

Empirical Evaluation

José I. Velarde Morales

September 7, 2022

1 Overview

This document describes the empirical approach used to evaluate the proposed method of designing agricultural index insurance. I describe the data sources used and how I combine them to create the final dataset that is used to evaluate the model. I conclude by describing my proposed method for evaluating our model using observational data.

2 Data Sources

2.1 NDVI Data

- The Normalized Difference Vegetation Index (NDVI) is a satellite-based indicator of the amount and health of vegetation.
- We have observations from 2000-present, and we have one observation for every 16 days.
- The data is available at a fairly high spatial resolution (250m).
- I downloaded the NDVI data for our area of interest from 2000-2015, the dataset is about 90GB large.
- This has NDVI values for different coordinates.

2.2 Marsabit Household Survey Data

- This survey was conducted as part of the development of the Index based livestock insurance (IBLI) program in northern Kenya. Marsabit is a district in Kenya with high drought risk and where pastoral livelihood systems are common.
- The data includes household location at the village level, there are 5 distinct areas where insurance was offered.
- Livestock loss information (date, number of livestock lost, cause of loss)
- The data is available for 2009-2013 and 2015.

2.3 Kenya Administrative Data

This data is available from the government of Kenya and includes geospatial data of the administrative boundaries of villages. It has the coordinates of the boundaries of the different villages.

3 Evaluation Dataset

This section describes the steps involved in making my dataset. **Note:** I am currently working on step 3.

1. Download NDVI data from NASA website.
2. Merge the NDVI data with the Kenya administrative data to get NDVI values for the villages in the Marsabit household survey data.
3. Use the merged NDVI and administrative data to calculate different features for each village, these features could include cumulative NDVI values over the past year, maximum and minimum NDVI values, and other features that could help predict herd mortality.
4. Use the household survey data to calculate average herd mortality in each village in each season, and merge with the data from the previous step to have a dataset with features (NDVI metrics) and outcomes (herd mortality).
5. Use data from the previous step to create a predictive model that predicts herd mortality based on NDVI. Create a final dataset that has predicted mortality and actual mortality for each village in each time period. I will only use data from the first 4 time periods to train the model.

4 Evaluation

With this dataset in hand, I will then use the method described Chantarat et al 2013 to design the status quo insurance. I will use the same data to design contracts based on our optimization model. I will enforce an identical budget constraint on the two methods, and then evaluate them on the last two years of data based on the following metrics:

- Probability that farmer wealth is below a certain threshold.
- Conditional value at risk of all payouts
- Average farmer wealth conditional on a loss event happening.

I am also considering doing a more advanced simulation procedure, but am still thinking about the details.