Pricing for Rice Crop:

A Vietnamese Example

Mayur Ankolekar

Institute of Actuaries of India,
Crop Insurance Workshop

August 14, 2013
IMC, Mumbai

Contents

I. Crop Insurance Challenges

Indemnity Basis

Although the indemnity basis crop insurance pays a more accurate compensation there are several disadvantages:

- Moral hazard can be higher for farmers, dependent on rainfall, once they have taken out loans.
- Anti-selection
- High expenses (surveyor visits, certification, etc.)
- Correlated risks with potentially large catastrophic risks

1. Crop Insurance Challenges (Contd.)

Index Based - two types:

a) **Weather Index based, and**

b) **Area Yield Index based**

Weather Index Based

A claim is paid depending on how a specified, measured event compares to specified thresholds or triggers.

Area Yield Index Based

A claim is assessed based on shortfall with the agreed yield parameter

1. Crop Insurance Challenges (Contd.)

Advantages:	Disadvantages:
–Negligible delays in reporting and settlement of claims –Claims are linked to an objective and independent source of information –Historic data on these events exists or can be easily collected –Insured event easily verifiable- easy reinsurance –Expenses in verifying claims reduced	–Basis risk might be high –Setting up and maintaining weather stations a challenging task –Constructing suitable indices and a suitable product design may be challenging and subject to more basis risk. –Indexed products may be more difficult to understand and explain esp. if the product design is made more complex to deal with basis risk.

1. Crop Insurance Challenges (contd.)

Technical Challenges, from an actuarial standpoint:

Basis Risk	Risk of choosing a wrong base for the settlement of the claim, resulting in a low correlation between the losses incurred and claims paid out.
Covariate Risk	Large losses from a single event due to geographically concentrated portfolio of risks e.g. poor rainfall throughout a particular region.
Heterogeneity in yields	Actual yields within farms could be heterogeneous and hence lead to difficulty in measuring impact, and also lead to basis risk.
Spatial issues	Rainfall measured at a particular location may not be a good indicator of the rainfall experienced by the farmer.

Contents

Pricing Approaches

1. **Historical Burn Rate**

2. **Pure Risk Premium Approach**

   A. **Predictive modeling using Binary or Ordinal Logistic Regression to the past weather and crop outcome data;**

   B. **Fitting 'full loss probabilities' based on past weather and crop outcome data.**

Preparing for Pricing

When historical burn rate approach i.e. loss ratio experience is not an option (little or no experience), pure risk premium approach would be acceptable.

Starting with the Pure Risk Premium approach:

- Check whether weather station data is correlated enough to aim for uniform pricing in a geography/ state/ country
- Understand the infrastructure of verifying weather data and crop yields
- Break down the farming lifecycle into various stages, with critical and optimum parameters identified for each stage
- Appreciate moral hazard issues and farmer response biases!

Preparing for Pricing: Reference Weather Stations Risk-ANOVA Test

Assuming weather stations do not offer a basis risk (say, each weather station is within 20-25 km of the insured area).

And assuming that weather data is available for long periods that have evidenced weather volatilities, say 20 yr +

Q. If one price is to be decided for the entire area, check if weather is consistent i.e. weather differences measured from weather stations are not significant.

Reference Weather Stations Risk-ANOVA Test

Use non–parametric ANOVA test because assuming normal distribution among weather patterns is not supported by data fitted to distributions.

e.g. Shapiro-Willis test to check if data is normal.

If data is not normal, use non-parametric Kruskal Wallis 1-way ANOVA.

We proceed with 'one price for all farms' once satisfied that the weather parameters are not significantly different.

Logistic Regression for modeling probability of crop losses

Binomial (or binary) logistic regression is a form of regression which is used when the dependent variable is a dichotomy **(For crop insurance- claim (1) or no claim(0))** and the independent variables are of categorical type.

$$\ln\left[\frac{p}{1-p}\right] = \beta_0 + \beta_1 x_1 + \cdots \beta_i x$$

p : Probability (Y=1 Vector x) or Probability (Claim Vector x)

Y : Dependent Variable

$x_1, x_2, \ldots, x_k$: Independent Variables

$\beta_0, \beta_1, \ldots, \beta_k$: Parameters of Model

Parameters of the model are estimated by Maximum Likelihood Method.

Likelihood Function

The Likelihood function is:

$$L(\beta_0 + \beta_1 + \beta_2 + \cdots \beta_i) = \prod p^{yi} (1 - p)^{1-yi}$$

And the statistical model:

$$\text{p} = \frac{e^{(\beta o + \beta 1 x 1 + \cdots .. \beta i x i)}}{1 + e^{(\beta o + \beta 1 x 1 + \cdots .. \beta i x i)}}$$

Building the Logistic Regression model

Quality of questionnaire administered to farmers is critical: how the respondents answer

- Whether there was a crop loss?
- Nature of loss: full or partial or nil?

These responses are considered as categorical or ordinal dependent variables, so logarithmic transformation is necessary.

And tabulated against the continuous weather variables (independent) like temperature, wind speed and rainfall.

Goodness of Fit: Hosmer-Lemeshow Test

H_0 : Model fit is good

H_1 : Model fit is not good

Ten groups are formed based on predicted probabilities.

For each group observed frequency and expected frequency is tabulated and value of χ^2 is calculated

Reject H_0 for large value of chi-square or if $P < 0.05$.

Hosmer Lemeshow test-Example

Partition for the Hosmer and Lemeshow Test					
		CLAIM = 1		NO CLAIM = 0	
Group	Total	Observed	Expected	Observed	Expected
1	56	0	0.31	56	55.69
2	56	1	1.29	55	54.71
3	56	5	2.86	51	53.14
4	56	6	5.19	50	50.81
5	56	7	8.16	49	47.84
6	56	10	12.10	46	43.90
7	56	15	16.88	41	39.12
8	56	24	22.54	32	33.46
9	56	31	30.50	25	25.50
10	56	44	43.17	12	12.83

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
3.4102	8	0.9060

Classification Table

Based on **cut-off value** of p, Y is estimated to be either 1 or zero

Example, if $p > 0.5$; $Y = 1$

$p \leq 0.5$; $Y = 0$

Cross tabulation of observed values of Y and estimated values of Y is called as Classification Table.

The predictive success of the logistic regression can be assessed by looking at the classification table, but classification table is not a good measure of goodness of fit since it **varies with the cut off value set**.

Classification Table-Terminology

Sensitivity	**% of occurrences correctly predicted**
Specificity	% of non occurrences correctly predicted
False Positive Rate	% of predicted occurrences which are incorrect
False Negative Rate	% of predicted non occurrences which are incorrect

Logistic Regression approach (contd.)

Binary model (Y/ N to claim) can be extended to Ordinal model (1/ 2/ 3 for full/ part/ no claim)

We need a large sample of cropping seasons, with answers 'Y/N/ or '1/2/3'

If the sample is not large, the model will be unstable and probability estimates would not stand the goodness of fit test.

Go for other options?

Another approach to pricing

Fitting 'full loss probabilities' based on past weather and crop outcome data.

How?

Understand the process

Plot the observed independent variables

Estimate the loss probability events

Preparing for Pricing:
Staging the Rice Farming Process

Rice Farming process is broken down into three stages.

Stage	No. of days
1. Germination to Rooting	30
2. Leaf Elongation to Anthesis	75
3. Ripening to Harvesting	15

Each stage has different levels of critical weather parameters.

Rainfall i.e. risk of drought and flood is indeed the most important parameter;

Temperature is the next critical whilst

Wind speed has a relatively minor impact

Preparing for Pricing: Approach to evaluate Weather Outliers in Rice Farming

Temperature parameters are as under:

	Critical Low Temperature (°C)	Critical High Temperature (°C)	Optimum Temperature (°C)
Germination	**16-19**	**45**	**18-40**
Seedling Emergence & Establishment	**12-16**	**35**	**25-30**
Rooting	**16**	**35**	**25-28**
Leaf Elongation	**7-12**	**45**	**31**
Tillering	**9-16**	**33**	**25-31**
Panicle Initiation	**15**	**NA**	**NA**
Panicle Differentiation	**15-20**	**30**	**NA**
Anthesis	**22**	**36**	**30-33**
Ripening	**12-18**	**>30**	**20-29**

Preparing for Pricing: Rainfall and Wind speed

Rainfall is perhaps the most important weather parameter in the second stage of leaf elongation and anthesis.

Scanty as well as excessive rainfall can result in losses.

From brainstorming sessions with agriculturists and field visits to rice farms, we decided to plot only those variables that purportedly influence rice farm outputs.

Wind speed came across as an insignificant variable in influencing rice farm outputs.

Data Handling

Data Cleaning	30 years' data for Meteorological Division was cleaned, formatted and arranged.
Season-wise data	Data was sliced per season and tabulated for two seasons of 120 days commencing 15Jan (winter crop) and 15May (spring crop).
Staging of the rice farming process	Rice farming process was broken down into 3 stages: viz. germination to rooting, leaf elongation to anthesis, and ripening to harvesting

Data Handling … Contd.

Brainstorming for critical values	Critical values of temperature, rainfall and wind speed for each stages were arrived after a brainstorming session comprising agronomists.
Commune responses	For the past 15 years i.e. 30 seasons, simple responses were elicited to the question: "What was the crop yield experience – no loss, partial loss, full loss?". These were further indicative for critical values of temperature, rainfall and wind speed.
Distribution fitting	Statistically, the data was fitted into an appropriate distribution with an acceptable goodness of fit.

Loss Probability Calculation Approach

Short listing Variables

Of the 37 random variables initially considered, we arrived at 4-5 important variables that would matter the most.

These are the defined variables for full loss

Stage 1: **NA**

Stage 2:

P (Average Temp) < '18°C'

P(Max Rainfall < '19mm' & (Number of days Temp < '10°C')) > 2

P (Two days of Rainfall) > 200mm)

P(Max Temperature) > 45°C

Stage 3:

P(Max Temperature) > 45°C

P(Max Rainfall) > 250mm

Loss Probability Calculation Approach (Contd.)

Could not use multiple linear regression due to limited full loss (1) and total observations (30)

Identified ‘full loss’ conditions

: possible only in stages 2 and 3

Fitted distributions to the critical weather data to calculate probability of full loss for each weather parameter

Full Loss Probability Results from Distribution Fitting

	Season 1 (Jan15 - Apr15)		Season 2 (Jun15 - Oct 15)	
	Distribution fitted	**Probability**	**Distribution fitted**	**Probability**
Critical Event Stage 2 (leaf elongation to anthesis)				
Average Temp < 18°C	Burr	**3.0480%**	GEV	**Nil**
Max rainfall < 19mm and At least 2 days temp < 10°C	Burr & Poisson	**0.0096%**	Dagum & Poisson	**Nil**
At least 2 days of rainfall > 200mm	Poisson	**Nil**	Poisson	**0.2430%**
Max Temp > 45°C	Beta	**Nil**	Log Logistic	**0.0290%**
Critical Event Stage 3 (ripening to harvesting)				
Max rainfall > 250mm	GEV	**Nil**	Log Pearson3	**3.6120%**
Max Temp > 45°C	Johnson SB	**Nil**	Gen Gamma	**Nil**
TOTAL		**3.0574%**		**3.8741%**

Intersections of the probabilities adjusted to arrive at the probability of a claim: note the final probability is not an addition of the tabular probabilities!

Extending full loss probabilities to partial losses

It is appropriate to corroborate the full loss probabilities using distribution fitting with simulation runs (min. 5,000 runs) from the past weather data

Full loss probabilities is only part of the problem to estimate the Expected Claims i.e. Summa (frequency x severity)

Frequencies of partial losses and their severity is a greater challenge

Use farmer responses on partial losses and "discretise the distribution." Some broad brush estimation is expected.

Expected claims will mostly be attributed to full, covariate losses!

Pure Risk Premium

Pure Risk Premium=

Expected Yields x Average Price x Probability of a claim

Non-stationary Random Variable

From Loss probability fitting

Irrigated	4-6 tonnes per hectare
Rain Fed Low land	2-3 tonnes per hectare
Deep Water	0.5 -2 tonnes per hectare
Upland	1-2 tonnes per hectare

Contents

Providing Margins of Adverse Deviations

Time Series Analysis

Bifurcating the sample yields into 5 year intervals with a representative average rainfall/ temperature

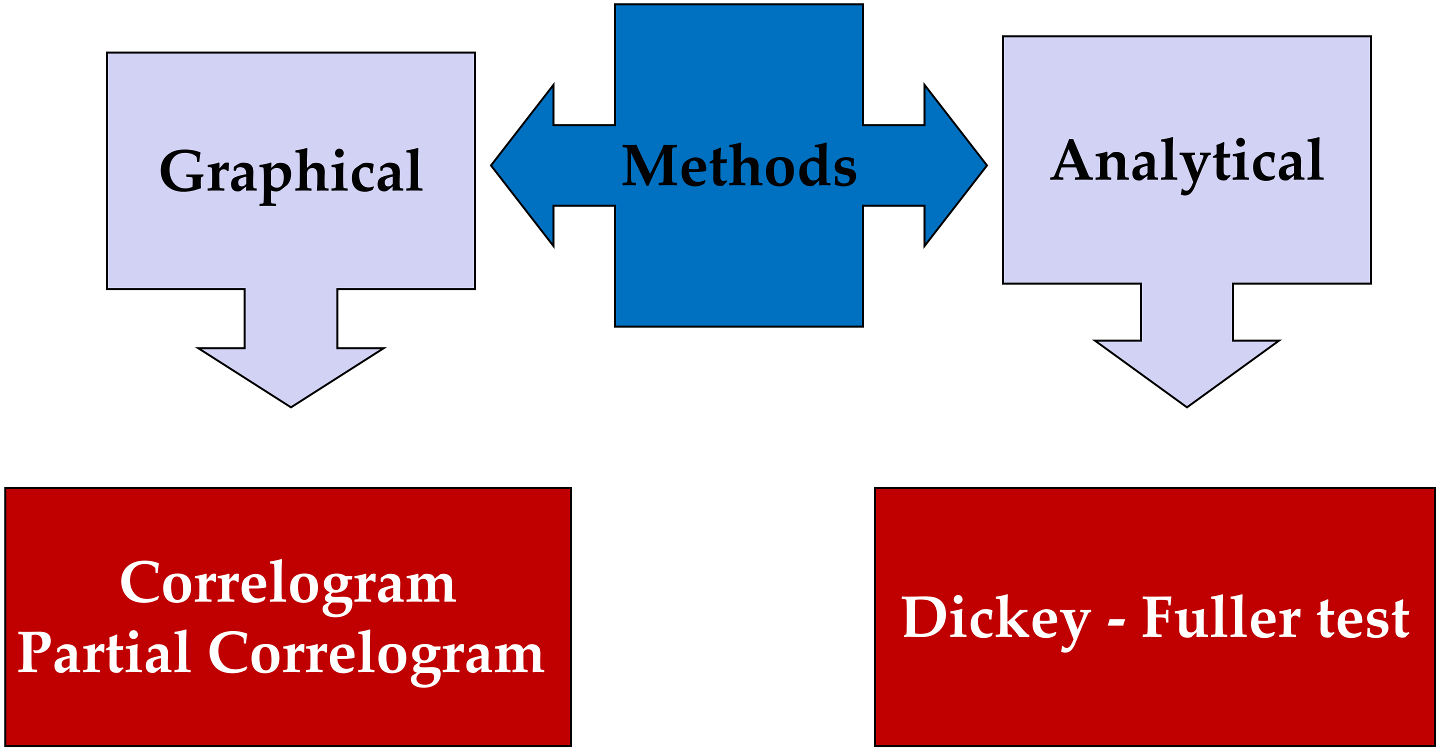
Considering the first order differences in rainfall/ temperature

Ensure that it is stationary (indicator of 'climate change trends' but only on rainfall – the critical weather for rice crop)

Note: Stress on observations of latest years!

Else, add to the pure risk premium

Identifying Stationary Time Series



Time Series Models

If a time-series is stationary, it can be modeled as

Autoregressive (AR) Process,

Moving Average (MA) Process,

Autoregressive Moving Average (ARMA) Process,

Integrated Process, and

Autoregressive Integrated Moving Average (ARIMA) Process

Weather data that points to Climate Change trends is intuitively autoregressive and not moving average (Source: own conclusion)

Contents

Covariate Risks' Impact on Cost of Capital

Test for multi-collinearity amongst weather parameters that are thought to be independent

Check impact on company/underwriter's portfolio, most likely will build a concentration risk

Greater covariate risks need higher economic capital – build a higher cost of capital

A virtual certainty for crop insurance

Contents

The Pricing Equation

Crop Insurance Premium =

Pure Risk Premium (using historical burning cost/ logistic regression) +

Margins for Adverse Deviations (time series analysis on climate change trends) +

Cost of capital for covariant risks (multi-collinearity of weather parameters & geographical concentration) +

Cost of capital (supervisory) +

Expenses +

Profit loading

Thank You