Incorporating model uncertainty into optimal insurance contract design

Georg Ch. Pflug[a,b,*], Anna Timonina[a], Stefan Hochrainer-Stigler[a]

[a]*International Institute for Applied Systems Analysis, Laxenburg, Austria*
[b]*Institute of Statistics and Operations Research, University of Vienna, Oskar Morgenstern-Platz 1, A-1090, Vienna, Austria*

Abstract

In stochastic optimization models, the optimal solution heavily depends on the chosen model for the scenarios. However, the scenario models are chosen on the basis of statistical estimates and are therefore subject to model error. We demonstrate here how the model uncertainty can be incorporated into the decision making process. We use a nonparametric approach for quantifying the model uncertainty and a minimax setup to find model-robust solutions. The method is illustrated by a risk management problem involving the optimal design of an insurance contract.

Keywords: insurance optimization, model error, minimax solution, distributional robustness

1. Introduction

Common approach in risk assessment and risk management is to base the risk estimates on observed data and to use the statistically obtained estimates for finding the optimal risk management strategies. However, the fact that statistical estimates can never give precise values of unknown parameters due to an estimation error, is quite often neglected. Moreover, the choice of the probability model, i.e. the class of possible distributions, is typically chosen by the statistician and is not further questioned.

In general, statistical estimation procedures do not allow to single out one specific probability model, but only a whole set of models can be determined, in which the true model lies with a prespecified probability. This confidence set can be taken as the set of models for a minimax decision, where the best decision under the worst model in the model set is sought for. We call such sets of models *ambiguity sets*. The minimax solution in this case is called *distributionally robust*.

Modeling uncertainty. Economic decisions are made under some assumptions of the decision relevant parameters. In deterministic optimization, the parameters are considered to be known and fixed. Already in the early days of optimization, this assumption was considered as too narrow. Two possible setups have then been developed: (i) in robust optimization, a set of possible parameters is determined, while (ii) in stochastic optimization the parameters are considered to follow a certain probability distribution. In robust optimization, the parameters are not weighted and one has to use minimax strategies (to find the best strategy under the worst case). Probability models come with a lot more of possible strategies: expected utility maximization or minimization of risk (shortfall risk, variance risk, tail risk, etc.). However, probability models depend heavily on the chosen probability model, which is typically based

---

[*]Corresponding author
Email addresses: `georg.pflug@univie.ac.at` (Georg Ch. Pflug), `timonina@iiasa.ac.at` (Anna Timonina), `hochrain@iiasa.ac.at` (Stefan Hochrainer-Stigler)

on some ad-hoc assumptions (e.g. parametric families of distributions) and on statistical estimation procedures, which may contain estimation error. This dependency has been ignored for quite a while, when most research was put into decision making under a given fixed probability model. However, if the model is completely fixed, only *aleatoric uncertainty*, i.e. the uncertainty about the realizations is to be considered.

As already noticed, the model choice is typically based on assumptions and estimations, i.e. an error in model choice cannot be excluded. The ambiguity in model selection is often referred to as *epistemic uncertainty*. By including the epistemic uncertainty into the decision making process, we can also to a certain extent reconcile this approach to the pure scenario approach. While in the latter all scenarios are in principle possible, stochastic ambiguity models would allow scenario probabilities to vary in order to accommodate for model uncertainty. In the extremal case that all probability distributions over the scenarios would be theoretically possible, the ambiguity modeling coincides with pure scenario analysis. In most application areas we know the importance or likelihood of scenarios at least to a certain extent, though we typically do not know the exact occurrence probabilities. This is why model ambiguity becomes an important issue in decision making. In ambiguous modeling, the possible model error is incorporated into the decision process, that allows to find robust decisions.

We summarize this as follows:

- If all parameters of an optimization problem are known, we call it *deterministic*.

- If some parameters are not exactly known, but known to lie in some set, then we have a *robust program*

- If for the unknown parameters a random distribution is specified, we call this a *stochastic program*.

- If the distribution of the random parameters is not known, but known to lie in some family of distribution, then the problem is called *ambiguity problem* and its minimax solution is called *distributionally robust*.

Bibliographic remarks. The idea of optimal decisions under several stochastic models (i.e., min-max solutions) appears for the first time in Scarf [15] in a linear inventory problem seeking the stockage policy, which maximizes the minimum profit considering all demand distributions with given mean and given standard deviation. More thorough studies of ambiguous decision problems than a minimax problems were initiated by Dupacová [3, 4] for the class of stochastic linear problems with recourse under general assumptions for the ambiguity set. The formulation is in a game theoretic setup, where the first player chooses the decision and the second player chooses the probability model. There are alternative names used in literature for the ambiguity problem, such as minimax stochastic optimization, model uncertainty problem or distributional robustness problem. Many proposals for ambiguity sets in the two-stage case have been made and analyzed. A list of popular classes of ambiguous models is presented by Dupacová [5].

Literature dealing with ambiguity either from theory or application point of view is growing rapidly. The situation when the ambiguity set consists of all probabilities with given first two moments was studied by Jagannathan [9] for the linear case. Shapiro and Kleywegt [16] define an ambiguity set as the convex hull of a finite collection of models; Ahmed and Shapiro [17] consider sets of models given by moment inequalities; a similar approach is adopted by Edirishinge. Calafiore [1] uses the Kullback-Leibler divergence to define neighborhoods of a baseline model as ambiguity sets. Thiele [18] considers the L_1 balls of densities as ambiguity sets. Delage and Ye [2] consider ambiguity set given by inequalities for the mean and the covariance matrix. Wozabal and Pflug [12] use for the first time ambiguity sets, which are balls with respect to the

Wasserstein distance, see also the recent book by Pflug and Pichler [13], which also deals with the multistage case. Hansen and Sargent consider in their 2007 book [8] alternative models of multistage stochastic optimization problems given by maximal deviation from a baseline model in Kullback-Leibler divergence. Goh and Sim [7] study multistage ambiguity sets which are defined by a mean, which must lie in some conical set, a given covariance matrix and some upper bounds on the exponential moments and extend this to multistage.

In this paper, we investigate the problem to determine an optimal insurance portfolio under model ambiguity. To simplify the approach, we consider only a single stage decision problem: A government has to decide about investment in and insurance for infrastructure for the next budget period. The infrastructure is subject to natural hazard such as earthquake, floods or tropical storms and the problem is to find the best mix between investment and insurance, i.e. between increased productivity and higher protection. The model is similar to the IIASA CATSIM model.

2. Model description

To determine the optimal design of an insurance scheme is a typical problem of stochastic optimization: Adoption of a robust approach would not make sense, as soon as it would mean that very rare events would be considered as important as quite frequent events and this would result in an overly pessimistic result. However, the probability model may be subject to error and this gives an argument for distributionally robust decisions.

The total insurable infrastructure stock of the country under consideration at time 0 is S_0. We assume that the country is under hight risk of natural hazards and we denote by L the annual loss variable. We further consider a stop-loss insurance with exit level Y, its payment function is $\min(L, Y)$. It is well known that stop-loss insurance contracts are "optimal" (see Raviv [14]). The stop-loss payment function is shown in Figure 1.
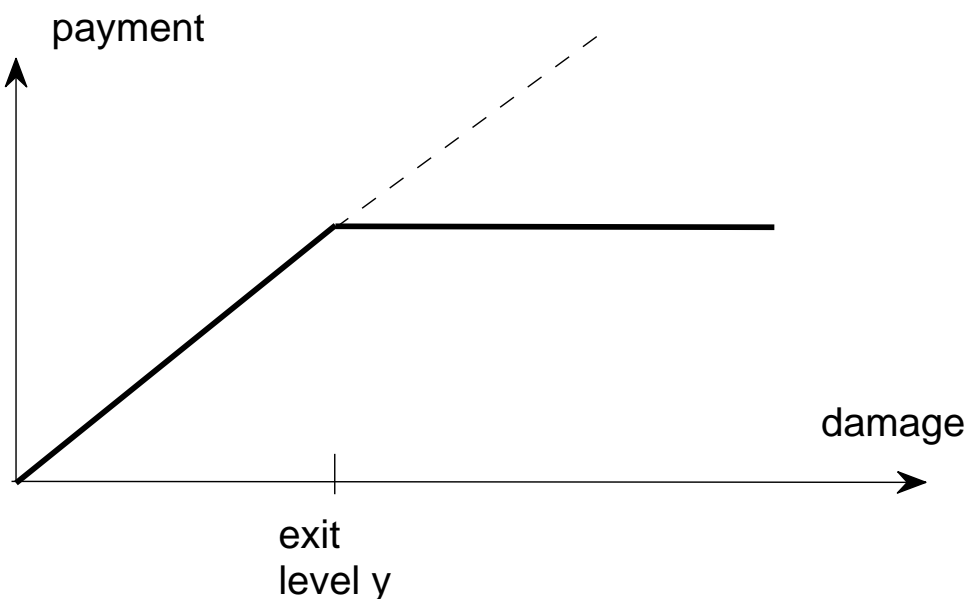

Figure 1: The payment function

The infrastructure S_1 at the end of the period (typically one year) is given by the previous amount S_1 minus the random damage and the amount obtained as compensation from the insurance plus the investment X.

$$S_1 = S_0 - L + \min(L, Y) + X$$

The premium for the stop-loss insurance is denoted by $\pi(L, Y)$. There is a budget B available, which may be used for investment and for infrastructure protection by insurance.

The decision problem considered here is to find the optimal mix between investment X and insurance with exit level Y for the given budget B, the objective to be maximised is the variance-corrected expectation of S_1. The complete (yet unambiguous) model is

$$\max_{Y} \mathbb{E}(S_1) - \lambda \mathbb{V}ar(S_1) \tag{2.1}$$
$$\text{s.t.} \quad S_1 = S_0 - L + \min(L, Y) + X$$
$$X + \pi(L, Y) = B$$
$$X \geq 0, Y \geq 0.$$

To summarize, we have introduced the following symbols.

S_0	the (deterministic) infrastructure value at the beginning of the period
S_1	the (stochastic) infrastructure value at the end of the period
L	the stochastic loss variable
B	the budget foreseen for investment and insurance
X	the investment in infrastructure
Y	the exit level for the insurance contract
λ	the penalty parameter for the variance
$\pi(L, Y)$	the premium of an insurance with relative loss variable ξ and exit level Y.

For further use we introduce some relative values

ξ	the relative loss variable, $\xi = L/S_0$
y	the relative exit level, $y = Y/S_0$

and

$$\pi_y = \pi(\min(L, yS_0))/S_0.$$

The insurance premium. We assume that the insurance premium is calculated according to a combination the distortion and the utility principle (see [10]): Suppose that F^L is the distribution function of the loss variable L. Using a *distortion function g* on [0,1], which satisfies $g(u) \geq u$, $g(0) = 0$, $g(1) = 1$, the loss distribution is distorted to the new loss distribution

$$F_g^L(u) = 1 - g(1 - F^L(u)).$$

Under the monotonic and convex (dis)utility function V, the premium for the distorted loss L is

$$\pi(L) = V^{-1} \left[\int_0^\infty V(u) \, d[1 - g(1 - F^L(u))] \right]$$

where V^{-1} is the inverse (dis)utility.

If the coverage is capped at Y, then the distribution function of the damage variable $\min(L, Y)$ is

$$F^{\min(L,Y)}(u) = \begin{cases} F^L(u) & \text{if } u \leq Y \\ 1 & \text{otherwise.} \end{cases}$$

Assuming that F_L has a density f^L, the premium for the loss capped at Y is

$$\pi(\min(L, Y)) = V^{-1}\left[\int_0^V (u)\, g'(1 - F^L(u)) f^L(u)\, du + V(Y) \cdot g(1 - F^L(Y))\right]$$

with $g'(u) = \frac{\partial}{\partial u} g(u)$.

The relative loss variable $\xi = L/S_0$ has distribution function $F^\xi(w) = F^L(w/S_0)$, $w \in [0, 1]$ with density $f^\xi(w) = 1/S_0\, f^L(w/S_0)$. For the relative exit level $y = Y/S_0$, one gets for the premium π_y of the contract with this exit level

$$
\begin{aligned}
\pi_y(\xi) = \pi(\min(yS_0, \xi S_0) &= V^{-1}\left[\int_0^y V(wS_0)g'(1 - F^\xi(w))f^\xi(w)\, du\right] + V(yS_0)g(1 - F^\xi(y))] \\
&= S_0 V_1^{-1}\left[\int_0^y V_1(w)g'(1 - F^\xi(w))f^\xi(w)\, dw\right]
\end{aligned}
$$

with $V_1(w) = V(S_0 w)$.

For the Example below, we have chosen the power distortion function $g(u) = u^r$ for $0 < r < 1$ and the (dis)utility function $V(u) = (a - u/S_0)^{-q} - a^{-q}$, i.e. $V_1(w) = (a - w)^{-q} - a^{-q}$. With this choice, the premium formula can be concretized to

$$\pi_y(\xi) = a - \left[\int_0^y [(a - w)^{-q} - a^{-q}]\, r(1 - F^\xi(w))^{r-1} f^\xi(w)\, dw + [(a - y)^{-q} - a^{-q}](1 - F^\xi(y))^r\right]^{-1/q}.$$
$$(2.2)$$

For the Madagascar Example (with a pointmass at 0 and a piecewise constant density, see below) and the parameter choice $a = 0.2$, $q = 1.1$, $r = 0.95$, the function $y \mapsto \pi_y$ is depicted in Figure 2.
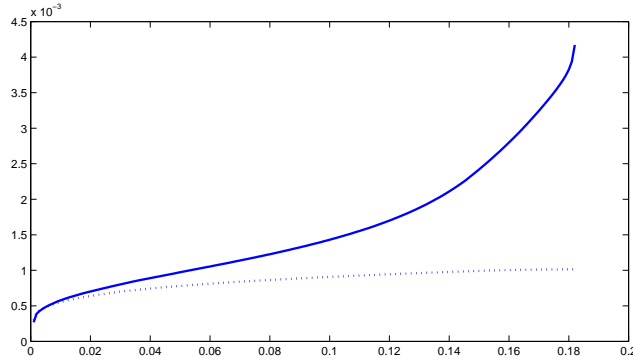

Figure 2: The relative premium π_y as function of the relative exit level y. The dashed curve is the expected relative loss to be covered by the stop-loss contract.

Notice that

$$S_1 = S_0(1 - \xi) + S_0 \min(\xi, y) - S_0 \pi_y(\xi) + B.$$

Since $\mathbb{E}[S_0(1 - \xi + \min(\xi, y))] = S_0 \mathbb{E}[1 - \xi + \min(\xi, y)]$ and $\mathbb{V}ar(S_1) = S_0^2 \mathbb{V}ar[1 - \xi + \min(\xi, y)]$ the following program is equivalent to (2.1)

$$\max_y \mathbb{E}[1 - \xi + \min(\xi, y)] - \pi_y(\xi) - \gamma \mathbb{V}ar[1 - \xi + \min(\xi, y)] \qquad (2.3)$$

$$\text{s.t. } \pi_y(\xi) \leq \frac{B}{S_0}$$

$$y \geq 0$$

with $\gamma = \lambda \cdot S_0$. Notice, that this is a one-dimensional optimization problem.

For the Madagascar Example, the objective function with the setting $\gamma = 100$ and $B/S_0 = 0.0182$ is shown in the Figure 3 below.
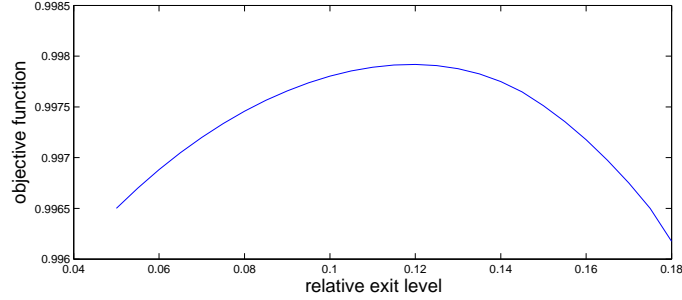

Figure 3: The objective as function of the relative exit level y

For the formulation of the distributionally robust problem, we assume that ξ has a point mass F_1 at $z_1 = 0$ and for $z \in (0, 1]$ a density which is constant in the intervals (z_i, z_{i+1}) for $i = 1, \ldots, k - 1$. The distribution function is a linear interpolation between the points (z_i, F_i), $i = 1, \ldots, k$, where F_i are the cumulative probabilities. Notice, that $F_k = 1$.

3. Alternative models

As we have introduced it, the baseline damage distribution is given by a discrete list of breakpoints $z_1, \ldots, z_k$ together with a list of cumulative probabilities $\hat{F}_1, \ldots, \hat{F}_k$. The alternative models will have the same interval boundaries z_i and differ only in the cumulative probabilities F_i. A simple ambiguity set $\mathcal{P}$ is given by

$$\mathcal{P} = \{F : \sum_i |F_i - F_i^*| \le \epsilon\}$$

or

$$\mathcal{P} = \{F : |F_i - F_i^*| \le \epsilon \ \forall i\}$$

However, such a neighborhoods do not take into account the values z_i and are therefore not appropriate.

We use the Wasserstein distance as the basic metric for loss distributions. It has not only the advantage of taking the values z_i into account, but it is based on a distance on the real line, which may be adapted to the needs of the problem at hand. For instance, we may use the basic distance
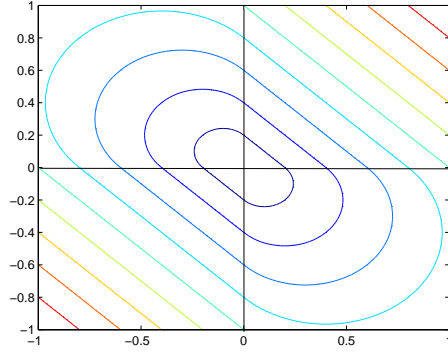
$$dist(z, w) = |z^s - w^s| \tag{3.1}$$

for some $s > 1$, meaning that the higher relative damages get higher distances, because they are more relevant for the insurance prices. However, in this paper we have set $s = 1$.

The Wasserstein distance between the two relative loss variables ξ resp. $\hat{\xi}$ with distribution functions F resp. $\hat{F}$ is defined as the optimal value of the (seemingly infinite) linear program

$$\min \mathbb{E}[dist(\xi, \hat{\xi})] \tag{3.2}$$
$$\text{s.t.} \quad \xi \sim F \tag{3.3}$$
$$\hat{\xi} \sim \hat{F} \tag{3.4}$$

6

The level sets of the function h.

The minimum is taken over all joint distributions with given marginals F resp. $\hat{F}$. Denote the optimal value of this program by $d(F, \hat{F})$. If the distance is chosen as in (3.1), then

$$d(F, \hat{F}) = \int |F(t) - \hat{F}(t)| s t^{s-1} \, dt$$

which is a slight extension of a result by Vallander (citeVall74). If the two distribution functions are piecewise linear with the same breakpoints z_i, then one finds after some calculation that the Wasserstein distance is given by

$$d(F, \hat{F}) = \frac{1}{2} \sum_{i=1}^{k-1} (z(i+1) - z(i)) \cdot h(F_i - \hat{F}_i, F_{i+1} - \hat{F}_{i+1})$$

where

$$h(a, b) = \begin{cases} a + b, & \text{if } a \geq 0, b \geq 0 \\ (a^2 + b^2)/(a - b), & \text{if } a \geq 0, b < 0 \\ (a^2 + b^2)/(b - a), & \text{if } a < 0, b \geq 0 \\ -a - b, & \text{if } a < 0, b < 0 \end{cases}$$

Notice that h is a convex function and that the sets $\{F_i : d(F, \hat{F}) \leq \epsilon\}$ are convex in the parameters F_i's.

The distributionally robust solution of our optimal insurance problem is the solution of

$$\max_y \min_{\{F : d(F, \hat{F}) \leq \epsilon\}} \mathbb{E}_F[1 - \xi + \min(\xi, y)] - \pi_y(\xi) - \gamma \mathbb{V}ar_F[1 - \xi + \min(\xi, y)] \qquad (3.5)$$

$$\text{s.t. } \xi \sim F \qquad (3.6)$$

$$F \text{ has a density in } (0, 1], \text{ which is piecewise constant in } [z_i, z_{i+1}] \qquad (3.7)$$

$$\pi(y) \leq B/S_0$$

$$y \geq 0.$$

Here $\mathbb{E}_F$ resp. $\mathbb{V}ar_F$ denote the expectation resp. the variance, when the distribution function of ξ is F.

4. A case study: Madagascar

Madagascar has one of the highest cyclone risks worldwide, especially the east coast, which is located in the path of destructive cyclones coming from the Indian Ocean (NOAA 2012).

The public sector plays an important role in financing losses after destructive cyclone events but usually falls short in providing adequate resources. Consequently, there is a keen interest in possible insurance mechanism (or a regional insurance pool) that could help financing disasters in a proactive manner. However, for such kind of assessment an annual loss distribution for cyclone events on the country level has to be estimated first. To estimate the damage potential of cyclones, different techniques can be used, e.g. stochastic or engineering approaches for estimating physical vulnerability of the assets exposed and combining them with hazard impacts and corresponding probabilities of given events (see Woo 2012 for a detailed discussion on catastrophe modeling approaches). However, as this kind of detailed information is not yet available yet historical losses have to be used for the risk assessment instead. There are two databases available that can be used for such kind of analysis. One is the open-source EMDAT disaster database (EM-DAT, 2012) maintained by the Centre for Research on the Epidemiology of Disasters at the Université Catholique de Louvain. EMDAT currently lists information on people killed, made homeless, and affected as well as overall financial losses for more than 16,000 sudden-onset (such as floods, storms, earthquakes) and slow-onset (drought) events from 1900 to present. Data are compiled from various sources, including UN agencies, non-governmental organizations, insurance companies, research institutes and press agencies. The second one is the newly produced time-series loss data by Malagasy officials based on the "Damage and losses assessment methodology" (from now on called MLoss). It consists of past public sector loss estimates for the Analanjirofo region from 1980 to 2012 separated into different sectorial impacts. Furthermore, the results for the Analanjirofo were upscaled to the national level by assuming the same exposure and vulnerability levels in other areas. The estimates are based on the assumption that losses belong to the maximum domain of attraction of an extreme value distribution (and as losses are always a downside risk) the Frechet type distribution was chosen as the basic loss distribution. For estimating the shape as well as the location parameter, a non-linear optimization model was built, which best fits the curve with the data at hand. Furthermore, to increase the robustness of the results, other models - such as the Generalized Pareto model - were tested and improved in a step-based manner to satisfactory levels (based on graphical tests such as P-P plots and Q-Q plots, see Embrechts et al. 1997 for more information on these techniques). The parameters obtained with this method were used to calculate annual loss return periods.

The same approach was used for the MLoss dataset. This dataset includes losses of the public sector, which are separated into different dimensions including damages to schools, hospitals, the telecommunications system, the environment, and transportation from 1980 to 2012. In total 4 different loss distributions were estimated all with different assumptions as well as different estimation techniques or datasets used, see Table 1.

For these loss distributions, extreme value distributions were fitted and then the relative losses were approximated by piecewise linear df's with knots at

$$
\begin{aligned}
z \quad = \quad & [0, 0.0018, 0.0027, 0.0036, 0.0055, 0.0091, 0.0137, 0.0182, 0.0365, 0.0547, \\
& 0.0730, 0.0912, 0.1095, 0.1460, 0.1825].
\end{aligned}
$$

Scenario	Maximum	Baseline	Minimum	MLoss
No loss	0.406	0.607	0.406	0.406
20	2655	114	372	149.0
50	3991	409	510	204.2
80	4773	775	581	232.5
100	5172	1046	614	245.9
150	5943	1802	676	270.4
200	6530	2646	719	287.7
250	7010	3562	752	301.1
300	7419	4541	780	312.1
400	8097	6657	823	329.4
500	8651	8954	857	342.9

Table 1: Source: Based on Hochrainer 2014. Estimated loss return periods for public cyclone risk based on different estimation methods. Losses (constant 2000) in million USD, public sector losses

5. Solving the Madagascar problem

If we consider the EMDAT model as the one and only model, we are faced with a standard stochastic optimization problem with just one real decision variable y. The pertaining objective function is shown in the Figure 4 below. The optimal exit level is $y^* = 0.1188$, i.e. to cap the insurance at 6510 mill. USD with a premium of 92.0 mill. The expected insured losses are 51.6 mill.
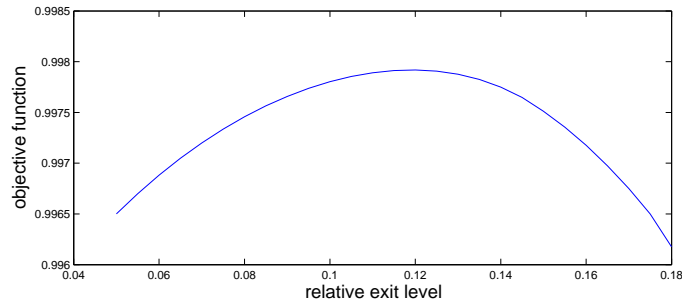

Figure 4: The objective function.

If $F^{(j)}, j = 1, \ldots, \ell$ is a finite set of different possible loss distributions, we solve

$$\max_{y} \min_{j=1,\ldots,\ell} \mathbb{E}_{F^{(j)}}[1 - \xi + \min(\xi, y)] - \pi_y(\xi) - \gamma \mathbb{V}ar_{F^{(j)}}[1 - \xi + \min(\xi, y)]$$
$$\text{s.t. } \pi(\xi_1, y) \leq B/S_0$$
$$y \geq 0.$$

Notice that $F^{(1)}$ is the basic model, on the basis of which all insurance premia are calculated. In the Madagascar case, there are 3 possible loss distribution which are plotted in Figure 5 below

It turns out that the inclusion of the alternative models does not change the optimal exit points ($y^* = 0.1188$), since the EMDAT loss distribution dominates the other models. The minimal objective function is the objective function of the EMDAT case.

However, we solved the full maximin problem for different radius of the ambiguity set, we found the following dependency of the optimal exit point in on the size of the ambiguity set.
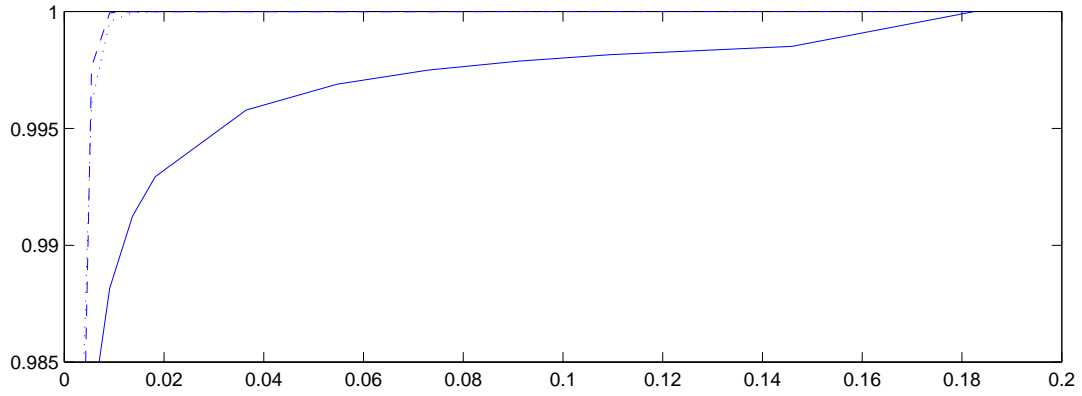
Figure 5: Loss distributions.

ambiguity radius ϵ	optimal exit level in mill.
0	6,510
0.001	6,788
0.005	6,943
0.01	7,073
0.03	7,471
0.05	7,774
0.07	8,044
0.1	8,711
0.15	9,221
0.2	9,331

The relative values of optimal exit points are shown in Figure 6 and the pertaining worst case models are depicted in Figure 7. Together with the optimal solutions, they describe the saddlepoints of our problem.
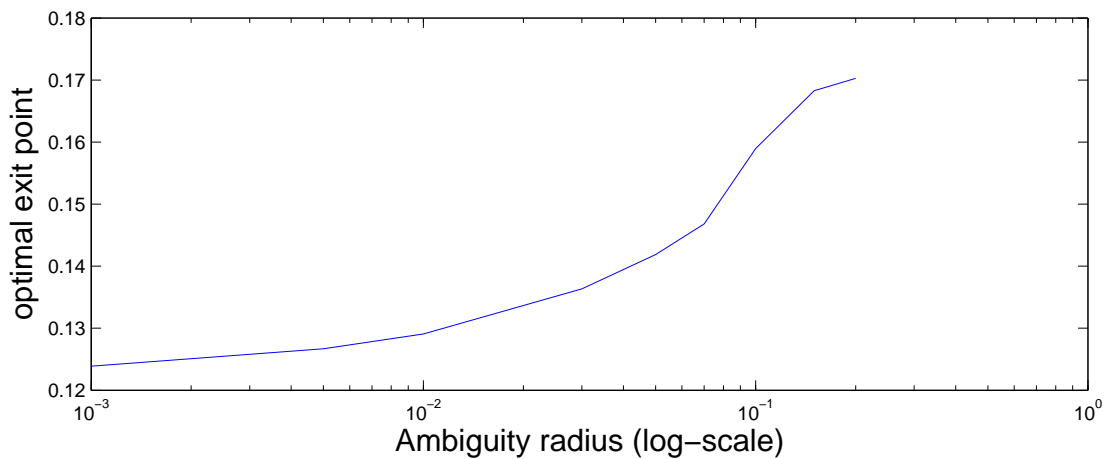

Figure 6: Dependency of the optimal exit point on the size of the ambiguity set.

The worst case models are ordered in the first order stochastic dominance sense.
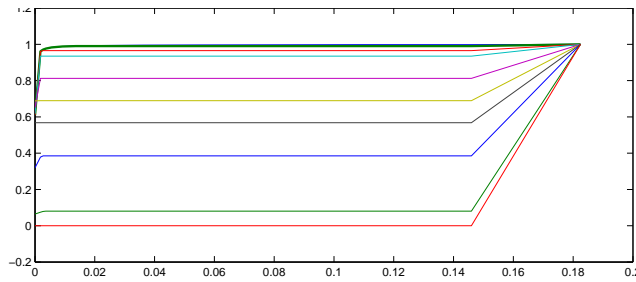
Figure 7: The worst case distribution functions for radii $\epsilon = 0.001, 0.005, 0.005, 0.01$

References

[1] G. Calafiore. Ambiguous risk measures and optimal robust portfolios. SIAM Journal of Control Optimization, 18 (3), pp. 853-877, 2007.

[2] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. Operations Research, 58, pp. 596-612, 2010

[3] J. Dupacová. On minimax decision rule in stochastic linear programing. Studies in Mathematical programming, pp. 47-60, 1980.

[4] J. Dupacová. The minimax approach to stochastic programming and an illustrative application. Stochastics, 20, pp. 73-88, 1987.

[5] J. Dupacová. Uncertainties in minimax stochastic programs. Optimization, 1, pp. 191-220, 2010.

[6] N. C. P. Edirishinge. Stochastic Programming Approximations using Limited Moment Information with Application to Asset Allocation. In: Stochastic Programming: The state of the Art in Honor of George B. Dantzig. International Series in Operations Research and Management Science. Springer, pp. 97-138, 2011.

[7] J. Goh and M. Sim. Distributionally robust optimization and its tractable approximations. Operations Research, 58, pp. 902-917, 2010.

[8] L. P. Hansen and T. J. Sargent. Robustness. Princeton University Press, 2007.

[9] R. Jagannathan. Minimax procedure for a class of linear programs under uncertainty. Operations Research, 25, pp. 173-177, 1977.

[10] Cuncun Luan. Insurance Premium Calculations with Anticipated Utility Theory. Astin Bulletin 31 (1), pp. 23-35, 2001.

[11] G. C. Pflug, A. Pichler and D. Wozabal. The 1/n investment strategy is optimal under high model ambiguity. J. Banking and Finance, 36, pp. 410-417.

[12] G. Ch. Pflug and D. Wozabal. Ambiguity in portfolio selection. Quantitative Finance, 7(4), pp. 435-442, 2007.

[13] G. Ch. Pflug and A. Pichler. Multistage Stochastic Optimization. Springer Verlag, 2014.

[14] A. Raviv. The Design of an Optimal Insurance Policy. The American Economic Review 69 (1), pp. 84-96

[15] H. E. Scarf. A min-max solution of an inventory problem. In Studies in the Mathematical Theory of Inventory and Production, Arrow KJ, Karlin S, Scarf H, (Eds.). Stanford University Press, 1958.

[16] A. Shapiro and A. J. Kleywegt. Minimax analysis of stochastic problems. Optimization Methods and Software, 17, pp. 523-542, 2002.

[17] A. Shapiro and Sh. Ahmed. On a class of minimax stochastic programs. SIAM Journal of Optimization, 14, pp. 1237-1249, 2004.

[18] A. Thiele. Robust stochastic programming with uncertain probabilities. IMA Journal of Management Mathematics, 19, pp. 289-321, 2008.

[19] S. S. Vallander, Calculation of the Wasserstein distance between probability distributions on the line. Theory of Probability and its Applications, 1974, 18 (4), pp. 784-786

Appendix

Convexity of h. If h is a nonnegative 1-homogeneous function on $\mathbb{R}^n$ (i.e. $h(\lambda x) = \lambda h(x)$ for $\lambda > 0$ such that its level sets are closed convex. Then h is convex.

Let $\mathcal{C}$ be the epigraph of h. $\mathcal{C}$ is a cone. For each element $(x, h(x)) \neq (0,0)$ on the boundary of $\mathcal{C}$, let $(\text{proj}(x), 1)$ be the element $(\frac{1}{h(x)}x, 1)$ on the 1-level set. Suppose that $(x, h(x))$ and $(y, h(y))$ are in the boundary of $\mathcal{C}$, but an element (z, α) on the line segment joining them is not in $\mathcal{C}$. Then also $(\text{proj}(z), 1)$ is not in $\mathcal{C}$, but it is on the line segment joining $(\text{proj}(x), 1)$ and $(\text{proj}(y), 1)$. But since these points are in the same level set, this is a contradiction to the assumption.

The solution algorithm.

Notice that the objective function

$$\mathbb{E}_P[1 - \xi + \min(\xi, y)] - \pi_y(\xi) - \gamma \mathbb{V}ar_P[1 - \xi + \min(\xi, y)]$$

is convex in the probability measure P and hence the inner problem

$$\min \mathbb{E}_F[1 - \xi + \min(\xi, y)] - \pi_y(\xi) - \gamma \mathbb{V}ar_F[1 - \xi + \min(\xi, y)]$$
$$F \text{ has a density in } (0, 1], \text{ which is piecewise constant in } [z_i, z_{i+1}]$$
$$d(F, \hat{(F)}) \leq \epsilon$$

is a convex optimization problem in $\mathbb{R}^k$.

The minimax problem is solved in a stepwise manner, for the details of see Algorithm 7.2 on page 221 in [13].