Agricultural Index Insurance Design: An Optimization Approach

José I. Velarde Morales

October 11, 2022

1 Introduction/Background

1.1 The Problem of Agricultural Risk

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

- Traditional agricultural 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 hard.

1.2 A Proposed Solution: Index Insurance

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

- Indices have been constructed using rainfall, weather, and satellite images. 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, moral hazard, and adverse selection, since the actions of individual farmers cannot affect the index.

1.3 Index Insurance in Practice

- Since it was first proposed, index insurance programs have been implemented in a variety of countries including India, Mexico, Tanzania, Malawi, Kenya, and many others.

- Today, tens of millions of farmers worldwide are covered by index insurance programs.

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

1.4 Literature Review

1.4.1 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)).

1.4.2 Optimization Literature

- We draw from the literature on chance constrained programs (Lagoa, Li, and Sznaier (2005); Charnes, Cooper, and Symonds (1958)).

- We also draw on the work on coherent risk measures (Artzner et al. (1999)), and work on the optimization of conditional value at risk by (Rockafellar, Uryasev, et al. (2000))

- Additionally, we use the results on convex approximations of chance constrained programs by (Nemirovski and Shapiro (2007)).

1.5 Project Overview

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

- Traditionally, the contract for each insured zone is designed independently of all other zones. This ignores valuable information about the correlation between zones that affects the cost of insuring the whole portfolio.

- Our proposed method improves the design of these contracts by simultaneously deciding the contract parameters for different areas, while taking into account the correlation between the areas.

- This allows it to make better trade offs between the benefits of more aggressive payouts, and the costs associated with capital requirements.

2 Optimization Approach

2.1 Index Insurance: Definition and Parameters

- Index insurance generally involves an easily observable signal, θ, that is used to predict the loss, $\hat{\ell}(\theta)$, of some agricultural product.

- Index insurance contracts normally have the form: $I(\hat{\ell}(\theta)) = \min\left\{\max\left\{0, a\hat{\ell}(\theta) + b\right\}, P\right\}$, where P is the maximum payout, and a, b are the contract parameters. For ease of notation, we will use $I(\theta)$ instead of $I(\hat{\ell}(\theta))$.

- The expected cost, $C(I(\theta))$ of an insurance contract, $I(\theta)$ for an insurer in a single period is: $C(I(\theta)) = \mathbb{E}[I(\theta)] + c_\kappa K(I(\theta))$, where c_κ is the cost of holding capital, and K is the amount of capital required to insure the contract.

- One commonly used formula for K is $K(I(\theta)) = CVaR_{1-\epsilon_P}\left(I(\hat{l}(\theta))\right) - \mathbb{E}[I(\theta)]$.

2.2 Risk Measures

We are interested in minimizing the risk faced by farmers, so we need a measure of this risk. We opt for the Conditional Value at Risk (CVaR) as a risk measure.

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

$$VaR_{1-\epsilon}(z) := \inf\left\{t : P(z \le t) \ge 1 - \epsilon\right\}$$

Definition 2 *For a random variable z, representing loss, the $(1 - \epsilon)$ Conditional Value at Risk (CVaR) is given by*

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

2.3 Idealized CVaR Model

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

$$\text{s.t. } I(\theta) = \min\{(a\hat{\ell}(\theta) + b)^+, P\} \tag{2}$$

$$\mathbb{E}\left[I(\theta)\right] + c_\kappa K \le B \tag{3}$$

$$K = \left(CVaR_{1-\epsilon}\left(I(\theta)\right) - \mathbb{E}[I(\theta)]\right) \tag{4}$$

2.4 Convex Approximations

We make the following convservative convex approximations to make the problem convex:

- We replace $I(\theta)$ in the objective with $\min\{a\hat{\ell}(\theta) + b, K\}$. This will give us an lower bound on the performance of our contracts, since $\ell - \min\{a\hat{\ell}(\theta) + b, K\} \ge \ell - \min\{(a\hat{\ell}(\theta) + b)^+, P\}$.

- We replace $I(\hat{\ell}(\theta))$ in the budget constraint with $\max\left\{0, a\hat{\ell}(\theta) + b\right\}$. This is a conservative approximation of the constraint, since $\min\{(a\hat{\ell}(\theta) + b)^+, P\} \le \max\left\{0, a\hat{\ell}(\theta) + b\right\}$.

- We also need approximations or proxies for $E[I(\theta)]$ in constraint 4. 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 4.

2.5 Model Parameters

- ϵ: This is the ϵ used for the $CVaR$ objective. $\epsilon = 0.1$ means that our objective is $E[\ell - I(\theta) | \ell - I(\theta) \ge VaR_{1-0.1}\left(\ell - I(\theta)\right)]$.

- ϵ_K: This is the epsilon used in the formula for required capital. Recall that the required capital $K(I(\theta)) = CVaR_{1-\epsilon_K}(I(\theta)) - E[I(\theta)]$.

- Z: number of insured zones.

- c_κ: cost of capital.

- P_z: maximum payout for zone z.

- π_{SQ}: This is the average payout made by the status quo contract, $E[I_{SQ}(\theta)]$.

2.6 Multiple Zone Model

$$\min_{a,b,K} \max_z \quad CVaR_{1-\epsilon}\left(\ell_z - \min\left\{(a_z\hat{\ell}_z(\theta_z) + b_z), P_z\right\}\right) \tag{5}$$

$$\text{s.t. } \mathbb{E}\left[\sum_z \max\left\{0, a_z\hat{\ell}_z(\theta_z) + b_z\right\}\right] + c_\kappa K \leq B \tag{6}$$

$$K + Z\pi_{SQ} \geq CVaR_{1-\epsilon_K}\left(\sum_z \max\left\{0, a_z\hat{\ell}_z(\theta_z) + b_z\right\}\right) \tag{7}$$

2.7 LP Reformulation

$$\min_{a,b,\alpha,\gamma,t,m,K^P} \quad m \tag{8}$$

$$\text{s.t.} \quad t_z + \frac{1}{\epsilon}\sum_j p^j\gamma_z^j \leq m, \forall z \tag{9}$$

$$\gamma_z^j \geq \ell^j - \min\left\{(a_z\hat{\ell}_z(\theta_z^j) + b_z), P_z\right\} - t_z, \forall j, \forall z \tag{10}$$

$$\gamma_z^j \geq 0, \forall j, \forall z \tag{11}$$

$$B \geq \frac{1}{N}\sum_j\sum_z \alpha_z^j + c_\kappa K \tag{12}$$

$$t_K + \frac{1}{\epsilon_K}\sum_j p^j\gamma_K^j \leq K + Z\pi_{SQ} \tag{13}$$

$$\gamma_K^j \geq \sum_z \alpha_z^j - t_K, \forall j \tag{14}$$

$$\gamma_K^j \geq 0, \forall j \tag{15}$$

$$\alpha_z^j \geq a_z\hat{\ell}_z(\theta_z^j) + b_z, \forall j, \forall z \tag{16}$$

$$\alpha_z^j \geq 0, \forall j, \forall z \tag{17}$$

3 Evaluation

3.1 Baseline Method

- This method is the most commonly used in academic publications (Chantarat et al. (2013),Flatnes, Carter, and Mercovich (2018),Jensen, Stoeffler, et al. (2019)) and is what is used in Kenya's IBLI program.

- Historical data is used to fit a linear regression model to predict losses in each insured area. A different model is estimated for each area. Each area normally has multiple villages in it.

- Contracts are of the form: $I(\theta) = \max(\hat{\ell}(\theta) - \ell^*, 0) \times TIU \times P_{IU}$ where $\hat{\ell}(\theta)$ is the predicted loss rate, ℓ^* is the strike value, TIU is the total number of insured agricultural units, and P_{AU} is the price per insured agricultural unit. In other words, their contract pays farmers for the full predicted loss beyond a threshold, ℓ^*. This threshold, ℓ^* is the contract's strike value. The strike value is chosen to minimize basis risk.

3.2 Synthetic Data Evaluation

3.2.1 Setup

We use the following DGP:

$$\ell = \beta\theta + \epsilon$$
$$\theta \sim \mathcal{N}((5,5), \Sigma), \beta = diag(1.5, 1.5)$$
$$\epsilon \sim \mathcal{N}(0, I)$$

And test the following scenarios

- **No correlation case:** $corr(\theta_1, \theta_2) = 0$

- **Positive correlation case:** $corr(\theta_1, \theta_2) = 0.8$

- **Negative correlation case:** $corr(\theta_1, \theta_2) = 0.8$

3.2.2 Evaluation Procedure

1. Draw training ($n = 300$) and test ($n = 100$) samples from model, samples will be of the form $\{\ell_1^i, \theta_1^i, \ell_2^i, \theta_2^i\}_{i=1}^N$ where ℓ is loss and θ is the predictor.

2. Train linear prediction model. We run $\ell = \beta_0 + \beta_1\theta + \epsilon$. Use model to generate predictions $\hat{\ell}(\theta)$ for training and test data.

3. Determine the parameters for baseline contracts using method described in section **??**.

4. Once the baseline contracts have been determined, use training data to determine cost of baseline method on the training data. This gives us B, π_{SQ} for our model.

5. Use $\hat{\ell}$ from step 2, training data, and B, π_{SQ} from step 4 as input into optimization model. Use optimization model to determine contract parameters.

6. Given test data, generate predictions and use predictions to calculate payouts from baseline and from optimal contract.

7. Calculate performance metrics on test data.

3.2.3 Results

- No corr results

- Pos corr results

- Neg corr results

3.3 Kenya Data Evaluation

3.3.1 Data Sources

- NDVI Data

- Marsabit Household Data

- Kenya Geospatial Data

3.3.2 Data Creation

- Calculate NDVI metrics for each village in each season.

- Use household data to calculate village seasonal mortality.

- Merge two datasets to create dataset for regression, train prediction model for each cluster.

- Add model predictions to household data.

3.3.3 Evaluation Procedure

1. Split into training/test set

2. Train prediction model on training data

3. Use model to calculate village level predicted losses, add this to household training data

4. Calculate baseline insurance contracts using training household data

5. Calculate optimal insurance contracts using training household data

6. Use prediction model to calculate village level predicted losses on the test set.

7. Calculate insurance payouts on test data.

3.3.4 Results

4 Conclusion and Next Steps

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

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

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