))dQ(\theta)(\boldsymbol{r}) \\ &= \theta \mathbb{E}_Q(\theta)(L) - \Lambda(\theta) = \theta\Lambda'(\theta) - \Lambda(\theta). \end{aligned} \tag{6}$$

If the identity (3) holds one can directly calculate $\mathrm{MaxLoss}(L, k)$:

$$\sup_{Q:D(Q||\nu)\le k} \mathbb{E}_Q(L) = \sup_{\theta:\theta\Lambda'(\theta)-\Lambda(\theta)\le k} \Lambda'(\theta) =: \Lambda'(\overline{\theta}),$$

where $\overline{\theta}$ is defined to be the solution of $\overline{\theta}\Lambda'(\overline{\theta}) - \Lambda(\overline{\theta}) = k$. (The last equality follows from the convexity of Λ.) This solution is illustrated in Fig. 2. MaxLoss is the slope of the tangent to the curve $\Lambda(\theta)$ passing through $(0, -k)$. $\overline{\theta}$ is the θ-coordinate of the tangent point. From the figure it is obvious that $\overline{\theta}\Lambda'(\overline{\theta}) - \Lambda(\overline{\theta}) = k$.

So far the intuition about the solution. It requires two important assumptions: Identity (3) should hold and the equation $\overline{\theta}\Lambda'(\overline{\theta}) - \Lambda(\overline{\theta}) = k$ should have a (unique) solution $\overline{\theta}$. The above argument makes it intuitively clear that MaxLoss can be calculated as $\Lambda'(\overline{\theta})$. The mathematical contribution of this paper is to give precise conditions under which the two assumptions hold and the solution is indeed of the generic form above (Theorem 1). Furthermore we give the solution also for 'pathological' cases where the conditions for the generic solution do not hold (Theorem 2).

The paper is structured as follows. In Section 2 we briefly discuss related literature. Section 3 gives the main mathematical results and Section 4 the proofs.

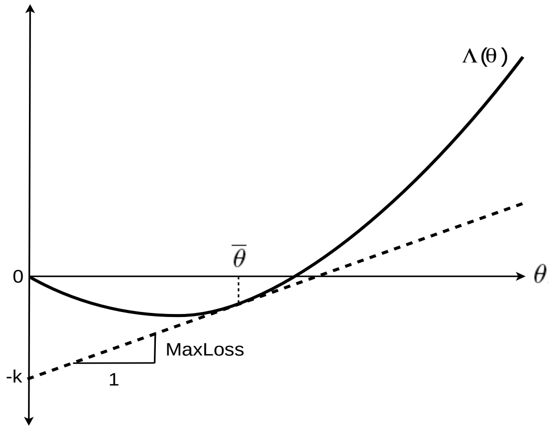

Figure 2: **Calculation of MaxLoss from Λ.** MaxLoss is the slope of the tangent to the curve $\Lambda(\theta)$ passing through $(0, -k)$. $\overline{\theta}$ is the θ-coordinate of the tangent point.

2 Relation to the literature

Stress tests During the recent crisis supervisors subjected financial institutions to stress tests. Stress testing started in market risk, see e.g. Basel Committee on Banking Supervision [1996], but in recent years it has been applied to credit risk and macro analysis as well. The Basel Committee on Banking Supervision [2005] requires banks to perform stress tests which meet three requirements: plausibility of stress scenarios (par. 718 (LXXIX)), severity of stress scenarios (par. 718 (LXXXIII)), and suggestiveness of risk reducing action (par. 718(LXXIX)). Principles of sound stress testing practices have been laid down by the Basel Committee on Banking Supervision [2009].

Stress tests examine the capital needs of institutions in economic downturn scenarios with increased credit losses. Thus, in effect, a scenario-based risk measure is being calculated. The stress scenarios are usually picked by hand on the basis of historical experience. The plausibility of scenarios is not always quantified and it remains unclear whether there are more severe scenarios of similar plausibility. If the scenarios considered are harmless, either because stress testers lack proficiency or wish to hide risks, stress tests convey a feeling of safety which might be false, see Berkowitz [2000].

A first step towards more objective stress tests was made by Studer [1997, 1999]. He proposed to perform stress tests *systematically.* Instead of considering just a few hand-picked scenarios Studer searches for the worst scenario among a set of plausible pure scenarios. In this way one ensures that no plausible scenario is missed and that only scenarios of sufficient plausibility are considered. Studer searches for the worst case over the ellipsoid of pure scenarios whose Mahalanobis distance is smaller than some threshold. He quantifies the harm done in a pure scenario $\boldsymbol{r}$ by the loss $L(\boldsymbol{r})$. MaxLoss as defined in (1) builds on an similar intuition but differs in important aspects: MaxLoss uses mixed scenarios instead of pure scenarios, measures their plausibility by the relative entropy instead of the Mahalanobis distance, and quantifies their harm by the expected loss instead of the loss. This circumvents the drawbacks of Studer's method pinpointed in Breuer [2008].

Relative entropy and other measures of plausibility *Every* coherent risk measure can be represented as $\sup_{Q\in\Gamma}\mathbb{E}_Q(L)$ for some closed convex set Γ of distributions (Delbaen [2002, Thm. 3.2]). Surely the existence of a representation of this kind is not sufficient for robustness. As a counterexample consider Expected Shortfall. For Expected Shortfall at level α, Γ is the set of all distributions whose density is bounded by $1/\alpha$, see Pflug and Römisch [2007, Theorem 2.34]. This Γ is not related to the question of robustness, since it does not even refer to the estimated distribution ν. For which choice of Γ can the resulting risk measure be interpreted as quantifying robustness?

Starting from the estimated distribution ν it is natural to include into Γ the distributions which are sufficiently plausible alternatives to ν. The plausibility of an alternative distribution is quantified by some distance from ν. In the literature, various distances[4] of probability distributions are used. One family of such distances, the f-divergences of Csiszár [1963], Ali and Silvey [1966], and Csiszár [1967], correspond to convex funtions f on the positive numbers. Relative entropy corresponds to $f(t) = t\log t$, several other choices of f also give distances often used in statistics. For details about f-divergences see Liese and Vajda [1987].

From the range of possible distances we have chosen relative entropy, which appears the most versatile one with many applications in statistics, information theory, statistical physics, see e.g. Kullback [1959], Csiszár and Körner [1981], Cover and Thomas [2006], Jaynes [1968, 1982]. Relative entropy has already been used in econometrics, see Golan et al. [1996] or Grechuk et al. [2009], or robust portfolio selection, see Calafiore [2007]. Using it also in the context of risk measurement looks certainly reasonable, though we do not claim that among the various distances of distributions this one is necessarily the best for this purpose.

In the context of inference the method of maximum entropy is distinguished by axiomatic considerations. Shore and Johnson [1980], Paris and Vencovská [1990], Jones and Byrne [1990] and Csiszár [1991] showed that it is

---

[4] Distance is meant in a broad sense, requiring neither symmetry nor the triangle inequality; those properties of distances in the narrow sense do not hold even for relative entropy.

the only method that satisfies certain intuitively desirable postulates. Still, as Uffink [1995, 1996] argued, relative entropy cannot be singled out as providing the only reasonable method of inference. Csiszár [1991] determined what alternatives come into account if some postulates are relaxed. Grunwald and Dawid [2004] argue that distances between distributions might be chosen in a utility dependent way. Relative entropy is natural only for decision makers with logarithmic utility. Picking up this idea, for decision makers with non-logarithmic utility one might define the radius of the scenario set in terms of some utility dependent distance. But this is not the approach of this paper. In our framework utility may enter into the loss function L but not into the scenario set.

Decision making under ambiguity aversion In many decision situations the distinction between Knightian ambiguity (uncertainty about the risk factor distribution) and risk (uncertainty about which risk factor values are realised) is prominent. The Ellsberg Paradox and related evidence have shown such a distinction to be behaviourally meaningful. The standard framework of subjective expected utility, however, excludes such a distinction.

A widely used class of preferences allowing for model ambiguity aversion are the multiple priors preferences axiomatized first by Gilboa and Schmeidler [1989]. Working in the setting of Anscombe and Aumann [1963] they use lottery acts, but Casadesus-Masanell et al. [2000] translated their approach to Savage acts, which is exactly our framework. Ambiguity averse agents prefer acts (portfolios with loss function L) with lower

$$\sup_{Q \in \Gamma} \mathbb{E}_Q(u \circ L), \tag{7}$$

where Γ is some closed convex set of finitely additive probabilities and u is a conitinuous strictly increasing utility function of the payoffs. The set Γ is interpreted as a set of priors held by the agent, and ambiguity is reflected by the multiplicity of the priors.

MaxLoss arises a special case of the Casadesus-Masanell et al. [2000] theory when utility is linear (or when it is absorbed into L) and the scenario set Γ is the Kullback-Leibler sphere. A decision maker who ranks portfolios by lower values of MaxLoss is ambiguity averse. Our results give analytical expressions of the Casadesus-Masanell et al. [2000] objective function in this case.

3 Main Results

We now explicitly calculate Maximum Loss (1) and the worst worst case scenario. The solution relies on techniques familiar in the theory of exponential families, see Barndorff-Nielsen [1978], and large deviations theory, see

Dembo and Zeitouni [1998]. Still, a self-contained development is provided, full proofs are given in Section 4.

The relative entropy of a probability distributions Q with respect to a reference distribution ν is defined as

$$D(Q||\nu) := \begin{cases} \int \log \frac{dQ}{d\nu}(\boldsymbol{r}) dQ(\boldsymbol{r}) & \text{if } Q \ll \nu \\ +\infty & \text{if } Q \not\ll \nu \end{cases}$$

where $Q \ll \nu$ denotes absolute continuity of the distribution Q with respect to the distribution ν.

Theorem 1. *(i) If* $\operatorname{ess\,sup}(L)$ *is finite, assume* k *is smaller than* $k_{\max} := -\log(\nu(\{\boldsymbol{r} : L(\boldsymbol{r}) = \operatorname{ess\,sup}(L)\}))$.

(ii) Assume $\theta_{\max} := \sup\{\theta : \Lambda(\theta) < +\infty\} > 0$,

(iii) If $\theta_{\max}, \Lambda(\theta_{\max})$, *and* $\Lambda'(\theta_{\max})$ *are all finite, assume* k *does not exceed* $k_{\max} := \theta_{\max}\Lambda'(\theta_{\max}) - \Lambda(\theta_{\max})$.

Under these assumptions the equation

$$\theta\Lambda'(\theta) - \Lambda(\theta) = k \tag{8}$$

has a unique positive solution $\overline{\theta}$. *The worst case scenario* $\overline{Q}$ *is the distribution with* ν*-density*

$$\frac{d\overline{Q}}{d\nu}(\boldsymbol{r}) := \frac{e^{\overline{\theta}L(\boldsymbol{r})}}{\int e^{\overline{\theta}L(\boldsymbol{r})} d\nu(\boldsymbol{r})} = e^{\overline{\theta}L(\boldsymbol{r}) - \Lambda(\overline{\theta})}. \tag{9}$$

The Maximum Loss achieved in the worst case scenario $\overline{Q}$ *is*

$$\text{MaxLoss}(L, k) = \mathbb{E}_{\overline{Q}}(L) = \Lambda'(\overline{\theta}). \tag{10}$$

This theorem gives conditions for the generic solution to apply and provides a practical procedure for calculating MaxLoss in the generic case.

1. Calculate $\Lambda(\theta)$ from (4). This involves the evaluation of an n-dimensional integral.
2. Starting from the point $(0, -k)$, lay a tangent to the curve $\Lambda(\theta)$.
3. MaxLoss is given by the slope of the tangent.

The procedure is illustrated in Fig. 2.

How should one choose the radius k of the Kullback-Leibler sphere? k is a parameter in Problem (1), in the same way as the confidence level is a parameter for Value at Risk or Expected Shortfall. Which choice of k is sensible? MaxLoss(k) dominates Tail-VaR at the level $\exp(-k)$:

$$\sup_{A:\nu(A)\geq e^{-k}} \int_A L(\boldsymbol{r}) d\nu(\boldsymbol{r})/\nu(A) \leq \sup_{Q:D(Q||\nu)\leq k} \mathbb{E}_Q(L).$$

(This is true because the distribution Q_A with density $dQ_A/d\nu := 1_A/\nu(A)$ satisfies $D(Q_A||\nu) \leq k$ if $\nu(A) \geq \exp(-k)$.) This inequality suggests reasonable orders of magnitude for k. For a 1%-tail the corresponding k is $-\log(0.01) = 4.6$. An alternative way to choose k would be to take k-values realised in historical crisis.

The second theorem deals with the pathological cases.

Theorem 2. *(i) If* $\operatorname{ess\,sup}(L)$ *is finite, and* $k \geq k_{\max}$, *then* $\mathrm{MaxLoss}(L, k) = \operatorname{ess\,sup}(L)$.

(ii) If $\theta_{\max} = 0$ *then* $\mathrm{MaxLoss}(L, k) = \infty$ *for all* $k > 0$.

(iii) If $0 < \theta_{\max} < +\infty$, *and both* $\Lambda(\theta_{\max})$ *and* $\Lambda'(\theta_{\max})$ *are finite, and additionally* $k > k_{\max}$, *then*

$$\mathrm{MaxLoss}(L, k) = (k + \Lambda(\theta_{\max}))/\theta_{\max}, \tag{11}$$

but there is no mixed scenario achieving $\mathrm{MaxLoss}(k)$*: the supremum in eq.* (1) *is not a maximum.*

The situation in the pathological cases (i) and (iii) is illustrated in Fig. 3.

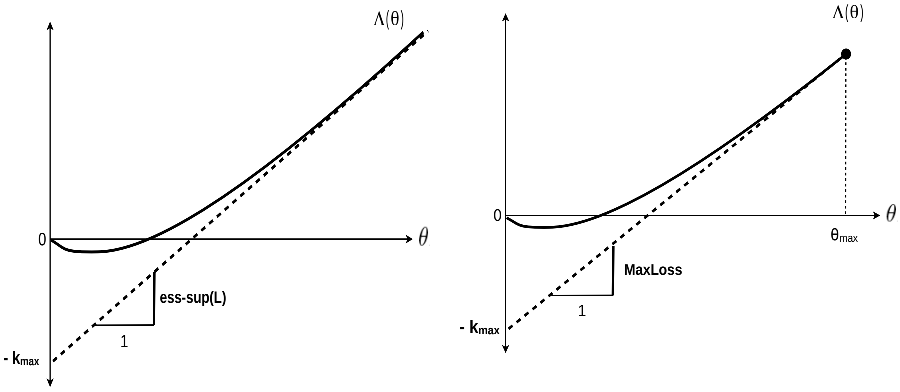

Figure 3: The pathological cases (i) and (iii).

Corollary 1. MaxLoss *is a law-invariant risk measure: If two portfolios* L_1, L_2 *have the same profit-loss distributions,* $\nu \circ L_1^{-1} = \nu \circ L_2^{-1}$, *then* $\mathrm{MaxLoss}(L_1, k) = \mathrm{MaxLoss}(L_2, k)$.

A robustness analysis of the risk measure MaxLoss emerges as a by-product. Increasing values of k represent increasing robustness requirements. The increasing slope of the Λ-curve describes how risk capital increases with increasing robustness requirements.

4 Proofs

In the proof of Theorems 1 and 2 we will use the following properties of the function $\Lambda(\theta)$ defined by eq. (4), which are standard and easy to check.

The function $\Lambda(\theta)$ is convex and lower semicontinuous on $\mathbb{R}$, its value is 0 at $\theta = 0$, and—excluding the trivial case when $\Lambda(\theta) = +\infty$ at all $\theta \neq 0$—its essential domain $D_\Lambda := \{\theta : \Lambda(\theta) < +\infty\}$ is a finite or infinite interval. In this interval, $\Lambda(\theta)$ is continuous and has derivative given by (5); when $\theta \in D_\Lambda$ is an endpoint of this interval, the derivative $\Lambda'(\theta)$ is understood as one-sided and is not necessarily finite.

The derivative $\Lambda'(\theta)$ equals $E_\nu(L)$ at $\theta = 0$, see (5), and is strictly increasing in D_Λ unless $L(\boldsymbol{r})$ is constant ν-almost everywhere. Moreover, as θ goes (increasingly) to $\theta_{\max} = \sup\{\theta : \Lambda(\theta) < +\infty\}$, the limit of $\Lambda'(\theta)$ equals $\operatorname{ess\,sup}(L)$ if $\theta_{\max} = +\infty$, while otherwise the limit equals $\Lambda'(\theta_{\max})$ or $+\infty$ according as $\theta_{\max}$ is in D_Λ or not.

The function

$$\Lambda^*(x) := \sup_\theta (\theta x - \Lambda(\theta)), \tag{12}$$

called the convex conjugate of $\Lambda(\theta)$, is also convex and lower semicontinuous on $\mathbb{R}$. Clearly,

$$\Lambda^*(x) = \theta x - \Lambda(\theta) \text{ if } x = \Lambda'(\theta). \tag{13}$$

However, for some x perhaps no θ satisfies $x = \Lambda'(\theta)$.

Lemma 1. *Excluding the cases (i) and (ii) of Theorem 2, there exists a unique x satisfying*

$$\Lambda^*(x) = k \text{ and } x > \mathbb{E}_\nu(L), \tag{14}$$

Except for the case (iii) of Theorem 2, this x equals $\Lambda'(\theta)$ for some (unique) $\theta > 0$.

Proof. If $\theta_{\max}$ is finite then $\Lambda^*(x)$ is finite for all $x > \Lambda'(0) = E_\nu(L)$. If $\theta_{\max} = \infty$ then $x \leq \lim_{\theta\to\infty} \Lambda'(\theta)$ is a necessary and $\Lambda'(0) < x < \lim_{\theta\to\infty} \Lambda'(\theta)$ is a sufficient condition for $\Lambda^*(x) < \infty$. In any case, $\sup\{x : \Lambda^*(x) < \infty\}$ is equal to $\operatorname{ess\,sup}(L)$.

Moreover, eqs. (13), (6) imply

$$\Lambda^*(\Lambda'(0)) = 0 < D(Q\|\nu) = \Lambda^*(\Lambda'(\theta))$$

if $0 < \theta < \theta_{\max}$. Hence, since the function Λ^* is convex, it is strictly increasing in the interval $[\Lambda'(0), \operatorname{ess\,sup}(L))$.

If $\operatorname{ess\,sup}(L) =: b < \infty$ then

$$\Lambda^*(b) = \sup_\theta(\theta b - \Lambda(\theta)) = \sup_\theta[-\log \int e^{\theta(L(\boldsymbol{r})-b)} d\nu(\boldsymbol{r})] = -\log \nu(\{\boldsymbol{r} : L(\boldsymbol{r}) = b\}),$$

and $\Lambda^*(x) \to \Lambda^*(b)$ as $x \uparrow b$. It follows, since the case $k \geq -\log \nu(\{\boldsymbol{r} : L(\boldsymbol{r}) = b\})$ has been excluded, that there exists a unique $x \in (\Lambda'(0), b)$

satisfying $\Lambda^*(x) = k$. Moreover, as $\Lambda'(\theta)$ approaches b for $\theta \to +\infty$, there exists $\theta \in (0, +\infty)$ with $x = \Lambda'(\theta)$.

If $\operatorname{ess\,sup}(L) = \infty$ then $\Lambda^*(x)$ is a strictly increasing convex function in the interval $[\Lambda'(0), \infty]$, hence it goes to $+\infty$ as $\theta \to \infty$. Thus, again, there exists a unique x satisfying (14). On account of (13), this x is equal to $\Lambda'(\theta)$ for some positive θ except for the case (iii) of Theorem 2. $\square$

4.1 Proof of Theorem 1

Equation (8) has a unique positive solution $\overline{\theta}$ due to Lemma 1 and (13). For $\overline{Q}$ as defined in (9) we have

$$\mathbb{E}_{\overline{Q}}(L) = \Lambda'(\overline{\theta}) \tag{15}$$

because of (5) as well as $D(\bar{Q}||\nu) = k$, see (6). We now show that this $\overline{Q}$ attains the maximum $\mathbb{E}_Q(L)$ among all mixed scenarios Q with $D(Q||\nu) \le k$. Take an arbitrary such Q. Then we have

$$\begin{aligned}
\overline{\theta}\Lambda'(\overline{\theta}) - \Lambda(\overline{\theta}) &= k \\
&\ge D(Q||\nu) \\
&= \int \log \frac{dQ}{d\nu}(\boldsymbol{r}) dQ(\boldsymbol{r}) \\
&= \int \left(\log \frac{dQ}{d\overline{Q}}(\boldsymbol{r}) + \log \frac{d\overline{Q}}{d\nu}(\boldsymbol{r}) \right) dQ(\boldsymbol{r}) \\
&= D(Q||\overline{Q}) + \int (\overline{\theta} L(\boldsymbol{r}) - \Lambda(\overline{\theta})) dQ(\boldsymbol{r}) \\
&= D(Q||\overline{Q}) + \overline{\theta}\mathbb{E}_Q(L) - \Lambda(\overline{\theta}),
\end{aligned}$$

implying $\overline{\theta}\mathbb{E}_Q(L) \le \overline{\theta}\Lambda'(\overline{\theta}) = \overline{\theta}\mathbb{E}_{\overline{Q}}(L)$. Since $\overline{\theta}$ is positive we conclude $\mathbb{E}_Q(L) \le \mathbb{E}_{\overline{Q}}(L)$.

4.2 Proof of Theorem 2

Proof. (i) First consider the case where $\operatorname{ess\,sup}(L) =: b$ is finite, and $0 < \nu(\{\boldsymbol{r} : L(\boldsymbol{r}) = b\}) =: \beta$. Then the measure $\nu_b \ll \nu$ with

$$\frac{d\nu_b}{d\nu}(\boldsymbol{r}) := \begin{cases} 1/\beta & \text{if } L(\boldsymbol{r}) = b \\ 0 & \text{otherwise} \end{cases}$$

satisfies

$$D(\nu_b||\nu) = \int \log \frac{d\nu_b}{d\nu} d\nu_b = -\log \beta,$$

hence $D(\nu_b||\nu) \le k$ if $k \ge -\log \beta$. Then $\text{MaxLoss}(k) \ge \mathbb{E}_{\nu_b}(L) = b$. Trivially $\text{MaxLoss}(k) \le \operatorname{ess\,sup}(L) = b$. The claim $\text{MaxLoss}(k) = b$ follows.

(ii) Next consider the case $\theta_{\max} = 0$. Let $\beta_{m,n} := \nu(\{\boldsymbol{r} : -m \leq L(\boldsymbol{r}) \leq n\})$ and consider the measures $\nu_{m,n} \ll \nu$ with

$$\frac{d\nu_{m,n}}{d\nu}(\boldsymbol{r}) := \begin{cases} 1/\beta_{m,n} & \text{if } -m \leq L(\boldsymbol{r}) \leq n \\ 0 & \text{otherwise.} \end{cases}$$

For any $Q \ll \nu_{m,n}$,

$$D(Q||\nu) = \int \log(\frac{dQ}{d\nu_{m,n}}\frac{d\nu_{m,n}}{d\nu})dQ = D(Q||\nu_{m,n}) - \log\beta_{m,n}$$

is arbitrarily close to $D(Q||\nu_{m,n})$ if m and n are sufficiently large. Hence to prove that $\mathrm{MaxLoss}(k) = +\infty$ for all $k > 0$, it suffices to find to any given m and sufficiently large n distributions $Q \ll \nu_{m,n}$ with $D(Q||\nu_{m,n})$ arbitrarily close to zero and $\mathbb{E}_Q(L)$ arbitrarily large.

In the rest of part (ii) of this proof, m is fixed and n will go to $+\infty$. Define Q and $\Lambda_{m,n}$ by

$$\frac{dQ}{d\nu_{m,n}}(\boldsymbol{r}) := \frac{e^{\theta L(\boldsymbol{r})}}{\int e^{\theta L(\boldsymbol{r})} d\nu_{m,n}(\boldsymbol{r})} =: e^{\theta L(\boldsymbol{r}) - \Lambda_{m,n}(\theta)}$$

for any $\theta > 0$. Q and $\Lambda_{m,n}$ depend on θ. As in the proof of Theorem 1, $\mathbb{E}_Q(L) = \Lambda'_{m,n}(\theta)$ and $D(Q\|\nu_{m,n}) = \theta\Lambda'_{m,n}(\theta) - \Lambda_{m,n}(\theta)$ for any $\theta > 0$. For each θ,

$$-\theta m \leq \Lambda_{m,n}(\theta) = \int_0^\theta \Lambda'_{m,n}(\xi)d\xi \leq \theta\Lambda'_{m,n}(\theta). \tag{16}$$

For fixed $\theta > 0$, $\Lambda_{m,n} \to \infty$ as $n \to \infty$ since $\Lambda(\theta) = \infty$ by assumption. By (16) it follows that $\Lambda'_{m,n}(\theta) \to \infty$ as $n \to \infty$, and hence there exists a sequence $\theta_n \downarrow 0$ such that $\Lambda'_{m,n}(\theta_n) \to \infty$ and $\theta_n\Lambda'_{m,n}(\theta_n) \to 0$ as $n \to \infty$. By inequality (16), this implies $|\Lambda_{m,n}(\theta_n)| \to 0$ and hence $D(Q\|\nu_{m,n}) \to 0$ as $n \to \infty$. This completes the proof that, for Q defined with $\theta = \theta_n$, $\mathbb{E}_Q(L)$ will be arbitrarily large and $D(Q||\nu_{m,n})$ arbitrarily small.

(iii) In case (iii) $0 < \theta_{\max} < +\infty$, and both $\Lambda(\theta_{\max})$ and $\Lambda'(\theta_{\max})$ are finite, and additionally $k > k_{\max} = \theta_{\max}\Lambda'(\theta_{\max}) - \Lambda(\theta_{\max})$. Define $\overline{Q}$ as in (9) but with $\theta_{\max}$ in the place of $\overline{\theta}$. Then for all $Q \ll \overline{Q}$ we have

$$\begin{aligned} D(Q\|\nu) &= \int \log(\frac{dQ}{d\overline{Q}}\frac{d\overline{Q}}{d\nu})dQ \\ &= D(Q\|\overline{Q}) + \int \log(\exp(\theta_{\max}L(\boldsymbol{r}) - \Lambda(\theta_{\max}))dQ(\boldsymbol{r}) \\ &= D(Q\|\overline{Q}) + \theta_{\max}\mathbb{E}_Q(L) - \Lambda(\theta_{\max}). \end{aligned} \tag{17}$$

Hence, if $D(Q||\nu) \leq k$ then

$$\mathbb{E}_Q(L) \leq (k + \Lambda(\theta_{\max}) - D(Q\|\overline{Q}))/\theta_{\max}, \tag{18}$$

proving that $\text{MaxLoss}(k) \leq (k+\Lambda(\theta_{\max}))/\theta_{\max}$. To show that equality holds, apply the result of (ii) to $\overline{Q}$ in the role of ν, then the role of $\Lambda(\theta)$ is played by

$$\overline{\Lambda}(\theta) := \log \int e^{\theta L(\boldsymbol{r})} d\overline{Q}(\boldsymbol{r}) = \Lambda(\theta + \theta_{\max}) - \Lambda(\theta_{\max}).$$

Clearly, $\overline{\Lambda}(\theta) = \infty$ for all $\theta > 0$, hence by the result of (ii) there exist distributions Q' with $D(Q' \| \overline{Q})$ arbitrarily small and $\mathbb{E}_{Q'}(L)$ arbitrarily large. Then, for any small $\epsilon > 0$, a suitable linear combination Q of Q' and $\overline{Q}$ satisfies $\mathbb{E}_Q(L) = (k+\Lambda(\theta_{\max})-\epsilon)/\theta_{\max}$ and $D(Q\|\overline{Q}) < \epsilon$. For this Q, eq. (17) implies that $D(Q||\nu) \leq k$ and the claim $\text{MaxLoss}(k) \leq (k+\Lambda(\theta_{\max}))/\theta_{\max}$ follows. It implies that $\text{MaxLoss}(k) = (k+\Lambda(\theta_{\max}))/\theta_{\max}$.

Eq. (18) implies that in eq. (1) the supremum is not attained because $D(Q\|\overline{Q})$ is strictly positive when $\mathbb{E}_Q(L) > \Lambda'(\theta_{\max}) = \mathbb{E}_{\overline{Q}}(L)$. □

Remark: By Theorem 1, under its hypotheses MaxLoss(k) is equal to $x = \Lambda'(\overline{\theta})$ for $\overline{\theta}$ satisfying (8), which x is the unique solution of (14). The last proof implies that MaxLoss always equals the solution of (14) when it exists, even if (8) does not have a solution.

4.3 Proof of Corollary 1

Proof. Assume two portfolios L_1, L_2 have the same profit-loss distributions, $\nu \circ L_1^{-1} = \nu \circ L_2^{-1}$, which we denote by μ_1, μ_2. Then

$$\begin{aligned}
\Lambda(\theta, L_1) &= \log\left(\int_\Omega e^{\theta L_1(\boldsymbol{r})} d\nu(\boldsymbol{r})\right) \\
&= \log\left(\int_\mathbb{R} e^{\theta s} d\mu_1(s)\right) = \log\left(\int_\mathbb{R} e^{\theta s} d\mu_2(s)\right) \\
&= \Lambda(\theta, L_2).
\end{aligned}$$

This holds for all θ for which Λ is defined. The equality of distributions also implies that $\operatorname{ess\,sup} L_1 = \operatorname{ess\,sup} L_2$. Since the Λ-function of L_1, L_2 agree, their derivatives also agree, and by Theorems 1 and 2 their MaxLoss is equal. □

5 Conclusion

Theorem 1 suggests a close analogy between stress testing with mixed scenarios and statistical mechanics. First, as pointed out above, the optimisation problem of finding the worst case scenario is ‘dual’ to the method of maximum entropy. Second, almost all quantities in general stress testing have counterparts in statistical mechanics: The worst case distribution (9) is the counterpart of the canonical distribution. θ is the counterpart of the temperature parameter $\beta = 1/kT$. The profit function $-L$ is the counterpart of the energy function E. The risk factor vector $\boldsymbol{r}$ is the counterpart of the phase space points $(\boldsymbol{p}, \boldsymbol{q})$. Λ is the counterpart of the logarithm of the partition function Z.

References

S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society Ser. B*, 28:131–142, 1966.

F. J. Anscombe and R. J. Aumann. A definition of subjective probability. *Annals of Mathematical Statistics*, 34:199–205, 1963.

P. Artzner, F. Delbaen, J.-M. Ebner, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

O. Barndorff-Nielsen. *Information and Exponential Families in Statistical Theory (Wiley series in probability & mathematical statistics).* Wiley, 1978.

Basel Committee on Banking Supervision. Principles for sound stress testing practices and supervision. Technical report, Bank for International Settlements, 2009. `http://www.bis.org/publ/bcbs155.pdf`.

Basel Committee on Banking Supervision. Amendment to the capital accord to incorporate market risks. Technical report, Bank for International Settlements, 1996. Also available as `http://www.bis.org/publ/cgfs18.htm`.

Basel Committee on Banking Supervision. International convergence of capital measurement and capital standards. a revised framework. Technical report, Bank for International Settlements, 2005.

J. Berkowitz. A coherent framework for stress testing. *Journal of Risk*, 2: 1–11., 2000.

T. Breuer. Overcoming dimensional dependence of worst case scenarios and maximum loss. *Journal of Risk*, 11(1):79–92, 2008.

G. C. Calafiore. Ambiguous risk measures and optimal robust portfolios. *SIAM Journal of Optimization*, 18:853–877, 2007.

R. Casadesus-Masanell, P. Klibanoff, and E. Ozdenorem. Maxmin expected utility of Savage acts with a set of priors. *Journal of Economic Theory*, 92:35–65, 2000.

T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley Series in Telecommunications and Signal Processing. Wiley, 2nd edition, 2006.

I. Csiszár. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Publ. Math. Inst. Hungar. Acad. Sci.*, 8:85–108, 1963.

I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2:299–318, 1967.

I. Csiszár. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Annals of Statistics*, 19 (4):2032–2066, 1991.

I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic Press, 1981.

F. Delbaen. Coherent risk measures on general probability spaces. In K. Sandmann and P. J. Schonbucher, editors, *Advances in Stochastics and Finance: Essays in Honour of Dieter Sondermann*, pages 1–37. Springer, 2002.

A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*, volume 38 of *Applications of Mathematics.* Springer, 2nd edition, 1998.

H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*, volume 27 of *de Gruyter Studies in Mathematics.* Walter de Gruyter, 2nd edition, 2004.

I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18:141–153, 1989.

A. Golan, G. G. Judge, and D. Miller. *Maximum Entropy Econometrics: Robust Estimation with Limited Data.* Wiley, 1996.

B. Grechuk, A. Molyboha, and M. Zabarankin. Maximum entropy principle with general deviation measures. *Mathematics of Operations Research*, 34:445–467, 2009.

P. D. Grunwald and A. P. Dawid. Game theory, maximum entropy, minimum discrepancy and robust bayesian decision theory. *Annals of Statistics*, 32 (4):1367–1433, 2004.

E. T. Jaynes. Prior probabilities. *IEEE Transactions on Systems Science and Kybernetics*, SSC-4:227–241, 1968.

E. T. Jaynes. On the rationale of maximum entropy methods. *Proceedings of the IEEE*, 70:939–952, 1982.

L. K. Jones and C. L. Byrne. General entropy criteria for inverse problems, with applications to data acompression, pattern classification and cluster analysis. *IEEE Transactions on Information Theory*, 36:23–30, 1990.

S. Kullback. *Information Theory and Statistics.* Wiley, 1959.

F. Liese and I. Vajda. *Convex Statistical Distances.* Teubner, 1987.

A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management.* Princeton University Press, 2005.

J. B. Paris and A. Vencovská. A note on the inevitability of maximum entropy. *International Journal in Inexact Reasoning*, 4:183–223, 1990.

G. Pflug and W. Römisch. *Modeling, Measuring and Managing Risk.* World Scientific, Singapure, 2007.

J. E. Shore and R. W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory*, IT-26:26–37, 1980. Correction IT-29 (1983), 942–943.

G. Studer. *Maximum Loss for Measurement of Market Risk.* Dissertation, ETH Zürich, Zürich, 1997. Also available as `http://www.gloriamundi.org/picsresources/gsmlm.pdf`.

G. Studer. Market risk computation for nonlinear portfolios. *Journal of Risk*, 1(4):33–53, 1999.

J. Uffink. Can the maximum entropy principle be explained as a consistency requirement? *Studies in History and Philosophy of Modern Physics*, 26B: 223–261, 1995.

J. Uffink. The constraint rule of the maximum entropy principle. *Studies in History and Philosophy of Modern Physics*, 27:47–79, 1996.

Weitere Arbeiten
Forschungszentrum Prozess- und Produkt-Engineering

ANWENDUNGEN
Kurzfristige Prognose des Stromverbrauchs in Vorarlberg auf Stunden- und Viertelstundenbasis
Thomas Steinberger, 2004

Weiterbildungs- und Qualifizierungsbedarf kleinerer und mittlerer Unternehmen in Vorarlberg bezüglich Prozess- und Projektmanagement, Führung, Strategie und Innovationsmanagement
Markus Reichart, Julia Schneider, Isabella Gratzer, 2004

Netzwerke für Innovationen
Martin Meusburger, Markus Reichart, Karin Feurstein, 2005

Neue Technologien im Produktinnovationsprozess
Julia Schneider, Markus Reichart, 2005

Bezug von externen Leistungen in der Produktentwicklung Aktueller Stand - Trends - Verbesserungspotenziale
Julia Schneider, 2005

project orientation [vorarlberg]
Martin Meusburger, Markus Reichart, Bratislav Veljovic, 2005

project orientation [vorarlberg II]
Martin Meusburger, Markus Reichart, Stefan Fink, 2006

Adverse Inter-Risk Diversification Effects for FX Forwards[1]
Thomas Breuer, Martin Jandacka, 2007

Optimierung eines Vertrages zum variablen Strombezug
Hans Vollbrecht, 2007

Szenarioanalyse mit unvollständiger Information: Beispiel Pflegekostenmodell Vorarlberg
Thomas Breuer, Martin Herburger, Manfred Hellrigl, Bertram Meusburger, Ruth Weiskopf, Falko Wilms, 2007

A Review and Redefinition of Knowledge Work from a Management-Oriented Perspective
Rainer Erne, Sonja Sackmann, 2006

Was bedeutet Produktivität in der Produktentwicklung und welche Prozessstandards sind dafür wirksam?
Thomas Breuer, Rainer Erne, 2007

Strategisches Management
Falko E. P. Wilms, 2008

Bericht zur eVORIS-Workshop-Reihe: Teilnahme von Menschen mit Behinderung am Arbeitsmarkt in Vorarlberg
Thomas Breuer, Oskar Müller, Martin Strele, 2007

Stress Tests: From Arts to Science
Thomas Breuer, Imre Csiszár, 2010

METHODEN
Identifying Worst Case Scenarios of Security Portfolios with Quasi-Random Search Algorithms
Thomas Breuer, Filip Pistovcak, 2004

A General Noise Model and Its Effects on Evolution Strategy Performance
Hans-Georg Beyer, Dirk V. Arnold, 2004

Using Quasi-Monte Carlo Scenarios in Risk Management
Thomas Breuer, Filip Pistovcak, 2004

An Explicit Characterization of Calogero-Systems
Fritz Gesztesy, Karl Unterkofler, Rudi Weikard, 2004

Reliability of old and new Ventricular Fibrillation Detection Algorithms for Automated External Defibrillators
Anton Amann, Robert Tratnig, Karl Unterkofler, 2005

Towards an Integrated Measurement of Credit and Market Risk
Thomas Breuer, Martin Jandacka, Gerald Krenn, 2005

Umgang mit Szenarien
Falko E. P. Wilms, 2005

Umgang mit unscharfen Informationen
Falko E. P. Wilms, 2005

A new ventricular fibrillation detection algorithm for automated external defibrillators
Anton Amann, Robert Tratnig, Karl Unterkofler, 2005

Removal of Resuscitation Artefacts from Ventricular Fibrillation ECG Signals Using Kalman Methods
Anton Amann, M. Baubin, Klaus Rheinberger, Karl Unterkofler, 2005

Detecting ventricular fibrillation by time-delay methods
Anton Amann, Robert Tratnig, Karl Unterkofler, 2005

Der Einsatz vagen Wissens bei Entscheidungsprozessen
Thomas Breuer, Hans Vollbrecht, Andreas Juen, 2005

Szenarien sind Systeme
Falko E. P. Wilms, 2006

Portfolio Selection with Transaction Costs under Expected Shortfall Constraints
Thomas Breuer, Martin Jandacka, 2006

An optimization model for storing and delivering spare parts
Hans-Georg Beyer, Stefan Röhl 2007

Folgenabschätzung von Massnahmen
Falko E. P. Wilms, 2007

Regulatory Capital for Market and Credit Risk
Interaction: Is Current Regulation Always Conservative?
Thomas Breuer, Martin Jandacka, Klaus Rheinberger, Martin Summer, 2007

An Intraday Spotmarket-Price Model based on Clustering
Hans Vollbrecht, 2008

DIALOG als gemeinsames Denken
Falko E. P. Wilms, 2008

Real-Parameter Black-Box Optimization Benchmarking 2009: Presentation of the Noiseless Functions
Steffen Finck, Nikolaus Hansen, Raymond Ros and Anne Auger, 2009

Real-Parameter Black-Box Optimization Benchmarking 2009: Presentation of the Noisy Functions
Steffen Finck, Nikolaus Hansen, Raymond Ros and Anne Auger, 2009

Management-Cockpits zur Entscheidungsvorbereitung
Falko E. P. Wilms, 2009

Robust Risk Measurement under Model Uncertainty
Thomas Breuer, Imre Csiszár, 2010

Fachhochschule Vorarlberg
Forschungszentrum
Prozess- und Produkt-Engineering
Hochschulstraße 1
A-6850 Dornbirn

T +43 (0)5572 792 7100
F +43 (0)5572 792 9510

www.fhv.at/res/ppe

Fachhochschule Vorarlberg
Forschungszentrum
Prozess- und Produkt-Engineering