

Matching Model

José I. Velarde Morales

April 25, 2023

1 Model for Incorporating Farmer Surveys

1.1 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.

We are interested in 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.

1.2 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. In the model below, the payouts are assumed to be in rates. The model will output values for a and b , and the final contracts will be $I(\eta) = \sum_t^T \min \{ \max \{ a_t \eta_t + b_t, 0 \}, w_t \}$, where η_t is the sum of rainfall in phase t . So if the total insured amount is s , the payout will be $sI(\eta)$.

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 $Y \in [20, 40]$.
- $\beta \in \{0, 1\}^Y$: β is a binary vector indicating bad years as identified by farmers. $\beta_i = 1$ if farmers identified year i as a bad year.
- T : number of phases in insurance contract.
- $\bar{\pi}$: budget
- \underline{f}, \bar{f} : minimum and maximum desired frequency for insurance
- s : is the total insured amount.
- m : the number of ordinal payout constraints we allow to be violated.

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 the i^{th} year of the historical data.
- $A \in \{0, 1\}^{T \times D}$: A is a binary matrix indicating which dekads are included in which phase. $A_{td} = 1$ if dekad d is included in phase t .
- $\eta \in \mathbb{R}^{Y \times T}$: η_t^y is the sum of rainfall in phase t of year y .
- $\alpha \in \mathbb{R}^{Y \times T}$: Upper bound on payout, $\alpha_{yt} = \max \{a_t \eta_t^y + b_t, 0\}$
- $\omega \in \mathbb{R}^{Y \times T}$: Lower bound on payout, $\omega_{yt} = \min \{a_t \eta_t^y + b_t, w_t\}$
- w_t is the weight assigned to phase t . We have that $\sum_t w_t = 1$
- $D \in \{0, 1\}^{T \times D-1}$: this is a helper variable to ensure that phases are continuous. For example, we wouldn't want phase 1 to be dekads 1, 2, 4 and for phase 2 to be dekads 3, 5, and 6.
- $d \in \{0, 1\}^J$: this is a binary vector indicating which of the ordinal payout constraints can be satisfied. Here, J is the number of constraints that specify that the payout in year i should be greater than the payout in year j if year i was ranked worse than year j . If $d_k = 1$, then the k^{th} ordinality constraint doesn't have to be satisfied.

$$\max_{A,D,z,\eta,a,b,d} \sum_{y=1}^Y \beta_y \sum_{t=1}^T \min \{a_t \eta_t^y + b_t, w_t\} \quad (1)$$

$$\text{s.t. } M z_y \geq \sum_{t=1}^T \max \{a_t \eta_t^y + b_t, 0\}, \quad \forall y \quad (2)$$

$$\underline{f} \leq \frac{1}{Y} \sum_{y=1}^Y z_y \leq \bar{f} \quad (3)$$

$$\frac{1}{Y} \sum_{y=1}^Y \sum_{t=1}^T \max \{a_t \eta_t^y + b_t, 0\} + \frac{1}{s} c_\kappa K \leq \bar{\pi} \quad (4)$$

$$K = CV a R_{1-\epsilon_K} \left(s \sum_{t=1}^T \max \{a_t \eta_t + b_t, 0\} \right) - \mathbb{E} \left[s \sum_{t=1}^T \min \{a_t \eta_t + b_t, w_t\} \right] \quad (5)$$

$$M d_k + \sum_{t=1}^T \max \{a_t \eta_t^{y_i} + b_t, 0\} \geq \sum_{t=1}^T \min \{a_t \eta_t^{y_j} + b_t, w_t\}, \forall i, j \quad \text{s.t. year } i \text{ worse than year } j \quad (6)$$

$$\sum_{k=1}^J d_k \leq m \quad (7)$$

$$\sum_{t=1}^T A_{t,d} \leq 1, \quad \forall d \quad (8)$$

$$D_{t,d} = |A_{t,d+1} - A_{t,d}|, \quad \forall t, \quad d = 1, \dots, D-1 \quad (9)$$

$$\sum_{d=1}^{D-1} D_{t,d} \leq 2, \quad \forall t \quad (10)$$

$$A_{t,1} + A_{t,D} \leq 1, \forall t \quad (11)$$

$$\eta_t^y = \sum_{d=1}^D \theta_{yd} A_{td}, \quad \forall t, d \quad (12)$$

$$\sum_{t=1}^T w_t = 1 \quad (13)$$

$$z_i, A_{ij} \in \{0, 1\} \quad (14)$$

Constraint (2) ensures that $z_i = 1$ if a payout was issued in year i , here M is a large number, say 1000, such that if $z_y = 1$ then the constraint is always satisfied. Constraint (3) is the payout frequency constraint. Constraints (4) and (5) are the budget constraints. Constraint (4) is our proxy for the premium, it includes the average payout and the cost of capital. It is a rate, and is supposed to represent the cost of the insurance per insured unit. Constraint (5) is the definition of the cost of capital. Constraint (6) ensures that if year i is worse than year j , then the payout in year i is greater than or equal to the payout in year j . Constraint (7) specifies that at least $J - m$ of the constraints specified in Constraint (6) need to be satisfied. This is because Constraint (6) will automatically be satisfied if $d_k = 1$. In other words, Constraint (6) can be violated if $d_k = 1$. As a result, Constraint (7) specifies that constraints in Constraint (6) can be violated at most m times. Constraint (8) ensures each dekad is assigned to at most one phase. D is a helper variable

we use to ensure that the phases are continuous. Let a_t be the t^{th} row of A , this vector will have $i = 1$ if dekad i is in phase t . We set $D_{t,d} = |A_{t,d+1} - A_{t,d}|$. Intuitively, D keeps tracks of the number of times a_t switches from 1 to 0 or vice versa. In order for a phase to be continuous, it can switch at most 2 times. The only non-continuous sequences that satisfy this are sequences starting and ending with ones, and having zeros in the middle: $(1, 1, 0, \dots, 0, 1, 1)$. We exclude these sequences with constraint (11). Constraint (12) defines η_t^y to be the sum of rainfall in phase t of year y . Constraint (13) specifies that the sum of the weights for all of the phases has to be equal to 1.

LP Reformulation Using the results from Rockafellar and Uryasev (2002), we can reformulate the above as a mixed integer linear program:

$$\begin{aligned}
& \max_{A,D,z,\gamma,a,b,\alpha} \sum_{y=1}^Y \sum_{t=1}^T \beta_y \omega_{yt} \\
& \text{s.t. } Mz_y \geq \sum_{t=1}^T \alpha_{yt}, & \forall y \\
& \underline{f} \leq \frac{1}{Y} \sum_{y=1}^Y z_y \leq \bar{f} \\
& \frac{1}{Y} \sum_{y=1}^Y \sum_{t=1}^T \alpha_{yt} + c_\kappa K \leq \bar{\pi} \\
& t_k + \frac{1}{\epsilon_K} \sum_{y=1}^Y p^y \gamma^y \leq K + \frac{1}{Y} \sum_y \sum_t \omega_{yt} \\
& \gamma^y \geq \sum_{t=1}^T \alpha_{yt} - t_K & \forall y \\
& \gamma^y \geq 0 & \forall y \\
& \alpha_{yt} \geq \sum_{d=1}^D h_{ytd} + b_t & \forall y, t \\
& \alpha_{yt} \geq 0 & \forall y, t \\
& \omega_{yt} \leq \sum_{d=1}^D h_{ytd} + b_t & \forall y, t \\
& \omega_{yt} \leq w_t & \forall y, t \\
& h_{ytd} \leq a_t \theta_{yd} & \forall y, t, d \\
& h_{ytd} \geq 0 & \forall y, t, d \\
& h_{ytd} \leq M A_{td} & \forall y, t, d \\
& h_{ytd} \geq a_t \theta_{yd} + M(A_{td} - 1) & \forall y, t, d \\
& \sum_{t=1}^T A_{t,d} \leq 1 & \forall d \\
& D_{t,d} \geq A_{t,d+1} - A_{t,d} & \forall t, \quad d = 1, \dots, D-1 \\
& D_{t,d} \geq -(A_{t,d+1} - A_{t,d}) & \forall t, \quad d = 1, \dots, D-1 \\
& \sum_{d=1}^{D-1} D_{t,d} \leq 2 & \forall t \\
& A_{t,1} + A_{t,D} \leq 1 & \forall t \\
& \eta_t^y = \sum_{d=1}^D \theta_{yd} A_{td} & \forall t, d \\
& \sum_{t=1}^T w_t = 1 \\
& z_i, A_{ij} \in \{0, 1\}
\end{aligned}$$