

Two Zone Evaluation

José I. Velarde Morales

August 19, 2022

1 Setup

Data Generating Process

In the two zone case, we will generate samples from the model: $\ell = \beta\theta + \epsilon$, with $\theta \sim \mathcal{N}((3, 5), \Sigma)$, $\beta = \text{diag}(1.5, 2)$, $\epsilon \sim \mathcal{N}(0, I)$. We will use the two following options:

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

I draw 100 training samples from the above model to train the prediction model and to use as input for the optimization program. I then evaluate both methods using 1000 samples drawn from the same model.

2 Status Quo

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:

- First, they use a clustering algorithm to group locations into clusters
- They then use historical data to fit a linear regression model to predict herd mortality rates in each cluster. They fit a different model for each cluster.
- Contracts are of the form: $I(\theta) = \max(\hat{M}(\theta) - M^*, 0) \times TLU \times P_{TLU}$ where $\hat{M}(\theta)$ is the predicted herd mortality rate, M^* is the strike value, TLU is the number of insured livestock units, and P_{TLU} is the price per insured livestock unit. In other words, their contract pays farmers for the full predicted loss beyond a threshold, M^* . This threshold, M^* is the contract's strike value.
- They choose the strike value that would explain the highest share of insurable losses in the historical data. Specifically, they run the following regression: $y_s = \beta_s \hat{y}_s + \epsilon$ where y_s is the actual insured losses at strike value s and \hat{y}_s is the predicted insured losses at strike value s . For example, suppose that $TLU = 100$ (ie there are 100 insured units), and that $P_{TLU} = 25$ (ie each unit is worth 25), and that $M^* = 0.25$ (ie contract starts paying out once the predicted mortality rate exceeds 25%). If the actual mortality rate is 0.5, then actual insured losses would be $y_{25} = \max(M - M^*, 0) \times TLU \times P_{TLU} = (0.5 - 0.25) \times (100) \times (25)$. If the predicted mortality rate in that scenario was 0.4, the predicted insured losses, $\hat{y}_{25} =$

$\max(\hat{M}(\theta) - M^*, 0) \times TLU \times P_{TLU} = (0.4 - 0.25) \times (100) \times (25)$. They use historical data to calculate y_s, \hat{y}_s , and then run the following regression: $y_s = \beta_s \hat{y}_s + \epsilon$. They choose the strike value $s = \arg \max_s \beta_s$. This takes into account the fact that the prediction model, $\hat{M}(\theta)$ might be better at predicting some losses better than others.

To mimick this in our toy example, we set the status quo contracts to be $I(\theta) = \max(\hat{l}(\theta) - l^*, 0)$, since we are already assuming that l is the total loss suffered. For the toy example, we fit a (correctly specified) linear regression model to predict losses: $l = \beta\theta + \epsilon \implies \hat{l}(\theta) = \hat{\beta}\theta$.

3 Optimization Approach

In order to separate the effect of contract design from the effect of prediction quality, we will be basing our contracts on the same predictions used by the status quo method. In other words, we will use the status quo method to estimate a model that predicts loss based on theta, $\hat{l}(\theta)$, and our payout function will use that as input instead of θ . In other words, our model will define payout functions $I(\hat{l}(\theta))$, where $\hat{l}(\theta)$ is the same prediction function used by the status quo method.

Minimum CV@R Model

This model minimizes the $CV@R$ of the farmer's net loss subject to a constraint on the premium. The premium constraints are expressed as a fraction of the full insured amount.

Model Parameters

- ϵ : This defines the $CV@R$ objective. $\epsilon = 0.1$ means that our objective is on the expected value of the loss given that it is above the 90th percentile.
- $\bar{\pi}$: This is the maximum value of the premium.
- $\underline{\pi}$: This is the minimum value of the premium.
- β_z : This is a measure of the relative riskiness of zone z .
- ϵ_P : This is the epsilon corresponding to the $CV@R$ of the entire portfolio. This is used to determine the required capital for the portfolio. Values I've seen used are $\epsilon_P = 0.01$ and $\epsilon_P = 0.05$.
- Z : number of insured zones.
- c_k : cost of capital
- K_z : maximum insured amount of zone z .

Multiple Zone Model

$$\min_{a,b,K^P} \max_z CV@R_{1-\epsilon}(\ell_z - \min \left\{ (a_z \hat{\ell}_z(\theta_z) + b_z), K_z \right\}) \quad (1)$$

$$\text{s.t. } K\bar{\pi} \geq a_z \mathbb{E}[\hat{\ell}_z(\theta_z^k)] + b_z + c_k \beta_z K^P, \forall k, \forall z \quad (2)$$

$$0 \leq a_z \hat{\ell}_z(\theta_z^k) + b_z, \forall k, \forall z \quad (3)$$

$$K^P + Z\underline{\pi} \geq CV@R_{1-\epsilon_P} \left(\sum_z a_z \hat{\ell}_z(\theta_z) + b_z \right) \quad (4)$$

The reformulation is:

$$\min_{a,b,\gamma,t,m,K^P} m \quad (5)$$

$$\text{s.t.} \quad t_z + \frac{1}{\epsilon} \sum_k p_k \gamma_z^k \leq m, \forall z \quad (6)$$

$$\gamma_z^k \geq \ell^k - \min \left\{ (a_z \hat{\ell}_z(\theta_z^k) + b_z), K_z \right\} - t_z, \forall k, \forall z \quad (7)$$

$$\gamma_z^k \geq 0, \forall k, \forall z \quad (8)$$

$$t_p + \frac{1}{\epsilon_p} \sum_k p_k \gamma_P^k \leq K^P + Z\pi \quad (9)$$

$$\gamma_P^k \geq \sum_z a_z \hat{\ell}_z(\theta_z^k) + b_z - t_p, \forall k \quad (10)$$

$$\gamma_P^k \geq 0, \forall k \quad (11)$$

$$K_z \bar{\pi} \geq a_z \mathbb{E}[\hat{\ell}_z(\theta_z)] + b_z + c_k \beta_z K^P, \forall z \quad (12)$$

$$0 \leq a_z \hat{\ell}_z(\theta_z^k) + b_z, \forall k, \forall z \quad (13)$$