Agricultural Index Insurance: An Optimization Approach

José I. Velarde Morales   Linwei Xin

University of Chicago
Booth School of Business

February 7, 2025

Project Overview

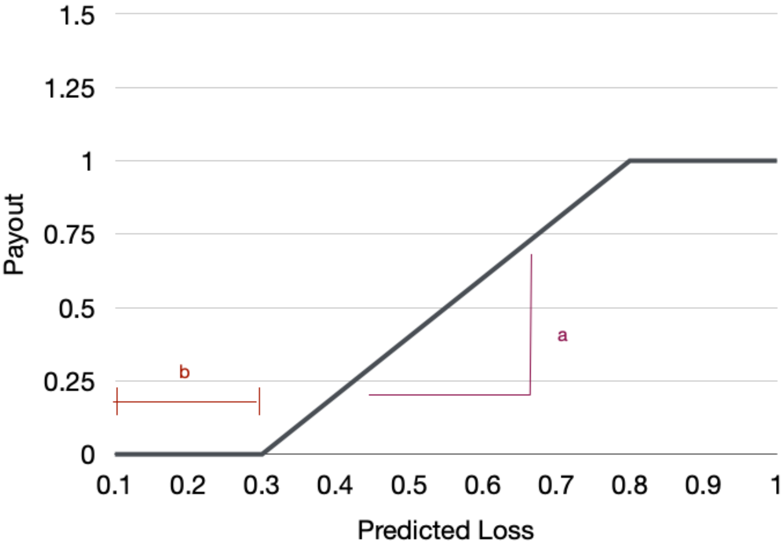
- The goal of this project is to improve the design of index insurance contracts. I am particularly interested in the developing country setting.
- The original motivation was to develop a method to simultaneously design contracts for all insured zones, in order to better manage risk. However, upon learning more about the context I realized that the optimization based approach could be an improvement even in the single zone case.
- We develop a method that tries to maximize farmer utility, incorporates different kinds of constraints, and yields interpretable contracts.

Content

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.

Example of Index Insurance Contract



Current Methods: Chen

- NN based approach, end-to-end, directly maps weather variables to payouts.
- **Pros:** directly maximizes utility, admits price constraints
- **Cons:** not interpretable, hard to adjust

Current Methods: Kenya ILRB

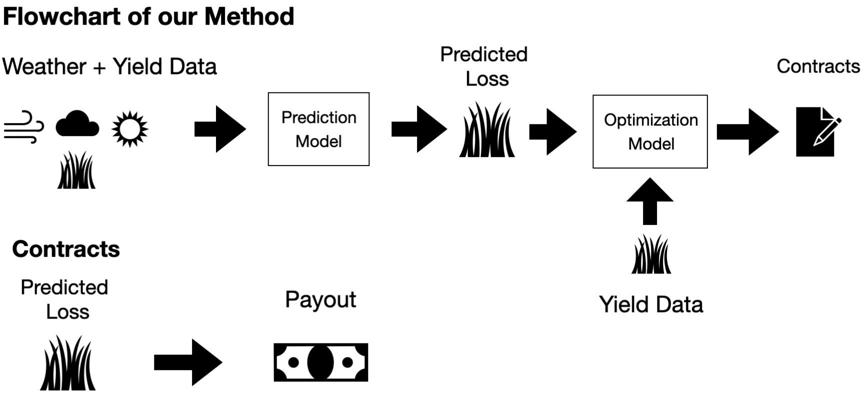
- Choose strike value that maximizes correlation between losses and payouts
- **Pros:** simple, interpretable
- **Cons:** does not admit price constraints.

Content

Overview

- We opt for a "predict-then-optimize" approach. We first train prediction models then design optimal contracts.
- **Pros:** interpretable results, maximizes utility, admits constraints.
- **Cons:** can overfit if there's insufficient data.

Our Method Flowchart



Model

Farmers start with wealth, w_0 and experience loss ℓ. There is an index insurance contract I, that is determined by a p dimensional vector of indices $\theta = (\theta_1, ..., \theta_p)$. Premium for contract I is $\pi(I)$. Farmer wealth is:

$$w = w_0 - \ell + I(\theta) - \pi(I)$$
$$\pi = \Pi(I)$$

Here, Π is the premium principle that is used to determine the insurance contract's price. There are several premium principles that are common in the actuarial literature, and our method is compatible with most of them:

Description	Definition
Expected value	$(1+\alpha)\mathbb{E}[I]$
Standard deviation	$\mathbb{E}[I] + \alpha\sigma(I)$
Variance	$\mathbb{E}[I] + \alpha\sigma^2(I)$
Exponential	$\frac{1}{\alpha}\log\mathbb{E}[e^{\alpha I}]$
Dutch	$\mathbb{E}[I] + \beta\mathbb{E}[(I - \alpha\mathbb{E}[I])_+]$

Idealized Model

$$
\begin{aligned}
\max_{a,b,\pi,K} \quad & \mathbb{E}\left[U\left(w_0 - \ell - \pi + I(\theta)\right)\right] && (1) \\
\text{s.t.} \quad & I(\theta) = \min\left\{\max\left\{0, a\hat{\ell}(\theta) + b\right\}, 1\right\} && (2) \\
& \pi = \mathbb{E}\left[I(\theta)\right] + c_\kappa\left[\mathrm{CVaR}_{1-\epsilon_K}\left(I(\theta)\right) - \mathbb{E}[I(\theta)]\right] && (3) \\
& \underline{f} \leq \mathbb{P}\left(I(\theta) > 0\right) \leq \overline{f} && (4) \\
& \pi \leq \overline{\pi}. && (5)
\end{aligned}
$$

Convex Relaxation

We use the following convex and concave relaxations of $I(\theta)$:

$$\overline{I(\theta)} \triangleq \max\left\{0, a\hat{\ell}(\theta) + b\right\}, \qquad \underline{I(\theta)} \triangleq \min\{a\hat{\ell}(\theta) + b, 1\}.$$

Note that $\overline{I(\theta)}$ is convex in $\hat{\ell}(\theta)$, $\underline{I(\theta)}$ is concave in $\hat{\ell}(\theta)$, and

$$\underline{I(\theta)} \leq I(\theta) \leq \overline{I(\theta)}.$$

We replace $I(\theta)$ in Problem (1) with either $\overline{I(\theta)}$ or $\underline{I(\theta)}$ where necessary to obtain a conservative and convex relaxation.

Convex Relaxation

$$
\begin{aligned}
\max_{a,b,K,\pi} \quad & \mathbb{E}\left[U\left(w_0 - \ell - \pi + \underline{I(\theta)}\right)\right] && (6)\\
\text{s.t.} \quad & \overline{I(\theta)} = \max\left\{0, a\hat{\ell}(\theta) + b\right\} \\
& \underline{I(\theta)} = \min\left\{a\hat{\ell}(\theta) + b, 1\right\} \\
& \pi = (1 - c_\kappa)\,\mathbb{E}\left[\overline{I(\theta)}\right] + c_\kappa \mathrm{CVaR}_{1-\epsilon_K}\left(\overline{I(\theta)}\right) && (7)\\
& a\hat{F}^{-1}(\overline{f}) \leq b \leq a\hat{F}^{-1}(\underline{f}) && (8)\\
& \pi \leq \overline{\pi}.
\end{aligned}
$$

Content

Data

Used two main data sources

- Illinois annual corn yield data from the National Agricultural Statistics Service (NASS). Data is available at the county level from 1925-2022. 84 counties.
- Weather data from the PRISM climate group. Has monthly data on several weather variables (temperature, precipitation, etc). Available 1895-present.

- We use a 70/15/15 train/val/test split. Data is kept in chronological order. Training data has older years and test data has the newest years.

- We use the training and validation data to design the contracts using both methods, apply the contracts to farmers in the test set, and compute performance metrics.

- We used a data shortening exercise to evaluate how the performance of both methods changed as more data became available.

Illinois Results: Short Datasets (Less than 40 years of data)

Method	DeltaCE	Premium	Insurer Cost
Our Method	0.05	62.49	68,890
Kenya ILRB	0.04	113.96	122,196
Chen uc	0.05	60.69	52,303

Illinois Results: Long Datasets (More than 40 years of data)

Method	DeltaCE	Premium	Insurer Cost
Our Method	0.04	41.75	45,494
Kenya ILRB	0.02	105.53	116,728
Chen uc	0.05	50.06	44,822

- Correlation of losses in the portfolio are important for insurers.
- We use a portfolio approach for premium pricing.
- We incorporate this premium pricing into the optmization model.

Multiple Zone Model

- Our method can be extended to design the contracts of multiple zones simultaneously.
- This allows it to take into account the correlation between the insured zones, allowing it to manage risk better.

Multiple Zone Model

$$\max_{a,b,K,\pi} \quad \mathbb{E}\left[\sum_z U\left(w_{0,z} - \ell_z^j - \pi_z + I_z(\theta_z^j)\right)\right] \tag{9}$$

$$\text{s.t.} \quad \pi_z = \mathbb{E}\left[\overline{I_z(\theta_z)}\right] + \frac{c_\kappa}{\sum_z s_z} K \tag{10}$$

$$K = \text{CVaR}_{1-\epsilon_K}\left(\sum_z s_z \overline{I_z(\theta_z)}\right) - \mathbb{E}\left[\sum_z s_z \underline{I_z(\theta_{z'})}\right] \tag{11}$$

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

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

$$\pi_z \leq \overline{\pi_z}.$$

Overview

- We evaluate our multiple zone model using data from Illinois, Indiana, Missouri, and Iowa. We examine the costs of insuring all four zones simultaneously using the three methods.
- Our multiple zone model adjusts contracts based on the correlation between the insured zones. In this case, it leads to contracts that pay out more frequently, but at a lower rate. This reduces the tail risk for the insurer and reduces the amount of capital needed.
- It outperforms Chen's model and the no insurance case consistently, and has lower costs and required capital than the Chen model.

Midwest Results: Short datasets (Less than 40 years)

Method	DeltaU	Insurer Cost	Required Capital
Our Method	0.04	150,414	6.6%
Kenya ILRB	0.06	397,474	4.4%
Chen uc	-0.09	159,928	12%

Midwest Results: Long datasets (More than 40 years)

Method	DeltaU	Insurer Cost	Required Capital
Our Method	0.03	99,769	5.6%
Kenya ILRB	0.02	389,678	4.3%
Chen uc	0.00	141,651	8.6%

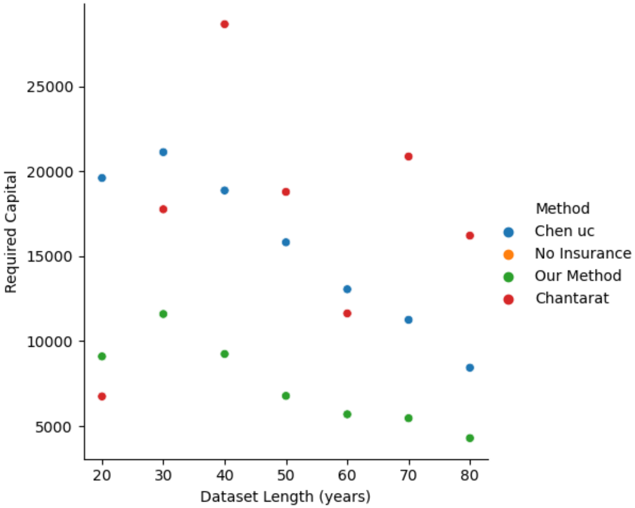
Data Description

- We use Tambon level loss data from the Department of Agricultural Extension and the BAAC
- Weather data from Google Earth Engine, we use rainfall, evapotranspiration, and temperature.

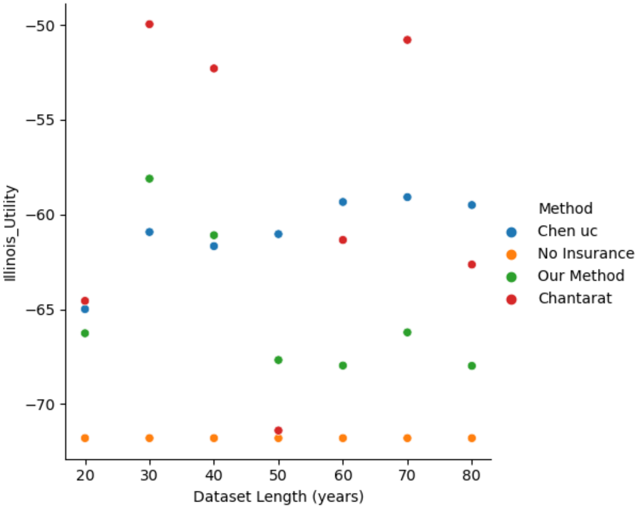
Results: Single Zone

Results: Multiple Zones

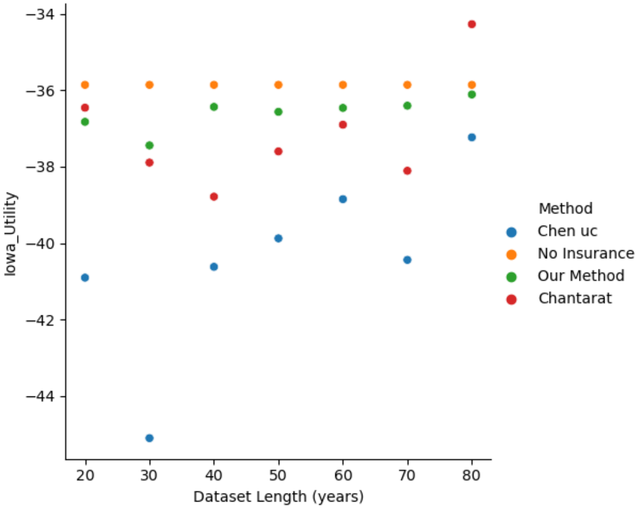
Midwest: Required Capital



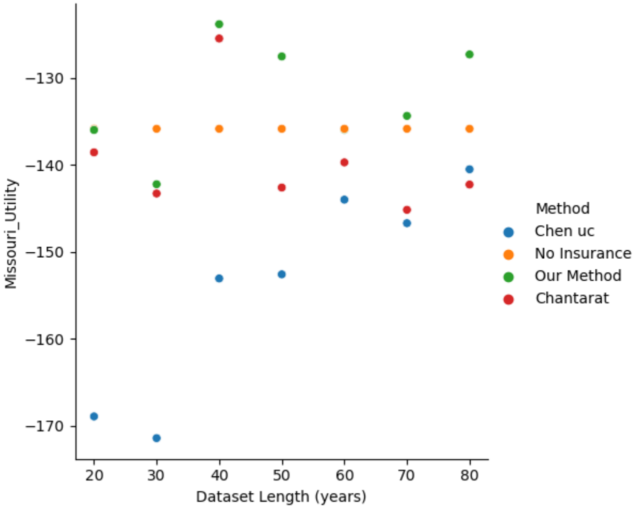
Illinois: Utility



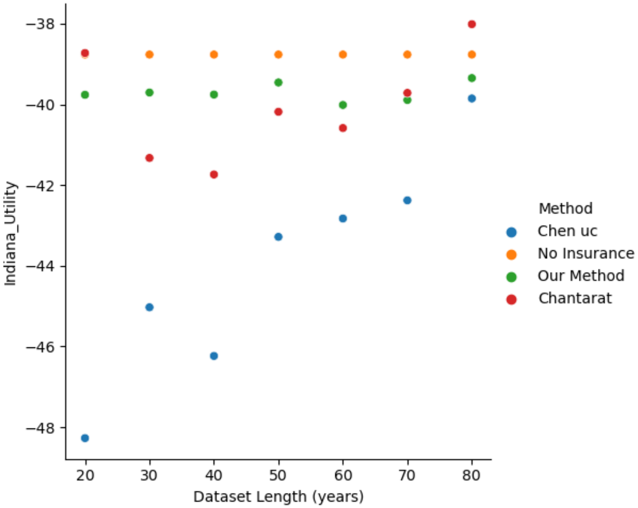
Iowa: Utility



Missouri Utility



Indiana Utility



Loss Definition

From what I can tell from the replication files, they seem to define loss in every year as:

$$\ell_{st} = R_s^* - R_{st}$$

where R_s^* is the maximum revenue observed in state s across all time periods, and R_{st} is the revenue in state s at time t.

Initial wealth

According to the paper, they set $w_0 = 389$. However, in the replication files, they set it to be $w_0 = 813 - 504 + 389$. According to the comments, 504 is the fixed cost of operating a farm, and there are no comments regarding the 813, I'm assuming it corresponds to R^*.

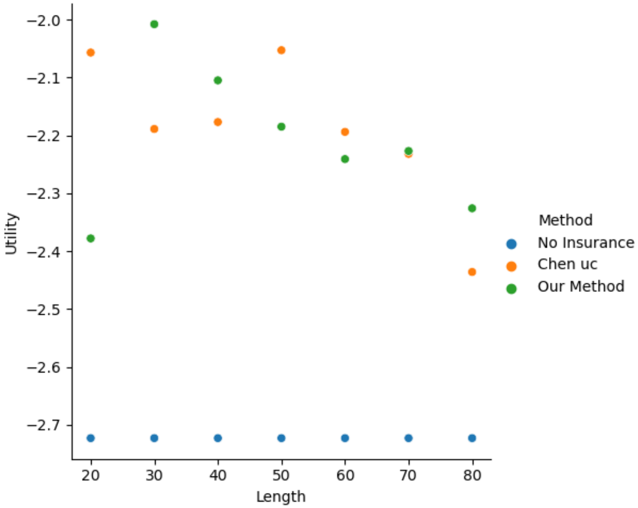
Detrending

- According to the paper, they detrend the county level yield data using a 2nd order polynomial fit with a "robust" regression method, but they don't specify what they use, and it's not in the replication files. They also don't specify if they remove the trend using additive or multiplicative decomposition model. Using an additive decomposition model yielded the most similar losses to what they provide in the replication files.
- There are a couple of papers showing that using locally weighted regression to detrend works better.

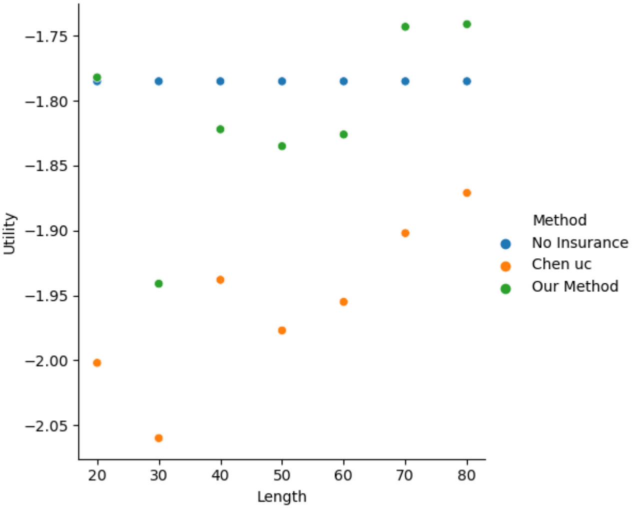
Questions

- Would it make sense to define loss as deviation from historical average? Allowing it to be positive in some years? In other words, we would first adjust all of the yield data to 2020 levels and then calculate the historical average. The loss in each year would be the deviation from this historical average.
- Should I simply follow their lead on detrending? Or should I try to improve on it?
- Do you think it's necessary to show results with both definitions of the premium?
- Do you think subsidy vs lump sum results would be interesting?

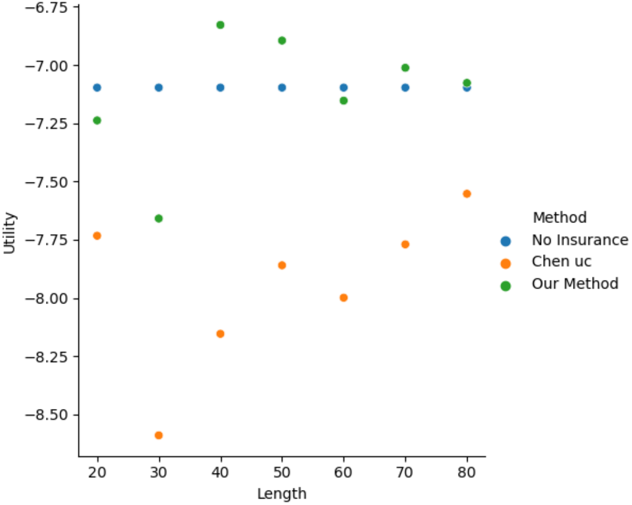
Their Defn of Premium: Illinois Utility



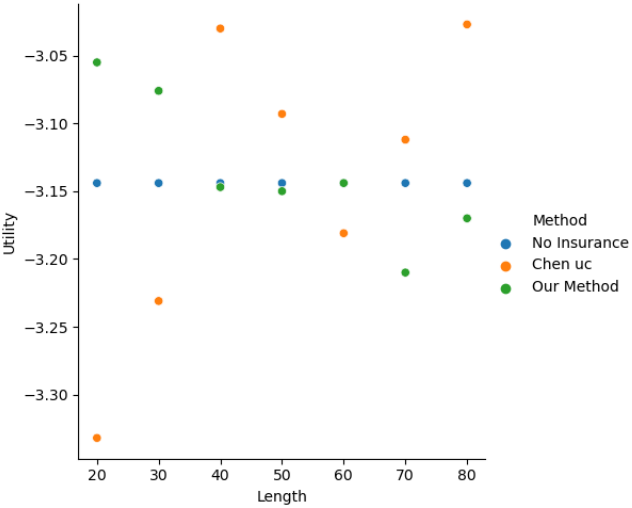
Their Defn of Premium: Iowa Utility



Their Defn of Premium: Missouri Utility



Their Defn of Premium: Indiana Utility



References

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.

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

$$\text{s.t. } I(\theta) = \min\{(a\hat{\ell}(\theta) + b)^{+}, P\} \tag{12}$$

$$\mathbb{E}\left[I(\theta)\right] + c_{\kappa}K \leq B \tag{13}$$

$$K = \left(CVaR_{1-\epsilon}\left(I(\theta)\right) - \mathbb{E}[I(\theta)]\right) \tag{14}$$

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 .