Meeting

José I. Velarde Morales

April 3, 2023

1 General Thoughts on the Project

I think that the best chance for this paper to be exceptional is if our methods end up being used in practice, especially if we can quantitatively show an improvement over the status quo. Right now, I think that the best opportunity for this to be implemented would be with Columbia's International Research Institute for Climate and Society. It seems that they are currently implementing several index insurance programs, and they seemed to think that with some modifications, our work could be really useful for them. Based on my conversations with them, I developed a model for some of the specific challenges they are facing, this model is described Section 3. Section 2 presents updates and questions on the robust optimization model.

2 Robust Optimization Model

We are interested in developing a robust version of our current program to better handle uncertainty. Here, $\mathcal{F}$ denotes the set of probability distributions we want to be robust against. $CVAR_{1-\epsilon}^F(x)$ means the expectation in the $CVAR$ is taken with respect to probability distribution F.

$$\min_{a,b,K,\pi} \sup_{F \in \mathcal{F}} \max_z \quad CVaR_{1-\epsilon}^F \left(s_z \left(\ell_z + \pi_z - \underline{I_z(\theta_z)} \right) \right) \tag{1}$$

$$\text{s.t. } \pi_z \leq \mathbb{E}\left[\overline{I_z(\theta_z)} \right] + \frac{1}{\sum_z s_z} c_\kappa K \tag{2}$$

$$K \leq CVaR_{1-\epsilon_K} \left(\sum_z s_z \overline{I_z(\theta_z)} \right) - \mathbb{E}\left[\sum_z s_z \underline{I_z(\theta_z)} \right] \tag{3}$$

$$\overline{I_z(\theta_z)} = \max\left\{ 0, a_z \hat{\ell}_z(\theta_z) + b_z \right\} \tag{4}$$

$$\underline{I_z(\theta_z)} = \min\left\{ a_z \hat{\ell}_z(\theta_z) + b_z, 1 \right\} \tag{5}$$

$$\pi_z \leq \overline{\pi_z} \tag{6}$$

- Thus far, I've been able to find the robust reformulation for the objective.

- Constraints (2) and (3) are coming from the insurance company. All quantities needed to calculate the premium, mainly $\mathbb{E}\left[\overline{I_z(\theta_z)} \right]$ and $K = CVaR_{1-\epsilon_K} \left(\sum_z s_z \overline{I_z(\theta_z)} \right) - \mathbb{E}\left[\sum_z s_z \underline{I_z(\theta_z)} \right]$ are calculated by the insurer. The insurer calculates these quantities by generating data from copulas and then calculating these quantities on the generated data. **Question:** do you think it would be OK to implement these constraints by generating data from a copula and calculating all of empirical expectations of the quantities of interest? In other words, we would

replace $\pi_z \leq \mathbb{E}\left[\overline{I_z(\theta_z)}\right] + \frac{1}{\sum_z s_z}c_\kappa K$ with $\pi_z \leq \frac{1}{N}\sum_{i=1}^{N}\overline{I_z(\theta_{iz})} + \frac{1}{\sum_z s_z}c_\kappa K$ and similarly for constraint (3). The θ_{iz}'s used in the sum would be generated by a copula. The motivation for this is that this is how these quantities are calculated in practice.

- **Question:** Thus far I've been treating ℓ_z and $\hat{\ell}_z$ as our uncertain variables. I think an intrinsic part of the problem is that we don't know what the true relationship is between ℓ_z and $\hat{\ell}_z$. I wanted to double check with you that modeling it this way makes sense. At the time the decision is made, we have past observations of $(\ell_z, \hat{\ell}_z)$. We then set a_z, b_z based on these observations. Afterwards, we see the realization of $\hat{\ell}_z$ and make payouts based on this, and we never observe ℓ_z. Given this timeline, does the modeling make sense?

3 Model for Incorporating Farmer Surveys

3.1 Background

I recently spoke to a researcher from the Institute Research Institute for Climate and Society (IRR) at Columbia University. They have implemented index insurance programs in several countries across Africa and are currently implementing programs in several other countries. They developed a more participatory approach to designing index insurance. Their approach involves conducting interviews with farmers to determine which were the worst years for the harvest. This participatory approach was motivated by several features of the context. First, household level loss data is usually not available in the countries this team works in. This makes it much harder to develop a model that directly predicts loss. Second, incorporating farmers into the design process makes it easier to ensure that the product being developed will be useful for the farmers. This increases the likelihood the farmers will buy the product. This approach has been used by the IRI team in several countries with success, however, the team expressed concerns that the approach they use to incorporate the farmer surveys is non-systematic.

The team is mainly focused on designing multi-phase insurance for single villages. In multi-phase insurance, the growing season is split up into multiple phases and there is a different payout function for each phase. The insurance designers must choose the exact dates to include in each phase, how to allocate the insured amount across phases (e.g. if there are two phases it could be 50/50 or 30/70), and what the payout rate should be in each phase. When designing the insurance, they can see in which past years their proposed insurance would have issued a payout, and they try to match as many of the bad years the farmers identified. The main constraints they are worried about is payout frequency and price. They want the insurance to payout approximately once every 5 years, but getting the frequency right becomes increasingly challenging as there are more phases.

More background and Thoughts In our conversation, the researcher from IRI also mentioned that there are three general ways in which index insurance is designed. The first is based on predictive models and assumes that household loss data is available. This is the setting that our current method addresses. The second approach is the one just described, mainly surveying farmers to determine the worst years, and then design the insurance to match these bad years as much as possible. The last approach is to use crop models. These are models that take rainfall and temperature as input and predict crop yields. The problem with these models is that they are often developed based on crop varieties used in rich countries and are sensitive to variables we don't have much information on, like soil type and farming practices. I think the participatory setting could be good, because it seems to be increasingly popular to incorporate feedback from

end users. I think there was recently a push to develop optimization models that can be augmented by expert/user knowledge, and this could contribute to that vein of research.

3.2   Model

In this model, we are maximizing the number of bad years as identified by farmers in which the insurance would have issued a payout. In other words, if year y was identified as farmers as a particularly bad year, we want the insurance we design to have issued out a payout in that year.

Data

- $\theta \in \mathbb{R}^{D \times Y}$: θ is remote sensing data where D is the number of dekads (10 day intervals), and Y is the number of years we have data for. In practice, θ is often rainfall, $D = 36$ because there are 365 days in the year, and $Y \in [20, 40]$.

- $b \in \{0, 1\}^Y$: b is a binary vector indicating bad years as identified by farmers. $b_i = 1$ if farmers identified year i as a bad year.

- T: number of phases in insurance contract.

- B: budget

- $\underline{f}, \overline{f}$: minimum and maximum desired frequency for insurance

Decision Variables

- $z \in \{0, 1\}^Y$: z is a binary variable indicating in which years the insurance would have issued a payout. $z_i = 1$ if the insurance would have issued out a payout in year i.

- $Y \in \{0, 1\}^{T \times D}$: Y is a binary matrix indicating which dekads are included in which phase. $y_{td} = 1$ if dekad d is included in phase t.

- w_t is the weight assigned to phase t. We have that $\sum_t w_t = 1$

- $\gamma \in \mathbb{R}^{T \times Y}$: γ_t^y is the sum of rainfall in phase t of year y.

- $\alpha \in \{0, 1\}^{T \times D}$: α is a helper variable to ensure that the phases are non overlapping.

$$\max_{y,z,\gamma,a,b,\alpha} \sum_{y=1}^{Y} b_y z_y \tag{7}$$

$$\text{s.t. } M z_y \geq \sum_{t=1}^{T} \max\left\{a_t \gamma_t^y + b_t, 0\right\} \tag{8}$$

$$\underline{f} \leq \frac{1}{Y} \sum_{y=1}^{Y} z_y \leq \overline{f} \tag{9}$$

$$\sum_{y=1}^{Y} \sum_{t=1}^{T} \max\left\{a_t \gamma_t + b_t, 0\right\} + c_\kappa K \leq B \tag{10}$$

$$K \leq CVaR_{1-\epsilon_P}\left(\sum_{t=1}^{T} \max\left\{a_t \gamma_t + b_t, 0\right\}\right) - \mathbb{E}\left[\min\left\{a_t \gamma^t + b_t, w_t\right\}\right] \tag{11}$$

$$\gamma_t^y = \sum_{d=1}^{D} \theta_{yd} Y_{td} \tag{12}$$

$$\alpha_1^t = Y_{t1} \tag{13}$$

$$\alpha_i^t \geq Y_{ti} \tag{14}$$

$$\alpha_i^t \geq \alpha_{i-1}^t \tag{15}$$

$$\sum_{t=1}^{T} w_t = 1 \tag{16}$$

$$z_i, \alpha_i^t, Y_{ij} \in \{0,1\} \tag{17}$$

Constraint (8) ensures that $z_i = 1$ if a payout was issued in year i. Constraint (9) is the payout frequency constraint. Constraint (10) is a budget constraint. Constraints 13-15 ensure that the phases are non-overlapping. **To do:** need to add constraints so that there are no 'gaps' in Y, for example that you don't have the situation that $Y_t = (0,0,1,1,0,0,1,1,0,0)$, and to have the payout sizes correspond to the severity of the years as identified by farmers. For example, if year i is identified as worse than year $i-1$, we want the payouts to reflect that.