Insurance: Mathematics and Economics 110 (2023) 106–124

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Empirical tail risk management with model-based annealing random search

Qi Fan [a], Ken Seng Tan [b], Jinggong Zhang [b,*]

[a] Barclays, New York, United States of America
[b] Division of Banking and Finance, Nanyang Technological University, Singapore

ARTICLE INFO

Article history:
Received August 2021
Received in revised form February 2023
Accepted 18 February 2023
Available online 28 February 2023

JEL classification:
C61
G22
Q14

Keywords:
Tail risk
Optimization
Random search
Weather index insurance
Value at risk
Conditional value at risk
Entropic value at risk

ABSTRACT

Tail risk measures such as Value at Risk (VaR) and Conditional Value at Risk (CVaR) are popularly accepted criteria for financial risk management, but are usually difficult to optimize. Especially for VaR, it generally leads to a non-convex NP-hard problem which is computationally challenging. In this paper we propose the use of model-based annealing random search (MARS) method in tail risk optimization problems. The MARS, which is a gradient-free and flexible method, can widely be applied to solving many financial and insurance problems under mild mathematical conditions. We use a weather index insurance design problem with tail risk measures including VaR, CVaR and Entropic Value at Risk (EVaR) as the objective function to demonstrate the viability and effectiveness of MARS. We conduct an empirical analysis in which we use a set of weather variables to hedge against corn production losses in Illinois. Numerical results show that the proposed optimization scheme effectively helps corn producers to manage their tail risk.

1. Introduction

Many important decisions and strategies in risk management involve formulating an optimization problem and obtaining its optimal solution under a certain criterion. Due to the asymmetric and heavy-tailed nature of financial and insurance risks, as well as regulatory constraints imposed by, for example, Basel III and Solvency II, tail risk measures such as Value at Risk (VaR) and Conditional Value at Risk (CVaR) have become the most popularly adopted risk criteria nowadays. In the literature of portfolio selection, there is a wide array of discussion extending the classic mean-variance framework of Markowitz (1952) to a mean-VaR model (see e.g., Gilli and Këllezi, 2002; Gaivoronski and Pflug, 2005; Dallagnol et al., 2009; Goh et al., 2012; Lwin et al., 2017), a mean-CVaR model (see e.g., Rockafellar and Uryasev, 2000; Andersson et al., 2001; Gilli and Këllezi, 2002; Lim et al., 2011), and to models with other tail risk measures such as Entropic Value at Risk (EVaR) (see e.g., Ahmadi-Javid and Fallah-Tafti, 2019; Mittal and Srivastava, 2021; Brandtner et al., 2018). These generalizations attempt to overcome the major critiques of adopting variance as a relevant measure of risk, as variance is symmetric, non-monotone, and insufficiently emphasizing extreme scenarios. In the actuarial-related literature, many researchers investigate the problem of optimal insurance/reinsurance design under the VaR and CVaR risk measures (see e.g., Cai et al., 2008; Bernard and Tian, 2009; Hu et al., 2015; Assa, 2015; Balbás et al., 2009; Cheung et al., 2014).

Despite being popularly adopted by both practitioners and regulators as their main risk objectives, the tail risk metrics generally introduce numerical challenges when it comes to optimization. For parametric problem formulations, unless the underlying probability distributions are assumed to be normal which is very light-tailed or lognormal which is moderately heavy-tailed, the optimal solutions are generally difficult to obtain. For non-parametric problem formulations, in which the risk measures are directly estimated from historical observations, there are also technical challenges to find the optimal solutions numerically. Rockafellar and Uryasev (2000) propose an

* Corresponding author.
E-mail addresses: qi.fan@barclays.com (Q. Fan), kenseng.tan@ntu.edu.sg (K.S. Tan), jgzhang@ntu.edu.sg (J. Zhang).

https://doi.org/10.1016/j.insmatheco.2023.02.005
0167-6687/

equivalent formulation to the CVaR-minimizing problem in the context of portfolio selection so that it can be transformed into a convex and linear program, and the method has also been applied to optimal insurance/reinsurance design and risk transfer problems (see e.g., Tan and Weng, 2014; Sun et al., 2017; Asimit et al., 2018). However, this method does not generally apply to other important tail risk measures such as VaR. In fact, optimizing VaR has the fundamental issue in the sense that it does not satisfy subadditivity and sometimes discourages diversification, as well as the computational issue that leads to non-convex NP-hard problems (Benati and Rizzi, 2007). Dallagnol et al. (2009) use a genetic algorithm and particle swarm optimization to solve for the mean-VaR portfolio. Lwin et al. (2017) use an efficient learning-guided hybrid multi-objective evolutionary algorithm to solve mean-VaR portfolio optimization problems with real-world constraints. After all, the literature on non-parametric tail risk minimization, including especially the important case of VaR, is still very limited.

Random search is a family of numerical optimization methods that is widely used to solve complex optimization problems because it only requires the function values rather than any structural information of the objective function (such as convexity or differentiability) to find improved solutions. Variants of random search include simulated annealing (Kirkpatrick et al., 1983), genetic algorithms (Goldberg, 1989), tabu search (Glover and Laguna, 1997), nested partition (Shi and Ólafsson, 2000), adaptive random search (Zabinsky, 2003) and model-based random search (Zlochin et al., 2004; Hu et al., 2008). Typical examples of the latter model-based search methods include ant colony optimization (Dorigo and Blum, 2005), estimation of distribution algorithms (Larrañaga and Lozano, 2002), the Cross-Entropy (CE) method (Rubinstein and Kroese, 2004), and Model Reference Adaptive Search (MRAS) (Hu et al., 2007). The model-based annealing random search (MARS) method of Hu and Hu (2011) extends both CE and MRAS by exploiting the idea from the annealing adaptive search algorithm of Romeijn and Smith (1994). As argued in Hu and Hu (2011), MARS can be interpreted as a gradient search on the parameter space of the surrogate probability distributions for solving a sequence of time-varying stochastic optimization problems with differentiable structures. They also show that the time-varying feature ensures MARS's global convergence with probability one, with the asymptotic convergence rate of at least $O(1/\sqrt{k})$ where k is the number of iterations. Furthermore, by applying to high-dimensional multi-extremal benchmark problems, they show that MARS yields high-quality solutions within a modest number of function evaluations and provides superior performance over some of the existing algorithms. The connection to gradient search provides novel and intuitive insights of how MARS addresses hard optimization problems with little structure. The theoretical convergence results along with the empirical analysis provide additional confidence on the effectiveness of MARS to find optimal solutions for those problems.

It is the theoretical property and the empirical efficiency that motivated us in adopting MARS to solve tail risk optimization problems in finance and insurance. In particular, we investigate the effectiveness of MARS by applying it to solve an optimal design problem associated with weather index insurance. An index insurance is an innovative tool to manage weather risk and to protect agricultural producers from extreme weather events (Chantarat et al., 2013; Conradt et al., 2015; Carter et al., 2016; Zhu et al., 2019). The feature distinguishing index-based insurance from conventional indemnity-based insurance is that the index insurance payout is determined exclusively by a set of pre-determined and transparent weather variables, instead of the actual loss incurred by the policyholder. Therefore, weather index insurance product is able to overcome several critical weaknesses of conventional agricultural insurance, and enjoys many attractive features including alleviating moral hazard and adverse selection, faster and more transparent payment structure, and lower administrative expenses (such as eliminating the need of onsite loss assessment). A central question that relates to weather index insurance is the choice of indemnity function linking weather variables to the index insurance payout. This, in turn, determines the severity of basis risk which quantifies the mismatch between the actual production loss and insurance payout. For a weather index insurance to be practical and effective, it is of paramount importance that the indemnity function should be chosen so as the resulting basis risk is as small as possible. For this reason, the choice of weather index and its indemnity function are an active area of research in the design of weather index insurance. Traditional approaches such as the regression methods, by relating crop yields (linearly or non-linearly) to weather indices, have been used to design weather index insurance (see, e.g. Shi and Jiang, 2016; Dalhaus et al., 2018). Alternatively, the problem of weather index insurance can also be addressed by formulating it as an optimization model. Zhang et al. (2019), for example, derive an optimal design of weather index insurance via a utility maximization framework.

While it may be reasonable to assume that some agricultural producers' preferences are governed by utility functions, it is also plausible that some other producers are more concerned with their exposure to losses in the extreme. As a result, these producers are more interested in an insurance product that is able to curtail their tail risk. Motivated by this argument, we propose a tail risk optimization model and seek a weather index insurance that optimally minimizes producers' exposure to tail risk using measures such as VaR, CVaR and EVaR. While using tail risk as a criterion to construct an optimal weather index insurance seems logical, there has yet been no discussion that explicitly exploits these risk measures in constructing weather index insurance. This, in part, may be due to the computational challenges, especially involving VaR minimization which is a non-convex problem.

Our empirical study provides an in-depth analysis on the relative effectiveness of weather index insurance constructed based on different tail risk measures. We assume a representative corn producer in Illinois, USA, who is keen on minimizing her exposure to tail risk via a weather index insurance policy. Our empirical results validate the effectiveness of the three insurance contracts (based on VaR, CVaR and EVaR objectives respectively) at reducing the producer's basis risk under extreme weather scenarios. Meanwhile, it is also observed that given the same confidence level, the EVaR-minimizing insurance has the highest endogenously determined price among the three while the VaR-minimizing one has the lowest, representing different levels of risk aversion. This finding reemphasizes the critical role of objective function that best characterizes the hedger's risk appetite, and thus highlights the merit of an optimization algorithm like MARS, which is flexibly and robustly applicable to a wide class of risk criteria. Lastly, to address the spatial dependence issue of agricultural data, we also investigate a spatial model to explicitly incorporate spatial spillover effects.

The contribution of this paper is threefold. First, this is the first article to formulate the problem of optimal index insurance design with tail risk objectives. In practice, it is generally important to choose an appropriate objective function that truly reflects the risk manager's or hedger's risk appetite. Second, this paper also represents the first attempt to apply model-based random search optimization methods in insurance risk management. Model-based random search does not require any structural information, such as convexity or differentiability, on the objective function and it usually has good theoretical results, making the method appropriate for tail risk optimization. Finally, we conduct comprehensive empirical analysis using various risk objectives and an extended model with spatial spillover terms. The results demonstrate the potential viability of using weather variables to hedge against agricultural production losses. Assuming a risk loading

factor of 20% and zero premium subsidy from the government, we show that all the proposed index insurance contracts could mitigate tail risk effectively and consistently.

The rest of this article is organized as follows. Section 2 describes the motivation and formulates the weather index insurance design problem as our illustrative example. Section 3 introduces the model-based annealing random search method. Section 4 conducts an empirical investigation on weather index insurance design using corn production and a set of weather variables in Illinois. Section 5 concludes the paper. The appendix collects the proofs of the propositions and additional graphical results.

2. Formulation and motivation

2.1. General problem

Consider a general risk management problem involving a finite number of (random) risk variables $\boldsymbol{X} = \left[X_1, X_2, ..., X_p\right]$ and a finite number of decision variables $\boldsymbol{\theta} = [\theta_0, \theta_1, ..., \theta_d]$ for integers p and d. The goal is to find an optimal decision such that the resulting portfolio is well aligned with the risk manager's appetite, usually represented by a risk measure. Mathematically, if the "true" underlying probabilistic model about $\boldsymbol{X}$ is known, the risk management task can be formulated as an optimization problem,

$$\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \rho\left\{J(\boldsymbol{X};\boldsymbol{\theta})\right\}, \tag{1}$$

where $\boldsymbol{\Theta}$ is the feasible set of possible decisions, $J(\cdot)$ models the loss variable as a function of decision variables, and $\rho(\cdot)$ is an appropriately chosen risk measure. In practice the most popular choices of $\rho(\cdot)$ are Value at Risk (VaR) and Conditional Value at Risk (CVaR), which for a given loss random variable Y and $\forall\ \alpha \in (0, 1)$, are defined respectively as

$$\mathrm{VaR}_\alpha(Y) = \inf\{y \in \mathbb{R} | \Pr(Y \le y) \ge \alpha\},$$

and

$$\mathrm{CVaR}_\alpha(Y) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_\beta(Y)\,\mathrm{d}\beta.$$

Another risk measure we consider in this paper is the Entropic Value at Risk (EVaR),[1] which is advocated by Ahmadi-Javid (2012) and takes the form of,

$$\mathrm{EVaR}_\alpha(Y) = \inf_{t>0}\left\{t^{-1}\ln\left[\frac{\mathrm{E}\left(\mathrm{e}^{tY}\right)}{1-\alpha}\right]\right\}. \tag{2}$$

In this paper, we are interested in finance and insurance related optimization problems as formulated in (1). Under some specific assumptions on the functional form J (such as linear), or on the joint distribution of $\boldsymbol{X}$ (such as normal or lognormal), there is a prolific literature that proposes elegant solutions to solving (1) (see e.g., Gaivoronski and Pflug, 2005; Uryasev, 2000; Rockafellar and Uryasev, 2000). These approaches, however, lack generality as they tend to be very specialized with restrictive applications in finance and insurance. It is the key objective of this paper to provide a flexible and robust method of tackling (1) so that risk managers will have more plausible strategic solutions involving a wider array of risk criteria. To demonstrate some of the inherent challenges associated with solving (1), we consider the problem of seeking an optimal design of weather index insurance that is of significant interest to agricultural producers, insurers, and the government. It should be emphasized that while our illustrations have adopted VaR, CVaR, and EVaR as the respective optimization criteria, the method used to solve the resulting optimization models is very general and is applicable to many desirable optimization frameworks.

2.2. The index insurance design problem

Index insurance is an innovative insurance provision, whose payouts to policyholders are exclusively determined by some transparent indices instead of the actual losses incurred by them. Compared with conventional indemnity-based insurance products, index insurance generally enjoys the benefit of minimal moral hazard and adverse selection, low administrative cost and easily processed claims. One popular application of index insurance is to use weather variables to hedge against agricultural production losses. To make it more effective in protecting policyholders' financial positions, the goal of index insurance design is to determine the payout structure of the index insurance such that their basis risk is minimized. Using the notation "$\vee$" and "$\wedge$" to denote, respectively, $\max(\cdot,\cdot)$ and $\min(\cdot,\cdot)$, the optimal weather index insurance design problem can be formulated mathematically by the following, as a special case of (1) with $d = p - 1$:

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \rho\left\{X_p - I(\boldsymbol{\theta}) + \pi(\boldsymbol{\theta})\right\}, \tag{3}$$

where $I(\boldsymbol{\theta}) = \left[\left(\theta_0 + \sum_{j=1}^{p-1}\theta_j X_j\right) \vee 0\right] \wedge M$ dictates the payout of the prescribed index, θ_j, $j = 0, 1, ..., p-1$ are the decision variables to be determined and $X_1, X_2, ..., X_{p-1}$ are a set of weather variables (together with $\theta_0, \theta_1, \ldots, \theta_{p-1}$) used to hedge against the production loss (represented by X_p). Hence, the payout $I(\boldsymbol{\theta})$ is non-negative and cannot be larger than the pre-specified M, and the difference between

[1] Some important properties of EVaR are summarized in Section 4.5. For more details, refer to Ahmadi-Javid (2012) and Ahmadi-Javid and Fallah-Tafti (2019).

X_p and $I(\boldsymbol{\theta})$ quantifies the basis risk in that it captures the discrepancy between the actual production loss and the insurance payout. Finally, $\pi(\boldsymbol{\theta}) = \gamma \mathrm{E}[I(\boldsymbol{\theta})]$ corresponds to the premium of the index insurance (i.e., hedging cost) where $\gamma - 1 \geq 0$ is the insurance policy's risk loading factor. The sum of basis risk and insurance premium represents the (random) loss position of the agricultural producer after the weather index insurance is purchased. Prudently, the above risk position should be kept as small as possible. This implies that a viable optimal weather index insurance can be defined as the one that minimizes the risk measure of the hedged risk position, which is precisely the optimization model (3).

Optimization model (3) involves a number of practical complexities, especially when a tail risk measure such as VaR, CVaR or EVaR is adopted. First, it is generally very challenging in practice to obtain a good probabilistic model on $\boldsymbol{X}$, which may have heavy-tailed marginal distributions and intricate dependence structure among its marginals. Although there are advanced statistical models such as extreme value theory to model tail behaviors and copulas to model tail dependence, due to data and dimensionality issues statistical modeling often remains a critical obstacle for financial and actuarial application, and misspecified models could potentially lead to substantial *model risk* and hence disastrous practical performance. Therefore in this paper we choose a simple yet useful alternative, the sample average approximation method (see e.g., Rockafellar and Uryasev, 2000; Ahmadi-Javid, 2012; Tan and Weng, 2014), to minimize the empirical tail risk based on an observed sample in a non-parametric framework. This entails reformulating (3) as

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \rho^{(\mathrm{e})}\left(\left\{x_{i,p} - \left[\left(\theta_0 + \sum_{j=1}^{p-1}\theta_j x_{i,j}\right)\vee 0\right]\wedge M + \pi^{(\mathrm{e})}(\boldsymbol{\theta})\right\}_{i=1,2,\ldots,N}\right), \tag{4}$$

where $\left\{[x_{i,1}, x_{i,2}, \ldots, x_{i,p}]\right\}_{i=1,2,\ldots,N}$ is a random sample of length N from population $\boldsymbol{X}$, and $\rho^{(\mathrm{e})}$ is an empirical tail risk measure. For any $\alpha \in (0,1)$, the empirical measures of VaR, CVaR, and EVaR are given, respectively, by

$$\mathrm{VaR}_\alpha^{(\mathrm{e})}\left(\{Y_i\}_{i=1,2,\ldots,N}\right) = Y_{k_N:N}, \tag{5}$$

$$\mathrm{CVaR}_\alpha^{(\mathrm{e})}\left(\{Y_i\}_{i=1,2,\ldots,N}\right) = \frac{(k_N - \alpha N)\, Y_{k_N:N}}{N(1-\alpha)} + \frac{1}{N(1-\alpha)}\sum_{i=k_N+1}^{N} Y_{i:N}, \tag{6}$$

and,

$$\mathrm{EVaR}_\alpha^{(\mathrm{e})}\left(\{Y_i\}_{i=1,2,\ldots,N}\right) = \inf_{t>0}\left\{t^{-1}\ln\left[\frac{\sum_{i=1}^N e^{tY_i}}{N(1-\alpha)}\right]\right\}, \tag{7}$$

where $k_N = \lceil \alpha N \rceil$ denotes the smallest integer greater than or equal to αN, and $Y_{i:N}$ represents the i-th order statistic from a random sample of size N from population Y. Analogously, the cost function $\pi(\boldsymbol{\theta})$ in (3) is also approximated by its empirical counterpart,

$$\pi^{(\mathrm{e})}(\boldsymbol{\theta}) = \frac{\gamma}{N}\sum_{i=1}^{N}\left\{\left[\left(\theta_0 + \sum_{j=1}^{p-1}\theta_j x_{i,j}\right)\vee 0\right]\wedge M\right\}. \tag{8}$$

The second major difficulty is introduced by the form of the risk measure functional ρ, which may be non-convex (e.g., the case of VaR), non-differentiable or may not even have an algebraic expression (e.g., the cases of VaR and CVaR). Non-convexity generally leads to an NP-hard problem which is computationally intractable, and not having an algebraic expression invalidates any gradient-based optimization techniques. For some special risk measures such as CVaR, the problem may stay convex and can be transformed into a more tractable problem (Rockafellar and Uryasev, 2000). However, this optimization technique involves two minimization programs, and more importantly, cannot be applied to other risk measures such as VaR.

To deal with the above mentioned challenges, in this paper we use the method of MARS to solve (4) with risk objectives (5)-(7). More specifically, we define three functions $F:\mathbb{R}^p \mapsto \mathbb{R}$, $G:\mathbb{R}^p \mapsto \mathbb{R}$ and $K:\mathbb{R}^p \mapsto \mathbb{R}$ by,

$$F(\boldsymbol{\theta}) = \mathrm{VaR}_\alpha^{(\mathrm{e})}\left(\left\{x_{i,p} - \left[\left(\theta_0 + \sum_{j=1}^{p-1}\theta_j x_{i,j}\right)\vee 0\right]\wedge M + \pi^{(\mathrm{e})}(\boldsymbol{\theta})\right\}_{i=1,2,\ldots,N}\right), \tag{9}$$

$$G(\boldsymbol{\theta}) = \mathrm{CVaR}_\alpha^{(\mathrm{e})}\left(\left\{x_{i,p} - \left[\left(\theta_0 + \sum_{j=1}^{p-1}\theta_j x_{i,j}\right)\vee 0\right]\wedge M + \pi^{(\mathrm{e})}(\boldsymbol{\theta})\right\}_{i=1,2,\ldots,N}\right), \tag{10}$$

and

$$K(\boldsymbol{\theta}) = \mathrm{EVaR}_\alpha^{(\mathrm{e})}\left(\left\{x_{i,p} - \left[\left(\theta_0 + \sum_{j=1}^{p-1}\theta_j x_{i,j}\right)\vee 0\right]\wedge M + \pi^{(\mathrm{e})}(\boldsymbol{\theta})\right\}_{i=1,2,\ldots,N}\right), \tag{11}$$

where $\mathrm{VaR}_\alpha^{(\mathrm{e})}$, $\mathrm{CVaR}_\alpha^{(\mathrm{e})}$, $\mathrm{EVaR}_\alpha^{(\mathrm{e})}$ and $\pi^{(\mathrm{e})}(\boldsymbol{\theta})$ are given by (5)-(8) respectively. By monotonicity of the above three risk measures, we prove the well-definedness of problem (4).

Proposition 1. *Given a fixed set of observations* $\left\{[x_{i,1}, x_{i,2}, \ldots, x_{i,p}]\right\}_{i=1,2,\ldots,N}$, *function values of* $F(\boldsymbol{\theta})$, $G(\boldsymbol{\theta})$ *and* $K(\boldsymbol{\theta})$ *defined in* (9)-(11), *respectively, are all bounded for all* $\boldsymbol{\theta} \in \mathbb{R}^p$. *Consequently the optimal value of problem* (4) *with any of* (5)-(7) *is finite.*

Proof. See Appendix A.1. □

The gradient of problem (4) is generally difficult to obtain and may not even exist, however, under many commonly adopted risk measures problem (4) is guaranteed to be continuous. For our optimization problems we can establish the continuity of objective functions (9)-(11) by the following proposition. These properties are essential for the use of the numerical scheme to be introduced in Section 3.

Proposition 2. *Functions* $F : \mathbb{R}^p \mapsto \mathbb{R}$, $G : \mathbb{R}^p \mapsto \mathbb{R}$, *and* $K : \mathbb{R}^p \mapsto \mathbb{R}$ *defined by* (9), (10) *and* (11) *are continuous.*

Proof. See Appendix A.2. □

In practice it is usually reasonable to further assume a compact feasible set $\boldsymbol{\Theta} \subset \mathbb{R}^p$, to ensure that the designed payout function is not too sensitive to one or a few weather variables. It is obvious that $F(\boldsymbol{\theta})$, $G(\boldsymbol{\theta})$ and $K(\boldsymbol{\theta})$ are also finite over a compact feasible set, and in the following discussion we assume the feasible set $\boldsymbol{\Theta}$ to be a hyperrectangle.

3. Model-based annealing random search (MARS)

This section provides a brief overview of random search methods, with greater emphasis on the model-based annealing random search (MARS) method proposed by Hu and Hu (2011). Our exposition, which is drawn largely from Hu and Hu (2011), focuses on practical implementation of MARS. Theoretical foundation of MARS can be found in Hu and Hu (2011) and the survey paper Hu et al. (2012).

Suppose we are interested in the global optimizer of the following maximization problem[2]:

$$s^* \in \underset{s\in\mathbb{S}}{\operatorname{argmax}}\, H(s), \tag{12}$$

where $\mathbb{S}$ is a nonempty compact solution space in $\mathbb{R}^p$ (can be either continuous or discrete), and $H(s)$ is a deterministic objective function. In Hu and Hu (2011) the following two assumptions are made:

1. For any $\epsilon < H(s^*)$, the set $\{s \in \mathbb{S} : H(s) \geq \epsilon\}$ has a positive Lebesgue or discrete measure.
2. For any $\delta > 0$, $\sup_{s\in A_\delta} H(s) < H(s^*)$, where $A_\delta := \{s \in \mathbb{S} : ||s - s^*|| \geq \delta\}$ with $||\cdot||$ being the Euclidean norm in $\mathbb{R}^p$, and we define the supremum over the empty set to be $-\infty$.

The first assumption is automatically satisfied if the objective function $H(s)$ is continuous, so are our tail risk objective functions $F(\boldsymbol{\theta}), G(\boldsymbol{\theta}), K(\boldsymbol{\theta})$ defined respectively by (9)-(11) (see Proposition 2 in Section 2). The second assumption ensures that the global optimal solution is unique in the feasible region. For our problem (4), the uniqueness of the global optimal solution relies not only on the chosen risk measure $\rho^{(e)}$ but also on the specific data sample $\left\{[x_{i,1}, x_{i,2}, ..., x_{i,p}]\right\}_{i=1,2,...,N}$. If the random sample is simulated from a known probabilistic model $\boldsymbol{X}$, the uniqueness may be established via the property of $\boldsymbol{X}$; however, if the sample contains real observations from practice, it is generally challenging to verify the uniqueness. Without the uniqueness assumption, Hu et al. (2011) show that the algorithm could also converge to certain set of stationary points to the problem and can therefore be used to find high-quality solutions in practice.

As pointed earlier, the global optimization problem (12), in general, is extremely difficult to solve exactly due to the possibility that the objective function is nonconvex, multimodal, or badly scaled. Driven by the complexity of the optimization problem and the desire of having a flexible numerical optimization method, in this paper we explore using a random search algorithm to solve (12). A random search algorithm is typically an iterative approach involving some kind of randomized mechanism to generate, iteratively, a sequence of candidate solutions or probabilistic models, in order to successively approximate the optimal solution. The random search algorithm has been shown to be useful for solving many ill-structured global optimization problems where the objective function may be non-convex and non-differentiable (Zabinsky, 2010).

Zlochin et al. (2004) classify random search algorithms as either instance-based or model-based. The instance-based method iteratively generates new candidate solution(s) based on the previously generated solutions. Evolutionary programming (Eiben and Smith, 2003), annealing (Kirkpatrick et al., 1983), genetic algorithms (Goldberg, 1989), tabu search (Glover and Laguna, 1997), and nested partition (Shi and Ólafsson, 2000) are popular examples of this category. The candidate solutions in the model-based random search method, on the other hand, are generated iteratively by sampling from an intermediate probability distribution model over the solution space that is updated iteratively. Established algorithms include ant colony optimization (Dorigo and Blum, 2005), estimation of distribution algorithms (Larrañaga and Lozano, 2002), the Cross-Entropy (CE) method (Rubinstein and Kroese, 2004), and Model Reference Adaptive Search (MRAS) (Hu et al., 2007). More specifically, if p_k represents a probability distribution on $\mathbb{S}$ for the k-th iteration, then the model-based random search method basically involves the following two steps (Hu et al., 2012):

Step 1. Randomly generate a set/population of candidate solutions $\mathbb{S}$ from p_k.
Step 2. Update p_k based on the sampled solutions in $\mathbb{S}$ to obtain a new distribution p_{k+1}; increase k by 1 and reiterate from Step 1.

The above algorithm shows that the probability distribution, not the candidate solutions (as in instance-based methods), will be updated at each iteration step. Furthermore, as $k \to \infty$, the intent is to have p_k converge to a limiting distribution that assigns most of its probability density/mass to the set of optimal solutions (Hu et al., 2012).

[2] A minimization problem can be formulated as a maximization problem by taking the negative of the objective function.

An important theoretical result inspired by the annealing adaptive search (AAS) algorithm of Romeijn and Smith (1994) is that the candidate solutions in the above algorithm are generated from a Boltzmann distribution g_k with a decreasing temperature series T_k at each iteration k:

$$g_k(s) = \frac{e^{H(s)/T_k}}{\int_{\mathbb{S}} e^{H(s)/T_k} \nu(ds)}, \tag{13}$$

where $\nu(\cdot)$ is the Lebesgue or discrete measure. As shown in Hu and Hu (2011), if the sequence T_k decreases to a positive number very close to zero, then g_k will concentrate the probability density/mass on the solution that maximizes $H(s)$ based on the construction of $g_k(s)$. For the class of Lipschitz optimization problems, it has been shown that the expected number of iterations for AAS to achieve a given level of precision increases at most linearly in the dimension of the problem (Zabinsky, 2003). While this result provides theoretical justification for using Boltzmann distribution as the sampling distribution, it has limited practical value due to the difficulty in sampling exactly from a Boltzmann distribution. Stipulated by this limitation, the model-based annealing random search (MARS) by Hu and Hu (2011), which extends CE and MRAS, addresses the computational challenge by using a parameterized distribution f_{ϕ_k} to approximate g_k at each iteration k. The candidate solutions are then sampled from the resulting approximated distributions. For this reason, Hu and Hu (2011) refer to f_{ϕ_k} with parameter space $\phi_k \in \Phi$ as the surrogate distribution and g_k as the reference distribution. For a given family of parameterized distributions $\{f_\phi, \phi \in \Phi \subseteq \mathbb{R}^n\}$, an optimal surrogate probability distribution f_{ϕ_k} at iteration k that best resembles the Boltzmann distribution g_k can be determined by solving the following minimization problem (Hu and Hu, 2011):

$$\phi_k = \underset{\phi \in \Phi}{\operatorname{argmin}}\, \mathscr{D}(g_k, f_\phi), \tag{14}$$

where $\mathscr{D}(g_k, f_\phi) = \mathrm{E}_{g_k}\left[\ln \frac{g_k(S)}{f_\phi(S)}\right]$ is the Kullback-Leibler divergence between g_k and f_ϕ, S denotes a generic random vector taking values in the solution space $\mathbb{S}$ and $\mathrm{E}_g[\cdot]$ denotes the expectation taken with respect to the probability density/mass function g. In other words, $\mathrm{E}_{g_k}\left[\ln \frac{g_k(S)}{f_\phi(S)}\right]$ is the expectation of $\ln \frac{g_k(S)}{f_\phi(S)}$ when the random vector $S \sim g_k$. This surrogate probability distribution approximation process is illustrated in Fig. 1.

For practical implementation of MARS, Hu and Hu (2011) recommend the following two additional refinements. First, at the $(k+1)$-th iteration, the optimal parameter ϕ_{k+1} for the surrogate distribution $f_{\phi_{k+1}}$ is determined by solving,

$$\phi_{k+1} = \underset{\phi \in \Phi}{\operatorname{argmin}}\, \mathscr{D}(\widetilde{g}_{k+1}(s), f_\phi), \tag{15}$$

where

$$\widetilde{g}_{k+1}(s) = \alpha_k g_{k+1}(s) + (1-\alpha_k) f_{\phi_k}(s), \tag{16}$$

with some smoothing parameter $\alpha_k \in (0, 1]$. This implies that the optimal surrogate distributions are not directly obtained from (14). Instead, they are determined from (15) using sequence $\{\widetilde{g}_k\}$ that is modified from the original sequence of reference distributions $\{g_k\}$ via (16). By construction, the modified $\widetilde{g}_{k+1}$ will remain close to f_{ϕ_k}. This has the advantage of reducing variability of f_{ϕ_k} across iterations.

The second refinement relates to the choice of surrogate distribution f_ϕ. It is ideal that the surrogate distribution is relatively simple so that the candidate solutions can be sampled easily. To this end, the Natural Exponential Family (NEF), which has the form $f_\phi(s) = \exp(\phi^\intercal \Gamma(s) - K(\phi))$, where $\Gamma(\cdot) : \mathbb{R}^p \mapsto \mathbb{R}^n$ is the sufficient statistic for NEF and $K(\phi) = \ln \int_{\mathbb{S}} \exp(\phi^\intercal \Gamma(s)) \nu(ds)$, is recommended (Hu and Hu, 2011). Based on the form of NEF, $m(\phi) := \mathrm{E}_{f_\phi}[\Gamma(S)]$ is strictly positive definite and invertible. In other words, once we obtain $m(\phi)$, we can obtain ϕ. In addition to these desirable properties, NEF is quite general in that it encompasses a wide range of commonly used distributions including normal distributions. For example, in the case of univariate normal distribution with mean μ and variance σ^2, we have $\phi = \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)^\intercal$, $\Gamma(s) = \left(s, s^2\right)^\intercal$ and $m(\phi) = \left(\mu, \mu^2 + \sigma^2\right)^\intercal$. Independent multivariate normal distributions are implemented in MARS and will also be used in the implementation for our problems in this paper.

Another useful result provided in Hu and Hu (2011) is that if the new parameter ϕ_{k+1} of the NEF-based surrogate distribution is optimally obtained from minimizing $\mathscr{D}(\widetilde{g}_{k+1}(s), f_\phi)$, then

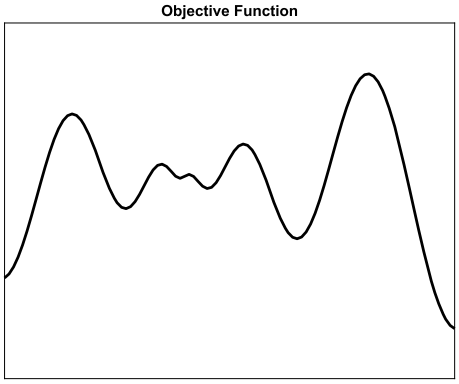
$$m(\phi_{k+1}) = m(\phi_k) - \alpha_k \nabla_\phi \mathscr{D}(g_{k+1}(s), f_\phi)|_{\phi=\phi_k}, \forall\, k = 0, 1, 2, \cdots. \tag{17}$$

This indicates that MARS is essentially a gradient search algorithm on the parameter space Φ of the surrogate probability distributions for solving a sequence of time-varying optimization problems $\mathscr{D}(g_{k+1}(s), f_\phi)$ with differentiable structures, where the solution to the limiting optimization problem is the optimal solution to the original objective function $H(s)$ as the limiting distribution of g_k concentrates the probability density/mass on the solution that maximizes $H(s)$. This also provides the intuition behind the effectiveness of MARS.

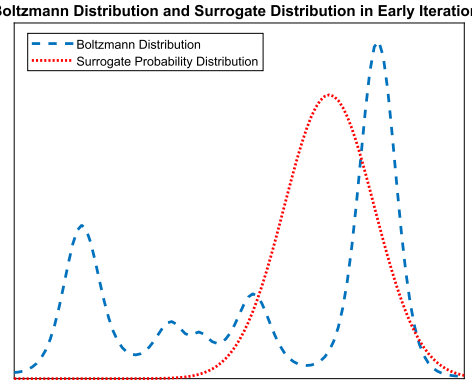
Following the property of NEF, Hu and Hu (2011) also show that updating (17) can be equivalently written as

$$\begin{aligned} m(\phi_{k+1}) &= (1-\alpha_k) m(\phi_k) + \alpha_k \mathrm{E}_{g_{k+1}}[\Gamma(S)] \\ &= (1-\alpha_k) m(\phi_k) + \alpha_k \frac{\int_{\mathbb{S}} e^{H(s)/T_{k+1}} \cdot \Gamma(s) \nu(ds)}{\int_{\mathbb{S}} e^{H(s)/T_{k+1}} \nu(ds)}. \end{aligned} \tag{18}$$

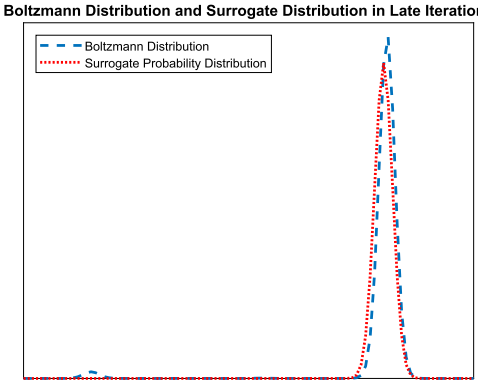
Because the above integrations cannot be evaluated exactly, they are evaluated with the solutions $s_k^1, s_k^2, \cdots, s_k^{N_k}$ generated from the previous surrogate probability distribution f_{ϕ_k}. As a result, Hu and Hu (2011) approximate the fraction in (18) as



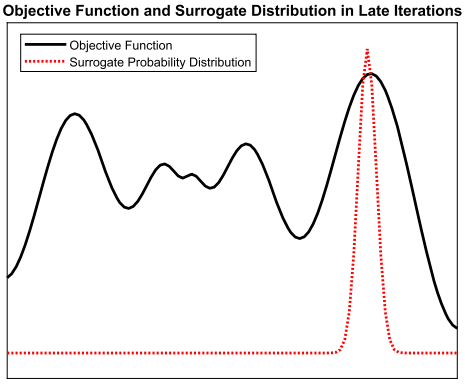
(a) The objective function, which contains multiple local maximal solutions



(b) The Boltzmann distribution and the surrogate probability distribution that "looks like" the Boltzmann distribution in early iterations



(c) The Boltzmann distribution and the surrogate probability distribution that "looks like" the Boltzmann distribution in late iterations



(d) The probability density concentrates around the global maximal solution in late iterations

Fig. 1. A schematic description of MARS.

$$\frac{\int_{\mathbb{S}} e^{H(s)/T_{k+1}} \cdot \Gamma(s)\nu(ds)}{\int_{\mathbb{S}} e^{H(s)/T_{k+1}} \nu(ds)} \approx \frac{\frac{1}{N_k}\sum_{i=1}^{N_k} \frac{e^{H\left(s_k^i\right)/T_{k+1}} \cdot \Gamma(s_k^i)}{f_{\phi_k}\left(s_k^i\right)}}{\frac{1}{N_k}\sum_{i=1}^{N_k} \frac{e^{H\left(s_k^i\right)/T_{k+1}}}{f_{\phi_k}\left(s_k^i\right)}}. \tag{19}$$

Intuitively, the division by $f_{\phi_k}\left(s_k^i\right)$ exploits the idea of importance sampling to compensate for solutions that are unlikely to be chosen, and guarantees that the sample averages in both the numerator and the denominator are unbiased estimators of the integrals in (19). Despite the errors introduced in approximating the integral, with mild assumptions on the sequences of $\{\alpha_k\}_{k\geq 1}$, $\{T_k\}_{k\geq 1}$, and $\{N_k\}_{k\geq 1}$, Hu and Hu (2011) show that $\lim_{k\to\infty} m(\phi_k) = \Gamma(s^*)$ with probability one, where s^* is the optimal solution to the original objective function $H(s)$. When independent multivariate normal distributions are used as the surrogate probability distributions, this result indicates that the mean vector of the multivariate normal distributions will converge to s^* and the variance matrix of the multivariate normal distributions will converge to a zero matrix with probability one.

For our implementation of MARS, we choose $\alpha_k = 1/(k+100)^{0.501}$, $N_k = \max(4, \lfloor k^{0.502} \rfloor)$ and $T_k = \left|H(s_k^*)\right| / \ln(1+k)$, where $\lfloor a \rfloor$ is the largest integer that is no greater than a and s_k^* is the current best solution found up to the k-th iteration of the algorithm. These choices are consistent with those in Hu and Hu (2011) and Hu et al. (2014). The details of MARS implemented for our problems are presented by Algorithm 1.

4. Numerical example: index insurance design

In this section we conduct an empirical study on weather index insurance design, by applying the MARS scheme to problem (4) with risk objectives (5)-(7). Section 4.1 describes the data and assumptions used in the empirical investigation. Section 4.2 discusses the model selection process. Sections 4.3-4.5 investigate the results and evaluate model performance under CVaR objective (6), VaR objective (5)

Algorithm 1: Model-based Annealing Random Search (MARS).

Output: Optimal solution s^*
Input : Objective function $H(s)$
Initialization: Choose an initial probability density/mass function $f_{\phi_0}(s)$ on $\mathbb{S}$.
Initialization: Specify the sequences for the smoothing parameter $\{\alpha_k\}_{k\geq 1}$, annealing temperature schedule $\{T_k\}_{k\geq 1}$ and sample size in each iteration $\{N_k\}_{k\geq 1}$.
Initialization: Specify a stopping condition.
Initialization: Set iteration counter $k = 0$.
while *The stopping condition is not met* **do**
  Generate N_k i.i.d samples $s_k^1, s_k^2, \cdots, s_k^{N_k}$ from the current probability density/mass function $f_{\phi_k}(s)$;
  Compute the new parameter $m(\phi_{k+1})$ according to (18) and (19), which has a one-to-one mapping to ϕ_{k+1};
  Update s^* with the current best solution: $s^* \leftarrow \underset{s\in\left(\{s^*\}\cup\left\{s_k^j\right\}_{1\leq j\leq N_k}\right)}{\operatorname{argmax}}\{H(s)\}$;
  Update $k \leftarrow k+1$;
return s^*

and EVaR objective (7), respectively, at $\alpha = 95\%$ level. Section 4.6 extends the model to a spatially lagged model to incorporate spatial dependence.

4.1. Data and assumptions

In our empirical analysis we consider an insurance provider who is interested in underwriting a weather index insurance for corn producers in Illinois, USA. The annual loss data is calculated from the county-level annual corn production experience over 1925-2017 collected from the National Agricultural Statistics Service (NASS),[3] together with corn price from the United States Department of Agriculture (USDA) Economic Research Service.[4] The crop yield data is first normalized to dollars per acre, and then detrended for each county to remove the impact of technological improvement, economic conditions, etc. The available weather indices contain six different weather variables of importance to corn production, including precipitation, maximum temperatures, minimum temperatures, maximum vapor pressure deficit (VPD), minimum VPD, and dew points, with their six monthly observations from May to October (i.e., the growing season of corn in Illinois). Hence for each annual agricultural cycle we have a total of $6 \times 6 = 36$ observed weather variables. Using the notations in Section 2 we denote a vector of weather variables by the 36-dimensional weather index $[X_1, X_2, ..., X_{36}]$, on which an index insurance payoff is constructed to hedge against the agricultural production loss X_{37}. The respective weather data is obtained from the PRISM Climate Group.[5]

The total length of data sample is $N = 6,789$, which corresponds to a 93-year observation period multiplied by 73 counties in Illinois. Due to the non-parametric nature of problem (4), the optimization must be interpreted in the statistical sense and the "bias-variance" trade-off must be considered. In other words, the ultimate goal is to find an insurance payout function which minimizes tail risk when evaluated on a test sample, rather than the training sample used to solve the model. To apply the insurance design framework (4) in a practically meaningful manner, we partition the data sample into two parts by preserving its temporal order: a training set containing all data within the period from 1925 to 2003 (about 85% of data) is used to obtain the optimal $\boldsymbol{\theta}$ in problem (4), and a test set with the remaining sample period 2004-2017 (about 15% of data) is used to evaluate the out-of-sample performance of our model.[6]

We then apply the "min-max feature scaling" to each dimension of $\boldsymbol{X}$ independently and normalize each value onto the range of $[0, 1]$. Because the transformation is linear, the normalization does not affect the interpretability of the results. The empirical loss distributions depicted in Fig. 2 for the respective samples exhibit positive skewness and a very heavy right tail. These empirical phenomena are common for financial and insurance risks and thus highlight the importance of tail risk management.

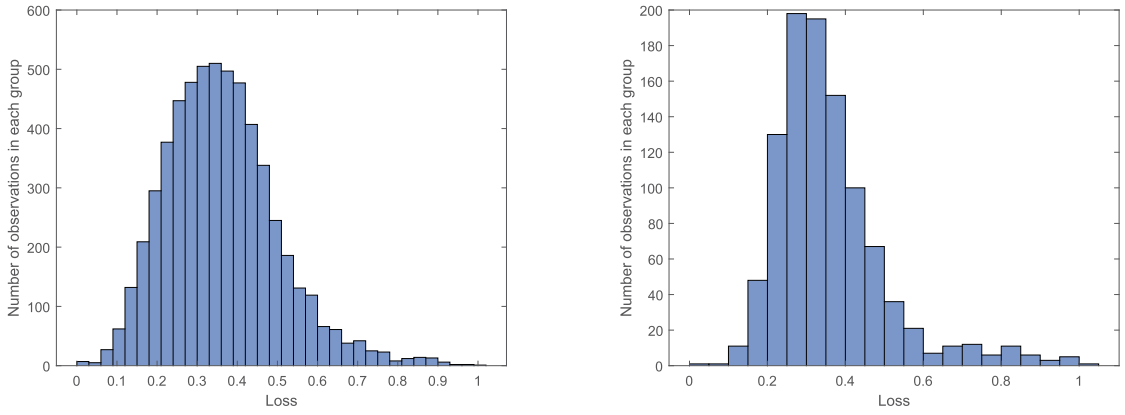
Additionally, in problem (4) we assume the largest insurance payout M to be 1 which corresponds to the largest historical loss, and we also assume a risk loading factor of $\gamma = 1.2$ which represents its cost efficiency compared with those conventional indemnity-based insurance products. Since $\gamma > 1$, the risk/reward tradeoff from the policyholder's perspective is directly considered within the optimization framework. Finally, we adopt a confidence level of $\alpha = 95\%$ in all risk objectives (9)-(11) throughout our numerical analysis.

4.2. Model selection

The preceding subsection has pointed out we have access to thirty six monthly weather variables that can be incorporated in our index insurance design. It is conceivable not all of these weather variables will be of (high) relevance. From the theory of statistical learning, a larger feasible set while, in theory, will produce a better solution to problem (4), it might lead to an overfitted solution with poor out-of-sample performance. Consequently in view of model parsimony, it is sensible to restrict the feasible set in (4). To this end, we proceed with the following steps.

First we apply the principal component analysis (PCA) to the 36-dimensional weather index to remove multicollinearity in the data. Fig. 3, which is the scree plot that shows how much variability is explained by each principal component (PC), indicates that the first ten PCs explain roughly 85% of the total variability.

[3] USDA National Agricultural Statistics Service, https://www.nass.usda.gov/.
[4] USDA Economic Research Service, https://www.ers.usda.gov/data-products/commodity-costs-and-returns/.
[5] PRISM Climate Group, Oregon State University, https://prism.oregonstate.edu/.
[6] We should acknowledge that it is extremely challenging to obtain an i.i.d. sample in this specific context, because spatial dependence inevitably exists among counties. However, our goal focuses on capturing the relationship between crop yield loss and weather conditions. There is generally a tradeoff between explicitly capturing spatial dependence and limiting model risk, and the non-parametric approach adopted here leans towards the latter. The way in which the data are split into training and test samples, together with the empirical results to be exhibited in Sections 4.3-4.5, indicate the practical significance of our empirical analysis. In Section 4.6 we discuss an extended model which incorporates spatial structure and find consistent results.



(a) Training sample (b) Test sample

Fig. 2. Histograms of the loss data.

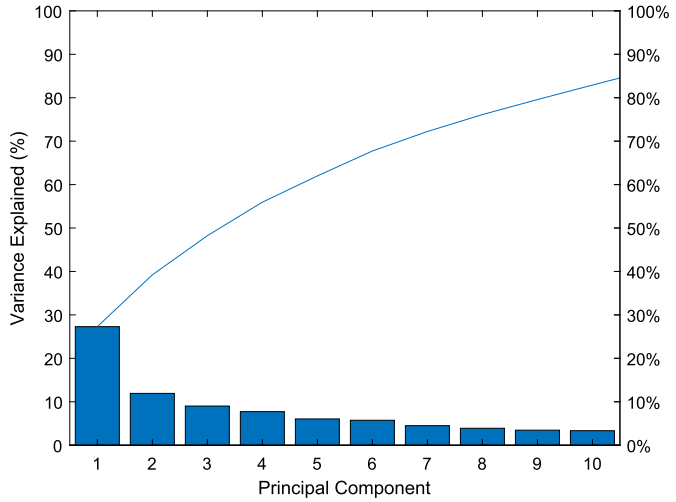


Fig. 3. Percent variability explained by each PC.

To further regularize the model, we extend the optimization problem (4) with the Least Absolute Shrinkage and Selection Operator (LASSO) of Tibshirani (1996) and it is given by

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^p} \rho^{(e)}\left(\left\{x_{i,p} - \left[\theta_0 + \left(\sum_{j=1}^{p-1}\theta_j x_{i,j}\right) \vee 0\right] \wedge M + \pi^{(e)}(\boldsymbol{\theta})\right\}_{i=1,2,\ldots,N}\right) + \lambda||\boldsymbol{\theta}||_1, \tag{20}$$

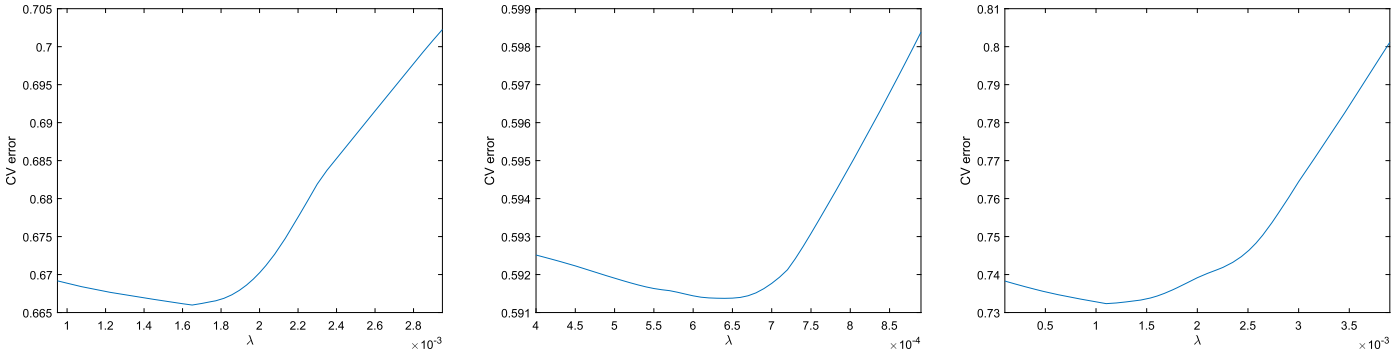
where $||\boldsymbol{\theta}||_1 := \sum_{j=0}^{p-1}|\theta_j|$ is the L1 norm, and λ is the tuning parameter to control the amount of shrinkage. In our analysis, the value of λ is determined by a 5-fold cross-validation (Picard and Cook, 1984). More specifically, we randomly partition the training sample into five equal sized subsamples (each representing 17% of the entire data set), use four of them each time to optimize our model (20) and the last one to estimate the prediction error. Fig. 4 plots the cross-validation errors as functions of λ and the λ that yields the lowest cross-validation error is then chosen to be the value for (20) with the corresponding risk metric. Via this process we strike a balance between overfitting and underfitting, and ensure the out-of-sample performance of our index insurance policy.

Once we identify the appropriate λ for model (20), we re-optimize the model using the whole training sample[7] (representing 85% of the entire data set), with their performance to be investigated in detail in Sections 4.3-4.5.

4.3. Results with the CVaR-minimizing insurance

In this subsection we investigate the hedging performance of the proposed CVaR-minimizing insurance. Fig. 2 reveals that the unhedged loss position for corn producers has a very heavy right tail. To show how the proposed insurance improves the policyholder's position, columns 2-4 of Tables 1 and 2 highlight the differences between the unhedged and the hedged positions.

[7] Because optimization Algorithm 1 involves randomness when sampling from the surrogate probability distribution f_{ϕ_k}, we apply model bagging with 40 random seeds to improve the stability and accuracy of results.



(a) CVaR minimization (b) VaR minimization (c) EVaR minimization

Fig. 4. Cross-validation (CV) errors as functions of the tuning parameter λ.

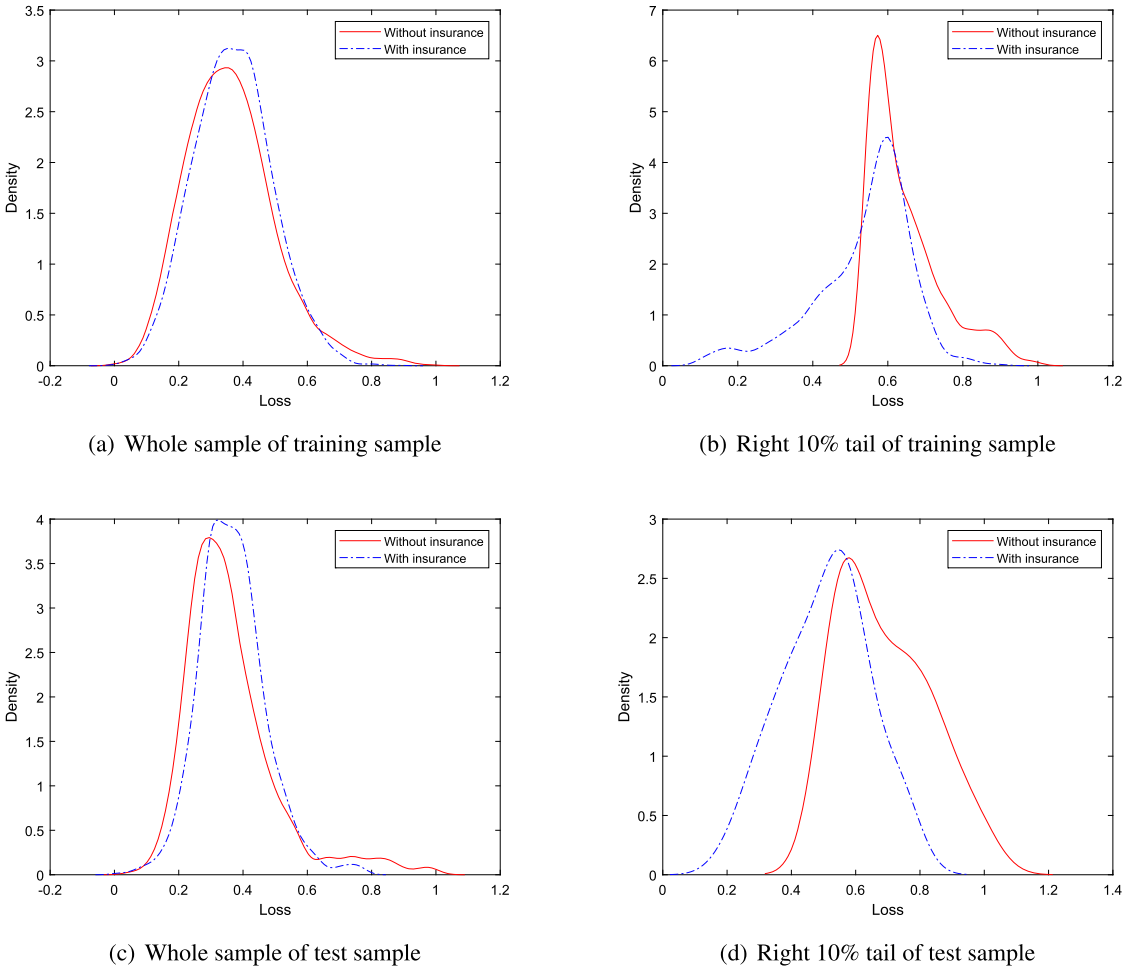
Table 1
Training-sample performance of the three proposed index insurance based on $\text{CVaR}_{95}^{(e)}$, $\text{VaR}_{95}^{(e)}$, and $\text{EVaR}_{95}^{(e)}$, respectively. Semi-deviation is defined as the expected downside deviation from the mean, $SD(Y) = \left(\text{E}\left[(Y - \text{E}(Y))^2 \cdot 1_{[Y \geq \text{E}(Y)]}\right]\right)^{1/2}$. For risk reduction, numbers in parentheses are standard deviations of the estimates obtained by bootstrapping.

		$\text{CVaR}_{95}^{(e)}$ minimizing		$\text{VaR}_{95}^{(e)}$ minimizing		$\text{EVaR}_{95}^{(e)}$ minimizing	
	Without insurance	With insurance	Risk reduction	With insurance	Risk reduction	With insurance	Risk reduction
Estimated price	–	0.0460	–	0.0327	–	0.0594	–
Mean	0.3615	0.3692	–	0.3670	–	0.3714	–
Standard deviation	0.1402	0.1213	0.135 (0.0084)	0.1183	0.1565 (0.0064)	0.1259	0.1022 (0.0104)
Skewness	0.7325	0.1809	–	0.3351	–	-0.0721	–
Kurtosis	4.0653	2.9969	–	3.1833	–	3.3257	–
Semi-deviation	0.1076	0.0876	0.1857 (0.0104)	0.0870	0.1916 (0.0084)	0.0883	0.1795 (0.0112)
$\text{VaR}_{95}^{(e)}$	0.6084	0.5744	0.0558 (0.0113)	0.5682	0.0662 (0.0103)	0.5784	0.0493 (0.0117)
$\text{CVaR}_{95}^{(e)}$	0.7157	0.6317	0.1174 (0.0113)	0.6318	0.1172 (0.0103)	0.6322	0.1167 (0.0118)
$\text{EVaR}_{95}^{(e)}$	0.7933	0.6875	0.1334 (0.0138)	0.6955	0.1233 (0.0105)	0.6863	0.1349 (0.0138)
$\text{VaR}_{99}^{(e)}$	0.7812	0.6661	0.1474 (0.0112)	0.6687	0.144 (0.0102)	0.6603	0.1548 (0.0116)
$\text{CVaR}_{99}^{(e)}$	0.8641	0.7122	0.1757 (0.0113)	0.7316	0.1533 (0.0103)	0.7104	0.1778 (0.0116)
$\text{EVaR}_{99}^{(e)}$	0.8966	0.7675	0.144 (0.0137)	0.7770	0.1334 (0.0103)	0.7642	0.1477 (0.0137)

Table 2
Test-sample performance of the three proposed index insurance based on $\text{CVaR}_{95}^{(e)}$, $\text{VaR}_{95}^{(e)}$, and $\text{EVaR}_{95}^{(e)}$, respectively. Semi-deviation is defined as the expected downside deviation from the mean, $SD(Y) = \left(\text{E}\left[(Y - \text{E}(Y))^2 \cdot 1_{[Y \geq \text{E}(Y)]}\right]\right)^{1/2}$. For risk reduction, numbers in parentheses are standard deviations of the estimates obtained by bootstrapping.

		$\text{CVaR}_{95}^{(e)}$ minimizing		$\text{VaR}_{95}^{(e)}$ minimizing		$\text{EVaR}_{95}^{(e)}$ minimizing	
	Without insurance	With insurance	Risk reduction	With insurance	Risk reduction	With insurance	Risk reduction
Estimated price	–	0.0489	–	0.0353	–	0.0628	–
Mean	0.3586	0.3668	–	0.3645	–	0.3691	–
Standard deviation	0.1442	0.1034	0.2829 (0.0215)	0.1057	0.267 (0.0164)	0.1118	0.225 (0.0269)
Skewness	1.6419	0.5379	–	0.8344	–	0.1381	–
Kurtosis	6.6568	4.2888	–	4.5424	–	4.7703	–
Semi-deviation	0.1207	0.0777	0.3559 (0.021)	0.0819	0.321 (0.017)	0.0806	0.3318 (0.0231)
$\text{VaR}_{95}^{(e)}$	0.6643	0.5419	0.1843 (0.0443)	0.5467	0.1771 (0.0392)	0.5560	0.163 (0.0458)
$\text{CVaR}_{95}^{(e)}$	0.8068	0.6197	0.2319 (0.0448)	0.6416	0.2048 (0.0399)	0.6283	0.2212 (0.0462)
$\text{EVaR}_{95}^{(e)}$	0.8671	0.6795	0.2164 (0.015)	0.6976	0.1955 (0.0121)	0.6919	0.2021 (0.0161)
$\text{VaR}_{99}^{(e)}$	0.8754	0.6999	0.2005 (0.0441)	0.7167	0.1812 (0.039)	0.6924	0.2091 (0.0455)
$\text{CVaR}_{99}^{(e)}$	0.9505	0.7389	0.2226 (0.0443)	0.7566	0.204 (0.0393)	0.7547	0.206 (0.0459)
$\text{EVaR}_{99}^{(e)}$	0.9688	0.7490	0.2268 (0.0152)	0.7709	0.2043 (0.0122)	0.7643	0.211 (0.0164)

Tables 1 and 2 show that this $\text{CVaR}_{95}^{(e)}$-minimizing insurance effectively reduces $\text{CVaR}_{95}^{(e)}$ by 11.7% (23.2%) on the training (test) sample. It is of interest to note that $\text{CVaR}_{99}^{(e)}$ is also significantly reduced, by 17.6% (22.3%) on the training (test) sample, even though the optimal index insurance has been designed to minimize CVaR at 95%, rather than at 99%. These results indicate that the optimal index insurance is robust to managing tail risk so that most severe losses in the right tail are effectively mitigated. In addition, hedging tail risk via the index insurance increases the mean of loss variable. This is to be expected in view of the additional cost incurred for the insurance premium. Because we assume the risk loading factor γ to be 1.2, on average the difference between the hedged and unhedged loss is equal to



(a) Whole sample of training sample

(b) Right 10% tail of training sample

(c) Whole sample of test sample

(d) Right 10% tail of test sample

Fig. 5. Loss distribution with CVaR minimization: with index insurance vs. without index insurance. The density curves are smoothed by normal kernel functions.

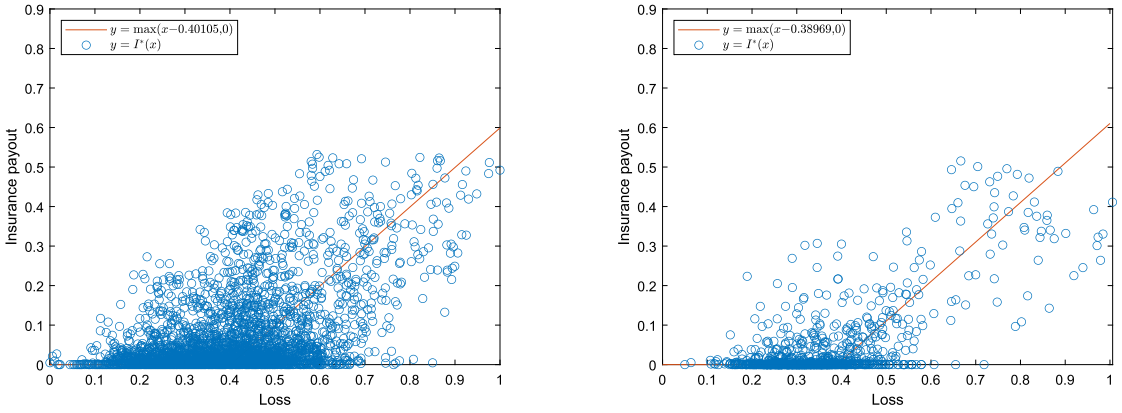
$\frac{\gamma-1}{\gamma} = \frac{1}{6}$ of the insurance's estimated price. We also note that the estimated price of the proposed index insurance is \$37.2 (\$39.5) on the training (test) sample when they are converted back to unnormalized data. Finally, we calculate the standard deviations of risk reductions for various risk metrics using bootstrapping and find that all these risk reductions are statistically significant.

Fig. 5 compares the distributions of loss variable with and without the optimal index insurance. Similar to the mean, the mode of density function increases in the presence of insurance, representing the risk premium paid to the insurer. We can see from Figs. 5(a) and 5(c) that the proposed index insurance significantly reduces the skewness of loss distribution, indicating its high effectiveness in managing tail risk. Figs. 5(b) and 5(d) enlarge the tail part of the distribution, by drawing the conditional density function given an extreme loss event in the upper 10% tail. It is clearly demonstrated that the largest losses are effectively managed. Lastly, the proposed insurance exhibits very similar pattern on both training (Figs. 5(a) and 5(b)) and test (Figs. 5(c) and 5(d)) samples, indicating that the overfitting issue is likely insignificant.

The scatter plots in Fig. 6 provide further comparison on the pointwise relationship between each observed loss and its corresponding insurance payout. When small losses occur, the chances for zero insurance payouts are high. Similar to conventional indemnity-based insurance, existence of this "insurance deductible" can be explained by a positive risk loading factor ($\gamma - 1 = 0.2$). Since the goal of this index insurance is to manage tail risk and the objective function (4) only penalizes a one-sided hedging error, in Fig. 6 we observe many more overpaid cases (top-left region in the scatter plots) than underpaid cases (bottom-right region in the scatter plots). It is worth mentioning that as long as the hedging objective (4) is achieved, the over-payment issue does not generally lead to moral hazard problems, because all the underlying weather variables are transparently observable and can hardly be manipulated by either insurers or policyholders. This is one of the major advantages of index insurance products.

4.4. Results with the VaR-minimizing insurance

We now move on to the investigation of the $\text{VaR}_{95}^{(e)}$-minimizing index insurance. Compared with CVaR optimization, VaR is generally more challenging due to its non-convex nature. Key summary statistics highlighting the model performance are given in columns 5-6 of Tables 1 and 2. The results are generally consistent with the CVaR model: the tail risk of policyholders is generally hedged very well, the mean of loss increases representing the hedging cost, and all model improvements are statistically significant across different risk measures. Additionally, we observe that the estimated price of the $\text{VaR}_{95}^{(e)}$-minimizing insurance is \$26.4 (\$28.5) on the training (test) sample, which is substantially lower than the $\text{CVaR}_{95}^{(e)}$-minimizing insurance premium. Compared with the $\text{CVaR}_{95}^{(e)}$-minimizing insurance,



(a) Training sample (b) Test sample

Fig. 6. Insurance payout against actual loss occurred. $y = I^*(x)$ represents the payouts of CVaR minimizing index insurance. $y = \max(x - d, 0)$ denotes the payoff function of a benchmarking indemnity-based insurance, with the deductible d chosen such that the two have the same expected payoff.

this $\text{VaR}_{95}^{(e)}$-minimizing insurance tends to be less likely to over-compensate those extremely large losses, which explains its lower price.[8] This observation is not surprising, since those $\text{VaR}_{95}^{(e)}$-minimizing hedgers would intentionally leave some of the extreme losses under-hedged to save their hedging cost.

4.5. Results with the EVaR-minimizing insurance

We then consider the problem of $\text{EVaR}_{95}^{(e)}$-minimizing. EVaR, defined by (2), has the following nice properties and is therefore a useful metric for financial tail risk management:

Coherence: EVaR_α is coherent for every $\alpha \in (0, 1)$.

Strictly monotonicity: For any pair of random variables X and Y such that (i) $X \geq Y$, (ii) $\Pr\{X > Y\} > 0$, and (iii) $\operatorname{ess\,sup}(X) > \operatorname{ess\,sup}(Y)$, we have $\text{EVaR}_\alpha(X) > \text{EVaR}_\alpha(Y)$. In contrast, VaR and CVaR are monotone but not strictly monotone.

Dual representation: For any $\alpha \in (0, 1)$, $\text{EVaR}_\alpha(X) = \sup_{Q \in \mathcal{Q}} \mathrm{E}^Q(X)$ where $\mathcal{Q} = \{Q \ll P | \mathscr{D}(Q, P) \leq -\ln(1-\alpha)\}$ and $\mathscr{D}(Q, P)$ is the Kullback-Leibler divergence from Q to P.

As an upper bound: For any $\alpha \in (0, 1)$, $\text{CVaR}_\alpha \leq \text{EVaR}_\alpha$. In other words, EVaR is more risk averse compared to VaR and CVaR at the same confidence level.

Performance of the $\text{EVaR}_{95}^{(e)}$-minimizing index insurance is summarized in the last two columns of Tables 1 and 2. As anticipated all listed risk metrics are significantly improved. The $\text{EVaR}_{95}^{(e)}$-minimizing model reduces the skewness of loss distribution more significantly than the previous two cases, by even turning its sign to negative on the training sample. This is also consistent with our observation that the $\text{EVaR}_{95}^{(e)}$-minimizing policy has the largest estimated price (or equivalently, the largest demanded insurance coverage from policyholders), which is explained by the fact that EVaR is more conservative than CVaR and thus VaR.[9]

4.6. Extension to spatial model

One typical feature about agricultural data is their spatial dependence.[10] This phenomenon is also empirically observed in our data as shown in Fig. 7 which plots the pairwise Kendall's τ of the crop yield loss as a function of distance between two county centroids. The presence of spatial dependence is indicated by the negative relationship between the correlation of crop yield loss (measured by the Kendall's τ) and the distance. To address the spatial dependence issue, in this subsection we employ the idea in spatial econometrics and use spatially lagged variables to model the spillover effects among different counties (see e.g., LeSage and Pace, 2009; Anselin et al., 2010; Elhorst, 2014). Specifically, we consider the following extension to our original model (4):

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^{2p-1}} \rho^{(e)}\Big(\big\{x_{i,j,p} - \big[\big(\theta_0 + [x_{i,j,1}, \ldots, x_{i,j,p-1}] \cdot [\theta_1, \ldots, \theta_{p-1}]^\intercal + W_j \cdot \boldsymbol{X}_i \cdot [\theta_p, \ldots, \theta_{2p-2}]^\intercal\big) \vee 0\big] \wedge M + \pi^{(e)}(\boldsymbol{\theta})\big\}_{i=1,2,\ldots,N_{year};\, j=1,2,\ldots,N_{county}}\Big), \tag{21}$$

[8] Graphical illustrations are provided in Appendix A.3.

[9] This is also graphically supported by Appendix A.3.

[10] It is well documented in the agricultural insurance literature that both crop yields and weather variables are in general spatially dependent. For example, Wang and Zhang (2003) take spatial statistics approach to examine the pattern of U.S. crop yield dependence. Xu et al. (2010) use copula models to measure weather dependence among different locations. Erhardt and Smith (2014) investigate the spatial dependence of weather extremes and its impact on weather derivative pricing. Zhu et al. (2018) propose hierarchical Archimedean copula to model the spatial dependence of weather risk. Hainaut (2019) proposes a continuous-time approach based on Gaussian field to model spatial dependence between temperatures.

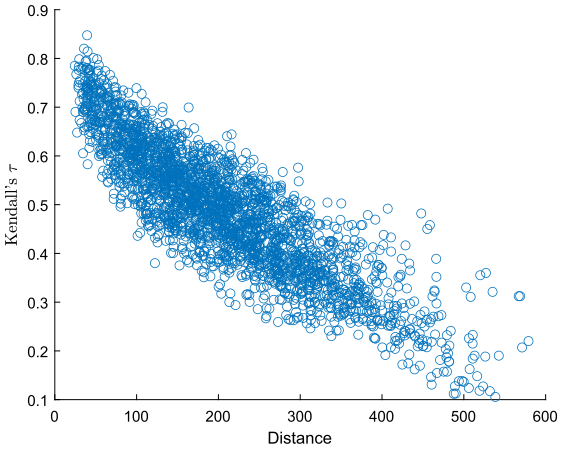


Fig. 7. Pairwise Kendall's τ of the crop yield loss as a function of distance between the county centroids. Distances are measured in kilometers.

Table 3
Performance of the index insurance based on the spatially lagged model (21) and $\text{CVaR}_{95}^{(e)}$ criterion. Semi-deviation is defined as the expected downside deviation from the mean, $SD(Y) = \left(\text{E}\left[(Y - \text{E}(Y))^2 \cdot 1_{[Y \geq \text{E}(Y)]}\right]\right)^{1/2}$. For risk reduction, numbers in parentheses are standard deviations of the estimates obtained by bootstrapping.

	Training			Test		
	Without insurance	With $\text{CVaR}_{95}^{(e)}$-minimizing insurance	Risk reduction	Without insurance	With $\text{CVaR}_{95}^{(e)}$-minimizing insurance	Risk reduction
Estimated price	–	0.0369	–	–	0.0392	–
Mean	0.3615	0.3677	–	0.3586	0.3652	–
Standard deviation	0.1402	0.1255	0.1045 (0.0084)	0.1442	0.1196	0.1706 (0.021)
Skewness	0.7325	0.1744	–	1.6419	0.6078	–
Kurtosis	4.0653	3.2838	–	6.6568	4.7168	–
Semi-deviation	0.1076	0.0908	0.1556 (0.0096)	0.1207	0.0906	0.2494 (0.0198)
$\text{VaR}_{95}^{(e)}$	0.6084	0.5839	0.0403 (0.011)	0.6643	0.5764	0.1324 (0.0425)
$\text{CVaR}_{95}^{(e)}$	0.7157	0.6448	0.099 (0.0109)	0.8068	0.6673	0.1729 (0.0425)
$\text{EVaR}_{95}^{(e)}$	0.7933	0.6993	0.1186 (0.0114)	0.8671	0.7357	0.1516 (0.0134)
$\text{VaR}_{99}^{(e)}$	0.7812	0.6811	0.1282 (0.0109)	0.8754	0.7578	0.1344 (0.0421)
$\text{CVaR}_{99}^{(e)}$	0.8641	0.7339	0.1506 (0.011)	0.9505	0.8024	0.1558 (0.042)
$\text{EVaR}_{99}^{(e)}$	0.8966	0.7785	0.1317 (0.0113)	0.9688	0.8184	0.1552 (0.0133)

where $\left\{[x_{i,j,1}, x_{i,j,2}, ..., x_{i,j,p}]\right\}_{i=1,2,...,N_{year};\, j=1,2,...,N_{county}}$ is a random sample with i being the time index, j being the location index, and N_{year} and N_{county} are the total number of years and the total number of countries in the sample, respectively; W_j is the j-th row of the Spatial Weights Matrix $W = \left(w_{ij}\right)_{i=1,2,...,N_{year};\, j=1,2,...,N_{county}}$,[11] representing the structure of dependence between locations in the sample (Halleck Vega and Elhorst, 2015); $\boldsymbol{X}_i$ is the matrix form of year-i data, defined by,

$$\boldsymbol{X}_i = \begin{pmatrix} x_{i,1,1} & x_{i,1,2} & \cdots & x_{i,1,p-1} \\ x_{i,2,1} & x_{i,2,2} & \cdots & x_{i,2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i,N_{county},1} & x_{i,N_{county},2} & \cdots & x_{i,N_{county},p-1} \end{pmatrix};$$

and $[\theta_p, ..., \theta_{2p-2}]^\intercal$ represents the local spatial spillover among counties, which also serves as the decision variable of optimization. In a nutshell, model (21) incorporates spatial interrelations by utilizing information of other counties in a spatially lagged manner, and the spatial spillover effect between two counties decreases as their distance increases, as indicated by Fig. 7. This is also aligned with the practice of weather index insurance to combine weather observations from different weather stations so as to achieve better model accuracy and robustness.

Table 3 exhibits the optimization results for extended problem (21) with $\text{CVaR}_{95}^{(e)}$ being the tail risk objective.[12] We observe congruous model performance with our baseline results: tail risk is effectively managed with all presented risk metrics for both training and test samples, and all the improvements are statistically significant. The risk reductions are slightly lower than those for the original model (4), while they also have smaller standard errors. It thus reveals a critical tradeoff for practical consideration in designing weather insurance: on the one hand, introducing the spatially lagged term explicitly captures the spatial spillover impact of weather conditions in neighboring

[11] In this analysis we use the quadratic kernel weights defined by, $w_{ij} = \frac{3}{4}\left[1 - \left(\frac{d_{ij}}{h_i}\right)^2\right]$ where d_{ij} is the distance from county i to j and h_i is the bandwidth (Anselin et al., 2010).

[12] Robustness checks with $\text{VaR}_{95}^{(e)}$ and $\text{EVaR}_{95}^{(e)}$ being objective functions are also conducted and similar results are obtained.

countries and thus can theoretically enhance model performance; on the other hand, this additional model component generally requires meticulous selection and prior knowledge of the Spatial Weights Matrix W and it approximately doubles the dimensions of optimization problem, and thus may significantly increase modeling and computational complexity.

5. Conclusion

Financial and insurance risks are typically heavy tailed, making tail risk metrics such as VaR and CVaR popular risk objectives. However, optimizations of these risk measures are usually computationally challenging due to the lack of structural information such as convexity and differentiability. In this study we introduce the model-based annealing random search (MARS) method (see e.g., Hu et al., 2008; Hu and Hu, 2011) to the actuarial literature, to solve for risk management problems with non-parametric tail risk objective functions including VaR, CVaR and EVaR. MARS has the benefit of not requiring mathematical structural information, and thus could be used as a powerful tool to address these questions. We illustrate the optimization scheme using the weather index insurance design problem as an example, but the method can be flexibly applied to many other financial and insurance applications such as portfolio selection, optimal hedging strategy, etc.

An empirical analysis, based on corn production and a set of 36-dimensional weather variables in Illinois, is conducted to show the performance of the MARS method. We solve three optimal index insurance contracts respectively based on risk objectives of VaR, CVaR and EVaR at the $\alpha = 95\%$ level, and apply LASSO and cross-validation to balance between goodness and smoothness of the optimal models. The three obtained index insurance policies are all demonstrated to effectively reduce the tail risk, while exhibiting some notable distinctions. The $\mathrm{EVaR}_{95}^{(\mathrm{e})}$-minimizing model has the highest endogenously determined insurance coverage level while the $\mathrm{VaR}_{95}^{(\mathrm{e})}$-minimizing model has the lowest, indicating that different risk metrics tend to require different levels of hedging cost. This is well aligned with our intuition that tail risk measures, especially those more risk averse ones, have highly asymmetric penalties between upside and downside errors. Lastly, our proposed insurance contracts perform well in managing tail risk with an out-of-sample test, indicating the practical viability and potential of weather index insurance in hedging against agricultural production losses.

Declaration of competing interest

The opinion, findings, conclusions in this paper are our own and do not necessarily reflect the views of Barclays.

Data availability

The authors do not have permission to share data.

Acknowledgements

The authors acknowledge helpful comments and suggestions from the Handling Editor and two anonymous referees. Tan acknowledges research funding from Nanyang Technological University, Singapore University Grant and President's Chair in Actuarial Risk Management. Zhang acknowledges the research funding support from the Nanyang Technological University Startup Grant (04INS000509C300) and the Singapore Ministry of Education Academic Research Fund Tier 1 Grant (RG55/20).

Appendix A

A.1. Proof of Proposition 1

Functions $\mathrm{VaR}_{\alpha}^{(\mathrm{e})}\left(\{y_i\}_{i=1,2,\ldots,N}\right)$, $\mathrm{CVaR}_{\alpha}^{(\mathrm{e})}\left(\{y_i\}_{i=1,2,\ldots,N}\right)$ and $\mathrm{EVaR}_{\alpha}^{(\mathrm{e})}\left(\{y_i\}_{i=1,2,\ldots,N}\right)$ are all non-decreasing functions with any increase in y_i, $\forall\ i = 1, 2, \ldots, N$. Therefore on the one hand,

$$
\begin{aligned}
&\min_{\boldsymbol{\theta}\in\mathbb{R}^p} F(\boldsymbol{\theta}) \\
&= \min_{\boldsymbol{\theta}\in\mathbb{R}^p} \mathrm{VaR}_{\alpha}^{(\mathrm{e})}\left(\left\{x_{i,p} - \left[\theta_0 + \left(\sum_{j=1}^{p-1}\theta_j x_{i,j}\right)\vee 0\right]\wedge M + \pi^{(\mathrm{e})}(\boldsymbol{\theta})\right\}_{i=1,2,\ldots,N}\right) \\
&= \min_{\boldsymbol{\theta}\in\mathbb{R}^p} \mathrm{VaR}_{\alpha}^{(\mathrm{e})}\left(\left\{x_{i,p} - \left[\theta_0 + \left(\sum_{j=1}^{p-1}\theta_j x_{i,j}\right)\vee 0\right]\wedge M\right\}_{i=1,2,\ldots,N}\right) + \pi^{(\mathrm{e})}(\boldsymbol{\theta}) \\
&\geq \mathrm{VaR}_{\alpha}^{(\mathrm{e})}\left(\left\{x_{i,p} - M\right\}_{i=1,2,\ldots,N}\right) + 0 \\
&= \mathrm{VaR}_{\alpha}^{(\mathrm{e})}\left(\left\{x_{i,p}\right\}_{i=1,2,\ldots,N}\right) - M,
\end{aligned}
$$

and $F(\boldsymbol{\theta})$ is bounded from below. On the other hand, similarly we have,

$$
\max_{\boldsymbol{\theta}\in\mathbb{R}^p} F(\boldsymbol{\theta}) \leq \mathrm{VaR}_{\alpha}^{(\mathrm{e})}\left(\left\{x_{i,p} - 0\right\}_{i=1,2,\ldots,N}\right) + \gamma M = \mathrm{VaR}_{\alpha}^{(\mathrm{e})}\left(\left\{x_{i,p}\right\}_{i=1,2,\ldots,N}\right) + \gamma M,
$$

and $F(\boldsymbol{\theta})$ is also bounded from above. Likewise we can show $G(\boldsymbol{\theta})$ and $K(\boldsymbol{\theta})$ are also bounded.

A.2. Proof of Proposition 2

A.2.1. Continuity of $F(\boldsymbol{\theta})$

Define $y_i(\boldsymbol{\theta}) = x_{i,p} - \left[\theta_0 + \left(\sum_{j=1}^{p-1}\theta_j x_{i,j}\right) \vee 0\right] \wedge M$, $i = 1, 2, ..., N$, and by definition $F(\boldsymbol{\theta}) = \mathrm{VaR}_\alpha^{(\mathrm{e})}\left(\{y_i(\boldsymbol{\theta})\}_{i=1,2,...,N} + \pi^{(\mathrm{e})}(\boldsymbol{\theta})\right) = y_{\lceil \alpha N\rceil:N}(\boldsymbol{\theta}) + \pi^{(\mathrm{e})}(\boldsymbol{\theta})$. It is obvious that $\pi^{(\mathrm{e})}(\boldsymbol{\theta})$ is continuous w.r.t. $\boldsymbol{\theta}$, and thus it is only necessary to show that $y_{\lceil \alpha N\rceil:N}(\boldsymbol{\theta})$ is also continuous w.r.t. $\boldsymbol{\theta}$. We use $y_{j:N}^{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_2)$ to denote the value of the j-th smallest element in $\{y_i(\boldsymbol{\theta}_1)\}_{i=1,2,...,N}$ valuated at $\boldsymbol{\theta}_2$, and still use the notation $y_{j:N}(\boldsymbol{\theta})$ for $y_{j:N}^{\boldsymbol{\theta}}(\boldsymbol{\theta})$ whenever it does not cause any ambiguity.

For any arbitrary but fixed $\boldsymbol{\theta}_0 \in \mathbb{R}^p$, we discuss the two following collectively exhaustive cases:

Case 1: if $y_{\lceil \alpha N\rceil-1:N}(\boldsymbol{\theta}_0) < y_{\lceil \alpha N\rceil:N}(\boldsymbol{\theta}_0)$.

For every $i = 1, 2, ..., N$, $y_i(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$, and hence for any $0 < \epsilon < y_{\lceil \alpha N\rceil:N}(\boldsymbol{\theta}_0) - y_{\lceil \alpha N\rceil-1:N}(\boldsymbol{\theta}_0)$ there exists $\delta_i > 0$ such that $|y_i(\boldsymbol{\theta}) - y_i(\boldsymbol{\theta}_0)| < \epsilon/2$ if $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| < \delta_i$, with $||\cdot||$ being the Euclidean norm in $\mathbb{R}^p$.

Then we denote $\delta = \min_{1\le i\le N}\delta_i$, and for any $\boldsymbol{\theta}$ such that $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| < \delta$,

$$\max_{1\le i\le \lceil \alpha N\rceil-1} y_{i:N}^{\boldsymbol{\theta}_0}(\boldsymbol{\theta}) < y_{\lceil \alpha N\rceil-1:N}(\boldsymbol{\theta}_0) + \frac{1}{2}\epsilon \le y_{\lceil \alpha N\rceil:N}(\boldsymbol{\theta}_0) - \frac{1}{2}\epsilon < \min_{\lceil \alpha N\rceil\le i\le N} y_{i:N}^{\boldsymbol{\theta}_0}(\boldsymbol{\theta}).$$

It implies that $\left\{y_{i:N}^{\boldsymbol{\theta}_0}(\boldsymbol{\theta})\right\}_{1\le i\le \lceil \alpha N\rceil-1} = \{y_{i:N}(\boldsymbol{\theta})\}_{1\le i\le \lceil \alpha N\rceil-1}$, and $\left\{y_{i:N}^{\boldsymbol{\theta}_0}(\boldsymbol{\theta})\right\}_{\lceil \alpha N\rceil\le i\le N} = \{y_{i:N}(\boldsymbol{\theta})\}_{\lceil \alpha N\rceil\le i\le N}$.

Hence, on the one hand,

$$y_{\lceil \alpha N\rceil:N}(\boldsymbol{\theta}) = \min_{\lceil \alpha N\rceil\le i\le N} y_{i:N}(\boldsymbol{\theta}) = \min_{\lceil \alpha N\rceil\le i\le N} y_{i:N}^{\boldsymbol{\theta}_0}(\boldsymbol{\theta}) > y_{\lceil \alpha N\rceil:N}(\boldsymbol{\theta}_0) - \frac{1}{2}\epsilon. \tag{A.1}$$

On the other hand,

$$y_{\lceil \alpha N\rceil:N}(\boldsymbol{\theta}) = \min_{\lceil \alpha N\rceil\le i\le N} y_{i:N}(\boldsymbol{\theta}) \le y_{\lceil \alpha N\rceil:N}^{\boldsymbol{\theta}_0}(\boldsymbol{\theta}) < y_{\lceil \alpha N\rceil:N}(\boldsymbol{\theta}_0) + \frac{1}{2}\epsilon. \tag{A.2}$$

Combining (A.1) and (A.2), we have $y_{\lceil \alpha N\rceil:N}(\boldsymbol{\theta})$ is continuous w.r.t. $\boldsymbol{\theta}$ and so is F.

Case 2: if $y_{\lceil \alpha N\rceil-1:N}(\boldsymbol{\theta}_0) = y_{\lceil \alpha N\rceil:N}(\boldsymbol{\theta}_0)$.

We assume $y_{\lceil \alpha N\rceil-a-1:N}(\boldsymbol{\theta}_0) < y_{\lceil \alpha N\rceil-a:N}(\boldsymbol{\theta}_0) = \cdots = y_{\lceil \alpha N\rceil-1:N}(\boldsymbol{\theta}_0) = y_{\lceil \alpha N\rceil:N}(\boldsymbol{\theta}_0) = \cdots = y_{\lceil \alpha N\rceil+b-1:N}(\boldsymbol{\theta}_0) < y_{\lceil \alpha N\rceil+b:N}(\boldsymbol{\theta}_0)$ for some positive integers a and b.

Due to continuity of $y_i(\boldsymbol{\theta})$, for any $i = 1, 2, ..., N$, and $0 < \epsilon < \min(\epsilon_1, \epsilon_2)$, where $\epsilon_1 = y_{\lceil \alpha N\rceil-a:N}(\boldsymbol{\theta}_0) - y_{\lceil \alpha N\rceil-a-1:N}(\boldsymbol{\theta}_0)$ and $\epsilon_2 = y_{\lceil \alpha N\rceil+b:N}(\boldsymbol{\theta}_0) - y_{\lceil \alpha N\rceil+b-1:N}(\boldsymbol{\theta}_0)$, there exists δ_i such that $|y_i(\boldsymbol{\theta}) - y_i(\boldsymbol{\theta}_0)| < \epsilon/2$ if $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| < \delta_i$.

Similarly to Case 1, we denote $\delta = \min_{1\le i\le N}\delta_i$. Then for any $\boldsymbol{\theta}$ such that $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| < \delta$, we have, $\left\{y_{i:N}^{\boldsymbol{\theta}_0}(\boldsymbol{\theta})\right\}_{1\le i\le \lceil \alpha N\rceil-a-1} = \{y_{i:N}(\boldsymbol{\theta})\}_{1\le i\le \lceil \alpha N\rceil-a-1}$, $\left\{y_{i:N}^{\boldsymbol{\theta}_0}(\boldsymbol{\theta})\right\}_{\lceil \alpha N\rceil-a\le i\le \lceil \alpha N\rceil+b-1} = \{y_{i:N}(\boldsymbol{\theta})\}_{\lceil \alpha N\rceil-a\le i\le \lceil \alpha N\rceil+b-1}$, and $\left\{y_{i:N}^{\boldsymbol{\theta}_0}(\boldsymbol{\theta})\right\}_{\lceil \alpha N\rceil+b\le i\le N} = \{y_{i:N}(\boldsymbol{\theta})\}_{\lceil \alpha N\rceil+b\le i\le N}$.

Hence, on the one hand,

$$y_{\lceil \alpha N\rceil:N}(\boldsymbol{\theta}) = \min_{\lceil \alpha N\rceil\le i\le N} y_{i:N}(\boldsymbol{\theta}) \ge \min_{\lceil \alpha N\rceil-a\le i\le \lceil \alpha N\rceil+b-1} y_{i:N}^{\boldsymbol{\theta}_0}(\boldsymbol{\theta}) > y_{\lceil \alpha N\rceil:N}(\boldsymbol{\theta}_0) - \frac{1}{2}\epsilon. \tag{A.3}$$

On the other hand,

$$y_{\lceil \alpha N\rceil:N}(\boldsymbol{\theta}) = \min_{\lceil \alpha N\rceil\le i\le N} y_{i:N}(\boldsymbol{\theta}) \le \max_{\lceil \alpha N\rceil-a\le i\le \lceil \alpha N\rceil+b-1} y_{i:N}^{\boldsymbol{\theta}_0}(\boldsymbol{\theta}) < y_{\lceil \alpha N\rceil:N}(\boldsymbol{\theta}_0) + \frac{1}{2}\epsilon. \tag{A.4}$$

Combining (A.3) and (A.4), we have $y_{\lceil \alpha N\rceil:N}(\boldsymbol{\theta})$ is continuous w.r.t. $\boldsymbol{\theta}$ and so is F.

A.2.2. Continuity of $G(\boldsymbol{\theta})$

It directly follows from the proof of continuity of $F(\boldsymbol{\theta})$. For all $i = \lceil \alpha N\rceil, \lceil \alpha N\rceil + 1, ..., N$, $y_{i:N}(\boldsymbol{\theta})$ are all continuous w.r.t. $\boldsymbol{\theta}$. Therefore $G(\boldsymbol{\theta})$ is also continuous as a linear combination of $\{y_{i:N}(\boldsymbol{\theta})\}_{\lceil \alpha N\rceil\le i\le N}$.

A.2.3. Continuity of $K(\boldsymbol{\theta})$

Define $z_i = x_{i,p} - \left[\theta_0 + \left(\sum_{j=1}^{p-1}\theta_j x_{i,j}\right) \vee 0\right] \wedge M + \pi^{(\mathrm{e})}(\boldsymbol{\theta})$, $i = 1, 2, ..., N$, and therefore by definition,

$$K(\boldsymbol{\theta}) = \inf_{t>0} k(\boldsymbol{\theta}, t),$$

where

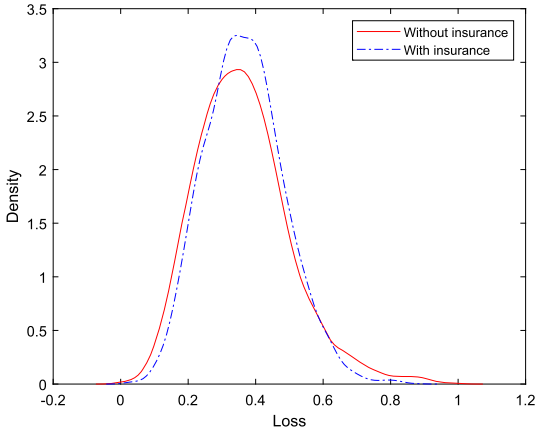
$$k(\boldsymbol{\theta}, t) = t \cdot \ln\left[\frac{1}{(1-\alpha)N}\sum_{i=1}^{N} \mathrm{e}^{t^{-1}\cdot z_i(\boldsymbol{\theta})}\right].$$

By Lemma 3.1 in Ahmadi-Javid(2012) $k(\boldsymbol{\theta}, t)$ is convex in t, and for any arbitrary but fixed $\boldsymbol{\theta}_0 \in \mathbb{R}^p$, we discuss the two following collectively exhaustive cases:

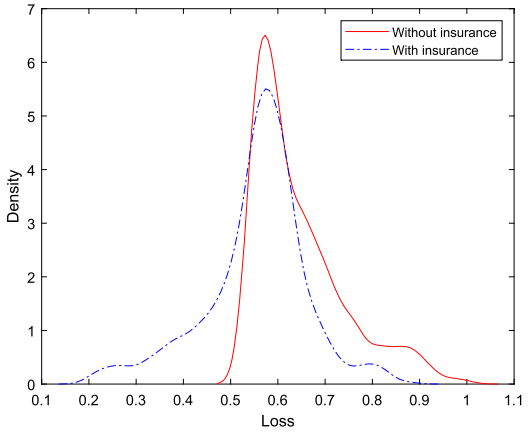
Case 1: if $\max_{1\le i\le N}\{z_i(\boldsymbol{\theta}_0)\} > \min_{1\le i\le N}\{z_i(\boldsymbol{\theta}_0)\}$.

Denote $t_0 = \mathrm{argmin}_{t>0} k(\boldsymbol{\theta}_0, t)$. By convexity, t_0 solves the first order condition, i.e.,

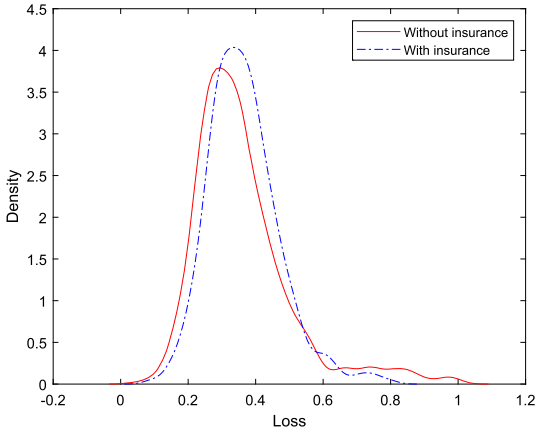
$$0 = \left.\frac{\mathrm{d}k(\boldsymbol{\theta}_0, t)}{\mathrm{d}t}\right|_{t=t_0} = L(\boldsymbol{\theta}_0, t_0), \tag{A.5}$$



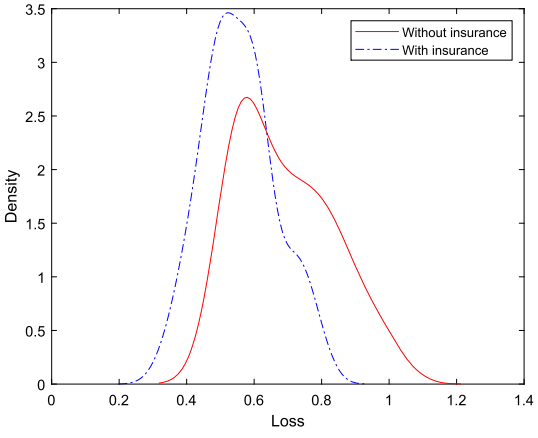
(a) Whole sample of training sample



(b) Right 10% tail of training sample

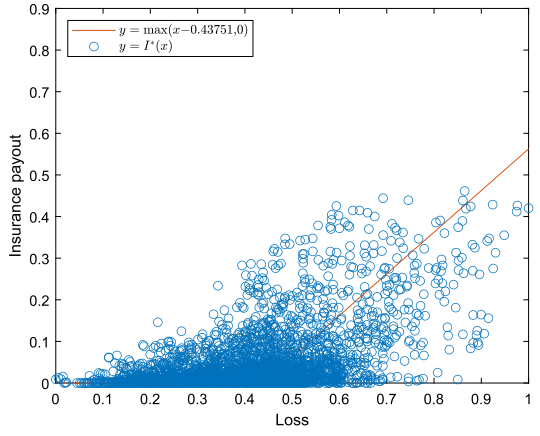


(c) Whole sample of test sample

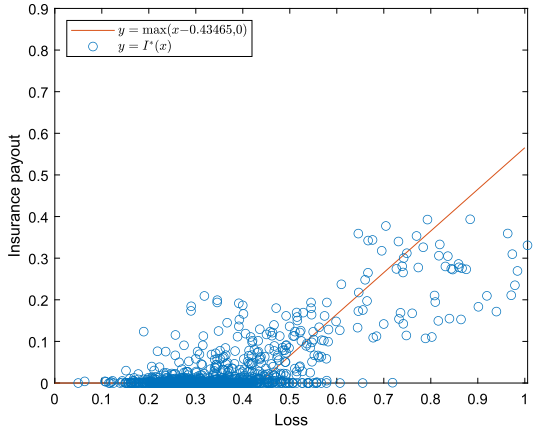


(d) Right 10% tail of test sample

Fig. A.1. Loss distribution with VaR-minimizing insurance vs. without insurance.



(a) Training sample

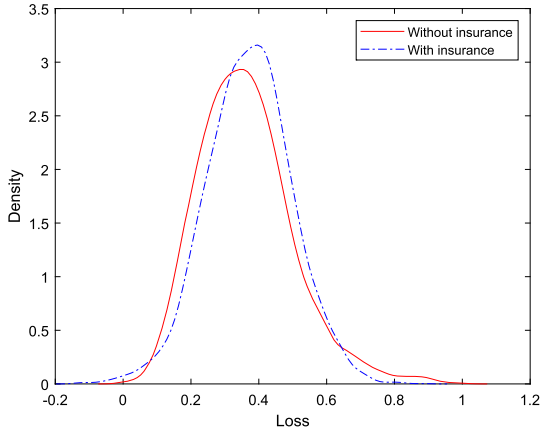


(b) Test sample

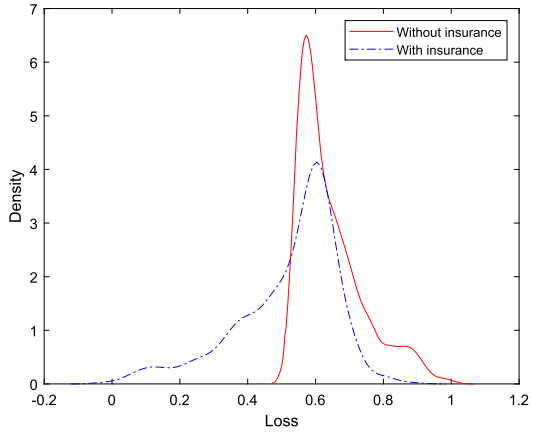
Fig. A.2. VaR-minimizing insurance payout against actual loss occurred.

where

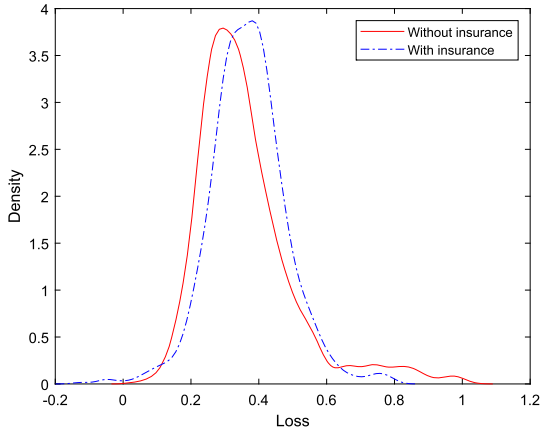
$$L(\boldsymbol{\theta}_0, t_0) = \ln\left[\frac{1}{(1-\alpha)N}\sum_{i=1}^{N} e^{t_0^{-1}\cdot z_i(\boldsymbol{\theta}_0)}\right] - \frac{t_0^{-1}\sum_{i=1}^{N} z_i(\boldsymbol{\theta}_0) e^{t_0^{-1}\cdot z_i(\boldsymbol{\theta}_0)}}{\sum_{i=1}^{N} e^{t_0^{-1}\cdot z_i(\boldsymbol{\theta}_0)}}.$$



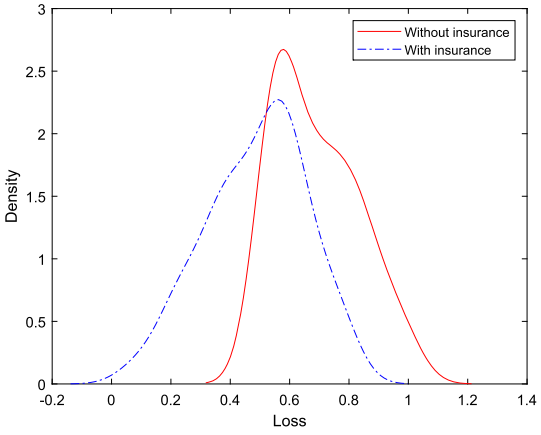
(a) Whole sample of training sample



(b) Right 10% tail of training sample

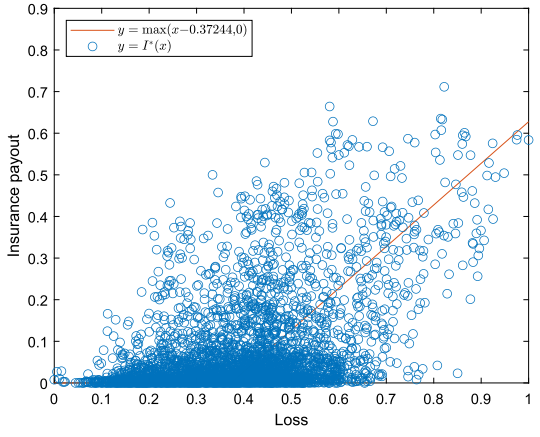


(c) Whole sample of test sample

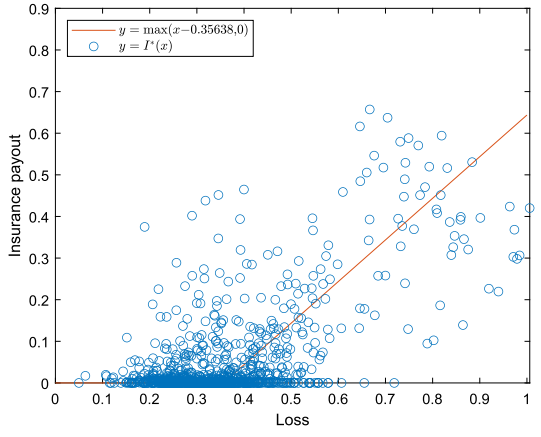


(d) Right 10% tail of test sample

Fig. A.3. Loss distribution with EVaR-minimizing insurance vs. without insurance.



(a) Training sample



(b) Test sample

Fig. A.4. EVaR-minimizing insurance payout against actual loss occurred.

As $z_i(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$, there exists $\delta > 0$ such that $\max_{1 \le i \le N} \{z_i(\boldsymbol{\theta})\} > \min_{1 \le i \le N} \{z_i(\boldsymbol{\theta})\}$ if $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| < \delta$. Therefore for any fixed $\boldsymbol{\theta}$ such that $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| < \delta$, we have,

$$\frac{\partial L(\boldsymbol{\theta},t)}{\partial t}=\frac{t^{-3}\left\{\sum_{i=1}^{N}z_i^2(\boldsymbol{\theta})\mathrm{e}^{t^{-1}\cdot z_i(\boldsymbol{\theta})}\cdot\sum_{i=1}^{N}\mathrm{e}^{t^{-1}\cdot z_i(\boldsymbol{\theta})}-\left[\sum_{i=1}^{N}z_i(\boldsymbol{\theta})\mathrm{e}^{t^{-1}\cdot z_i(\boldsymbol{\theta})}\right]^2\right\}}{\left[\sum_{i=1}^{N}\mathrm{e}^{t^{-1}\cdot z_i(\boldsymbol{\theta})}\right]^2}>0,$$

due to the Cauchy–Schwarz inequality and equality cannot be obtained because not all $z_i(\boldsymbol{\theta})$'s are the same. The by the implicit function theorem, $t_0(\boldsymbol{\theta}_0)$ determined by (A.5) is continuous in $\boldsymbol{\theta}_0$ within a neighborhood of $\boldsymbol{\theta}_0$. Finally, $K(\boldsymbol{\theta}_0)=k(\boldsymbol{\theta}_0,t_0(\boldsymbol{\theta}_0))$ is continuous w.r.t. $\boldsymbol{\theta}_0$ for all $\boldsymbol{\theta}_0\in\mathbb{R}^p$ as well.

Case 2: if $\max_{1\le i\le N}\{z_i(\boldsymbol{\theta}_0)\}=\min_{1\le i\le N}\{z_i(\boldsymbol{\theta}_0)\}$.

It implies $z_1(\boldsymbol{\theta}_0)=z_2(\boldsymbol{\theta}_0)=\ldots=z_N(\boldsymbol{\theta}_0)$. Therefore,

$$K(\boldsymbol{\theta})=\inf_{t>0}\ t\cdot\ln\left[\frac{1}{(1-\alpha)N}\sum_{i=1}^{N}\mathrm{e}^{t^{-1}\cdot z_i(\boldsymbol{\theta})}\right]=\inf_{t>0}\ \frac{t\cdot z_1(\boldsymbol{\theta})+\ln\left(\frac{1}{1-\alpha}\right)}{t}=z_1(\boldsymbol{\theta}),$$

and thus it is continuous in $\boldsymbol{\theta}$.

A.3. Graphical results for VaR- and EVaR-minimizing insurance

This appendix graphically illustrates the results for VaR-minimizing insurance (Figs. A.1 and A.2), and VaR-minimizing insurance (Figs. A.3 and A.4) discussed in Sections 4.4 and 4.5, all of which share similar patterns to those in Section 4.3.

References

Ahmadi-Javid, A., 2012. Entropic value-at-risk: a new coherent risk measure. Journal of Optimization Theory and Applications 155 (3), 1105–1123.
Ahmadi-Javid, A., Fallah-Tafti, M., 2019. Portfolio optimization with entropic value-at-risk. European Journal of Operational Research 279 (1), 225–241.
Andersson, F., Mausser, H., Rosen, D., Uryasev, S., 2001. Credit risk optimization with conditional value-at-risk criterion. Mathematical Programming 89 (2), 273–291.
Anselin, L., Syabri, I., Kho, Y., 2010. Geoda: an introduction to spatial data analysis. In: Fischer, M.M., Getis, A. (Eds.), Handbook of Applied Spatial Analysis. Springer, Berlin, Heidelberg, pp. 73–89.
Asimit, A.V., Gao, T., Hu, J., Kim, E.-S., 2018. Optimal risk transfer: a numerical optimization approach. North American Actuarial Journal 22 (3), 341–364.
Assa, H., 2015. On optimal reinsurance policy with distortion risk measures and premiums. Insurance. Mathematics & Economics 61, 70–75.
Balbás, A., Balbás, B., Heras, A., 2009. Optimal reinsurance with general risk measures. Insurance. Mathematics & Economics 44 (3), 374–384.
Benati, S., Rizzi, R., 2007. A mixed integer linear programming formulation of the optimal mean/value-at-risk portfolio problem. European Journal of Operational Research 176 (1), 423–434.
Bernard, C., Tian, W., 2009. Optimal reinsurance arrangements under tail risk measures. The Journal of Risk and Insurance 76 (3), 709–725.
Brandtner, M., Kürsten, W., Rischau, R., 2018. Entropic risk measures and their comparative statics in portfolio selection: coherence vs. convexity. European Journal of Operational Research 264 (2), 707–716.
Cai, J., Tan, K.S., Weng, C., Zhang, Y., 2008. Optimal reinsurance under VaR and CTE risk measures. Insurance. Mathematics & Economics 43 (1), 185–196.
Carter, M.R., Cheng, L., Sarris, A., 2016. Where and how index insurance can boost the adoption of improved agricultural technologies. Journal of Development Economics 118, 59–71.
Chantarat, S., Mude, A.G., Barrett, C.B., Carter, M.R., 2013. Designing index-based livestock insurance for managing asset risk in northern Kenya. The Journal of Risk and Insurance 80 (1), 205–237.
Cheung, K., Sung, K., Yam, S., Yung, S., 2014. Optimal reinsurance under general law-invariant risk measures. Scandinavian Actuarial Journal 2014 (1), 72–91.
Conradt, S., Finger, R., Spörri, M., 2015. Flexible weather index-based insurance design. Climate Risk Management 10, 106–117.
Dalhaus, T., Musshoff, O., Finger, R., 2018. Phenology information contributes to reduce temporal basis risk in agricultural weather index insurance. Scientific Reports 8 (46).
Dallagnol, V., van den Berg, J., Mous, L., 2009. Portfolio management using value at risk: a comparison between genetic algorithms and particle swarm optimization. International Journal of Intelligent Systems 24 (7), 766–792.
Dorigo, M., Blum, C., 2005. Ant colony optimization theory: a survey. Theoretical Computer Science 344, 243–278.
Eiben, A., Smith, J., 2003. Introduction to Evolutionary Computing. Natural Computing Series. Springer.
Elhorst, J.P., 2014. Spatial Econometrics from Cross-Sectional Data to Spatial Panels. Springer, Berlin, Heidelberg.
Erhardt, R.J., Smith, R.L., 2014. Weather derivative risk measures for extreme events. North American Actuarial Journal 18 (3), 379–393.
Gaivoronski, A.A., Pflug, G., 2005. Value-at-risk in portfolio optimization: properties and computational approach. The Journal of Risk 7 (2), 1–31.
Gilli, M., Këllezi, E., 2002. A global optimization heuristic for portfolio choice with VaR and expected shortfall. In: Computational Methods in Decision-Making, Economics and Finance. Springer, pp. 167–183.
Glover, F., Laguna, M., 1997. Tabu Search. Kluwer Academic Publishers, Norwell, MA, USA.
Goh, J.W., Lim, K.G., Sim, M., Zhang, W., 2012. Portfolio value-at-risk optimization for asymmetrically distributed asset returns. European Journal of Operational Research 221 (2), 397–406.
Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
Hainaut, D., 2019. Hedging of crop harvest with derivatives on temperature. Insurance. Mathematics & Economics 84, 98–114.
Halleck Vega, S., Elhorst, J.P., 2015. The SLX model. Journal of Regional Science 55 (3), 339–363.
Hu, J., Fu, M.C., Marcus, S.I., 2007. A model reference adaptive search method for global optimization. Operations Research 55 (3), 549–568.
Hu, J., Fu, M.C., Marcus, S.I., 2008. A model reference adaptive search method for stochastic global optimization. Communications in Information and Systems 8, 245–276.
Hu, J., Hu, P., 2011. Annealing adaptive search, cross-entropy, and stochastic approximation in global optimization. Naval Research Logistics (NRL) 58 (5), 457–477.
Hu, J., Hu, P., Chang, H.S., 2011. A stochastic approximation framework for a class of randomized optimization algorithms. IEEE Transactions on Automatic Control 57 (1), 165–178.
Hu, J., Wang, Y., Zhou, E., Fu, M.C., Marcus, S.I., 2012. A survey of some model-based methods for global optimization. In: Optimization, Control, and Applications of Stochastic Systems. Springer, pp. 157–179.
Hu, J., Zhou, E., Fan, Q., 2014. Model-based annealing random search with stochastic averaging. ACM Transactions on Modeling and Computer Simulation 24, 1–23.
Hu, X., Yang, H., Zhang, L., 2015. Optimal retention for a stop-loss reinsurance with incomplete information. Insurance. Mathematics & Economics 65, 15–21.
Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. Science 220 (4598), 671–680.
Larrañaga, P., Lozano, J.A., 2002. Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation. Kluwer Academic Publisher, Boston, MA, USA.
LeSage, J., Pace, R.K., 2009. Introduction to Spatial Econometrics. Chapman and Hall/CRC.
Lim, A.E., Shanthikumar, J.G., Vahn, G.-Y., 2011. Conditional value-at-risk in portfolio optimization: coherent but fragile. Operations Research Letters 39 (3), 163–171.
Lwin, K.T., Qu, R., MacCarthy, B.L., 2017. Mean-VaR portfolio optimization: a nonparametric approach. European Journal of Operational Research 260 (2), 751–766.
Markowitz, H., 1952. Portfolio selection. The Journal of Finance 7 (1), 77–91.
Mittal, S.K., Srivastava, N., 2021. Gated neural network-based mean–evar–skewness portfolio optimization under uncertain environment. Journal of Circuits, Systems, and Computers 30 (06), 2150100.

Picard, R.R., Cook, R.D., 1984. Cross-validation of regression models. Journal of the American Statistical Association 79 (387), 575–583.
Rockafellar, R.T., Uryasev, S., 2000. Optimization of conditional value-at-risk. The Journal of Risk 2, 21–42.
Romeijn, H.E., Smith, R.L., 1994. Simulated annealing and adaptive search in global optimization. Probability in the Engineering and Informational Sciences 8 (4), 571–590.
Rubinstein, R.Y., Kroese, D.P., 2004. The Cross Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation (Information Science and Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA.
Shi, H., Jiang, Z., 2016. The efficiency of composite weather index insurance in hedging rice yield risk: evidence from China. Agricultural Economics 47, 319–328.
Shi, L., Ólafsson, S., 2000. Nested partitions method for global optimization. Operations Research 48 (3), 390–407.
Sun, H., Weng, C., Zhang, Y., 2017. Optimal multivariate quota-share reinsurance: a nonparametric mean-cvar framework. Insurance. Mathematics & Economics 72, 197–214.
Tan, K.S., Weng, C., 2014. Empirical-based approach to optimal reinsurance. North American Actuarial Journal 18 (2), 315–342.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B, Methodological 58 (1), 267–288.
Uryasev, S., 2000. Conditional value-at-risk: optimization algorithms and applications. In: Proceedings of the IEEE/IAFE/INFORMS 2000 Conference on Computational Intelligence for Financial Engineering (CIFEr)(Cat. No. 00TH8520). IEEE, pp. 49–57.
Wang, H.H., Zhang, H., 2003. On the possibility of a private crop insurance market: a spatial statistics approach. The Journal of Risk and Insurance 70 (1), 111–124.
Xu, W., Filler, G., Odening, M., Okhrin, O., 2010. On the systemic nature of weather risk. Agricultural Finance Review 70 (3), 267–284.
Zabinsky, Z.B., 2003. Introduction to Stochastic Search and Optimization. Kluwer Academic Publishers.
Zabinsky, Z.B., 2010. Random search algorithms. In: Cochran, J.J., Cox Jr., L.A., Keskinocak, P., Kharoufeh, J.P., Smith, J.C. (Eds.), Wiley Encyclopedia of Operations Research and Management Science. Wiley Online Library.
Zhang, J., Tan, K.S., Weng, C., 2019. Index insurance design. ASTIN Bulletin 49 (2), 491–523.
Zhu, W., Porth, L., Tan, K.S., 2019. A credibility-based yield forecasting model for crop reinsurance pricing and weather risk management. Agricultural Finance Review 79 (1), 2–26.
Zhu, W., Tan, K.S., Porth, L., Wang, C.-W., 2018. Spatial dependence and aggregation in weather risk hedging: a Lévy subordinated hierarchical Archimedean copulas (LSHAC) approach. ASTIN Bulletin 48 (2), 779–815.
Zlochin, M., Birattari, M., Meuleau, N., Dorigo, M., 2004. Model-based search for combinatorial optimization: a critical survey. Annals of Operations Research 131, 373–394.