Agricultural Index Insurance: An Optimization Approach

José I. Velarde Morales   Linwei Xin

University of Chicago
Booth School of Business

May 22, 2023

The Problem of Agricultural Risk

- Farmers face a lot of risk, and the lack of risk management tools forces them to use coping strategies that hurt their long term welfare.

- Traditional insurance is prohibitively costly in most developing countries due to lack of data and high verification costs.

- Moral hazard, adverse selection, and the presence of large covariate shocks make the problem of agricultural insurance especially hard.

A Proposed Solution: Index Insurance

- In index insurance, an index (or statistic) is created using easily observable quantities (e.g. rainfall), and it is used to determine whether the insured party suffered an adverse event.

- If the index falls below a pre-determined threshold, the insurance company automatically issues out payments to the insured.

- This allows the insurance company to circumvent the issue of verification and moral hazard, since the actions of individual farmers cannot affect the index.

Index Insurance in Practice

- Since it was first proposed, index insurance programs have been implemented in many countries including India, Mexico, Tanzania, Malawi, Kenya, and many others (Jensen and Barrett (2017)).

- Today, tens of millions of farmers worldwide are covered by index insurance programs (Greatrex et al. (2015)).

- However, in most of these cases, the insurance has to be heavily subsidized by governments due to high cost and low demand (Greatrex et al. (2015)).

Project Overview

- Traditionally, the contract for each insured zone is designed independently of all other zones.

- The goal of this project is to make insurance less costly by improving the design of the insurance contracts.

- Our method simultaneously determines the contract parameters for different areas, while taking into account the correlation between the areas, reducing risk for the insurer.

Index Insurance Literature

- **Impact of Index Insurance:** Overall, there is evidence that index insurance reduces reliance on detrimental risk-coping strategies, increases investment, and leads to riskier, but more profitable production decisions (Jensen and Barrett (2017); Cole et al. (2013); Mobarak and Rosenzweig (2013); Karlan et al. (2014)).

- **Demand for Index Insurance:** Demand for index insurance tends to be low and highly price sensitive (Jensen and Barrett (2017); Cole et al. (2013); Cai, De Janvry, and Sadoulet (2020),Casaburi and Willis (2018)).

- **Design of Index Insurance:** There has been relatively little research studying the design of index insurance. The method developed by Chantarat et al. (2013) is the most commonly used in academic publications (Jensen, Stoeffler, et al. (2019); Flatnes, Carter, and Mercovich (2018)).

Content

Index Insurance: Design

1. Prediction: building a model to predict loss.

2. Contract design: designing contracts specifying payouts based on model predictions.

3. Pricing: product pricing

Index Insurance: Definition and Parameters

- Index insurance uses a signal, θ, to predict agricultural loss, $\hat{\ell}(\theta)$

- Contract form: $I(\theta) = \min\left\{\max\left\{a\hat{\ell}(\theta) - b, 0\right\}, 1\right\}$, a, b are the contract parameters.

- Expected cost for insurer: $C(I(\theta)) = \mathbb{E}[I(\theta)] + c_\kappa K(I(\theta))$, where c_κ is the cost of capital, and K is required capital.

- $K(I(\theta)) = CVaR_{1-\epsilon_P}(I(\theta)) - \mathbb{E}[I(\theta)]$.

Example of Index Insurance Contract
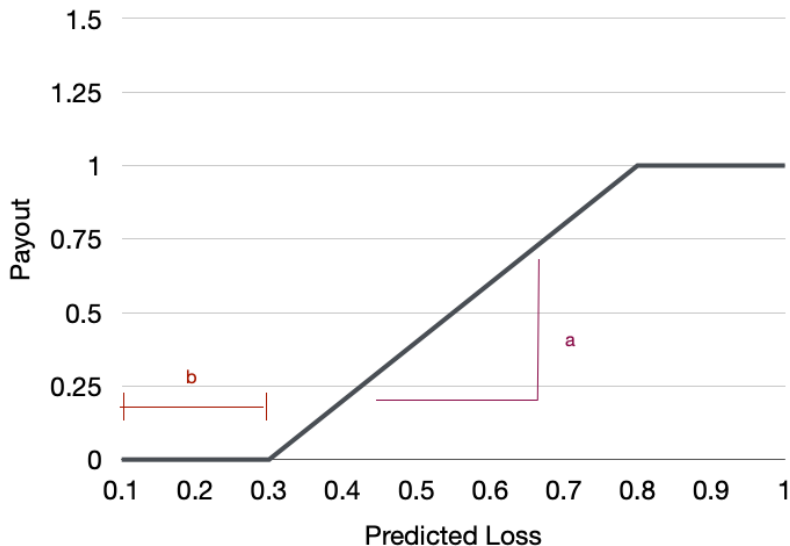
Practitioner Interviews

- We conducted interviews with researchers and practitioners that had implemented index insurance programs in several countries (Malawi, Kenya, Senegal, Thailand, among others) to learn more about the context.

- **Objective:** minimize risk faced by farmers

- **Constraints:** Price constraints and payout frequency

Risk Measures

Definition

For a random variable z, representing loss, the $(1-\epsilon)$ Value at Risk (VaR) is given by

$$VaR_{1-\epsilon}(z) := \inf\{t : P(z \leq t) \geq 1 - \epsilon\}$$

Definition

For a random variable z, representing loss, the $(1-\epsilon)$ Conditional Value at Risk (CVaR) is given by

$$CVaR_{1-\epsilon}(z) := \mathbb{E}[z | z \geq VaR_{1-\epsilon}(z)]$$

$$\min_{a,b,\pi,K} \quad \text{CVaR}_{1-\epsilon}\left(\ell + \pi - I(\theta)\right)$$

$$\text{s.t.} \quad I(\theta) = \min\left\{\max\left\{0, a\hat{\ell}(\theta) - b\right\}, 1\right\} \tag{1}$$

$$\pi = \mathbb{E}\left[I(\theta)\right] + c_\kappa K \tag{2}$$

$$K = \text{CVaR}_{1-\epsilon_K}\left(I(\theta)\right) - \mathbb{E}[I(\theta)] \tag{3}$$

$$\underline{f} \leq \mathbb{P}\left(I(\theta) > 0\right) \leq \overline{f} \tag{4}$$

$$\pi \leq \overline{\pi}. \tag{5}$$

Idealized CVaR Model

$$\min_{a,b,\pi,K} \quad \text{CVaR}_{1-\epsilon}\left(\ell + \pi - I(\theta)\right)$$

$$\text{s.t.} \quad I(\theta) = \min\left\{\max\left\{0, a\hat{\ell}(\theta) - b\right\}, 1\right\} \tag{6}$$

$$\pi = \mathbb{E}\left[I(\theta)\right] + c_\kappa K \tag{7}$$

$$K = \text{CVaR}_{1-\epsilon_K}\left(I(\theta)\right) - \mathbb{E}[I(\theta)] \tag{8}$$

$$\underline{f} \le \mathbb{P}\left(I(\theta) > 0\right) \le \overline{f} \tag{9}$$

$$\pi \le \overline{\pi}. \tag{10}$$

Multiple Zone

$$\min_{a,b,K} \max_{z} \quad \mathrm{CVaR}_{1-\epsilon}\left(\ell_z + \pi_z - I_z(\theta_z)\right) \tag{11}$$

$$\text{s.t.} \quad I_z(\theta_z) = \min\left\{\max\left\{a_z\hat{\ell}_z(\theta_z) + b_z, 0\right\}, 1\right\} \tag{12}$$

$$\pi_z = \mathbb{E}\left[I_z(\theta_z)\right] + \frac{c_\kappa}{\sum_z s_z}K \tag{13}$$

$$K = \mathrm{CVaR}_{1-\epsilon_K}\left(\sum_z s_z I_z(\theta_z)\right) - \mathbb{E}\left[\sum_z s_z I_z(\theta_z)\right] \tag{14}$$

$$\underline{f} \le \mathbb{P}\left(I_z(\theta_z) > 0\right) \le \overline{f} \tag{15}$$

$$\pi_z \le \overline{\pi_z} \tag{16}$$

Multiple Zone

$$\min_{a,b,K} \max_{z} \quad \text{CVaR}_{1-\epsilon} \left(\ell_z + \pi_z - I_z(\theta_z) \right) \tag{17}$$

$$\text{s.t.} \quad I_z(\theta_z) = \min \left\{ \max \left\{ a_z \hat{\ell}_z(\theta_z) + b_z, 0 \right\}, 1 \right\} \tag{18}$$

$$\pi_z = \mathbb{E}\left[I_z(\theta_z) \right] + \frac{c_\kappa}{\sum_z s_z} K \tag{19}$$

$$K = \text{CVaR}_{1-\epsilon_K} \left(\sum_z s_z I_z(\theta_z) \right) - \mathbb{E}\left[\sum_z s_z I_z(\theta_z) \right] \tag{20}$$

$$\underline{f} \leq \mathbb{P}\left(I_z(\theta_z) > 0 \right) \leq \overline{f} \tag{21}$$

$$\pi_z \leq \overline{\pi_z} \tag{22}$$

Convex Relaxation

$$\min_{a,b,K,\pi} \max_z \quad \text{CVaR}_{1-\epsilon}\left(\ell_z + \pi_z - \underline{I_z(\theta_z)}\right) \tag{23}$$

$$\text{s.t.} \quad \pi_z = \mathbb{E}\left[\overline{I_z(\theta_z)}\right] + \frac{c_\kappa}{\sum_z s_z} K \tag{24}$$

$$K = \text{CVaR}_{1-\epsilon_K}\left(\sum_z s_z \overline{I_z(\theta_z)}\right) - \mathbb{E}\left[\sum_z s_z \underline{I_z(\theta_z)}\right] \tag{25}$$

$$a\hat{\ell}_{\underline{f}} \le b \le a\hat{\ell}_{\overline{f}} \tag{26}$$

$$\overline{I_z(\theta_z)} = \max\left\{0, a_z\hat{\ell}_z(\theta_z) + b_z\right\} \tag{27}$$

$$\underline{I_z(\theta_z)} = \min\left\{a_z\hat{\ell}_z(\theta_z) + b_z, 1\right\} \tag{28}$$

$$\pi_z \le \overline{\pi_z}.$$

LP Reformulation

Using the results from Rockafellar and Uryasev (2002), we get:

$$\min_{a,b,\alpha,\omega,\gamma,t,m,K,\pi} \quad m$$

$$\text{s.t.} \quad t_z + \frac{1}{\epsilon} \sum_j p^j \gamma_z^j \leq m, \forall z$$

$$\gamma_z^j \geq s_z \left(\ell^j + \pi_z - \omega_z^j \right) - t_z, \forall j, \forall z$$

$$\gamma_z^j \geq 0, \forall j, \forall z$$

$$\pi_z = \frac{1}{N} \sum_j \alpha_z^j + \frac{1}{\sum_z s_z} c_\kappa K$$

$$t_K + \frac{1}{\epsilon_K} \sum_j p^j \gamma_K^j \leq K + \frac{1}{N} \sum_j \sum_z s_z \omega_z^j$$

$$\gamma_K^j \geq \sum_z s_z \alpha_z^j - t_K, \forall j$$

$$\gamma_K^j \geq 0, \forall j$$

$$\alpha_z^j \geq a_z \hat{\ell}_z(\theta_z^j) + b_z, \forall j, \forall z$$

$$\alpha_z^j \geq 0, \forall j, \forall z$$

$$\omega_z^j \leq a_z \hat{\ell}_z(\theta_z^j) + b_z, \forall j, \forall z$$

$$\omega_z^j \leq 1, \forall j, \forall z$$

$$\pi_z \leq \overline{\pi_z}, \forall z.$$

Content

Baseline Method

Baseline method: Chantarat et al. (2013). Common in academic publications, used to design Kenya's IBLI program. Method consists of:

1. Use linear model to predict loss.

2. Contracts are of the form: $I(\theta) = \times \min\left\{\max\left\{0, \hat{\ell}(\theta) - b\right\}, 1\right\}$. b is chosen to maximize the correlation of payouts and losses.

Content

What we wanted to test

- How is our method's performance affected by errors in the prediction model?

- How are costs affected by correlation between zones?

Setup

We use the following data generating process $\ell_z = \frac{1}{1+e^{f(\theta_z)}}$ with $f(\theta_z)$:

- Linear case: $f(\theta) = \beta\theta + \epsilon$
- Nonlinear case: $f(\theta) = \beta_0 + \beta_1\theta + \beta_2\theta^2 + \ldots + \beta_n\theta^n + \epsilon$

In both cases, $\theta \sim \mathcal{N}((0,0), \Sigma), \epsilon \sim \mathcal{N}(0, \rho I)$. In both cases, β_i is drawn randomly. Vary ρ to test effect of correlation.

Simulation Details

1. Generate training and test samples.

2. Use training data and model predictions to design contracts using both methods.

3. Apply insurance contracts designed by the two methods to farmers in the test set and compare outcomes.

We run this 200 times for every scenario we test.

Performance Metrics

For each sample in the test set, we calculate the net loss. Then we calculate the following metrics:

Farmer Welfare Measures	Insurer Cost/Risk Measures
Conditional Value at Risk (CVaR)	Average Cost
Value at Risk (VaR)	Required Capital
Semi-Variance	

Summary of Results: (Almost) Correctly Specified Model

Our model offers slightly worse protection than the baseline, but at a much lower cost and risk for the insurer. This is achieved through both, better risk management, and smarter allocation of payouts.

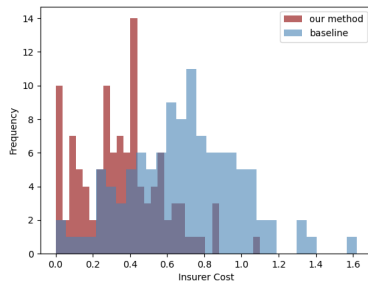- **Metrics on which our model does better:** Maximum Semi-Variance, Required Capital, Average Cost
- **Metrics on which baseline model does better:** Maximum CVaR
- **Metrics on which we tie:** Maximum VaR

Histograms: Model Correctly Specified

- Distribution of farmer loss is similar under both
- Cost for insurer is radically different
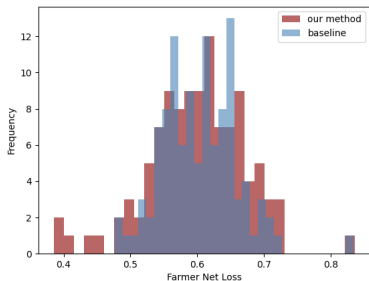
(a) Farmer net loss

(b) Insurer cost

Figure: Distribution of outcomes

Results: Correctly Specified Prediction Model

Model	Max CVaR	Max VaR	Max SemiVar	$|VaR_2 - VaR_1|$	Required Capital	Average Cost
Baseline	0.60	0.62	0.02	0.01	0.64	0.81
	[0.58, 0.63]	[0.6, 0.65]	[0.01, 0.04]	[0.0, 0.04]	[0.55, 0.79]	[0.65, 0.92]
Opt	0.64	0.63	0.01	0.01	0.55	0.48
	[0.61, 0.67]	[0.61, 0.66]	[0.01, 0.03]	[0.0, 0.03]	[0.43, 0.69]	[0.43, 0.54]
Diff	-0.04	-0.01	0.01	0.00	0.10	0.33
	[-0.07, -0.01]	[-0.02, 0.0]	[0.0, 0.02]	[-0.02, 0.03]	[0.05, 0.16]	[0.15, 0.43]

(a) No correlation

Model	Max CVaR	Max VaR	Max SemiVar	$|VaR_2 - VaR_1|$	Required Capital	Average Cost
Baseline	0.61	0.64	0.02	0.01	0.88	0.82
	[0.59, 0.65]	[0.62, 0.66]	[0.01, 0.03]	[0.0, 0.03]	[0.72, 1.1]	[0.68, 0.96]
Opt	0.65	0.65	0.01	0.01	0.74	0.52
	[0.63, 0.68]	[0.62, 0.67]	[0.0, 0.02]	[0.0, 0.03]	[0.6, 0.98]	[0.44, 0.59]
Diff	-0.03	-0.01	0.01	0.00	0.13	0.31
	[-0.06, -0.02]	[-0.02, 0.0]	[0.0, 0.01]	[-0.02, 0.02]	[0.09, 0.17]	[0.18, 0.43]

(b) Positive correlation

Results: Correctly Specified Prediction Model

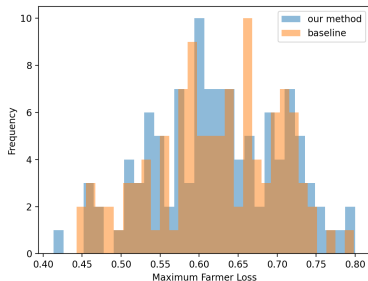| Model	Max CVaR	Max VaR	Max SemiVar	$|VaR_2 - VaR_1|$	Required Capital	Average Cost
Baseline	6.19	6.3	19.72	0.85	3.33	5.56
	[6.13, 6.26]	[6.24, 6.36]	[19.27, 20.17]	[0.79, 0.9]	[3.28, 3.38]	[5.48, 5.64]
Opt	6.11	5.98	15.09	0.22	2.91	5.53
	[6.06, 6.17]	[5.93, 6.02]	[14.68, 15.5]	[0.2, 0.23]	[2.86, 2.95]	[5.45, 5.61]
Diff	0.08	0.32	4.63	0.63	0.42	0.04
	[0.05, 0.11]	[0.29, 0.35]	[4.35, 4.9]	[0.57, 0.69]	[0.39, 0.45]	[0.03, 0.04]

Table: Negative correlation

Summary of Results: Misspecified Model

In this case, our results are even stronger. We match the performance of the basline method on most metrics, and the cost is significantly lower.

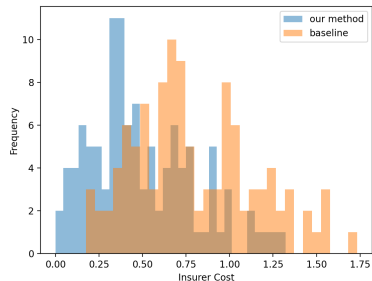- **Metrics on which our model does better:** Maximum Semi-Variance, Required Capital

- **Metrics on which baseline model does better:** Maximum CVaR

- **Metrics on which we tie:** Maximum VaR, Average Cost

Histograms: Model Misspecified

- Distribution of farmer loss is similar under both
- Cost for insurer is radically different

(a) Farmer net loss

(b) Insurer cost

Figure: Distribution of outcomes

Results: Misspecified Prediction Model

Model	Max CVaR	Max VaR	Max SemiVar	$|VaR_2 - VaR_1|$	Required Capital	Average Cost
Baseline	0.65	0.67	0.02	0.02	0.70	0.91
	[0.6, 0.84]	[0.61, 0.77]	[0.01, 0.08]	[0.0, 0.07]	[0.21, 0.93]	[0.7, 0.97]
Opt	0.71	0.67	0.03	0.02	0.61	0.56
	[0.64, 0.86]	[0.61, 0.77]	[0.01, 0.08]	[0.0, 0.08]	[0.0, 0.86]	[0.0, 0.74]
Diff	-0.04	-0.00	-0.00	-0.00	0.10	0.36
	[-0.1, 0.02]	[-0.02, 0.06]	[-0.03, 0.02]	[-0.03, 0.03]	[-0.03, 0.26]	[0.18, 0.81]

(a) No correlation

Model	Max CVaR	Max VaR	Max SemiVar	$|VaR_2 - VaR_1|$	Required Capital	Average Cost
Baseline	0.68	0.69	0.02	0.01	0.98	0.95
	[0.62, 0.85]	[0.63, 0.79]	[0.01, 0.05]	[0.0, 0.04]	[0.26, 1.21]	[0.59, 1.03]
Opt	0.73	0.69	0.03	0.02	0.83	0.57
	[0.66, 0.86]	[0.62, 0.76]	[0.01, 0.08]	[0.0, 0.05]	[0.0, 1.19]	[0.0, 0.8]
Diff	-0.04	0.00	-0.00	-0.00	0.13	0.36
	[-0.09, 0.01]	[-0.02, 0.07]	[-0.04, 0.01]	[-0.03, 0.02]	[-0.07, 0.38]	[0.14, 0.81]

(b) Positive correlation

Content

Data Sources

- **NDVI Data:** The Normalized Difference Vegetation Index (NDVI) is a satellite-based indicator of the amount and health of vegetation. We use NDVI data for Kenya between 2000-2015.

- **Kenya Household Survey Data:** This survey was conducted as part of the development of the Index based livestock insurance (IBLI) program in northern Kenya. This dataset has information on household location, livestock levels and changes for each month in 2010-2013. There are 900 households in this dataset.

Data Creation

1. Calculate NDVI metrics for each village in each season. These metrics will be the features for our predictive model.
2. Use household data to calculate average livestock mortality in each village in each season.
3. Merge the two resulting datasets to create a dataset for regression.
4. Train prediction model for each cluster and use this model to predict herd mortality at the village level.
5. Add model predictions to household data.

Evaluation Procedure

We use leave one out cross validation to evaluate our method. In each iteration, we leave out one year of data for testing.

1. Split data into training and test sets
2. Train prediction model on training data, create predictions for the test set
3. Use training data and model predictions to design contracts using both methods.
4. Apply insurance contracts designed by the two methods to farmers in the test set and compare outcomes.

Results

The insurance contracts developed by our model provide slightly better protection at a much lower cost. The cost of our contracts is 9% lower, and the cost of capital is 13% lower.

Model	Max CVaR	Max VaR	$|VaR_2 - VaR_1|$	Average Cost
Baseline	0.69	0.52	0.13	5263.64
Opt	0.65	0.53	0.10	3794.80

Table: Results using Kenya household data

Conclusions

- The contracts designed by our model are able to offer better protection at a similar costs, or comparable protection at lower costs than the baseline method.

- It outperforms the baseline when the prediction model is incorrectly specified and on the Kenyan pastoralist data.

- Our method is more cost effective because it takes into account spatial correlations between areas and the costs of capital requirements. Thus, the model makes better trade offs between costs and coverage than the baseline method.

Next Steps

- We are working with practitioners to improve the model and possibly test it in practice.

- We are working with the Bank of Thailand on the implementation of their satellite-based index insurance program.

- We are also talking to the International Research Institute for Climate and Society at Columbia, they have worked on the implementation of numerous index insurance programs in Africa.

References

Rockafellar, R Tyrrell and Stanislav Uryasev (2002). "Conditional value-at-risk for general loss distributions". In: *Journal of banking & finance* 26(7), pp. 1443–1471.

Chantarat, Sommarat et al. (2013). "Designing index-based livestock insurance for managing asset risk in northern Kenya". In: *Journal of Risk and Insurance* 80(1), pp. 205–237.

Cole, Shawn et al. (2013). "Barriers to household risk management: Evidence from India". In: *American Economic Journal: Applied Economics* 5(1), pp. 104–35.

Mobarak, Ahmed Mushfiq and Mark R Rosenzweig (2013). "Informal risk sharing, index insurance, and risk taking in developing countries". In: *American Economic Review* 103(3), pp. 375–80.

Karlan, Dean et al. (2014). "Agricultural decisions after relaxing credit and risk constraints". In: *The Quarterly Journal of Economics* 129(2), pp. 597–652.

Greatrex, Helen et al. (2015). "Scaling up index insurance for smallholder farmers: Recent evidence and insights". In: *CCAFS Report.*

Jensen, Nathaniel and Christopher Barrett (2017). "Agricultural index

Idealized CVaR Model

- **Objective:** conditional value at risk of the farmers' loss net of insurance.
- **Constraint 1:** piecewise linear structure of the contract.
- **Constraint 2:** budget constraint.
- **Constraint 3:** definition of required capital.

$$\min_{a,b,\pi,K} \quad CVaR_{1-\epsilon}\left(\ell - I(\theta)\right)$$

$$\text{s.t. } I(\theta) = \min\{(a\hat{\ell}(\theta) + b)^+, P\} \tag{29}$$

$$\mathbb{E}\left[I(\theta)\right] + c_\kappa K \leq B \tag{30}$$

$$K = \left(CVaR_{1-\epsilon}\left(I(\theta)\right) - \mathbb{E}[I(\theta)]\right) \tag{31}$$

The problem is non-convex, so we need convex approximations

We use the following approximations of $I(\theta)$ to make the problem convex:

$$\overline{I(\theta)} \triangleq \max\left\{0, a\hat{\ell}(\theta) + b\right\}$$
$$\underline{I(\theta)} \triangleq \min\{a\hat{\ell}(\theta) + b, K\}$$

- Note that $\overline{I(\theta)} \geq I(\theta)$ and $\overline{I(\theta)}$ is convex. Conversely, $\underline{I(\theta)} \leq I(\theta)$ and $\underline{I(\theta)}$ is concave.
- We replace $I(\theta)$ with either $\overline{I(\theta)}$ or $\underline{I(\theta)}$ where necessary to obtain conservative and convex approximations.
- We also need approximations or proxies for $E[I(\theta)]$ in constraint . We use $\pi_{SQ} = E[I_{SQ}(\theta)]$, where I_{SQ} is the contract designed using the status quo method, as a proxy for $E[I(\theta)]$ in constraint .

As ϵ gets smaller, the slope increases and the function shifts to the right.
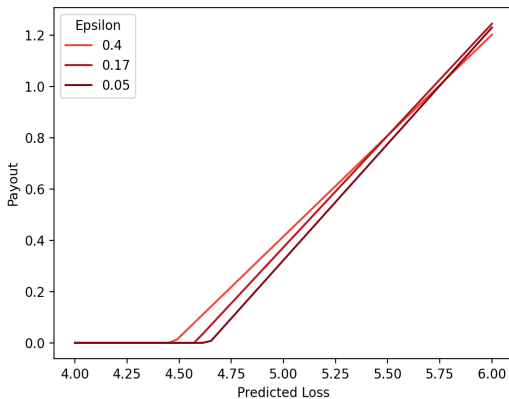
Results: Misspecified Prediction Model

Model	Max VaR	$|VaR_2 - VaR_1|$	Required Capital	Average Cost
Baseline	27.42	1.65	53.85	40.73
	[25.13, 29.57]	[0.16, 5.56]	[44.81, 59.88]	[36.06, 45.83]
Opt	27.41	1.96	49.97	40.36
	[24.53, 29.83]	[0.15, 5.6]	[42.87, 58.53]	[35.52, 45.5]

(a) No correlation

Model	Max VaR	$|VaR_2 - VaR_1|$	Required Capital	Average Cost
Baseline	27.16	1.23	57.1	41.36
	[24.7, 29.62]	[0.1, 5.05]	[51.16, 62.9]	[36.24, 46.52]
Opt	27.71	1.1	56.51	41.31
	[24.91, 30.31]	[0.09, 3.62]	[50.92, 62.26]	[35.98, 46.82]

(b) Positive correlation

Model	Max VaR	$|VaR_2 - VaR_1|$	Required Capital	Average Cost
Baseline	27.52	1.9	25.2	36.42
	[25.09, 29.69]	[0.17, 6.0]	[17.99, 36.35]	[33.13, 41.88]
Opt	27.39	1.84	26.77	36.67
	[24.33, 29.71]	[0.16, 5.49]	[18.17, 37.5]	[33.21, 42.03]

(c) Negative correlation