

Correlated Losses in Federal Crop Insurance

Jonathon Siegle, Victor Gardner

Overview

Crop failures are typically highly correlated across space and time. This project explores possible models to identify key associations between crop indemnification in the US Federal Crop Insurance Program. While identified pro-cyclical and anti-cyclical pairings would be useful for several purposes, these models sparsity of results and performance against a regional-average baseline implies that further investigation to correlations between crops should consider an alternative route.

Stakeholders

USDA RMA : Group plan pricing, balancing loss ratio experience across system

Firms: Correlated Crops Losses help balance portfolio of farms. Identifying correlated crops could predict the experience of migrating planting to new areas as the climate changes. Underwriters could use correlations to extract rent from the FCIP system.

Approach

The variety of commodities and insurance contracts available in the U.S. food system is too vast to enable interpretable analysis. This analysis prunes that list first by data wrangling to create a meaningfully overlapping dataset, and secondly by using machine learning methods to endogenously identify 'signal' commodities from the noise.

In our first approach, we create a binary tree model in order to understand the roles of the independent variables in determining whether the loss ratio (indemnity / total premium) rose above a certain threshold; we chose a binary tree for its interpretability and for its ability to compute the relative importance of feature variables.

Analysis was limited to the Continental United States. Commodities were chosen based on the most frequently overlapping in insurance records as well as those with the greatest gross liabilities insured by the FCIP and restricting commodities to annual plants. The final list includes Wheat, Cotton, Corn, Grain Sorghum, Soybeans, Pasture and Barley.

Results

Our binary tree model achieved an accuracy of 0.703, with the most important variables being Total Premium, Liability, and Net Planted Quantity [total insured acres]; other competing features included crop and geographic hot coded variables. When coupled with how the tree splits on Total Premium, this tell us that while we may expect insurance companies to price risk in the premium, it seems that large farms with good insurance sometimes take risks due to having good insurance.

From the LASSO analysis, we see that Wheat, Soybeans, Grain Sorghum and Corn more commonly survive the regularization than the other commodities. Shock losses to these commodities plausibly correlate with the substantial majority of considered Crop-Coverage levels we considered. Losses among the different coverage levels were highly correlated for most crops, with a std. deviation in log loss ratio for one associated with 30-50% of a standard deviation in the log loss ratio for the other. However, other

coefficients of interest are typically small even when non-zero. Of the many pairings tested, Corn-Soybeans and Wheat-Barley (for low deductible) associations that large. Furthermore, for substantial alphas, the LASSO typically failed to meaningfully improve on the performance of the baseline model.