

# Next Steps

José I. Velarde Morales

October 24, 2022

## 1 John Birge Meeting Summary

- Thoughts on CVaR
  - Maybe estimating it empirically isn't that bad, since in this setting losses are limited (because we are insuring a fixed amount), and thus we don't have to worry as much about fat tails.
  - He proposed we use the expected loss beyond a fixed threshold instead of the CVaR.
- When estimating the distribution of weather and losses, we should consider using a model that takes into account serial autocorrelation of the data. We can look into the time-series literature for this.
- We should think about an imputation method to use for the cases where there's missing data.
- He said that we should figure out how we will deal with uncertainty before making the evaluation more rigorous.

## 2 Multi-phase Model

### 2.1 Motivation

In practice, weather index insurance contracts specify a coverage period which is split up into different phases. Each of these phases corresponds to a different stage of growth in the crop cycle. The stages are generally establishment/vegetative, flowering/reproductive, and the ripening stage, but it can differ by crops. There are some crops that have 5 different stages. Each stage has different climate requirements, and thus a different payout function is calculated for each phase. This makes the insurance additionally flexible, because it allows farmers to only insure certain parts of the season, depending on what they're more worried about. For example, a farmer might choose to only insure for the early part of the season, if he feels that that's where most of the risk is coming from. The team at Columbia mentioned that they are looking for a way to decide how to split up the coverage between the different phases.

### 2.2 Full Data Model

This model assumes that we have observations of average losses for the smallest unit being insured (e.g. village level). The model minimizes the conditional value at risk of farmers' net loss (i.e. loss net of insurance), subject to an overall budget constraint. The total cost of the insurance consists of payouts and of costs associated with holding capital. We first describe the model parameters,

and then describe the model. The model will output values for  $a$  and  $b$ , and the final insurance contracts will be  $I_t(\theta_t) = \min \left\{ \max \left\{ 0, a_t \hat{\ell}_t(\theta_t) + b \right\}, P \rho_t \right\}$ .

### Model Parameters

- $\epsilon$ : This is the  $\epsilon$  used for the *CVaR* objective.  $\epsilon = 0.1$  means that our objective is  $E[\ell - I(\theta) | \ell - I(\theta) \geq VaR_{1-0.1}(\ell - I(\theta))]$ .
- $\epsilon_K$ : This is the epsilon used in the formula for required capital. Recall that the required capital  $K = CVaR_{1-\epsilon_K}(I(\theta)) - E[I(\theta)]$ .
- $B$ : This is the budget constraint for the total cost of the insurance.
- $\pi_{SQ}$ : This is the average payout made by the status quo contract,  $E[ISQ(\theta)]$ . This is also called the premium of the status quo contract.
- $P$ : This is the complete insured amount across all periods.
- $\rho_t$ : This is the share of the overall coverage that will be used to insure phase  $t$ .
- $c_K$ : This is the cost of capital.

**Model** In the model below, our objective is the conditional value at risk of the farmer's loss net of insurance. The first constraint is the budget constraint and the last constraint is the definition of the required capital.

$$\min_{a,b,K,\rho} CVaR_{1-\epsilon} \left( \ell - \sum_{t=1}^T \min \left\{ (a_t \hat{\ell}_t(\theta_t) + b_t), P \rho_t \right\} \right) \quad (1)$$

$$\text{s.t. } \mathbb{E} \left[ \sum_{t=1}^T \max \left\{ 0, a_t \hat{\ell}_t(\theta_t) + b_t \right\} \right] + c_K K \leq B \quad (2)$$

$$K = CVaR_{1-\epsilon_P} \left( \sum_{t=1}^T \max \{ 0, a_t \hat{\ell}_t(\theta_t) + b_t \} \right) - \mathbb{E} \left[ \sum_{t=1}^T \min \left\{ (a_t \hat{\ell}_t(\theta_t) + b_t), P \rho_t \right\} \right] \quad (3)$$

$$\sum_t \rho_t = 1 \quad (4)$$

$$0 \leq \rho_t \leq 1, \forall t \quad (5)$$

We reformulated the problem as a linear program using the results from Rockafellar, Uryasev, et al. (2000). In the model below,  $p^j$  is the probability of event  $j$ , and  $j$  indexes the possible realizations of  $\theta, \ell$ .  $N$  is the total number of samples.

$$\min_{a,b,\gamma,\gamma_K,\alpha,t,t_K} t + \frac{1}{\epsilon} \sum_j p^j \gamma^j \quad (6)$$

$$\text{s.t. } \gamma^j \geq \ell^j - \sum_{t=1}^T \min \left\{ (a_t \hat{\ell}_t(\theta_t^j) + b_t), P \rho_t \right\} - t, \forall j \quad (7)$$

$$\gamma^j \geq 0, \forall j \quad (8)$$

$$B \geq \frac{1}{N} \sum_{j=1}^N \sum_t \alpha_t^j + c_\kappa K \quad (9)$$

$$t_K + \frac{1}{\epsilon_K} \sum_{j=1}^N p^j \gamma_K^j \leq K + \frac{1}{N} \sum_{j=1}^N \sum_{t=1}^T \min \left\{ (a_t \hat{\ell}_t(\theta_t^j) + b_t), P \rho_t \right\} \quad (10)$$

$$\gamma_K^j \geq \sum_{t=1}^T \alpha_t^j - t_K, \forall j \quad (11)$$

$$\gamma_K^j \geq 0, \forall j \quad (12)$$

$$\alpha_t^j \geq a_t \hat{\ell}_t(\theta_t^j) + b_t, \forall j \quad (13)$$

$$\alpha_t^j \geq 0, \forall j \quad (14)$$

$$\sum_t \rho_t = 1 \quad (15)$$

$$0 \leq \rho_t \leq 1, \forall t \quad (16)$$

### 2.3 Limited Data Setting

This setting more accurately reflects the data available to the International Research Institute for Climate and Society in Columbia. Suppose we have a set of villages  $A$  that we are interested in insuring. These villages belong to a state,  $S$ , and we have that  $A \subset S$ . We also have a set of years  $Y$ . Let  $\ell_v(y)$  be a mapping of the average loss of in village  $v$  in year  $y$ .  $\ell_v(y)$  is unobserved by us. For every village,  $v \in A$ , we also have an ordering  $P_v$  which ranks each year  $y \in Y$  in terms of severity. In other words, for  $y, y' \in Y$  and  $v \in A$ , we know if  $\ell_v(y) > \ell_v(y')$  or vice versa, but we don't know  $\ell_v(y) - \ell_v(y')$ . Let  $f_v(\theta_1^v, \dots, \theta_T^v)$  denote the average output of village  $v$ . Here,  $\theta_t$  is our weather signal for phase  $t$  of the crop's growing schedule. Let  $f_S(\theta_1, \dots, \theta_T)$  be the total output of state  $S$ . We observe  $f_S(\theta_1, \dots, \theta_T)$ , and we know that  $f_S(\theta_1, \dots, \theta_T) = \sum_{v \in A} f_v(\theta_1^v, \dots, \theta_T^v) + \sum_{v \in S \setminus A} f_v(\theta_1^v, \dots, \theta_T^v)$ . We observe  $f_S(y)$ , have the ordering  $P_v$ , and we observe  $\theta_1^v, \dots, \theta_T^v$  for every  $v \in A$  and every  $y \in Y$ . We want to design the insurance contracts based on this information.

**Note:** I was thinking of two approaches here. One option would be to try to impute the average losses of farmers in each village for the years in question, in that case we could just feed the imputed data into the model specified in section 2.2. Another option would be to use this information to try to construct an uncertainty set for the realizations of  $(\ell, \theta)$  and use robust optimization. The third option would be to use this information to create an uncertainty set for the joint distribution of  $(\ell, \theta)$  and use DRO.