

Paper Outline

November 29, 2023

1 Contract Design Method

In this section, we introduce our method for designing index insurance contracts.

1.1 Objective and Constraints

We currently have a brief discussion of how we came up with the objective and constraints for our method. We mention practitioner interviews. Alternatively, this could be more of a "problem setting" section, where we discuss the key features of the problem that guided our modeling assumptions. For example, we could talk about the nonlinear relationship between weather variables and losses, the importance of correlated losses and tail risk, and data scarcity.

1.2 Model

Here, we can introduce how we are modeling the problem. We can define the relevant variables $\ell, \hat{\ell}, I(\theta), \pi(I(\theta))$ and the assumptions we make about them. We can highlight the differences between our model and Chen et al's model, mainly the definition of the premium. We can also introduce our objective here.

1.3 Single Zone Model

Here, we can show the idealized (non-convex) version of the problem in the single-zone case, as well as our convex relaxation of the problem and its reformulation as a convex program.

1.4 Multiple Zone Model

Here, we can just show the convex relaxation of the multiple-zone program and its reformulation as a convex program.

2 Prediction Models

This section can describe the different prediction models that we will test. We want to test models that work well with different amounts of data. More specifically, we want to include some models that don't need as much data to be trained.

2.1 Traditional ML Models

This would include the typical ML models: logistic regression, random forests, gradient boosting machines, and maybe even simpler heuristic methods.

2.2 Deep learning models

This would include Chen’s model as well as a fully convolutional neural network that I think will work better.

2.3 Time-series classification models

This would include models that have been designed specifically for time-series. Time-series classification is a subfield of machine learning that focuses on developing methods to make predictions on whole time-series. The prediction problem here seems to be exactly that. We have a time-series of remote sensing observations (e.g. rainfall, temperature, etc) and we want to use it to predict loss. Even though there are specialized algorithms for this problem, I haven’t seen them used in this context. I think that these methods might work well in the data scarce setting.

3 Evaluation

3.1 Data

In this section we would describe the different data sources we use. In this case, it will be the same ones used by Chen et al 2023. We use yield data from the United States Department of Agriculture and weather data from the PRISM climate group at Oregon State.

3.2 Baseline Evaluation

3.2.1 Performance Metrics

Here, we would introduce the performance metrics we use to compare the two methods. The metrics will be average farmer utility, insurer costs/profits, and risk for the insurer. We could potentially also include VaR and CVaR.

3.2.2 Comparison with Chen method

Here, we would describe how we run the comparison. We will use the same training, validation, and test data for both methods, and we will present the results for the baseline case. The baseline case is using the full data set, i.e. the same thing Chen et al 2023 did.

3.3 Analysis

3.3.1 Gains from better prediction model

Here, we will preform an analysis to show that the improvements from Chen’s method are mostly from using a more complex prediction model. To show this, we will train Chen’s neural network to just predict loss, and then we will use the method used in Chantararat et al 2013 to design the contracts. I suspect this will have a similar performance to Chen’s method.

3.3.2 Pitfalls of ignoring risk

In this section, we could go into more detail about the differences in risk for the insurer implied by both methods. We could maybe show the distribution of insurer losses or profits under both methods and the probability of insurer ruin (i.e. probability of negative profits).

4 Robustness

4.1 Effects of shorter dataset

Here, we will artificially shorten the dataset. We can create a graph with dataset length on the x-axis and method performance on the y-axis. The method performance could be all of our previously mentioned performance metrics: farmer average utility, insurer profits, and insurer risk.

4.2 Cost of insuring the US midwest

We could also present a result where we study the scenario of insuring several states in the US midwest. We can design contracts using both methods, and present the results. We could also try adding states one by one and plot how the performance changes as we add more states. Then, we could plot number of states on the x-axis, and model performance on the y-axis. We would just need to figure out a good rule for figuring out the order in which to add states, because it might matter.

5 Extensions

5.1 Interpretability

We could try some methods to make the predictions more interpretable. For example, we could try training a decision tree and see what the performance costs are.

5.2 Bank insuring loan portfolio

We could study the scenario where a bank makes loans to farmers in several states, and wants to purchase insurance for the default of these loans.

5.3 Changes in take-up

We could use estimates from the literature to estimate the change in take-up of index insurance from our method's cost reduction.