Catastrophe Insurance Pricing: A Robust Optimization Approach

Dimitris Bertsimas[1,2] and Cynthia Zeng[1]

[1]Operations Research Center, Massachusetts Institute of Technology
[2]Sloan School of Management, Massachusetts Institute of Technology

Abstract

The escalating frequency and severity of natural disasters, exacerbated by climate change, underscore the critical role of insurance in facilitating recovery and promoting investments in risk reduction. This work introduces a novel Robust Optimization (RO) framework tailored for the calculation of catastrophe insurance premiums, with a case study applied to the United States National Flood Insurance Program (NFIP). To the best of our knowledge, it is the first time a RO approach has been applied to for disaster insurance pricing. Our methodology is designed to protect against both historical and emerging risks, the latter predicted by advanced machine learning models, thus directly incorporating amplified risks induced by climate change. The framework offers three key contributions: first, using the US flood insurance data as an example, our model demonstrates a robust capacity to cover losses and produce surpluses, with a smooth balance transition through parameter fine-tuning; second, it introduces an equitable premium pricing structure, aligning higher premiums with regions of heightened risk; third, the framework offers versatility and generalizability, making it adaptable to a variety of natural disaster scenarios, such as wildfires, droughts, etc. This work not only advances the field of insurance premium modeling but also serves as a vital tool for policymakers and stakeholders in building resilience to the growing risks of natural catastrophes.

---

1 Introduction

Global climate change causes serious consequences in climate variability and weather extremes, which could lead to more frequent and costly natural disasters worldwide [28]. As shown in Figure 1, the number of disasters worldwide has increased tremendously during the last decades. Enhanced risks for catastrophic events, such as tropical cyclones, draughts, floods, heatwaves, can inflict severe damage and losses on individuals, businesses, communities, and the entire society. It is crucial to mitigate disaster risks and facilitate climate change adaptation [22]. The International Panel on Climate Change (IPCC) has emphasized the

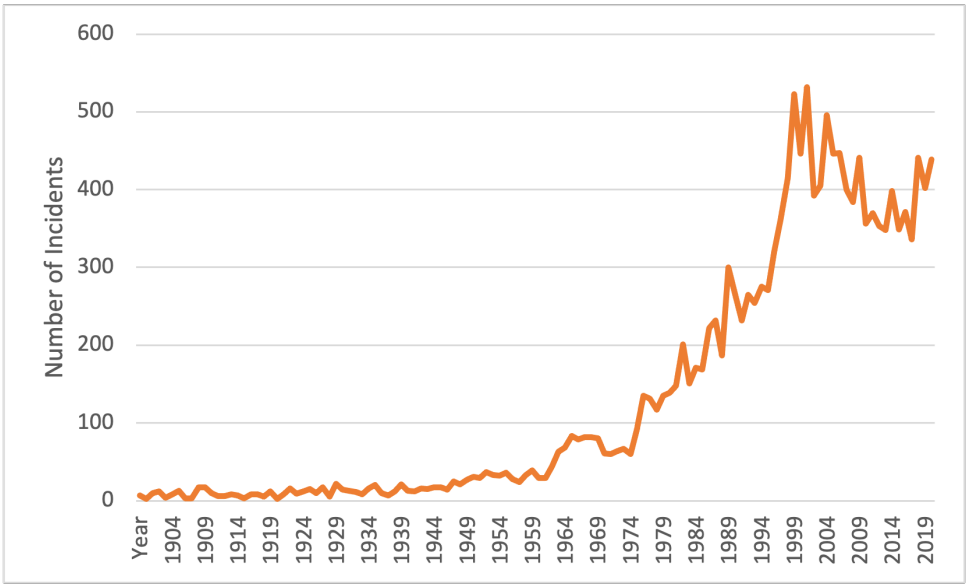


Figure 1: Number of major disasters globally since 1900, maintained by the EM-DAT database [1]. A disaster is defined as an event which overwhelms local capacity, necessitating a request to the national or international level for external assistance. Disasters include: flood, storm, earthquake, drought, landslide, extreme temperature, wildfire, volcanic activity, mass movement (dry), glacial lake outburst, fog, etc.

need for financial instruments for disaster risk management and climate change adaptation [20]. Catastrophe insurance emerges as a crucial risk management tool, offering financial support in recovery and incentivizing investments for mitigating efforts.

Catastrophe insurance, also known as disaster insurance, focuses on large-scale, low-frequency events that have the potential to cause widespread damage. This form of insurance typically covers disasters such as hurricanes, earthquakes, floods, terrorist acts, and pandemics. The rarity of such catastrophic events complicates the insurance process, as traditional actuarial methods fall short, often due to a lack of comprehensive historical data. Compounding this challenge, climate change is leading to more regular and destructive climate-related catastrophes, which traditional reliance on historical data alone tends to underestimate. In contrast to conventional insurance policies that spread risks across insured individuals, catastrophe insurance confronts a temporal problem of matching of regular influx of annual premiums with the irregular and unforeseeable distribution of payouts for losses. In the United States, catastrophe insurance has historically been managed predominantly through national programs. The National Flood Insurance Program (NFIP), for example, is the principal provider for flood insurance in the country, covering more than 95% of the underwriting risks [24]. However, the NFIP's actuarial effectiveness has been the subject of scrutiny, with the program operating at a significant deficit—19 billion US dollars as of 2023—highlighting the need for reform in the structuring of such insurance schemes [17]. The entry of private insurers into the catastrophe

coverage market has been deemed crucial, yet the inherent complexities associated with rare events have led to minimal participation from private entities [23]. Indeed, many private insurers are withdrawing from regions deemed "uninsurable", due to escalating risks from climate change [12].

In this work, we present a Robust Optimization (RO) framework for catastrophe insurance premium pricing designed to protect against uncertain losses. To the best of our knowledge, this is the first work using an RO approach to set disaster insurance premiums. We develop the framework and implement it to flood insurance using the National Flood Insurance Program (NFIP) data. The main contributions are three-fold:

- We present a Robust Optimization (RO) framework for catastrophe insurance premium pricing against uncertain losses. This framework encompasses two distinct uncertainty sets for potential losses: one retrospective, grounded in historical loss distributions, and the other prospective, utilizing risk estimates derived from machine learning predictions.
- We apply our RO framework to US flood insurance using data from the National Flood Insurance Program. We employ historical data from 1975 to 2012 to construct the uncertainty sets and train machine learning models, and we evaluate model performance using data from 2013 to 2022. We benchmark against two baselines: the actual NFIP policies and premiums derived from cumulative moving average (CMA) rules. Our evaluations indicate that the RO model outperforms in effectively covering losses and does so equitably, ensuring high-risk states bear appropriate premiums.
- We highlight the adaptability and generalizability of our framework, suggesting the potential application of an RO approach to pricing a wide range of catastrophic events, such as wildfires, droughts, extreme weather events.

The structure of the paper is as follows. In Section 2, we review the relevant literature. In Section 3 we introduce the problem and outline the Robust Optimization (RO) framework. In Section 4, we demonstrate the application of our framework through a case study on flood insurance in the United States. We explain the model parameter estimation and details on the machine learning risk prediction model. We compare our model results against two baselines: historical NFIP premiums and Cumulative Moving Average (CMA) scheme. Finally, we draw conclusions in Section 5.

2 Literature Review

This paper is related to the catastrophe modeling literature, often abbreviated at CAT modeling. It is a pivotal method for assessing and managing risks associated with extreme events employed by insurance companies. [14] offers a comprehensive analysis of how catastrophe models are employed for risk assessment and

management purposes. Most CAT models are proprietary, with AIR Worldwide, Risk Management Solutions (RMS), and EQECAT being the major players in the private sector; and the open-source HAZUS model developed by FEMA [27]. One major challenge of the CAT modeling approach lies in the lack of historical data due to the rarity of the events, and thus standard actuarial techniques fail to capture tail-event risks. In addition, catastrophe modeling depends on scenario simulations, which can be both numerous and lead to largely varying outcomes. How to combine different"what-if" scenarios remains a challenge to decision makers [19]. Our paper illustrates the promise of an optimization-based approach to model uncertain losses and climate risks directly, thus offering transparency and offer greater robustness against rare events.

Our work is also related to the literature of disaster management. Many works have focused the problem of general resource allocations to different programs or regions under a given budget constraint [29, 2, 26]. [30] discusses fund allocation for flash flood reduction, and [21] further incorporates the use of insurance premiums as a source of funding. As highlighted by [19], the critical issue in insurance policy lies in the need for decision-making robustness in the face of climate change's uncertainties. There is also growing work on using robust optimization in disaster relief management to deal with uncertainties [4, 31]. Few works discuss the use of optimization for catastrophe insurance pricing directly, [8] applies stochastic optimization to the Dutch Flood insurance scheme, and [9] discusses an integrated catastrophe management approach by incorporating CAT models with stochastic optimization methods. Our study fills in the gap by proposing a robust optimization framework by modeling uncertain losses and integrating machine learning forecast risks to address the increasing unpredictability of weather events driven by climate change.

Finally, this work broadly belongs to the climate finance literature, and see [16] for a comprehensive overview. [7, 15] examine the public policy implications in funding climate change adapation. It is critical to design new tools to financially manage weather risks. [25] discusses the role of insurance sector in decreasing the vulnerability of human and natural systems. [3, 13] propose the use of municipal bonds to finance natural disasters. [18] considers insuring climate change induced risks across broad business spectrum. Our paper adds to the literature by proposing a new design of the weather-related insurance contract to manage weather risk.

3 Optimization Framework for Catastrophe Insurance

3.1 Robust Optimization

Robust Optimization (RO) is a useful methodology for handling optimization problems with uncertain data [5]. It has been applied to address uncertainties in various fields, including operations research, engineering,

and finance. Consider a generic linear programming problem

$$\min_{\mathbf{x}} \left\{ \mathbf{c}^T \mathbf{x} \mid \mathbf{A}\mathbf{x} \leq \mathbf{b} \right\}$$

where $\mathbf{c} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{R}^m$ and $\mathbf{A} \in \mathbb{R}^m \times \mathbb{R}^n$. Robust Optimization addresses the problem where data (c, $\mathbf{A}, \mathbf{b}$) are uncertain, but are known to reside in an uncertainty set $\mathcal{U}$. Based on prior information or assumptions, we construct the uncertainty set $\mathcal{U}$ to express the uncertainties in data. We are addressing a family of problems for each realization of $(\mathbf{c}, \mathbf{A}, \mathbf{b}) \in \mathcal{U}$. Therefore, we can reformulate the problem into its Robust Counter part

$$\min_{\mathbf{x}} \left\{ \mathbf{c}^T \mathbf{x} \mid \mathbf{A}\mathbf{x} \leq \mathbf{b} \quad \forall (\mathbf{c}, \mathbf{A}, \mathbf{b}) \in \mathcal{U} \right\}.$$

3.2 Nominal Formulation

We consider setting insurance premiums for N locations for the insurance period of T years, which we denote with variable $p_{i,t}$, where $i = 1 \dots N$, $t = 1 \dots T$. We are given historical losses for each location for each year in the past T_0 years, which we denote with $\bar{l}_{i,t}$, where $i = 1 \dots N$, $t = 1 \dots T_0$. We assume that the future losses for each location i n the insurance period of T years with $l_{i,t}$, $i = 1 \dots N$, $t = 1 \dots T$. Note that this quantity is unknown, but we assume the knowledge of it to introduce a simple deterministic LP formulation to introduce the basic requirements and set up the generic framework. In the next subsections, we expand on how to model this uncertain quantity through Robust Optimization framework.

We formulate an LP model to set insurance premium price. The objective function is to minimize the overall premium collected, or the least required premium needed:

$$\min_{p_{i,t}} \sum_{i=1}^{N} \sum_{t=1}^{T} f(p_{i,t}) * p_{i,t}. \tag{1}$$

We require the premium price to cover projected losses with an additional buffer amount, denoted by δ, which is set at constant for each location. In addition, to model the consumer behavior that as we increase premium price less consumers are willing to purchase the insurance product, we introduce a damping function $f : \mathbb{R} \to \{0, 1\}$, a monotonically decreasing function representing decline in demand due to higher premiums. Details of the choice of the damping function will be discussed later. Thus we require:

$$\sum_{t=1}^{T} f(p_{i,t}) * p_{i,t} - \sum_{t=1}^{T} f(p_{i,t}) * l_{i,t} \geq \delta \quad i \in [N]. \tag{2}$$

In addition, we impose a constraint to require premiums collected over consecutive years to vary slowly, in

order to prevent drastic changes in insurance premiums:

$$|p_{i,t} - p_{i,t-1}| \leq \gamma_1 \quad i \in [N], t \in [T]. \tag{3}$$

Variables $p_{i,t}$ should be positive, for each location i for each period t:

$$p_{i,t} \in \mathbb{R}^+ \quad i \in [N], t \in [T]. \tag{4}$$

3.3 Robust Optimization Formulation

In the nominal formulation, we assume the knowledge of projected loss for each of the future period. However, this quantity is unknown and highly uncertain. In this subsection, we expand on how to construct uncertainty sets to describe this quantity.

The overall optimization formulation is the same as before, except for constraint 3, where we require the inequality to hold for all losses in the uncertainty set:

$$\sum_{t=1}^{T} f(p_{i,t}) * p_{i,t} - \sum_{t=1}^{T} f(p_{i,t}) * l_{i,t} \geq \delta \quad i \in [N], \quad \forall l_{i,t} \in \mathcal{U}. \tag{5}$$

We propose two uncertainty sets to model the uncertainties of the future losses: with Central Limit Theorem (CLT) and with Machine Learning driven risks.

3.3.1 Uncertainty Set from Central Limit Theorem

Based on the assumption that for each specific location, future flooding losses follow the distribution of historical losses. We adopt the central limit theorem (CLT) to form the uncertainty set. Formally, for a fixed location i, $L_{i,t}, t \in [T]$ are independent, identically distributed random variables with mean $\bar{l_i}$ and standard deviation $\bar{\sigma_i}$, where $\bar{l_i}, \bar{\sigma_i}$ are historical mean and standard deviation for location i. Then as proposed in [6], we can assume the quantities $L_{i,t}$ take values such that

$$\left| \sum_{t=1}^{T} L_{i,t} - T \cdot \bar{l_i} \right| \leq \gamma_2 \cdot \sigma_i \sqrt{n}, \tag{6}$$

where γ_2 is a small constant to denote how close future losses distribution should deviation from the normal distribution. In other words, we describe the uncertain quantities $L_{i,t}$ as values in the uncertainty set

$$\mathcal{U}_i^{CLT} = \{(l_{i,1}, \ldots, l_{i,T}) : \frac{|\sum_{t=1}^{T} l_{i,t} - \bar{l_i}|}{\sigma_i \sqrt{T}} \leq \gamma_2\}, \tag{7}$$

where $\bar{l}_i, \sigma_i$ can be computed for each location i using historical data. The larger we set γ_2, the more conservative the optimization model, and higher premiums will be. As a remark, we derive one uncertainty set for each location using the historical mean and standard deviation for that location to acomodate different flooding risk profiles. We derive the uncertainty sets separately for each location $\mathcal{U}_i^{CLT}$.

3.3.2 Uncertainty Set from (Machine Learning) Risk Models

In addition, suppose we have some information on how future losses should be, which could be informed by risk models. We can formulate such model predictions as uncertainty sets. One way of obtaining such risk prediction is through machine learning models, as recent advances in ML models demonstrate capabilities for models to predict accurate multi-year forecasts. In this work, we build machine learning models to obtain flooding risks, see greater details in the next section. Nevertheless, one can obtain such risk predictions from physical-based models, or other alternative approaches.

We define a major flood event to be flood incurring loss over a threshold loss level Θ. Suppose we have model predictions for the risk that at location i, the probability of having one flood event in the next k years to be $q_{i,k}$. Then we can express the model prediction as

$$\mathbb{P}(\sum_{t=1}^{k} z_{i,t} = 1) = q_{i,k}, \tag{8}$$

where $z_{i,t} \in \{0,1\}$ are binary variables denoting if there is a major flood at location i at time t. By modeling such, we assume having more than one major event is negligible for reasonable k. Since model predictions are probabilistic predictions which can have errors, and we want to be conservative and protect against suffering potential huge losses. Thus, assuming actual incidence of having a major flood is close to the model predictions, we can express $z_{i,t}$ as random variables taking values in an uncertainty set as

$$\mathcal{U} = \{z_{i,t} : |\sum_{t=1}^{k} z_{i,t} - q_{i,k}| \leq \epsilon, \quad \sum_{t=1}^{k} z_{i,t} \leq 1\}, \tag{9}$$

where the constant ϵ is a parameter to indicate how close we believe the model predictions are to actual probabilities. The larger the value, the less confident and more conservative our model will be. Therefore, linking variables $z_{i,t}$ to $l_{i,t}$, and assuming the future losses will be bounded by the expected value coming from suffering a major flood, we model the uncertainty set of the future loss for each location i as follows

$$\mathcal{U}_i^{ML} = \{z_{i,t}, l_{i,t} : \sum_{t=1}^{k} l_{i,t} \leq \Theta \cdot \sum_{t=1}^{k} z_{i,t}, \quad |\sum_{t=1}^{k} z_{i,t} - q_{i,k}| \leq \epsilon, \quad \sum_{t=1}^{k} z_{i,t} \leq 1\}. \tag{10}$$

3.3.3 The Robust Counterpart

Combining uncertainty sets from Central Limit Theorem and from Machine Learning risk models, the robust optimization formulation of the problem is as follows

$$\begin{aligned}
&\min_{p_{i,t}} \sum_{i=1}^{N}\sum_{t=1}^{T} f(p_{i,t}) * p_{i,t} \\
&\quad \sum_{t=1}^{T} f(p_{i,t}) * p_{i,t} - \sum_{t=1}^{T} f(p_{i,t}) * l_{i,t} \geq \delta, \quad \forall l_{i,t}, z_{i,t} \in \mathcal{U}_i^{CLT}, l_{i,t} \in \mathcal{U}_i^{ML}, \quad i \in [N] \\
&\quad ||p_{i,t} - p_{i,t-1}|| \leq \gamma_1, \quad i \in [N], t \in [T] \\
&\quad p_{i,t} \in \mathbb{R}^+, z_{i,t} \in \{0,1\}, \quad i \in [N], t \in [T].
\end{aligned} \tag{11}$$

Since the uncertain occurs only in one constraint, we can write Problem 11 as a min-max problem

$$\begin{aligned}
\min_{p_{i,t}} \max_{l_{i,t}, z_{i,t} \in \mathcal{U}} & \sum_{i=1}^{N}\sum_{t=1}^{T} f(p_{i,t}) * p_{i,t} \\
& \sum_{t=1}^{T} f(p_{i,t}) * p_{i,t} - \sum_{t=1}^{T} f(p_{i,t}) * l_{i,t} \geq \delta, \quad i \in [N] \\
& ||p_{i,t} - p_{i,t-1}|| \leq \gamma_1, \quad i \in [N], t \in [T] \\
& l_{i,t} \in \mathcal{U}_i^{CLT}, \quad i \in [N] \\
& l_{i,t}, z_{i,t} \in \mathcal{U}_i^{ML}, \quad i \in [N] \\
& p_{i,t} \in \mathbb{R}^+, z_{i,t} \in \{0,1\} \quad i \in [N], t \in [T].
\end{aligned} \tag{12}$$

Next, we consider the inner problem

$$\begin{aligned}
\max_{l_{i,t}, z_{i,t} \in \mathcal{U}} & \sum_{i=1}^{N}\sum_{t=1}^{T} f(p_{i,t}) * p_{i,t} \\
& \sum_{t=1}^{T} f(p_{i,t}) * p_{i,t} - \sum_{t=1}^{T} f(p_{i,t}) * l_{i,t} \geq \delta, \quad i \in [N] \\
& l_{i,t} \in \mathcal{U}^{CLT} \\
& l_{i,t}, z_{i,t} \in \mathcal{U}^{ML} \\
& p_{i,t} \in \mathbb{R}^+, z_{i,t} \in \{0,1\} \quad i \in [N], t \in [T].
\end{aligned} \tag{13}$$

For the inner problem, we can treat $p_{i,t}$ as constants, thus the inenr problem can be simplified to

$$\begin{aligned} \max_{l_{i,t}, z_{i,t} \in \mathcal{U}} & \sum_{i=1}^{N} \sum_{t=1}^{T} l_{i,t} \\ & l_{i,t} \in \mathcal{U}^{CLT} \\ & l_{i,t}, z_{i,t} \in \mathcal{U}^{ML} \\ & p_{i,t} \in \mathbb{R}^{+}, z_{i,t} \in \{0,1\} \quad i \in [N], t \in [T]. \end{aligned} \tag{14}$$

Note that the uncertainty set is composed of two separate uncertainty sets $\mathcal{U}^{CLT}$ and $\mathcal{U}^{ML}$. We can thus decompose the inner problem into two subproblems, and the objective of the original problem takes the maximum of the two subproblems. Solving each subproblem separately, we can then plug back the analytical solution from each subproblem back to the original problem.

Proposition. *The overall min-max problem is equivalent to*

$$\begin{aligned} & \min_{p_{i,t}} \sum_{i,t}^{T} f(p_{i,t}) * p_{i,t} \\ & \sum_{t=1}^{T} f(p_{i,t}) * p_{i,t} - \frac{1}{T} \sum_{t=1}^{T} f(p_{i,t}) * L_i^{CLT} \geq \delta \quad \forall i \in [N] \\ & \sum_{t=1}^{k} f(p_{i,t}) * p_{i,t} - \frac{1}{k} \sum_{t=1}^{k} f(p_{i,t}) * L_i^{ML} \geq \delta \quad \forall i \in [N] \\ & ||p_{i,t} - p_{i,t-1}|| \leq \gamma_1 \quad \forall i \in [N], t \in [T] \end{aligned} \tag{15}$$

where

$$L_i^{CLT} = T \cdot \bar{l}_i + \gamma_2 \cdot \sigma_i \sqrt{T} \tag{16}$$

$$L_i^{ML} = \Theta \cdot \min\{1, q_{i,k} + \epsilon\}. \tag{17}$$

As a remark, we take the average "damping" over losses $l_{i,t}$ because we solve for the optimal $\sum_t l_{i,t}$.

Proof. Consider the first subproblem

$$\begin{aligned} \max_{l_{i,t}} & \sum_{i=1}^{N} \sum_{t=1}^{T} l_{i,t} \\ & |\sum_{t=1}^{T} l_{i,t} - T \cdot \bar{l}_i| \leq \gamma_2 \cdot \sigma_i \sqrt{T} \quad \forall i \in [N] \\ & l_{i,t} \in \mathbb{R}^{+}, \end{aligned} \tag{18}$$

looking at the constraint, we can take out $|\cdot|$ since we are maximizing over $l_{i,t}$, and rearranging terms

$$\sum_{t=1}^{T} l_{i,t} \leq T \cdot \bar{l}_i + \gamma_2 \cdot \sigma_i \sqrt{T} \quad \forall i \in [N] \tag{19}$$

and optimality is achieved at equality, we can solve for i.e., $l^*_{i,t}$ s.t. :

$$\sum_{t=1}^{T} l_{i,t} = T \cdot \bar{l}_i + \gamma_2 \cdot \sigma_i \sqrt{T} \quad \forall i \in [N]. \tag{20}$$

Hence we can compute for all location $i \in [N]$, and denote the analytical solution to the first subproblem as L_i^{CLT} :

$$L_i^{CLT} := \sum_{t=1}^{T} l^*_{i,t} = T \cdot \bar{l}_i + \gamma_2 \cdot \sigma_i \sqrt{T} \sqrt{(T_1 - T_0)} \tag{21}$$

Consider now the second subproblem

$$\begin{aligned} \max_{l_{i,t} z_{i,t}} & \sum_{i=1}^{N} \sum_{t=1}^{T} l_{i,t} \\ & \sum_{t=1}^{k} l_{i,t} \leq \Theta \cdot \sum_{t=1}^{k} z_{i,t} \\ & |\sum_{t=1}^{k} z_{i,t} - q_{i,k}| \leq \epsilon \\ & \sum_{t=1}^{k} z_{i,t} \leq 1 \\ & l_{i,t} \in \mathbb{R}^+, z_{i,t} \in \{0,1\}. \end{aligned} \tag{22}$$

Without loss of generality, we relax $z_{i,t}$ to take continuous value $z_{i,t} \in [0,1]$, because we can treat $\sum_i^k z_{i,t}$ as one variable taking continuous values in $[0,1]$. Thus looking at the constraints concerning $z_{i,t}$, and take out $|\cdot|$ since we are maximizing over $z_{i,t}$, we get $z^*_{i,t}$ achieve optimality at

$$\sum_{t=1}^{k} z^*_{i,t} = \min\{1, q_{i,k} + \epsilon\} \tag{23}$$

which gives $l^*_{i,t}$ at optimality at

$$\sum_{t=1}^{k} l^*_{i,t} = \Theta \cdot \sum_{t=1}^{k} z^*_{i,t}. \tag{24}$$

Combining with the solution from the first subproblem given by equation 21 we have the sum of future loss

for each location given as follows

$$\sum_{t=1}^{T} l_{i,t}^{*} = T \cdot \bar{l}_i + \gamma_2 \cdot \sigma_i \sqrt{T} \tag{25}$$

$$\sum_{t=1}^{k} l_{i,t}^{*} = \Theta \cdot \min\{1, q_{i,k} + \epsilon\}, \tag{26}$$

where the first equality bounds the entire future period T, the second equality bounds the period depending on machine learning model's risk forecast horizon k. As a remark, we can have multiple risk models for different forecasting horizons. □

3.4 Choices of demand damping function f(p)

Recall that to model the behavioral aspect that as insurance premium increase there is less demand for it, we introduce the damping function $f(p) : \mathbb{R} \to \{0, 1\}$ into the constraint. To preserve the convexity property of the overall problem, we model the behavior using piece-wise linear functions under the scope of this work. In the next section, we explain in greater detail on the estimation using NFIP data, as well as sensitivity analysis on the choice of the function.

As a remark, as we damp demand, we correspondingly damp the covered loss in the constraint. Since the inner problem gives the the overall loss over the forecasting period T, i.e., L_i^{CLT} gives the maximum loss deduced from the uncertainty set over period T, we have taken the corresponding damping term to be the average over T periods, i.e., $\frac{1}{T}\sum_{t=1}^{T} f(p_{i,t})$.

4 Case Study for US National Flood Insurance

4.1 Data

We used two redacted datasets from the National Flood Insurance Program (NFIP) on claims and policies respectively. Both data sets are created and maintained by Federal Emergency Management Agency (FEMA). The claims transaction data provide details on NFIP claims transactions across from all states in the United States [10]. This dataset consist of 2570089 lines of claim transactions ranging, dated from 1970. Due to limited data availability in the early years, we included data from 1975 to 2022. For this study, we aggregate data into state level on an annual basis, and have used the following features: date ('dateOfLoss'), state, claim amount ('amountPaidOnBuildingClaim').

The policy premium data contains 228664 lines of data, covering 54 states and 6704 unique zip codes, and

contains policies from 2009 to 2022 [11]. We use the policy data as a benchmark to compare model performance between last ten years of testing period from 2013 to 2022. Similar to the claims data, we aggregated data into state level on an annual basis, and have used the following features: state ('propertyState'), date ('policyTeminationDate') and premium ('totalInsurancePremiumOfThePolicy').

In this work, we consider setting insurance premium at state level on an annual basis. We consider setting the premium for the last 10 years, from 2013 to 2022. We trained machine learning models using data from 1975 to 2012, to derive risks in the testing period from 2013 to 2022. We used machine learning derived risks as parameter input for the optimization model, and drive premiums using the RO framework.

4.2 Optimization Model Parameter Estimation

4.2.1 Computing L_i^{CLT}

Recall from equation 16, to compute L_i^{CLT} we need to compute the historical mean and variance for each state. Table 1 exhibits the historical mean and standard deviation for the top 10 most costly states, computed on the annual basis.

State	Max	Mean	Std	Median
LA	3,763,390	45,084	58,467	20,583
TX	8,973,270	44,046	65,420	19,801
NJ	4,022,518	32,727	58,000	13,042
NY	9,467,720	35,725	74,510	13,306
FL	9,100,033	25,351	63,809	8,052
MS	10,000,000	46,603	94,924	14,823
NC	1,294,678	21,725	41,345	7,992
PA	1,889,793	19,040	41,943	6,939
AL	4,900,000	28,059	95,474	8,466
SC	1,764,000	25,574	43,752	10,088

Table 1: Statistics of top 10 most costly states in the US, including maximum annual claim loss, mean, standard deviation and median.

4.2.2 Demand Damping

In this work, we model the demand sensitivity to insurance premium through a piece-wise linear demand function. We estimate the decline rate using historical data from several states. Figures 2 and 3 below shows the scatter plot of number of policy holders in a year against the mean policy premium of that state at that year. Different states have different degrees of sensitivity to price, but in general we observe a downward trend of decline in policy holder number as a function of increased price. For the illustrative purpose of this work, we do not specify different sensitivity in different states, but use the same demand damping function

across all states.

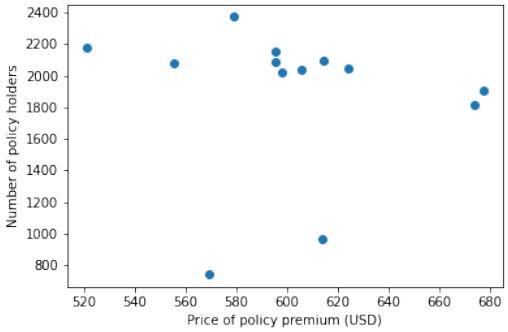


Figure 2: (a) Demand damping estimation for Louisiana state (LA).

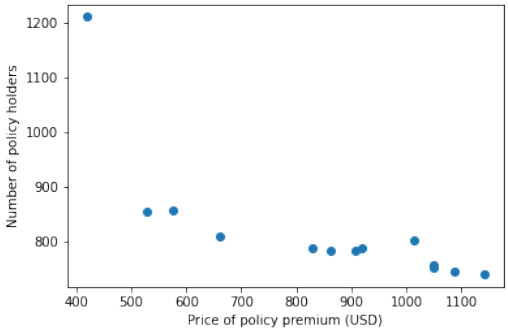


Figure 3: (b) Demand damping estimation for New York state (NY).

We use the following piece-wise linear demand damping function to model demand damping

$$f(p) = \begin{cases} 1 & \text{if p} \leq P_0 \\ 1 - m \cdot (p - P_0) & \text{if p} \geq P_0 \text{ and } f(p) \geq c_{min} \\ \text{c}_{min} & \text{otherwise} \end{cases} \tag{27}$$

where P_0 is the minimum premium at which demand damping starts to occur, and c_{min} is the minimum fraction of demand. In this work, we choose P_0 to be $\frac{1}{10} \cdot P_{hist}^{max}$, a fraction of maximum historical premium ever charged. We choose c_{min} to be 0.2 representing at least 20% of the total demand is preserved regardless of price. We experimented with different demand damping rate: $m = 1/P_{hist}^{max}$. Figure 4 illlustrate several choices of the demand damping curve.

4.3 Machine Learning Model

The goal of the machine learning model is to predict the risks of total loss surpassing a certain threshold for each state in the next 1-K years.

4.3.1 Data Processing

First we kept data from 1975 to 2022 due to limited availability in prior years. We have dropped claims where claim amount is missing. We have dropped data from the following states, 'MP'(Northern Mariana Islands), 'AS' (American Samoa), 'GU' (Guam), 'DC' (District of Columbia), due to limited availability. In total, we have data from 53 states across 48 years.

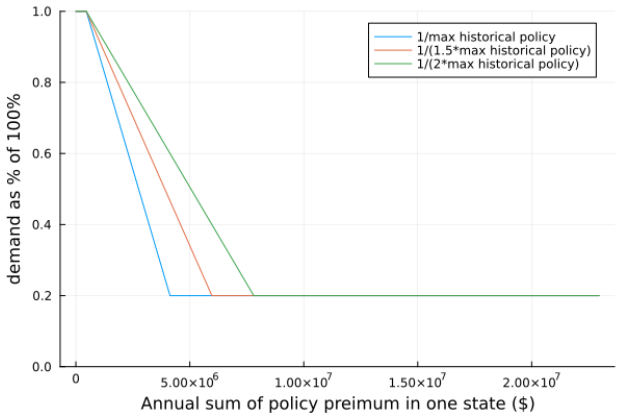


Figure 4: Different piece-wise linear demand damping curves corresponding to different rate of decline.

To construct the machine learning model, first, we aggregate data to state and annual level. The index of the data is two levels: state and year. For missing data, we performed linear interpolation within each state using previous and later years. Then for each state at a particular year, we construct the following features: state (categorical), current year annual loss, past 1-5 years annual loss. Current and past year losses are numerical features, where as state (categorical) feature is treated with one-hot encoding.

We train binary classification models to predict the target, with 1 indicating for a particular state at a particular year, the state will suffer an annual loss passing through the threshold Θ in the next 1 to K years. We have experimented with three threshold values, corresponding to 90th, 95th and 99th percentile annual claim amount values across all states over all training data, corresponding to USD 18,558,788, USD 50,688,672 and USD 321,903,271. In addition, we have experimented with three K values, corresponding to 3, 5, 10 years respectively.

4.3.2 Training and Testing Protocols

We split the data set chronologically into training period from 1975 to 2011, and testing period from 2012 to 2022. We experiment with two standard machine learning models, logistic regression and XGBoost. We employed 3-fold cross validation to search for the best parameters for each type of models. The search space for hyperparameter tuning can be found in Table 2 below.

Model	Hyperparameters	Values
XGBoost	number of estimators	100, 150
	maximum tree depth	4, 6
	learning rate	0.1, 0.3
Logistic Regression	C	0, 0.2, 0.4, 0.6, 0.8, 1
	penalty	L1, L2

Table 2: Hyperparameters searched for our models.

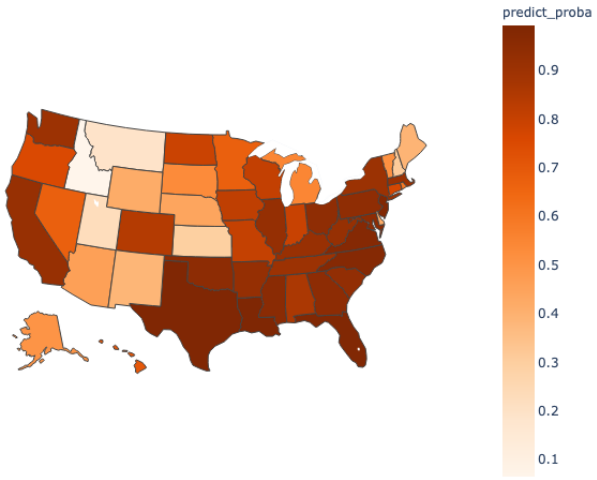


Figure 5: Map illustrating machine learning (ML) predicted risks for all states surpassing 90th-percentile flooding risk within a 5-year time frame from 2016. Regions are color-coded, with darker shades indicating a higher probability of predicted risks.

4.3.3 Prediction Results

Figure 5 illustrates machine learning predicted risks for all states surpassing 90th percentile flooding risk within the next 5-year time frame on 2016. For each state, we produce one prediction for each year over the testing period, for each threshold, and for 3-year, 5-year, 10-year time frames. Table 3 4 5 record out of sample prediction results using testing data, corresponding to data between 2012 to 2022. We treat data in the testing period on a rolling basis, and we drop the years where we do not have target data, i.e., in year 2019, we predict for K=3 but not for K = 5 or 10. We remark that accuracy is generally higher for longer forecasting horizons. This is likely due to the following reasons: first, longer forecasting horizon lead to higher probability of flood, which leads to more balanced data; second, we have less testing samples. We use the probabilistic prediction results for each state at each testing year $q_{i,k}$ as input to construct uncertainty sets for the robust optimization model as given by equation 17.

Model	n_pred = 3		n_pred = 5		n_pred = 10	
	logreg	xgb	logreg	xgb	logreg	xgb
auc	0.743	0.776	0.763	0.808	0.767	0.912
f1	0.630	0.665	0.593	0.735	0.621	0.817
accu	0.693	0.675	0.632	0.736	0.623	0.830
accu_bl	0.625	0.706	0.605	0.753	0.641	0.809
precision	0.407	0.457	0.520	0.625	0.658	0.777
recall	0.437	0.793	0.362	0.901	0.500	0.967

Table 3: Out of sample accuracy for 90% percentile threshold prediction, i.e., annual loss passing 18558788. We experiment two families of models, logistic regression (logreg) and XGBoost (xgb). We predict for 3 time horizons, for 3, 5, 10 years respectively.

Model	n_pred = 3		n_pred = 5		n_pred = 10	
	logreg	xgb	logreg	xgb	logreg	xgb
auc	0.818	0.871	0.818	0.897	0.794	0.921
f1	0.684	0.673	0.700	0.744	0.764	0.763
accu	0.856	0.781	0.808	0.792	0.830	0.774
accu_bl	0.665	0.742	0.699	0.829	0.736	0.820
precision	0.300	0.306	0.368	0.460	0.595	0.560
recall	0.391	0.688	0.516	0.891	0.500	0.938

Table 4: Out of sample accuracy for 95% percentile threshold prediction, i.e., annual loss passing 50688672.

4.4 Computational resources

The source codes for this study, implemented in Julia 1.7 and Python 3.9, are publicly accessible at [repository link]. The convex optimization problems were solved using Ipopt and Gurobi solvers, with machine learning models trained on a local machine with 4 Intel CPUs on a Macbook Pro personal computer. Comprehensive documentation detailing the methodology and specific parameters can be found in the repository's code comments.

4.5 Results

We implement the robust optimization with linearly decreasing demand decline rate and risk forecasts obtained from the trained machine learning models. We benchmark our outcomes against two established baseline policies and undertake a sensitivity analysis to assess the impact of our model's parameter selections. As argued in [**?**], we evaluate model performance on two key criteria:

- Effectiveness, measured by the capability to offset losses, a criterion inspired by the existing deficit in the NFIP program.
- Equity of premium distribution, a criterion based upon the observation that states with higher risks account for the majority share of losses.

Model	n_pred = 3		n_pred = 5		n_pred = 10	
	logreg	xgb	logreg	xgb	logreg	xgb
auc	0.920	0.945	0.922	0.956	0.952	0.992
f1	0.704	0.687	0.737	0.771	0.835	0.906
accu	0.910	0.950	0.909	0.943	0.925	0.962
accu_bl	0.809	0.646	0.786	0.731	0.958	0.979
precision	0.253	0.215	0.313	0.380	0.556	0.714
recall	0.696	0.304	0.640	0.480	1.000	1.000

Table 5: Out of sample accuracy for 99% percentile threshold prediction, i.e., annual loss passing 321903271.

We evaluate performance during the testing period, the last ten years of available data in the NFIP data set between 2013 to 2022. For the rest of the section, we choose γ_1 to be 50000, and δ to be 10000.

4.5.1 Baseline Models

We compare results against two baseline policy premiums. First, we use historical premiums charged for each state, which is referred to as 'hist'. Second, we set the premium to be the cumulative moving average loss up to that year, which we refer to as 'CMA'. We compute the cumulative moving average loss as follows

$$p_{i,t}^{CMA} = \frac{1}{t}\sum_{t'=0}^{t} l_{i,t'}. \tag{28}$$

4.5.2 Sensitivity Analysis

We examine the model sensitivity to parameter choices in our model. Specifically, we examine the effect of the value of γ_2 and the demand damping rate m. Recall that γ_2 is the parameter controlling how conservative the CLT uncertainty set given by equation 16, and the demand damping rate controls how fast demand declines in response to increase in price given by equation 27. To examine the overall performance of the premium, we compute the the cumulative surplus S across all states over the testing period as follows

$$S(\gamma_2) = \sum_{i=1}^{N}\sum_{t=1}^{T1} p_{i,t} - \sum_{i=1}^{N}\sum_{t=1}^{T1} l_{i,t}^{act}. \tag{29}$$

Figure 6 shows the level of surplus as a function of different γ_2, with different demand damping rate. The two dotted line shows the constant surplus computed by two baselines: using the actual premiums collected during this period and the CMA rule. We observe that both baselines incur a loss over the testing period, with historical premiums resulting in about 20 billion loss, and CMA rule resulting in 8 million loss.

We let γ_2 to take values between 0 and 1.5 at a stepsize of 0.1, and resolve the optimization model at each γ_2 value, and compute the sum of surplus across all states across testing years. $\gamma_2 = 0$ corresponding to convex

optimization without uncertainty, and $\gamma_2 = 1.5$ with the maximum degree of uncertainty. As we increase the value of γ_2, the size of the uncertainty set increases, and the model becomes more conservative resulting in surplus as expected. We remark that the surplus breaks even when γ_2 takes value between 0.6 and 0.7. And we observe a smooth increase in surplus as γ_2 increases.

In addition, we experiment with three demand damping rates: no damping, $m_1 = 1/P_{hist}^{max}$, $m_2 = 1/(2 * P_{hist}^{max})$, with m_2 damps twice as fast as m_1. Similar as above, we experiment with varying γ_2 corresponding to the different demand damping rates. We observe that the choice of demand damping is less significant compared with the variation of γ_2.

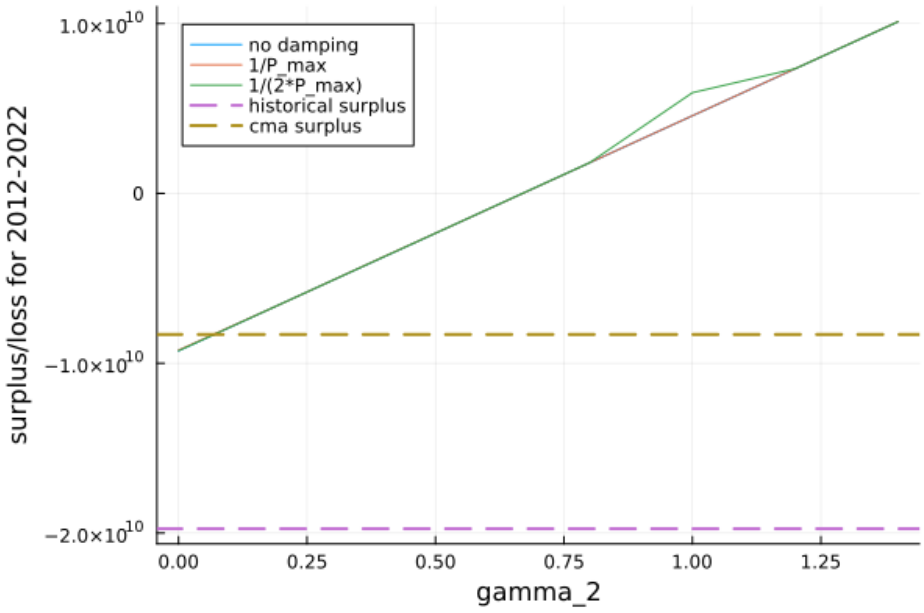


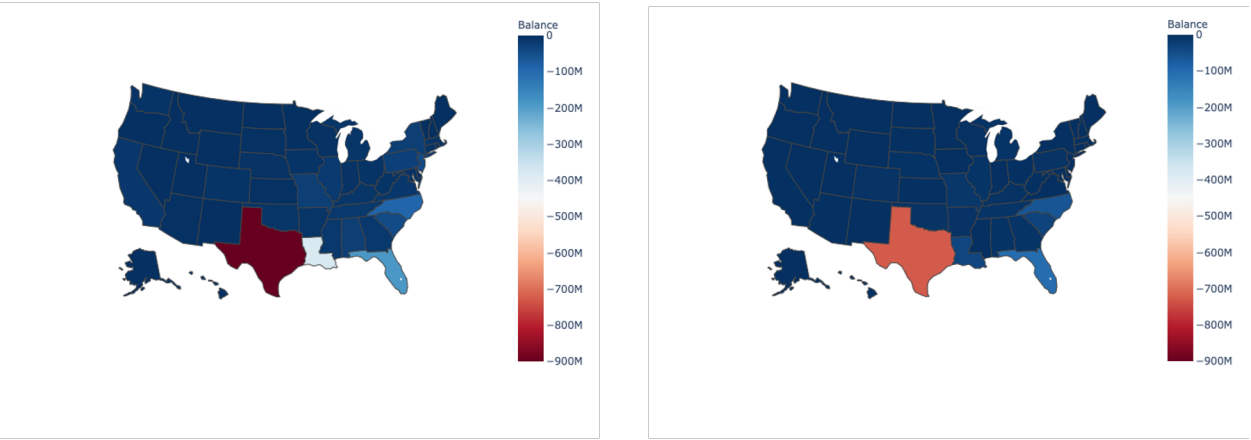
Figure 6: Surplus (or loss) computed during 2012 to 2022 across all states when vary different levels of γ_2. Two dotted lines demonstrate the level of surplus from two baseline models: historical surplus calculated from the actual premiums collected, cma surplus is computed using the cumulative moving average.

4.5.3 Effectiveness in Covering Losses

In figures 7a and 7b, we illustrate respectively the financial balance for two baseline policy schemes. The balances are calculated at a state level as the difference between cumulative premiums and cumulative losses, over the testing period from 2013 to 2022. Our analysis reveals that the NFIP premium consistently operates at a deficit across all states, averaging a shortfall of approximately 1.9 billion dollars annually. This underscores the fact that the NFIP considerably undercharges in every state. In a parallel observation, the CMA premium scheme also results in a deficit, albeit at a lower average of 830 million dollars per year. These findings underscore the limitations of relying on historical averages as indicators for insurance pricing,

especially in the context of a shifting climate landscape.

On the other hand, recall that the balance outcome of the RO approach is contingent upon the selection of the parameter γ_2, which orchestrates a gradual transition between deficit and surplus scenarios. Figure 8a and 8b showcase two distinct selections for γ_2. With $\gamma_2 = 0.6$, the model incurs an annual deficit of 96 million. When $\gamma_2 = 0.8$, it yields an annual surplus of 180 million. These observations underscore the inherent flexibility and potential advantage of the RO approach, granting policy providers the discretion to determine the desired surplus level. Detailed state level losses and premiums can be found in the appendix.



(a) NFIP premium scheme

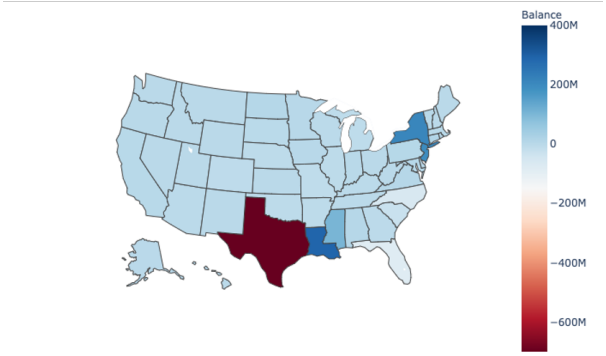
(b) CMA premium scheme

Figure 7: Average balance, computed as premium charged subtracting losses incurred, during the the testing period 2013-2022 by two baseline policy schemes: historical NFIP premiums on the left, and CMA rule on the right. For both schemes, the policy runs deficit across all states during the testing years. NFIP scheme runs at deficit on average 1.9 billion per year; CMA scheme runs on average at 830 million per year.

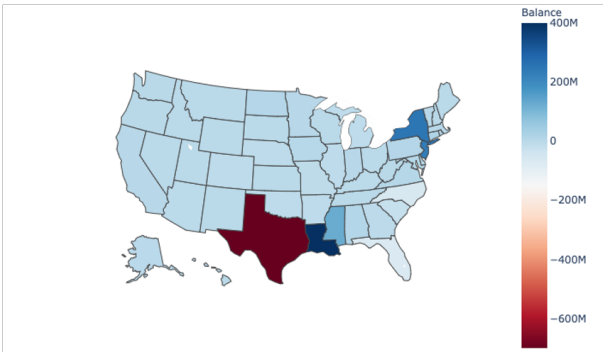
4.5.4 Equity in Premium Attribution

Figure 9 illustrates the total claimed losses since 1975. We observe that the total loss peaks whenever a major flooding event occurs in a certain state. In the US, most of the claim losses can be attributed to the high risk states, thus arising the equity concern in measuring the quality of insurance premium schemes. As displayed in Table 6, the eight most expensive states contribute to a cumulative 77% of the total loss during the testing period between 2013-2022, with Texas, Louisiana and Florida contributing 41%, 17% and 9% respectively. These observations highlight the need to achieve equity in premium distribution, aligning high risk states with high premium requirements.

Figure 10 presents the cumulative claim losses alongside various insurance schemes for the top 8 most costly states during the testing period. It is evident that the NFIP scheme consistently undercharges these high-cost states. Conversely, the CMA rule tends to overcharge states with significant historical losses. For instance,



(a) RO scheme with $\gamma_2 = 0.6$



(b) RO scheme with $\gamma_2 = 0.8$

Figure 8: Average balance, computed as premium charged subtracting losses incurred, during the the testing period 2013-2022 by Robust Optimization (RO) policy schemes using two values of γ_2 at 0.6 on the left and 0.8 on the right respectively. Recall that γ_2 parameter ensures a smooth transition between deficit to surplus in the RO scheme. For $\gamma_2 = 0.6$, the policy runs at a deficit at 96 million per year on average; for $\gamma_2 = 0.8$, the policy runs at a surplus at 180 million per year on average.

State	**Sum**	**Percentage**
TX	9.277e9	41.1%
LA	3.790e9	16.8%
FL	1.931e9	8.6%
NC	8.943e8	4.0%
NJ	4.610e8	2.0%
SC	4.543e8	2.0%
PA	2.742e8	1.2%
NY	2.494e8	1.1%
Rest States	5.214e9	23.1%

Table 6: Table showing the cumulative loss by top 8 costly states and their respective percentages during the testing period of 2013-2022.

the aftermath of Hurricane Katrina in 2005 inflated the CMA premium for Louisiana, even though the state did not experience comparably high losses in the subsequent testing period. In comparison, the RO scheme offers premiums that more closely align with actual losses. A smiliarly trend is noticeable for other states like New Jersey and New York. Finally, it is worth noting that none of the schemes adequately accounted for the substantial losses experienced in Florida. One contributing factor may be the occurrence of several major floods in Florida during the testing period. Our robust optimization (RO) formulation's machine learning (ML) uncertainty set is designed to protect against only a single major flood event over the entire period.

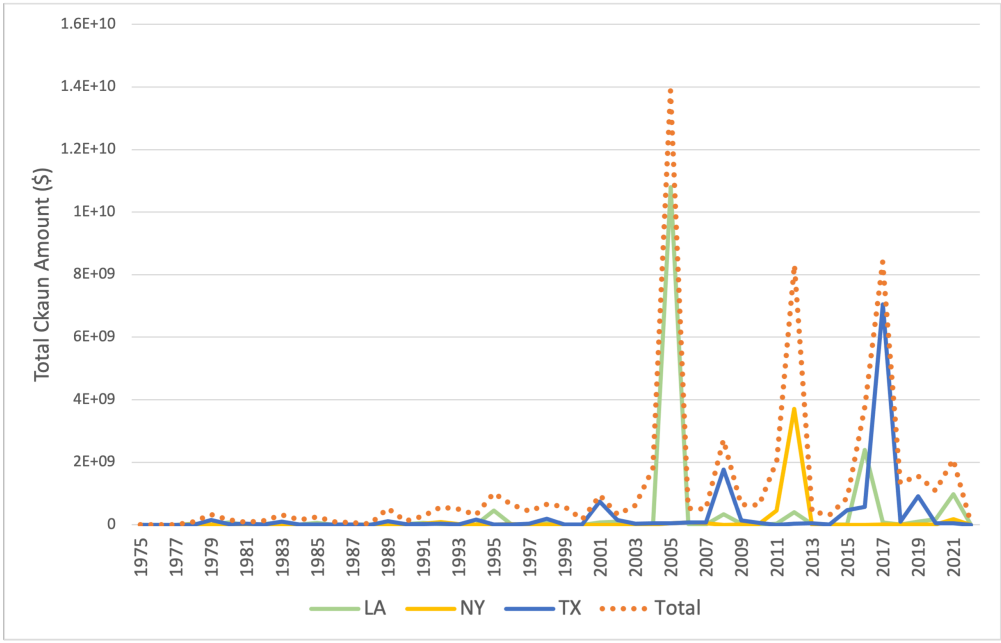


Figure 9: Trend in claim loss since 1975. Dotted line displays the total claimed loss in each year across all states. Gree, yellow and blue solid lines display the claimed loss among three high risk states.

5 Conclusion

In conclusion, we present a Robust Optimization (RO) framework to catastrophe insurance premium pricing. We first present a nominal linear optimization formulation to introduce the problem of setting insurance prices for rare catastrophe events, and present a robust optimization formulation with two distinct uncertainty sets. The Central Limit Theorem (CLT) uncertainty set protects against deviations from historical losses, and the Machine Learning (ML) uncertainty set to incorporate predicted risks. We derive the robust counter part by solving the inner problem in closed form, and present a convex optimization re-formulation.

We applied the framework to the US flood insurance and evaluate our performance against two baseline benchmarks: the US National Flood Insurnace Program (NFIP) premiums and CMA premiums. We employed historical data from 1975 to 2012 to construct the uncertainty sets and train machine learning models, and we used the last ten years of available NFIP data to evaluate RO insurance scheme against the two benchmarks: the actual historical premium policies and premiums derived from cumulative moving average (CMA) rules. We demonstrate the superiority of an RO approach in two metrics: effectiveness and equity. First, RO model is able to effectively cover losses, achieving a smooth transition from deficit to surplus balance depending on model parameter value, thus granting policy providers the discretion to determine the desired surplus level. Second, RO model is able to distribute risk in a more equitable manner, penalizing high risk states and charging premiums that more closely align with actual losses.

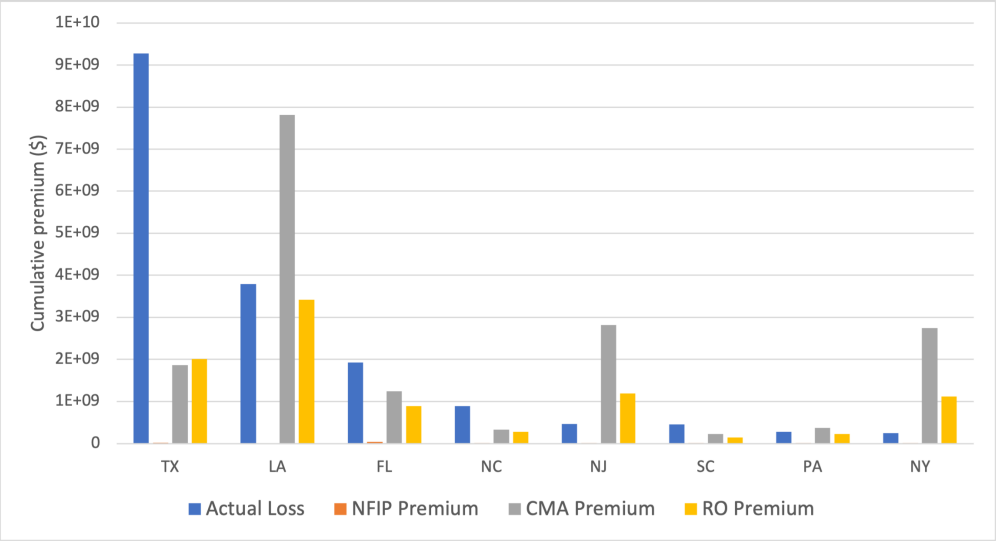


Figure 10: Cumulative claim losses and various insurance schemes across top 8 costly states during the testing period of 2013-2022. Blue bars represent the actual loss incurred, orange bars represent the NFIP premiums, gray bar represent CMA premiums, and yellow bars represent RO premiums with $\gamma_2 = 0.8$.

Finally, we emphasize the versatility and broad applicability of our framework. This underscores the possibility of employing the RO approach not just in the context of flood insurance, but also in pricing various catastrophic events such as wildfires, droughts, and other extreme weather conditions. The adaptability of our model suggests it could be a valuable tool in diverse scenarios, offering insights into risk assessment and pricing strategies across different disaster types.

References

[1] EM-DAT The International Disaster Dataset, 2021.

[2] Douglas Alem, Alistair Clark, and Alfredo Moreno. Stochastic network models for logistics planning in disaster relief. *European Journal of Operational Research*, 255(1):187–206, 2016.

[3] Jun Kyung Auh, Jaewon Choi, Tatyana Deryugina, and Tim Park. Natural Disasters and Municipal Bonds, July 2022.

[4] Aharon Ben-Tal, Byung Do Chung, Supreet Reddy Mandala, and Tao Yao. Robust optimization for emergency logistics planning: Risk mitigation in humanitarian relief supply chains. *Transportation research part B: methodological*, 45(8):1177–1189, 2011.

[5] Dimitris Bertsimas, David B. Brown, and Constantine Caramanis. *Robust Optimization*. Princeton University Press, 2011.

[6] Dimitris Bertsimas and Dick den Hertog. *Robust and Adaptive Optimization*. 2022.

[7] Laurens M. Bouwer and Jeroen C.J.H. Aerts. Financing climate change adaptation. *Disasters*, 30(1):49–63, 2006. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9523.2006.00306.x.

[8] T. Ermolieva, T. Filatova, Y. Ermoliev, M. Obersteiner, K. M. de Bruijn, and A. Jeuken. Flood Catastrophe Model for Designing Optimal Flood Insurance Program: Estimating Location-Specific Premiums in the Netherlands. *Risk Analysis*, 37(1):82–98, 2017. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/risa.12589.

[9] Tatiana Ermolieva and Yuri Ermoliev. Modeling catastrophe risk for designing insurance systems. *Integrated Catastrophe Risk Modeling: Supporting Policy Processes*, pages 29–52, 2013.

[10] Federal Emergency Management Agency (FEMA). OpenFEMA Dataset: FEMA NFIP Redacted Claims - v2, 2023. data retrieved on March 1 2023, `https://www.fema.gov/openfema-data-page/fima-nfip-redacted-claims-v2`.

[11] Federal Emergency Management Agency (FEMA). OpenFEMA Dataset: FIMA NFIP Redacted Policies - v1, 2023. data retrieved on March 1 2023, `https://www.fema.gov/openfema-data-page/fima-nfip-redacted-policies-v1`.

[12] Christopher Flavelle, Jill Cowan, and Ivan Penn. Climate Shocks Are Making Parts of America Uninsurable. It Just Got Worse. *The New York Times*, May 2023.

[13] Jacob Fowles, Gao Liu, and Cezar Brian Mamaril. Accounting for Natural Disasters: The Impact of Earthquake Risk on California Municipal Bond Pricing. *Public Budgeting & Finance*, 29(1):68–83, 2009. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-5850.2009.00924.x.

[14] Patricia Grossi. *Catastrophe modeling: a new approach to managing risk*, volume 25. Springer Science & Business Media, 2005.

[15] Stefan Hochrainer-Stigler, Reinhard Mechler, Georg Pflug, and Keith Williges. Funding public adaptation to climate-related disasters. Estimates for a global fund. *Global Environmental Change*, 25:87–96, March 2014.

[16] Harrison Hong, G Andrew Karolyi, and José A Scheinkman. Climate Finance. *The Review of Financial Studies*, 33(3):1011–1023, March 2020.

[17] Diane P Horn and Jared T Brown. *Introduction to the national flood insurance program (nfip)*. Congressional Research Service Washington DC, USA, 2017.

[18] Razmig Keucheyan. Insuring climate change: new risks and the financialization of nature. *Development and Change*, 49(2):484–501, 2018.

[19] Howard Kunreuther, Geoffrey Heal, Myles Allen, Ottmar Edenhofer, Christopher B. Field, and Gary Yohe. Risk management and climate change. *Nature Climate Change*, 3(5):447–450, May 2013. Number: 5 Publisher: Nature Publishing Group.

[20] Joanne Linnerooth-Bayer and Stefan Hochrainer-Stigler. Financial instruments for disaster risk management and climate change adaptation. *Climatic Change*, 133:85–100, 2015.

[21] Tongxin Liu, Jianfang Shao, and Xihui Wang. Funding allocations for disaster preparation considering catastrophe insurance. *Socio-Economic Planning Sciences*, 84:101413, 2022.

[22] Jessica Mercer. Disaster risk reduction or climate change adaptation: Are we reinventing the wheel? *Journal of International Development: The Journal of the Development Studies Association*, 22(2):247–264, 2010.

[23] Erwann Michel-Kerjan and Howard Kunreuther. Redesigning Flood Insurance. *Science*, 333(6041):408–409, July 2011. Publisher: American Association for the Advancement of Science.

[24] Erwann O Michel-Kerjan. Catastrophe economics: the national flood insurance program. *Journal of economic perspectives*, 24(4):165–186, 2010.

[25] Evan Mills. Synergisms between climate change mitigation and adaptation: an insurance perspective. *Mitigation and Adaptation Strategies for Global Change*, 12(5):809–842, June 2007.

[26] Javier Salmerón and Aruna Apte. Stochastic optimization for natural disaster asset prepositioning. *Production and operations management*, 19(5):561–574, 2010.

[27] Philip J Schneider and Barbara A Schauer. Hazus—its development and its future. *Natural Hazards Review*, 7(2):40–44, 2006.

[28] Maarten K Van Aalst. The impacts of climate change on the risk of natural disasters. *Disasters*, 30(1):5–18, 2006.

[29] Duo Wang, Kai Yang, and Lixing Yang. Risk-averse two-stage distributionally robust optimisation for logistics planning in disaster relief management. *International Journal of Production Research*, 61(2):668–691, 2023.

[30] Weichao Yang, Kui Xu, Chao Ma, Jijian Lian, Xuelian Jiang, Yadong Zhou, and Lingling Bin. A novel multi-objective optimization framework to allocate support funds for flash flood reduction based on multiple vulnerability assessment. *Journal of Hydrology*, 603:127144, 2021.

[31] Shiva Zokaee, Ali Bozorgi-Amiri, and Seyed Jafar Sadjadi. A robust optimization model for humanitarian relief chain design under uncertainty. *Applied Mathematical Modelling*, 40(17-18):7996–8016, 2016.

A Detailed Losses at State Level

States	Actual Loss	Historical	CMA	RO
TX	9.277e9	2.573e7	2.013e9	1.869e9
LA	3.790e9	1.472e7	3.423e9	7.809e9
FL	1.931e9	4.797e7	8.891e8	1.243e9
NC	8.943e8	7.608e6	2.805e8	3.351e8
NJ	4.610e8	3.892e6	1.187e9	2.819e9
SC	4.543e8	2.298e6	1.454e8	2.277e8
PA	2.742e8	5.312e6	2.225e8	3.736e8
NY	2.494e8	8.809e6	1.113e9	2.749e9
MO	2.456e8	8.216e5	1.318e8	1.858e8
AL	2.441e8	9.455e5	1.991e8	3.622e8
IL	1.751e8	5.969e6	1.017e8	1.275e8
GA	1.418e8	1.796e6	7.459e7	1.161e8
TN	1.376e8	3.703e5	6.861e7	2.688e8
MS	1.349e8	2.587e6	5.460e8	1.385e9
VA	1.263e8	5.028e6	1.338e8	2.206e8
CA	1.223e8	9.417e6	1.109e8	1.647e8
KY	9.890e7	1.335e6	6.048e7	8.834e7
OK	8.995e7	3.414e5	3.730e7	4.490e7
MI	8.985e7	9.781e5	1.683e7	2.400e7
MA	8.966e7	1.909e6	7.601e7	1.199e8
AR	8.253e7	3.807e5	2.796e7	3.477e7
OH	7.341e7	9.350e5	5.855e7	7.911e7
CO	6.864e7	3.352e5	1.646e7	3.121e7
WA	6.507e7	8.077e5	5.142e7	8.058e7
WV	6.214e7	2.652e5	5.563e7	7.825e7
IA	6.199e7	3.615e6	5.754e7	1.129e8
IN	5.729e7	3.964e5	5.163e7	8.820e7
NE	5.376e7	4.980e5	1.070e7	1.553e7
MD	5.120e7	1.184e6	5.531e7	1.146e8
CT	4.416e7	1.571e6	1.038e8	2.159e8
HI	4.082e7	1.194e6	2.016e7	3.140e7
WI	3.647e7	2.852e5	1.842e7	3.221e7
PR	2.513e7	6.745e4	1.386e7	1.852e7
KS	2.335e7	2.326e5	1.710e7	2.435e7
AZ	2.241e7	1.788e6	9.509e6	1.073e7
SD	1.955e7	3.187e4	9.313e6	1.468e7
VI	1.655e7	6.708e4	7.506e6	9.855e6
DE	1.236e7	9.825e5	1.401e7	2.327e7
MN	1.127e7	3.031e5	2.943e7	5.342e7
ME	9.955e6	2.126e5	5.587e6	8.210e6
NM	9.583e6	6.783e5	1.396e6	2.235e6
RI	8.369e6	3.438e5	1.791e7	3.378e7
MT	7.249e6	1.210e5	1.759e6	2.832e6
AK	5.741e6	3.131e5	6.964e6	1.167e7
ND	5.695e6	5.272e5	2.635e6	4.303e6
NV	5.675e6	3.960e5	5.248e6	8.800e6
UT	5.536e6	3.584e5	6.061e6	9.523e6
NH	5.270e6	1.757e5	9.513e6	1.564e7
VT	3.510e6	1.014e5	2.708e6	4.024e6
ID	3.413e6	1.386e5	3.045e6	4.747e6
DC	3.019e6	2.712e5	2.157e6	3.180e6
GU	2.693e6	1.899e4	1.083e6	1.464e6
WY	9.151e5	3.269e4	4.766e5	6.040e6
Total	1.991e10	1.669e8	1.161e10	2.193e10

Table 7: Cumulative premium collected based on different regimes of the 2013-2022 testing period. The table is sorted by the total actual loss from the most experienced to the least in the period. "Actual Loss" represents the actual loss during those years, "Historical" represents the historical policy premium collected, "CMA" shows results from using the CMA baseline policy, and "RO" displays results from implementing

States	Actual Loss	Histocial	CMA	RO	RO+ML
TX	9.277e9	-9.251e9	-7.264e9	-7.408e9	-7.408e9
LA	3.790e9	-3.775e9	-3.669e8	4.019e9	4.019e9
FL	1.931e9	-1.883e9	-1.042e9	-6.887e8	-6.886e8
NC	8.943e8	-8.867e8	-6.138e8	-5.592e8	-5.592e8
NJ	4.610e8	-4.571e8	7.260e8	2.358e9	2.358e9
SC	4.543e8	-4.520e8	-3.089e8	-2.266e8	-2.266e8
PA	2.742e8	-2.689e8	-5.165e7	9.943e7	9.945e7
NY	2.494e8	-2.406e8	8.634e8	2.499e9	2.499e9
MO	2.456e8	-2.448e8	-1.138e8	-5.981e7	-5.979e7
AL	2.441e8	-2.432e8	-4.498e7	1.180e8	1.180e8
IL	1.751e8	-1.691e8	-7.341e7	-4.761e7	-4.759e7
GA	1.418e8	-1.400e8	-6.723e7	-2.575e7	-2.572e7
TN	1.376e8	-1.372e8	-6.896e7	7.531e6	1.312e8
MS	1.349e8	-1.324e8	4.111e8	1.250e9	1.250e9
VA	1.263e8	-1.213e8	7.494e6	9.425e7	9.428e7
CA	1.223e8	-1.129e8	-1.142e7	4.239e7	4.242e7
KY	9.890e7	-9.756e7	-3.841e7	-1.058e7	-1.055e7
OK	8.995e7	-8.961e7	-5.265e7	-4.508e7	-4.505e7
MI	8.985e7	-8.887e7	-7.302e7	-7.602e7	-6.585e7
MA	8.966e7	-8.776e7	-1.365e7	3.019e7	3.021e7
AR	8.253e7	-8.215e7	-5.457e7	-4.784e7	-4.776e7
OH	7.341e7	-7.248e7	-1.486e7	5.672e6	5.698e6
CO	6.864e7	-6.830e7	-5.218e7	-6.506e7	-3.743e7
WA	6.507e7	-6.427e7	-1.366e7	1.548e7	1.551e7
WV	6.214e7	-6.187e7	-6.508e6	1.609e7	1.612e7
IA	6.198e7	-5.837e7	-4.441e6	5.086e7	5.088e7
IN	5.729e7	-5.689e7	-5.663e6	3.088e7	3.091e7
NE	5.376e7	-5.327e7	-4.306e7	-4.212e7	-3.823e7
MD	5.120e7	-5.001e7	4.115e6	6.341e7	6.344e7
CT	4.416e7	-4.259e7	5.960e7	1.717e8	1.717e8
HI	4.082e7	-3.962e7	-2.066e7	-9.461e6	-9.420e6
WI	3.647e7	-3.618e7	-1.804e7	-6.703e6	-4.262e6
PR	2.513e7	-2.506e7	-1.126e7	-6.646e6	-6.611e6
KS	2.335e7	-2.311e7	-6.249e6	9.614e5	9.983e5
AZ	2.241e7	-2.062e7	-1.290e7	-1.176e7	-1.168e7
SD	1.955e7	-1.952e7	-1.024e7	-4.915e6	-4.868e6
VI	1.655e7	-1.648e7	-9.041e6	-6.723e6	-6.692e6
DE	1.236e7	-1.138e7	1.645e6	1.088e7	1.091e7
MN	1.127e7	-1.097e7	1.816e7	4.209e7	4.212e7
OR	1.063e7	-8.655e6	7.707e6	2.158e7	2.162e7
NV	5.718e6	-5.595e6	1.454e6	7.741e6	7.773e6
NM	5.298e6	-5.012e6	-2.121e6	-3.276e5	8.948e6
NH	5.060e6	-4.842e6	4.723e6	1.135e7	1.138e7
AK	4.877e6	-4.816e6	-2.819e6	-2.836e6	1.481e7
ME	4.615e6	-4.444e6	3.594e6	8.166e6	8.197e6
ID	4.202e6	-3.980e6	-2.626e6	-2.073e6	7.200e5
VT	3.482e6	-3.401e6	9.952e6	3.107e7	3.112e7
RI	3.384e6	-2.855e6	1.924e7	4.146e7	4.148e7
ND	2.975e6	-2.753e6	5.273e7	1.087e8	1.087e8
MT	2.026e6	-1.899e6	3.045e5	2.222e6	9.335e6
UT	1.604e6	-1.573e6	-3.656e5	6.087e5	4.930e6
DC	1.586e6	-1.574e6	-7.791e5	-1.206e5	-1.889e5
WY	9.151e5	-8.824e5	-4.385e5	-1.475e5	5.125e6
Total	1.991e10	-1.975e10	-8.306e9	1.805e9	2.020e9

Table 8: Table displaying the surplus (or loss) over the testing period across all states. Surplus is computed using different premiums collected subtracted by the actual loss. Table is ranked from the highest cumulative cost to the lowest in the state level.