Matching Model Updates

José I. Velarde Morales

May 3, 2023

1 Multi-Phase Model

1.1 Problems

I was able to implement the model, however, it doesn't work as one would hope. It seems to be having the problems you predicted, mainly, the formulation doesn't always force the decision variables to take on useful values. Im listing the problems and some solutions I tried below:

1. z_y wasn't tracking if there was a payout in year y, we would sometimes get solutions with $z_y = 1$ even if there was no payout in year y.

2. Dekads would only be assigned to the most important phase.

3. All of the weight is allocated to the most important phase

4. All of the payout is allocated to the most important phase

Attempted Solutions

1. Add constraint that $\epsilon - M(1-z) \leq \sum_{t=1}^{T} \min \left\{ a_t \eta_t^y + b_t, w_t \right\}$. We want $z = 1$ if there was a payout and $z = 0$ if there was no payout.

2. Add constraints that each phase is assigned at least one dekad: $\sum_{d=1}^{D} A_{td} \geq 1, \forall d$.

3. Add constraint that $w_t \in [\underline{w_t}, \overline{w_t}]$. **Note:** This still just assigns as much weight as possible to the most important phase.

4. Replace premium constraint with budget constraint $\sum_y \sum_t \max \left\{ a_t \eta_t^y + b_t, 0 \right\} \leq B$ and another constraint that $B\underline{w_t} \leq \sum_y \max \left\{ a_t \eta_t^y + b_t, 0 \right\} \leq B\overline{w_t}$. **Note:** This did not fix the problem, because we use α_{yt} to represent $\max \left\{ a_t \eta_t^y + b_t, 0 \right\}$, and the program would set $\alpha_{yt} = B\underline{w_t}$ for one year and just leave $a_t, b_t = 0$.

1.2 Data Generating Process

The original data generating process for the toy example was: $Y = w_1 Y_1 + w_2 Y_2 + w_3 Y_3$. Here, Y represents total output, Y_t represents output in phase t, and w_t is the weight, or relative importance of phase t. $Y_t = \min \left\{ \max \left\{ \theta_t - \theta^*, 0 \right\} 1 \right\}$. Here, $\theta_t \sim N(3, 1)$ and $\theta^* = 2$. Here, θ_t was meant to represent rainfall in phase t, and the intuition behind the data generating process is that output is starts dropping if rainfall goes below a certain threshold. Here, output would start to drop if rainfall was one standard deviation below the mean, and there would be full loss of phase t's output if rainfall was two standard deviations below the mean.

I tried this original data generating process, but the contract parameters, a, b would always be set to 0. However, I realized that under this DGP, the optimal value for a, b would be $a = -1, b = 2$. This is a problem because we use a reformulation to express the bilinear term: $a_t \eta_t^y$ (since $\eta_t^y = \sum_{d=1}^{D} \theta_{yd} A_{td}$, the term $a_t \eta_t^y$ is bilinear). The reformulation works for a binary variable, z and a continuous variable x, but requires that $x \geq 0$. As a result, I changed the DGP to reflect predicted loss instead of rainfall. The new DGP is $L = w_1 \ell_1 + w_2 \ell_2 + w_3 \ell_3$, with $\ell_t \sim N(3, 1)$. This was kind of a quick change, and I will work on improving it. However, this did help the problem, and a, b were no longer always 0.

1.3 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.

- π: budget

- $\underline{f}, \overline{f}$: 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 ith year of the historical data.

- $A \in \{0, 1\}^{T \times D}$: Y 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 kth ordinality constraint doesn't have to be satisfied.

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

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

$$\underline{f} \leq \frac{1}{Y} \sum_{y=1}^{Y} z_y \leq \overline{f} \tag{3}$$

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

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

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

$$\sum_{k=1}^{J} d_k \leq m \tag{7}$$

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

$$D_{t,d} = |A_{t,d+1} - A_{t,d}|, \quad \forall t, \quad d = 1, ..., D-1 \tag{9}$$

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

$$A_{t,1} + A_{t,D} \leq 1, \forall t \tag{11}$$

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

$$\sum_{t=1}^{T} w_t = 1 \tag{13}$$

$$z_i, A_{ij} \in \{0,1\} \tag{14}$$

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 tth 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: $(1, 1, 0, ..., 0, 1, 1)$. 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:

$$\max_{A,D,z,\gamma,a,b,\alpha} \sum_{y=1}^{Y} \sum_{t=1}^{T} \beta_y \omega_{yt}$$

$$\text{s.t. } Mz_y \geq \sum_{t=1}^{T} \alpha_{yt}, \qquad\qquad \forall y$$

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

$$\frac{1}{Y} \sum_{y=1}^{Y} \sum_{t=1}^{T} \alpha_{yt} + c_\kappa K \leq \overline{\pi}$$

$$t_k + \frac{1}{\epsilon_K} \sum_{y=1}^{Y} p^y \gamma^y \leq K + \frac{1}{Y} \sum_{y} \sum_{t} \omega_{yt}$$

$$\gamma^y \geq \sum_{t=1}^{T} \alpha_{yt} - t_K \qquad\qquad \forall y$$

$$\gamma^y \geq 0 \qquad\qquad \forall y$$

$$\alpha_{yt} \geq \sum_{d=1}^{D} h_{ytd} + b_t \qquad\qquad \forall y, t$$

$$\alpha_{yt} \geq 0 \qquad\qquad \forall y, t$$

$$\omega_{yt} \leq \sum_{d=1}^{D} h_{ytd} + b_t \qquad\qquad \forall y, t$$

$$\omega_{yt} \leq w_t \qquad\qquad \forall y, t$$
$$h_{ytd} \leq a_t \theta_{yd} \qquad\qquad \forall y, t, d$$
$$h_{ytd} \geq 0 \qquad\qquad \forall y, t, d$$
$$h_{ytd} \leq MA_{td} \qquad\qquad \forall y, t, d$$
$$h_{ytd} \geq a_t \theta_{yd} + M(A_{td} - 1) \qquad\qquad \forall y, t, d$$

$$\sum_{t=1}^{T} A_{t,d} \leq 1 \qquad\qquad \forall d$$

$$D_{t,d} \geq A_{t,d+1} - A_{t,d} \qquad\qquad \forall t, \quad d = 1, ..., D-1$$
$$D_{t,d} \geq -(A_{t,d+1} - A_{t,d}) \qquad\qquad \forall t, \quad d = 1, ..., D-1$$

$$\sum_{d=1}^{D-1} D_{t,d} \leq 2 \qquad\qquad \forall t$$

$$A_{t,1} + A_{t,D} \leq 1 \qquad\qquad \forall t$$

$$\eta_t^y = \sum_{d=1}^{D} \theta_{yd} A_{td} \qquad\qquad \forall t, d$$

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

$$z_i, A_{ij} \in \{0, 1\}$$