Evaluation Proposal

José I. Velarde Morales

August 10, 2022

Baseline Approach

We will be comparing our proposed approach to the method developed by Chantarat et al. (2013), which, to the extent of our knowledge is what is currently being used for Kenya's Index Based Livestock Insurance (IBLI) program. The method is as follows:

- Cluster locations into two clusters
- Fit a separate herd mortality function for each cluster. They use a linear regression model to predict herd mortality.
- Contracts are offered at the cluster level, but payouts are at the location level.
- They choose the strike level that would explain the highest share of insurable losses in the historical data.

Toy Examples

Data Generating Process

For the toy examples, we will use the following data generating processes:

- $w = \beta\theta + \epsilon$ where $w, \theta \in \mathbb{R}^z$ where z is the number of zones, and $\theta \sim \mathcal{N}(\mu, \Sigma), \epsilon \sim \mathcal{N}(0, \sigma I)$.
- $w = f(\theta) + \epsilon$ where $\theta \sim \mathcal{N}(\mu, \Sigma)$, f is nonlinear, and $\epsilon \sim \mathcal{N}(0, \sigma I)$.

Performance Metrics

We will be using the following performance metrics to compare the two approaches.

- Accuracy: percentage of correct decisions (i.e. giving a payout when a covered loss occured, and not giving a payout when a covered loss does not occur.)
- Insurer Basis Risk: probability that the contract will pay out when a covered loss did not occur.
- Insured Party Basis Risk: probability that the contract will not pay out when a covered loss occurs.
- Probability of farmer ruin: probability that wealth net of insurance falls below a pre-specified threhsold.
- Cost to insurer: average cost to insurer, should also include extreme cases.
- Share of losses covered

Parameter Values

Since we will be using a simplified model in the toy examples, we will only need to set three parameters. We describe how we set each one below.

- $\bar{w}$: This is the wealth threshold we want farmers to be above. We will set this to be the 10^{th} percentile of the distribution of w for now. In the future, this should probably be determined by the setting, and what would be considered catastrophic in each setting.
- c_K: This is the cost of capital, and there are estimates available (e.g. here). According to this source, the average cost of capital for general insurance companies was 4.66%.
- β_z: This is the relative riskiness of each zone. For now, we will set $\beta_z = \frac{1-R_z^2}{\sum_z 1-R_z^2}$.

1 Zone Case

Data Generating Process

In the one zone case, we will generate 1,100 samples from the model: $w = \beta\theta + \epsilon$, with $\theta \sim \mathcal{N}(5, 10), \beta = 2, \epsilon \sim \mathcal{N}(0, 1)$. 80 samples will be used for training, 20 for model selection, and 1000 for evaluation.

Baseline Approach for 1 zone case

In the one zone case, we will replicate the baseline approach as follows:

1. We will fit a linear model to the simulated data of w, θ. This model will generate predictions, $\hat{w}(\theta)$
2. The insurance contract will be of the form $(\hat{w}(\theta) - w^*)^+$, where w^* will be decided according to the share of losses covered in the hold out set.

Optimization Approach for 1 zone case

For the single zone case, we will be solving the following model, with $\bar{w}$ set to the 20^{th} percentile of the empirical distribution of w. In order to separate the effect from the design of the contract from the design of the index, we will use the same index, $\hat{w}(\theta)$ as in the baseline approach.

$$\begin{aligned}
&\min_{a,d,\pi} E[I^k] && (1)\\
\text{s.t. } & I^k = a\hat{w}(\theta^k) + d, \forall k && (2)\\
& 0 \leq I^k, \forall k && (3)\\
& t + \frac{1}{\epsilon}\sum_k p_k \gamma_k \leq 0 && (4)\\
& \gamma_k \geq \bar{w} - w^k + E[I] - I^k - t, \forall k && (5)\\
& \gamma_k \geq 0, \forall k && (6)
\end{aligned}$$

2 Zone Case

Data Generating Process

In the two zone case, we will generate 1,100 samples from the model: $w = \beta\theta + \epsilon$, with $\theta \sim \mathcal{N}((3,5),\Sigma), \beta = (1.5, 2), \epsilon \sim \mathcal{N}(0, I)$. 80 samples will be used for training, 20 for model selection, and 1000 for evaluation. We will use the two following options:

- Σ s.t. $Corr(\theta_1, \theta_2) = 0.6$
- Σ s.t. $Corr(\theta_1, \theta_2) = -0.6$

Baseline Approach for 2 zone case

In the one zone case, we will replicate the baseline approach as follows:

1. We will fit a linear model to the simulated data of w, θ. This model will generate predictions, $\hat{w}(\theta)$
2. The insurance contract will be of the form $(\hat{w}(\theta) - w^*)^+$, where w^* will be decided according to the share of losses covered in the hold out set.

Optimization Approach for 2 zone case

For the single zone case, we will be solving the following model, with $\bar{w}$ set to the 20^{th} percentile of the empirical distribution of w. In order to separate the effect from the design of the contract from the design of the index, we will use the same index, $\hat{w}(\theta)$ as in the baseline approach.

$$\min_{a,b,\pi} \max_{z} \pi_z \tag{7}$$

$$\text{s.t. } I_z^k = a_z \hat{w}(\theta_z^k) + b_z, \forall k, \forall z \tag{8}$$

$$0 \leq I_z^k \leq y_z, \forall k, \forall z \tag{9}$$

$$t_z + \frac{1}{\epsilon}\sum_{k=1}^{K} p_k \gamma_z^k \leq 0, \forall z \tag{10}$$

$$\gamma_z^k \geq \bar{w} - w_z^k + \pi_z - I_z^k - t_z, \forall k, \forall z \tag{11}$$

$$\gamma_z^k \geq 0, \forall k, \forall z \tag{12}$$

$$t_P + \frac{1}{\epsilon_P}\sum_{k=1}^{K} p_k \gamma_P^k \leq K^P + Z\bar{\pi} \tag{13}$$

$$\gamma_P^k \geq \sum_z I_z^k - t_P \tag{14}$$

$$\gamma_P^k \geq 0, \forall k \tag{15}$$

$$\pi_z = E[I_z] + c_K K^P \beta_z \tag{16}$$

Data-Based Evaluation

Data Sources

We will be using three main data sources for the empirical evaluation of our method. These sources are:

- NDVI Data: Normalized Difference Vegetation Index data. These are satellite images of the vegetation in the insured area.

Simulation Procedure

Performance Metrics