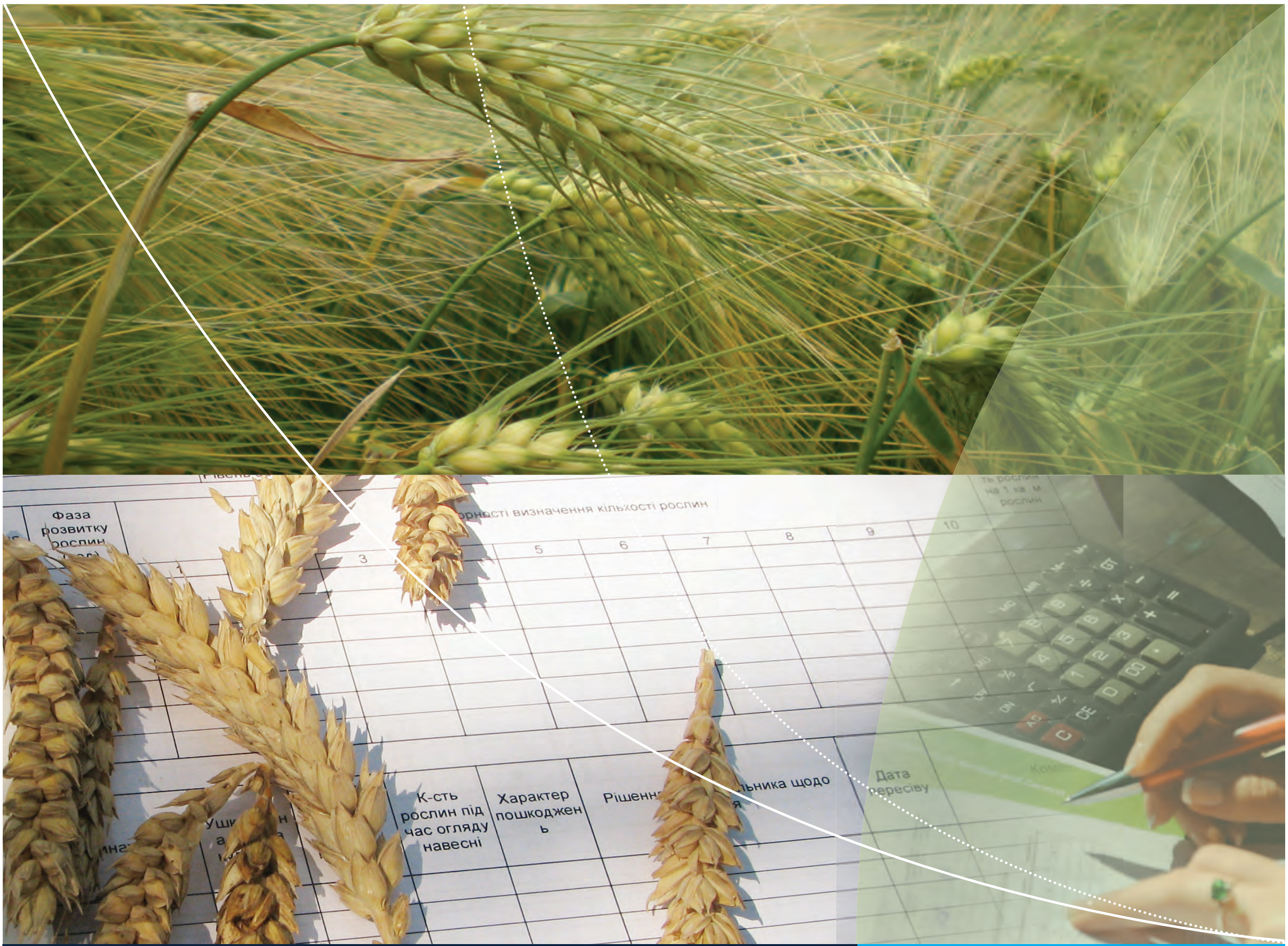


ACTUARIAL BASICS

Manual 2

WORLD BANK GROUP

2121 Pennsylvania Avenue, N.W.
Washington, D.C. 20433
www.ifc.org

ACTUARIAL BASICS

Manual 2

IN PARTNERSHIP WITH DONORS

WORLD BANK GROUP

Table of Contents

Preface and Acknowledgements

Five manuals were prepared by IFC for the development of agri-insurance markets where the public and private sectors work together in a partnership (PPP). The manuals are designed to strengthen the capacity of the government and market players to effectively design agri-insurance products, both traditional indemnity and index, introduce them to the market, and build sales. The manuals are designed to be succinct yet at the same time sufficient to create the technical and administrative foundation for a modern agri-insurance system, and to allow programs in early stages of development to properly plan the required system. Finally the manuals are designed to train practitioners, to build local capacity for skills that are required to start the program, and to enable the program to grow over time.

The principle author of the manuals is Professor Myles Watts, University Professor, Lead Actuary at Watts & Associates, Member of the Board at the Federal Agricultural Mortgage Corporation, and 5th Generation Montana Farmer. Watts and Associates designed and launched numerable agri-insurance products in North America, frequently consults for the major reinsurers, and supports insurance programs around the world. They have established their own index insurance company, eWeatherRisk. The manuals incorporate practical lessons learned over the past 40 years.

The development of the manuals was a joint activity of the Ukraine Agri-Insurance Project (2007-2015), IFC's Global Agri-Finance Team, and the Global Index Insurance Facility (GIIF) (2009 to present). Dr. Gary Reusche led the Ukrainian project, served as a technical specialist on the global agri-finance team, and as a member of the GIIF technical committee and core management team. Agri-insurance development is closely linked to agricultural finance and value chains and they are effectively developed in unison.

The manuals result from training workshops developed by the agri-insurance project in Ukraine and globally by GIIF technical experts. The entire agri-insurance team in Ukraine made practical contributions to the manuals, with special recognition due to Victoria Yakubovich for collecting, organizing and preparing the initial drafts and Andrey Zaripov a member of the GIIF team for helping to develop the reinsurance and cash flow models. The project team included experts from the Alberta (Canada) provincial agri-insurance program, in particular Richard McConnell, who contributed his experience and expertise to the training activities.

Peer review and Spanish language translations of the manual resulted from IBRD consultants in Central and South America, especially Pablo R. Valdivia Zelaya and Roberto Dario Bacchini.

The team is grateful to Professor Gary Brester for his practical contributions and editing of the manuals and Olesya Zhuchenko for coordinating the publication of the manuals, working with designers to plan the layout, and printing.

Finally support for the manuals was provided by the Canadian government, and the Global Index Insurance Facility (GIIF) lead by Gilles Jacques Galludec (Program Manager) funded by the European Union, Japan and the Netherlands.

Acronyms

IFC – International Finance Corporation

GIIF – Global Index Insurance Facility

IIARM – International Institute for Agricultural Risk Management

DFATD – Foreign Affairs, Trade and Development Canada

CDF – Cumulative Distribution Function

GHCND - Global History Climate Network – Daily

GSOD – Global Summary of Day

HA – Hectares

PDF – Probability Distribution Function

UAH – Ukrainian Hryvnia, money unit of Ukraine

1.0. Introduction

This manual presents theoretical concepts and applied methods commonly used for actuarial processes in agriculture insurance. Actuarial processes refer to the activities that establish insurance premium rates and related quantitative analyses. Key considerations for designing agriculture insurance programs are presented. In addition, the influence of program design on data requirements and the use of statistical methods for establishing insurance premiums are illustrated.

The presentation presumes that readers have a solid understanding of basic statistics principles. The level of knowledge would be equivalent to material contained in an introductory university-level statistics textbook. Readers are encouraged to familiarize themselves with basic statistical principles, including descriptive statistics, hypothesis tests, sampling, regression analysis, and distribution theory.

Although most of this document is devoted to actuarial issues, program design features are noted throughout. For conciseness, the manual will discuss actuarial processes associated with a specific agricultural insurance product: yield insurance. The principles and methods presented are, in general, applicable to other crop insurance products, but every insurance product has unique features that further complicate actuarial processes.

1.1. Stakeholder Interest

The major stakeholder groups associated with agriculture insurance are agricultural producers, insurance firms, and governments. Each has different, and sometimes opposing, interests.

Agricultural Producers want to generate net income, recover production costs, gain access to credit, improve expected profitability, and reduce risk.

Insurance Firms want to generate profit for shareholders, diversify risk portfolios, and avoid losses. This sector includes both issuing agencies (those firms that market and often service insurance contracts) and reinsurers (firms that accept most of the risk acquired by issuing agencies in exchange for fees).

Governments often seek to stabilize rural incomes, stabilize and expand agricultural output, and provide disaster relief. Government roles can range from providing all aspects of an agricultural insurance program to providing regulatory oversight and legal structures.

2.0. Crop Insurance Basics

Crop insurance shares many elements associated with other forms of insurance. The purchaser of an insurance contract pays a premium to an issuing agency to transfer undesirable outcome risks. Actuarially sound premiums are established such that expected indemnities (payouts in the case of insured events) and costs of providing insurance are offset by premium collections. An issuing agency often pays a reinsurer to accept much of the risk that has been acquired through the sale of insurance contracts.

2.1. Definitions

Indemnity Payments represent a transfer of funds from an insurer to an insured party in order to partially or fully compensate for insured losses.

Deductibles are the portions of a claim that are not offset by an insurance product in the case of a loss. An insurance purchaser must absorb the deductible before an insurer will provide an indemnity payment.

Coverage Level is one minus the deductible:

$$\text{Coverage Level} = 1 - \text{Deductible}$$

Expected Yield is the weighted (usually by probability of occurrence) average of all possible yields. Often, it is estimated by the simple average of historical yields.

Trigger Yield: A yield below the trigger yield justifies an indemnity payment.

$$\text{Trigger Yield} = \text{Expected Yield x Coverage}$$

Liability represents the maximum possible indemnity payment. This occurs if the insured experiences a complete loss (i.e., the harvested yield is zero):

$$\text{Liability} = \text{Expected Yield x Coverage x Commodity Price}$$

Liability increases as deductibles decline (i.e., as coverage levels increase).

Pure Risk Premium is the expected indemnity. The expected indemnity is the weighted (usually by probability of occurrence) average of all possible indemnity payments.

Pure Risk Premium Rate is calculated by dividing the Pure Risk Premium by the Liability:

$$\text{Pure Risk Premium Rate} = \text{Pure Risk Premium/Liability}$$

Load is the cost of servicing an insurance contract.

Load Rate is calculated by dividing the Load by the Liability.

$$\text{Load Rate} = \text{Load/Liability}$$

Total Premium represents the price that an insured pays for insurance contract and is the sum of the pure risk premium and the load.

$$\text{Total Premium} = \text{Pure Risk Premium} + \text{Load}$$

Total Premium Rates are often stated as a proportion of liability.

Total Premium Rate = Total Premium/Liability

A Subsidy is the proportion of a total premium rate that is paid by a government or an entity other than the insured party.

Subsidy Rate is calculated as a proportion of the total premium.

Subsidy Rate = Subsidy/Total Premium

Producer Premiums are calculated as the total premium less any applicable subsidies. It represents a producer's cost of insurance:

Producer Premium = (1 – Subsidy Rate) x (Total Premium)

2.1.1. Yield Insurance Example: Suppose that a producer wants to insure a wheat crop against potential weather-related perils. Assume the following values:

Deductible = 40%

Pure Risk Premium Rate = 6%

Load Rate = 3%

Subsidy Rate = 25%

Price of Output (wheat) = 1.0

1 hectare is insured

Liability is based on average yields. Suppose that the producer has the following production history:

Year	Historical Yield (tons per hectare)
1	2.7
2	3.6
3	2.4
4	3.3

Expected (or average) yield = 3.0 tons/ha

Liability = (Expected Yield) x (Coverage) x (Price) x (Area) =

(3 tons / ha) x (1 – 0.4) *x* ($1 / *ton) x* (1.0 *ha)* = $1.80

Total Premium Rate = (Load Rate + Pure Risk Premium Rate) = (3% + 6%) = 9%

Total Premium = (Liability x Total Premium Rate) = ($1.80) *x* (0.09) = $0.162

Subsidy = (Subsidy Rate) x (Total Premium) = (0.25) *x* ($0.162) = $0.0405

Producer Premium = (1 – Subsidy Rate) x (Total Premium) = (1-0.25) *x* ($0.162) = $0.1215

Trigger Yield = (Coverage) x (Expected Yield) = (1 – 0.4) x (3 tons/ha) = 1.8 tons/ha

If actual yield is greater than the indemnity trigger, no indemnity is paid. For example, if the actual yield is 2 tons/ha, then an indemnity is not generated. If actual yield is less than the indemnity trigger, however, then an indemnity is due. If the actual harvest yield totals 1 ton/ha, then the indemnity payment is calculated as:

Indemnity payment = (Trigger Yield – Yield Outcome) x (Price) x (Area) =
(1.8 *tons/ha* – 1 *tons/ha) x* ($1/*ton) x* (1.0 *ha)* = $0.80

3.0. Data

Insurance premium rates are generally established using statistical estimation methods applied to historical data. That is, the probability of the occurrence of an insured event is determined by the historical frequency of its occurrence. Consequently, appropriate estimation methods must be used to accurately forecast the probability of future occurrence. In addition, the quality and quantity of historical data greatly influence the accuracy of premium rating. Data must be carefully reviewed and analyzed. These reviews can range from simple graphs and plots to complex statistical analyses. If possible, data should be cross-referenced with other sources, evaluated by well-informed market participants, and compared to other variables that are correlated over time. Data collection and sampling techniques must be scrutinized. Finally, data generation and analyses must be documented so that future analysis, reinsurance, and updating can be conducted appropriately.

3.1. Data Quality

Data quality depends on sampling, availability, collection, entry, and accuracy. Analysts must recognize that data quality varies for many reasons. Hence, it is important that data be evaluated for accuracy and relevancy. Special care must be taken to identify systematic anomalies. For example, if missing farm-level yield data occur most frequently when regional yields or growing season participation were low, then missing yields may signal a systematic problem. Systematic problems must be addressed prior to developing premium rates. The determination of such systematic problems often requires actuarial expertise.

3.1.1. Data Cleaning refers to the practice of identifying and managing outlier observations and other irregularities that result from collection, transcription, coding, entry, or other errors. For example, a producer's yield data could contain an observation several times greater than the mean of the observations obtained by a group of producers from the same region (e.g., an observation of 30 tons/hectare when the mean of the data is 3.0 tons/hectare). After careful review of all internal data management processes, several options exist to manage such anomalies including: (1) deleting the outlier observation, (2) replacing the outlier observation with a more "reasonable" estimate, or (3) revisiting the initial reporting and coding activities of individuals.

The last option is often prohibitively expensive because of temporal, spatial, and volume considerations. Consequently, average regional yields are often used as replacements for outliers in the case of individual yield data. Nonetheless, insurance indemnities are triggered by low yields. While the preceding example of a 30 ton/hectare yield is implausible, a very small yield (e.g., 0.30 tons/hectare) is a possibility. Plausible high yields are also important for insurance rating because premiums and indemnity payments are based on expected yields.

When using relatively small data sets, graphs often help identify outliers, missing observations, and other anomalies. For larger data sets, software must often be used to uncover such anomalies. After review, certain observations may need to be deleted from a data set so that the rating process remains unbiased.

Often, judgment must be used when cleaning data. Consequently, the process is somewhat subjective. Therefore, it is important to include experienced personnel throughout the process. The removal or replacement of specific data points can have a substantial influence on the rating process.

3.1.2. Missing Data points result from a variety of causes. In some cases, a data point may not have been created because a crop was not produced. In other cases, records may not be available. In addition, simple data entry errors can occur.

A missing data point creates two problems. First, information about the actual value of that observation (if any) is lost, causing problems for rating the process.

A second problem is created because electronic databases often require that a data entry point exist, even if it is unavailable. For example, if a zero is entered for a data point that was actually missing, then the rating process assumes that the yield for that year was actually zero. In fact, only seldom do we encounter situations that result in an actual zero yields. Some data management software considers a blank to be a missing observation. If a value must be entered, one needs to enter an implausible number as an indicator of a missing observation. For example, a missing observation may be indicated by entering a number such as -999.

Every reasonable effort should be made to obtain correct values for missing data. If the missing values cannot be obtained, then analyses may be necessary to evaluate whether the missing data is systematic, which usually requires sophisticated techniques. For example, if farm data appear to be more frequently missing when regional yields or precipitation has been low, then further investigation is warranted. If missing data are deemed to be highly systematic, then the data may have to be abandoned.

4.0. Actuarial Methods

Actuarial rating methods include activities used to determine liability, liability triggers, pure risk premiums, and loads.

4.1. Determining Liability and Indemnity Triggers

A variety of methods can be used to determine liability and indemnity triggers, including average yields, indexing, regional yields, and transition yields.

4.1.1. Farmer Average Yield. The most common way to determine a **Yield Indemnity Trigger** is to multiply an individual farmer's historical average (expected yield) by a selected coverage level:

$$t = c\overline{X},$$

where t is the indemnity trigger, c is the coverage level, and $\overline{X}$ is a farmer's historical arithmetic mean or average yield. This approach is easily understandable and commonly used. However, if the sample used to determine $\overline{X}$ is relatively small, then the variance of the expected yield may be large. For example, let s_X^2 be the variance of yield, X. Then, the variance of the average yield is:

$$s_{\overline{X}}^2 = \frac{s_X^2}{n}.$$

As n declines, the estimated variability of $\overline{X}$ increases.

4.1.2. Indexing procedures determine liability by comparing individual farm yields to other information (such as regional yields) to ascertain if abnormally high or low individual yields were consistent with regional experiences. This approach may also be used to account for changes in technology causing yields to trend upwards over time.

4.1.3. Regional Yields. It is also possible to use regional expected yields as a proxy for individual expected yields. This method is simple and, because it does not require individual farm-level data, is less data intensive. Nonetheless, this approach masks actual differences in yields among farms. Consequently, the approach often increases problems related to adverse selection.

4.1.4. Transition Yields. If farm-level data are used to establish liability, provisions may be necessary for those farmers who have no historical data but wish to purchase crop insurance. Usually, regional yields (or a proportion thereof) may be used for this purpose. Actual farm yields are substituted for transition yields as data become available.

4.2. Commodity Pricing

Establishing a commodity price for each insured agricultural product is necessary for calculating liability, premiums, and indemnity payments. For yield insurance, a single price is stipulated in insurance contracts prior to purchase. This price is used to calculate the value of yield losses should an indemnity be triggered. The price is also a component of calculating liability and premiums.

In the above yield insurance example, a commodity price was needed to calculate liability. In turn, the selected price affects premium and subsidy calculations as well as indemnity payments.

Expected commodity prices used in insurance contracts can be estimated in a variety of ways including:

- Historical average prices;
- Price forecasts from experts;
- Costs of production;
- Futures markets.

Where such markets exist, using futures markets is the most common approach. In cases where domestic futures markets do not exist, foreign futures contract prices may need to be adjusted by exchange rates. However, such processes may increase the price basis embedded in insurance contracts.

If actual commodity prices are lower than the previously agreed upon insured price, farmers have an incentive to seek indemnity payments rather than market their production. To avoid the associated moral hazard, commodity price selections for indemnity calculations should be sufficiently low to encourage crop stewardship and harvesting rather than the pursuit of indemnity payments.

4.3. Farm-Level Actuarial Processes

Farm-level actuarial processes begin with estimates of pure risk premium rates. Rating is conducted using an empirical process, a parametric process, or a combination of the two. Empirical rating methods use historical data to calculate premiums directly. Parametric rating methods use specific probability distributions to establish premium rates. Distributions are selected based on their similarity to the believed (although unknown) underlying yield generating function. A selected distribution's parameters are usually estimated from historical data. Parametric rating methods are often used in cases where historical data are not of sufficient quantity or quality to allow the use of empirical rating methods. The most common probability distributions used to rate insurance products include normal, triangular, uniform, log-normal, beta, and extreme value distributions. In general, the quantity and quality of historical data dictate the choice of rating method.

4.3.1. Pure Risk Premium Rate Calculation. Assume that:

(1) I is the indemnity payment associated with a random event;

(2) X is a random variable that quantifies the outcome of a random event. This could be crop yields (for yield insurance), or precipitation (for weather index insurance products);

(3) t is an indemnity trigger (calculated by multiplying the coverage level by the expected yield).

(4) Price is equal to 1 for simplicity of exposition (hence, it does not appear in the following calculations).

Indemnity payments are calculated as:

$$I = t - X \;\; if \;\; X < t$$

$$I = 0 \;\; if \;\; X \geq t$$

Let $f(X)$ be the probability density function of X. Then, the expected indemnity is:

$$E(I) = \int_0^\infty I \; f(X)dX \; = \int_0^t I \; f(X)dX$$

Because yields are truncated at zero and indemnities are likewise zero at levels above t, integration only needs to occur for values of X between 0 and t such that:

$$E(I) = \int_0^t If(X)dX = \int_0^t (t - X)f(X)dX = \int_0^t tf(X)dX - \int_0^t Xf(X)dX$$

$$= t\int_0^t f(X)dX - \frac{\int_0^t Xf(X)dX}{F(t)}F(t) = tF(t) - E(X/X < t)F(t) = [t - E(X/X < t)]F(t),$$

where $F(t)$ is the cumulative density of X at t.

The term $[t - E(X/X < t)]$ is the expected indemnity when an indemnity is paid. This term represents the severity of indemnity payments. The term $F(t)$ is the probability that an indemnity payment will be paid and is called the **frequency** of indemnity payments.

Insurance companies need to understand both the severity and frequency of indemnity payments to determine the level of required financial reserves and servicing costs. For example, loss adjusting is heavily dependent on frequency of losses while necessary financial reserves are dependent on both frequency and severity of losses.

The expected indemnity is: $E(I) = Severity \; x \; Frequency$

The **Pure Risk Premium Rate** (w) is equal to the expected indemnity payment divided by the liability (or trigger):

$$w = \frac{E(I)}{Liability} = \frac{E(I)}{t} = \frac{[t - E(X/X < t)]F(t)}{t}$$

Notice the liability and trigger are equal because price has been set equal to 1. Even if price is not equal to 1, the pure risk premium rate is unaffected by price because price would appear in both the numerator and denominator of the pure risk premium rate. However, premiums (as opposed to the premium rate) and indemnity payments are affected by price.

For a discrete case:

$$E(I) = \sum_{j=1}^{K} (t - X_j)f(X_j) = [t - E(X/X \leq t)]F(X_K),$$

where X_j is ordered so that the first K values are those below the indemnity trigger t. Then,

$$w = \frac{E(I)}{t} = \frac{\sum_{j=1}^{K}(t - X_j)f(X_j)}{t}$$

4.3.2. Empirical Rating Processes involve calculating the size of an indemnity payment each time it would have been triggered within an historical set of data or using actual loss histories from an existing insurance program. The frequency of occurrence is calculated by dividing the number of indemnity occurrences by the total number of observations in the data set. Severity is calculated by summing the total value of indemnity payments over the data set and dividing it by the number of indemnity occurrences.

Multiplying frequency by severity provides the expected indemnity payment.

4.3.2.1. Empirical Rating Example. Assume that the following situation occurs for a producer:

$$Acreage = 1\ hectare$$

$$Price = 1\ UAH\ per\ ton$$

$$Deductible = 40\%$$

In addition, the producer has the following yield history, for which the indemnity trigger and payments are:

Table 4.1. Simple Rating Example for Yield Insurance

Year	Yield Outcome	Indemnity Trigger	Indemnity Payment
1	2.70	1.8	0.00
2	1.72	1.8	0.08
3	3.24	1.8	0.00
4	4.28	1.8	0.00
5	4.20	1.8	0.00
6	4.73	1.8	0.00
7	0.32	1.8	1.48
8	2.77	1.8	0.00
9	4.10	1.8	0.00
10	1.92	1.8	0.00

Therefore,

$$Expected\ Yield = Mean\ Yield = 3.0$$

and,

$$Coverage = (1 - deductible) = 60\%.$$

The Indemnity Trigger yield for this example is calculated as:

$$Indemnity\ Trigger$$
$$= (Coverage)\ x\ (Expected\ Yield) = (0.60)x(3) = 1.8\ tons/ha.$$

An indemnity is paid in any year in which actual yield outcome is less than 1.8 tons/ha. The actual yield outcome is lower than the indemnity trigger in years 2 and 7. Indemnity payments are calculated as:

$$Indemnity\ Payment_{year\ 2}$$
$$= (Indemnity\ Trigger\ - Yield\ Outcome_{year\ 2})\ x\ (Price)$$
$$= (1.80\ - 1.72)x(1) = 0.08\ UAH/ha$$

$$Indemnity\ Payment_{year\ 7}$$
$$= (Indemnity\ Trigger\ - Yield\ Outcome_{year\ 7})\ x\ (Price)$$
$$= (1.80\ - 0.32)(1) = 1.48\ UAH/ha.$$

The severity is:

$$Severity = (0.08 + 1.48) / 2 = 0.78.$$

The frequency is:

$$Frequency = 2/10 = 0.20.$$

The expected indemnity is:

$$E(I) = severity\ x\ frequency = (0.78)(0.20) = 0.156.$$

The expected indemnity payment can also be calculated as:

$$Expected\ Indemnity\ Payment = \frac{Sum\ of\ Indemnity\ Payments}{Number\ of\ Observations}$$

$$= \frac{0.08 + 1.48}{10} = 0.156$$

In this case, the liability is calculated as:

$$Liability = Indemnity\ Trigger\ x\ Price.$$

$$Given\ that\ Price = 1.0,$$

$$Liability = 1.8\ \text{x}\ 1.0.$$

The pure risk rate is calculated as the quotient of the expected indemnity payment and the liability:

$$Pure\ Risk\ Rate = \frac{Expected\ Indemnity\ Payment}{Liability} = \frac{0.156}{1.8} = 0.087\ .$$

Assuming a load of 3%, the total premium rate is 11.87% or (0.1187). The total premium is 0.1187 times the liability of 1.8, or 0.2136 UAH/ha. This example is presented for illustration purposes only. Ten observations are not sufficient for actuarially sound rating processes. In addition, the price of a commodity greatly influences total liability.

4.3.3. Parametric Rating. A variety of parametric rating approaches can be used. The most common distributions used for this purpose are the uniform, triangular, normal, log-normal, beta, and extreme value functions. The general approach is to estimate or assume a specific distribution of yields and then calculate pure risk rates, premiums, and indemnities based on the selected distribution. The parameters of the distribution may be calculated from historical data or may be based on agronomic or other information.

4.3.3.1. The Uniform Distribution is the simplest to use and provides an upper bound for premium rates. The uniform distribution is bounded by a minimum and a maximum, and every outcome between the two is presumed to have equal probability of occurrence. The uniform distribution has two parameters – a minimum value (*a*) and maximum value (*b*).

The **Cumulative Distribution Function** (cdf) for the uniform distribution (figure 4.1) is:

$$F(\underline{X}) = \frac{\underline{X} - a}{b - a} \quad \text{for } a \le X < b$$

$$F(\underline{X}) = 0 \quad \text{for } \underline{X} < a$$

$$F(\underline{X}) = 1 \quad \text{for } \underline{X} \ge b$$

The **Probability Density Function** (pdf) is (figure 4.2):

$$f(\underline{X}) = \frac{1}{b - a} \quad \text{for } a \le \underline{X} < b$$

$$f(\underline{X}) = 0 \quad \text{for } \underline{X} < a$$

$$f(X) = 0 \quad \text{for } \underline{X} \ge b.$$

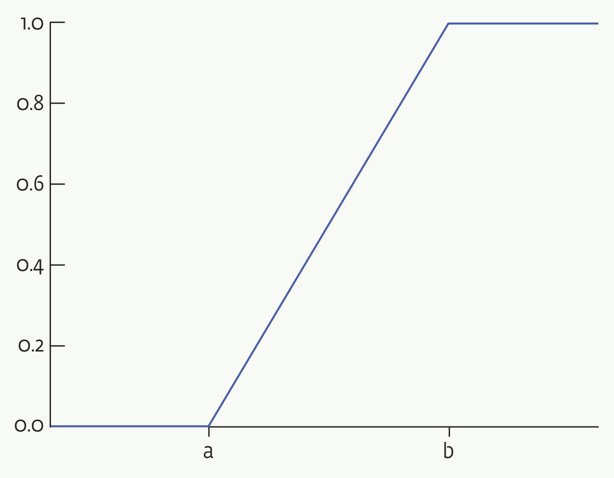


Figure 4.1. Cumulative Distribution Function

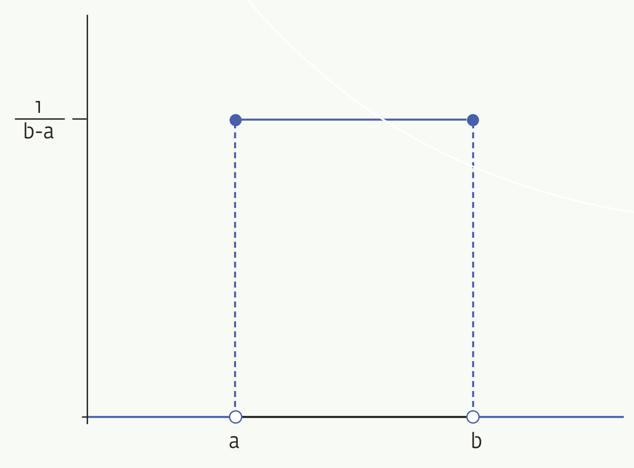


Figure 4.2. Probability Density Function

Mean ($\overline{X}$) = **Median** = $\dfrac{a+b}{2}$

Mode does not exist because each outcome has an identical probability of occurrence.

Variance (s^2): = $\dfrac{(b-a)^2}{12}$, **Standard Deviation** (s): = $\dfrac{b-a}{\sqrt{12}}$.

Coefficient of Variation = $\dfrac{b-a}{(b+a)\sqrt{3}}$.

Coefficient of Skewness: = 0 because the distribution is symmetric.

Coefficient of Kurtosis: = -9/5.

Recall that:

$$E(I) = severity\ x\ frequency.$$

The frequency for the uniform distribution is:

$$F(t) = \frac{t-a}{b-a}.$$

The severity is:

$$t - E(X / X \leq t).$$

Where:

$$[E(X / X \leq t)] = \frac{\int_a^t Xf(X)dx}{F(t)}.$$

However,

$$\int_a^t Xf(X)dX = \frac{1}{b-a}\int_a^t XdX = \frac{t^2-a^2}{2(b-a)}.$$

So,

$$[E(X / X \leq t)] = \frac{t^2-a^2}{2(b-a)\dfrac{(t-a)}{(b-a)}} = \frac{1}{2}(t+a).$$

Severity is:

$$t - ½(t+a) = ½(t-a);$$

$$E(I) = \frac{1}{2}(t-a)\frac{(t-a)}{(b-a)} = \frac{(t-a)^2}{2(b-a)}.$$

The **Pure Risk Premium Rate** is

$$w = \frac{E(I)}{t} = \frac{(t-a)^2}{2(b-a)t}.$$

An interesting situation occurs if $a = 0$ (lowest crop yield =0).

Then,

$$\frac{b}{2} = \overline{X}, \text{ so } b = 2\overline{X}.$$

Therefore,

$$E(I) = \frac{t^2}{2(2\overline{X})} = \frac{t^2}{4\overline{X}}, \text{ and}$$

$$w = \frac{t^2}{4\overline{X}t} = \frac{t}{4\overline{X}}.$$

However, $t = c\overline{X}$, where c is the coverage level, then

$$E(I) = \frac{(c\overline{X})^2}{4\overline{X}} = \frac{c^2\overline{X}}{4}, \text{ and } w = \frac{c^2\overline{X}}{4t} = \frac{c^2\overline{X}}{4c\overline{X}} = \frac{c}{4}.$$

If the first two moments of the distribution ($\overline{X}$ and s) are available, then

$$\overline{X} = \frac{a+b}{2} \text{ and } s = \frac{b-a}{\sqrt{12}}.$$

It can be shown that

$$a = \overline{X} - \sqrt{3}\, s \text{ and } b = \overline{X} + \sqrt{3}\, s.$$

Pros and Cons of using the Uniform Distribution:

Pros:

1. The uniform distribution is simple and easily understood. If $a=0$, then the premium rate is the coverage level divided by 4 which can be quickly calculated.
2. The process results in conservative estimates of premium rates (upper bounds) relative to more realistic distributions.
3. If empirical premium rates obtained from actual loss histories are larger than those obtained from a uniform distribution, then the underlying data set may reflect problems such as substantial fraud influences.

Cons:

1. Most yield distributions are unimodal and the mode does not exist for the uniform distribution.
2. Premium rates are often higher than those obtained from more realistic distributions.
3. The uniform distribution is very inflexible. One can obtain unrealistic values for the *a* and *b* parameters even if reasonable estimates of mean and variance are used.

Note: *When a distribution for a random yield is unknown, there is some justification from statistical theory for using a "diffuse prior," such as a uniform distribution.*

4.3.3.2. The Triangular Distributions characterized by three parameters: a minimum, a maximum and a mode.

Assume that:

a = minimum value;

b = maximum value;

d = mode.

The triangular distribution may or may not be symmetric and may exhibit either negative or positive skewness.

The Triangular Cumulative Distribution Function (figure 4.3) is:

$$F(\underline{X}) = 0, \text{ if } \underline{X} < a$$

$$F(\underline{X}) = \frac{(\underline{X}-a)^2}{(b-a)(d-a)}, \text{ if } a \le \underline{X} \le d$$

$$F(\underline{X}) = 1 - \frac{(b-\underline{X})^2}{(b-a)(b-d)}, \text{ if } d < \underline{X} \le b$$

$$F(\underline{X}) = 1, \text{ if } \underline{X} > b$$

The Probability Density Function (figure 4.4) is:

$$f(\underline{X}) = 0, \text{ if } \underline{X} < a$$

$$f(\underline{X}) = \frac{2(\underline{X}-a)}{(b-a)(d-a)}, \text{ if } a \le \underline{X} \le d;$$

$$f(\underline{X}) = \frac{2(b-\underline{X})}{(b-a)(b-d)}, \text{ if } d < \underline{X} \le b$$

$$f(X) = 0, \text{ if } \underline{X} > b$$

Figure 4.3. Cumulative Distribution Function

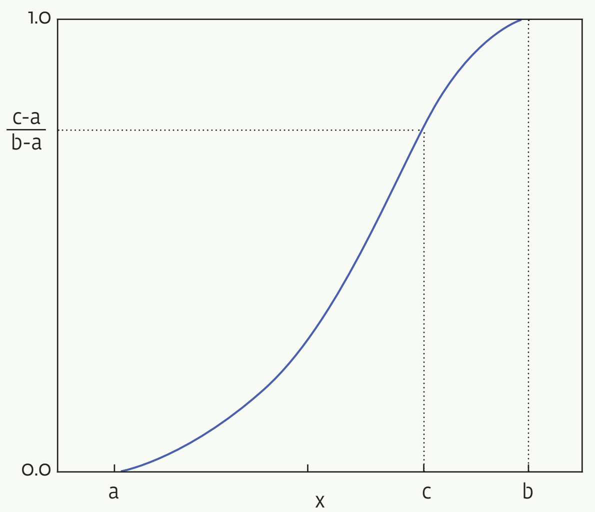
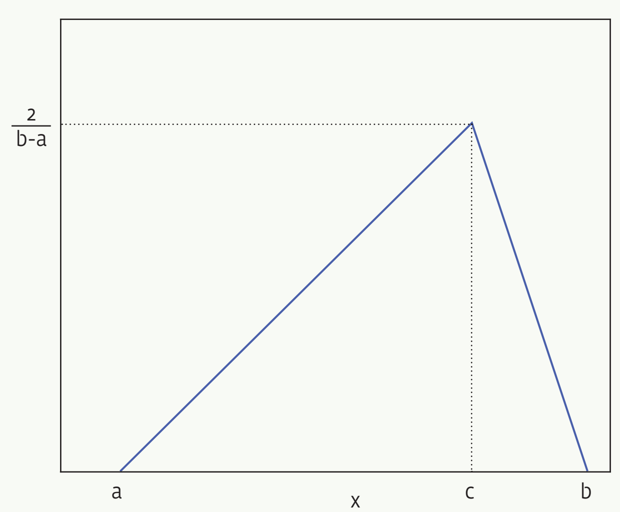


Figure 4.4. Probability Density Function



Mean: $\overline{X} = \dfrac{a+b+d}{3}$

Variance: $s^2 = \dfrac{a^2+b^2+d^2-ab-ad-bd}{18}$

Mode: d

Median:

$$= a + \frac{\sqrt{(b-a)(d-a)}}{\sqrt{2}}, \text{ if } \frac{b+a}{2} < d, \text{ and}$$

$$= d, \quad if \ \frac{b+a}{2} = d$$

$$= b - \frac{\sqrt{(b-a)(b-d)}}{\sqrt{2}}, \text{ if } \frac{b+a}{2} > d$$

Skewness: $= \dfrac{\sqrt{2}(a+b-2d)(2a-b-d)(a-2b-d)}{5(a^2+b^2+d^2-ab-ad-bd)^{3/2}}$

If d equals $(a + b)/2$, the distribution is symmetric so the skewness is equal to zero.

Kurtosis: =-3/5

The kurtosis coefficient is a constant regardless of the values of *a*, *b*, and *d*.

Usually, observations occurring above the mode are not insured, as the frequency and severity of indemnity payments would be very large. Consequently, premium rates would have to be set at a level that makes the program unmarketable. For practical insurance purposes, we are only interested in the area that lies to the left of the mode or where $t < d$. Therefore, for $t < d$:

Frequency: $F(t) = \dfrac{(t-a)^2}{(b-a)(d-a)}$

Severity: $= t - E(X / X < t)$

$$E(X / X < t) = \frac{\int_a^t Xf(X)dX}{F(t)}$$

$$\int_a^t Xf(X)dX = \int_a^t X\frac{2(X-a)}{(b-a)(d-a)}dX = \frac{2}{(b-a)(d-a)}(\int_a^t X^2dX - a\int_a^t XdX)$$

$$= \frac{2}{(b-a)(d-a)}(\frac{t^3-a^3}{3} - a\frac{t^2-a^2}{2})$$

$$= \frac{1}{(b-a)(d-a)}\left(\frac{2t^3-2a^3-3at^2+3a^3}{3}\right)$$

$$= \frac{2t^3+a^3-3at^2}{3(b-a)(d-a)} = \frac{(t-a)^2(a+2t)}{3(b-a)(d-a)}$$

So,

$$E(X/X<t)=\left[\frac{(t-a)^2(a+2t)}{3(b-a)(d-a)}\Big/\frac{(t-a)^2}{(b-a)(d-a)}\right]=\frac{a+2t}{3},$$

and

Severity = $t-\frac{a+2t}{3}=\frac{t-a}{3}$.

Expected Indemnity: $E(I)=Severity\times Frequency$.

So, $E(I)=\frac{(t-a)^2}{(b-a)(d-a)}\times\frac{t-a}{3}=\frac{(t-a)^3}{3(b-a)(d-a)}$.

Pure Risk Premium Rate: w $=\frac{(t-a)^3}{3(b-a)(d-a)t}$.

Several interesting situations can arise with the use of the triangular distribution. Assume that $a=0$, which is consistent with the lowest possible crop yield being equal to zero, and again focusing on $t<d$,

then $f(X)=\frac{2X}{bd}$,

$$F(X)=\frac{X^2}{bd},$$

Mean: $\overline{X}=\frac{b+d}{3}$,

Mode: $=d$,

Median: $=\sqrt{\frac{bd}{2}}$, if $\frac{b}{2}\le d$, and

$$=b-\sqrt{\frac{bd}{2}},\ \text{if}\ \frac{b}{2}>d.$$

Variance: $=\frac{b^2+d^2-bd}{18}$.

Frequency: $F(t)=\frac{t^2}{bd}=\frac{\frac{c^2(b+d)^2}{9}}{bd}$.

Severity: $=\frac{t}{3}=\frac{c(b+d)}{9}$.

Expected Indemnity: $E(I)=\frac{t^3}{3bd}=\frac{c^3(b+d)^3}{81bd}$.

Rate: $w=\frac{t^2}{3bd}=\frac{c^2(b+d)^2}{27bd}$.

Another important situation arises if one assumes that $a = 0$ and $d = \frac{b}{2}$ (i.e., symmetry). Thus, d must be halfway between a and b.

If $a = 0$ and $d = \frac{b}{2}$, then

$$f(X) = \frac{4X}{b^2} \text{ (symmetric)},$$
$$F(X) = \frac{2X^2}{b^2},$$

Mean: $\overline{X} = \frac{2d + d}{3} = d = \frac{b}{2}$,

Mode: $= d$,

Median: $= d$,

Variance: $= \frac{b^2 + (\frac{b}{2})^2 - b(\frac{b}{2})}{18} = \frac{b^2}{24} = \frac{d^2}{6}$,

Frequency: $F(t) = \frac{t^2}{bd} = \frac{t^2}{b\frac{b}{2}} = \frac{2t^2}{b^2} = \frac{c^2}{2}$ because $t = \frac{cb}{2}$,

Severity: $= \frac{t}{3} = \frac{cb}{6}$,

Expected Indemnity: $E(I) = \frac{t^3}{3bd} = \frac{t^3}{3b\frac{b}{2}} = \frac{2t^3}{3b^2} = \frac{c^3 b}{12}$,

Rate: $w = \frac{c^3 b}{12t} = \frac{c^2}{6}$.

If the mean and standard deviation are known or estimated from data, the parameters b and d can be calculated assuming $a = 0$ as:

$$b = \frac{3\overline{X} + \sqrt{-3\overline{X}^2 + 24s^2}}{2};$$

$$d = \frac{3\overline{X} - \sqrt{-3\overline{X}^2 + 24s^2}}{2}.$$

An alternative approach is to calculate the parameters b and d from the median (m) and mean. Then,

$$if \quad d \geq \frac{b}{2},$$

$$b = \frac{3\overline{X} + \sqrt{9\overline{X}^2 - 8m^2}}{2};$$

$$d = \frac{3\bar{X} - \sqrt{9\bar{X} \;\; - 8m^2}}{2};$$

$$if \quad d < \frac{b}{2};$$

$$b = \frac{3\bar{X} + 4m + \sqrt{9\bar{X}^2 \; + 24\bar{X}m - 8m^2}}{6};$$

$$d = \frac{15\bar{X} - 4m - \sqrt{9\bar{X}^2 \; + 24\bar{X}m - 8m^2}}{6}.$$

Furthermore, if the median is less than the mean, then the mode, *d*, must be less than *b*/2. For example, if the median = 0.866 and mean =1, then $d < b/2$ so the second set of equations is used to find *b* = 2.366 and *d*=0.634.

Pros and Cons of Using the Triangular Distribution:

Pros:

1. It is simple and often provides reasonable approximations to other unimodal distributions.
2. Usually, rates obtained from the triangular distribution are close (although slightly larger) than those obtained from a normal distribution if the data are symmetric. But, the rates obtained from the triangular distribution are easier to calculate.

Cons:

1. In some situations, the triangular distribution may be too restrictive. For example, selected standard deviations and means (perhaps obtained from empirical data) may result in unrealistic values for the parameters *a*, *b*, and *d*. For example, assume that *a* is believed to be equal to 0. The smallest variance that can be modeled with the triangular distribution is where

$$\frac{\partial(s^2)}{\partial d} = 0,$$

since, $s^2 = \frac{b^2 + d^2 - bd}{18}$ then $\frac{\partial(s^2)}{\partial d} = 2d - b = 0.$

Or where *d*=*b*/2 (symmetry) is the value of *d* that yields the smallest variance. Therefore, the smallest possible variance is

$$d = \frac{b}{2} \text{ and } s^2 = \frac{b^2 + (\frac{b}{2})^2 - b(\frac{b}{2})}{18} = \frac{b^2}{24}.$$

Assuming that a=0, the largest variance is where d=0 or where b=d since $0 \le d \le b$. Both result in

$$s^2 = \frac{b^2}{18}.$$

Therefore, the triangular distribution, when a=0, must meet the requirement that

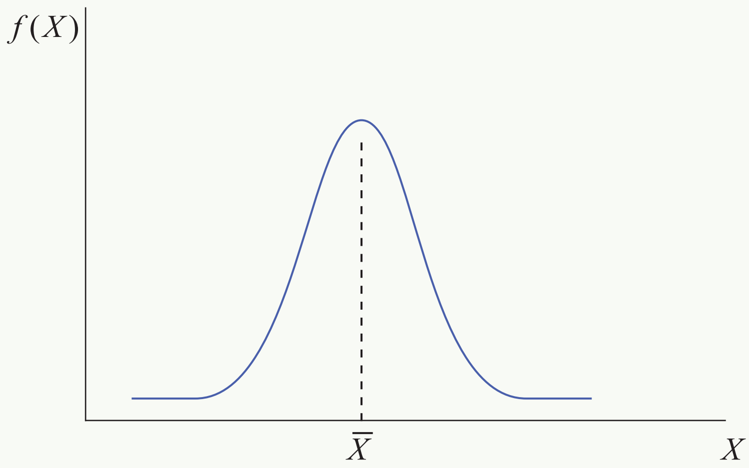
$$\frac{b^2}{24} < s^2 < \frac{b^2}{18}.$$

Note that all of the change in variance when a=0 occurs because of changes in d or skewness. The variance is not affected by changes in kurtosis since it is a constant. Thus, a drawback of this distribution is that the range of variance is limited and kurtosis is predetermined.

Because of the simplicity of the triangular distribution, it is commonly used to rate insurance products. However, practitioners need to be aware of its limits. In particular, the kurtosis in the data should be compared to -3/5 and the yield variance needs to be in the required range.

4.3.3.3. The Normal Distribution (figure 4.5) is the most widely used distribution in statistical analyses and has many desirable properties. However, it is complex and has no closed-formed solution for its cumulative distribution function.

Figure 4.5. The Normal Distribution Probability Density Function



Range: $X \in (-\infty; +\infty)$

Probability Density Function: $f(X) = \dfrac{e^{-\frac{1}{2}\left(\frac{X-\overline{X}}{s}\right)^2}}{s\sqrt{2\pi}}$

Cumulative Distribution Function: $F(X) = \int_{-\infty}^{X} f(X)dX$

Mean = Median = Mode = $\overline{X}$

Skewness: $= 0$

Kurtosis: 3

The normal distribution has several interesting properties, including:

1. Any linear transformation of normally distributed variables is also normally distributed.
2. The Central Limit Theorem states that the mean of a sufficiently large number of independent random variables, each with finite mean and variance, will be approximately normally distributed.

Let $f(z)$ be the standard normal distribution ($\overline{z} = 0$, $s = 1$), so that

$$f(z) = \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}}$$

then:

$$df(z)/dz = f'(z) = -zf(z).$$

Let $g(X)$ be the normal distribution where x is not standardized (i.e., so that $\overline{X} \neq 0, s \neq 1$) or

$$g(X) = \frac{e^{-\frac{1}{2}\left(\frac{X-\overline{X}}{s}\right)^2}}{s\sqrt{2\pi}}$$

then:

$$\frac{dg(X)}{dX} = g'(X) = -\frac{X-\overline{X}}{s^2} g(X)$$

and:

$$Xg(X) = \overline{X}g(x) - s^2 g'(X).$$

Frequency: $G(t) = F(\tilde{t})$, where $\tilde{t}$ is the standardized trigger and defined as $\tilde{t} = \frac{t-\overline{X}}{s}$

where $G(t) = \int_0^t g(X)\,dX$ and $F(\tilde{t}) = \int_0^{\tilde{t}} f(z)dz$

Severity: $= t - E(X / X < t)$

So, $\int_{-\infty}^{t} Xg(X)dX = \int_{-\infty}^{t}\left(\overline{X}g(X) - s^2 g'(X)\right)dX = \overline{X}\int_{-\infty}^{t} g(X)dX - s^2\int_{-\infty}^{t} g'(X)dX =$

$$= \overline{X}G(t) - s^2 g(t).$$

However, $G(t) = F(\tilde{t})$ and $g(t) = \frac{f(\tilde{t})}{s}$

So that $E(X \mid X < t) = \overline{X} - s\frac{f(\tilde{t})}{F(\tilde{t})}$

Thus, **Severity** $= t - \left(\overline{X} - s\frac{f(\tilde{t})}{F(\tilde{t})}\right) = s\frac{f(\tilde{t})}{F(\tilde{t})} - (\overline{X} - t)$

Expected Indemnity:

$$E(I) = Severity \times Frequency = \left(s\frac{f(\tilde{t})}{F(\tilde{t})} - (\overline{X} - t)\right)F(\tilde{t}) = sf(\tilde{t}) - (\overline{X} - t)F(\tilde{t})$$

Pure Risk Premium Rate for the Normal Distribution:

$$w = \frac{E(I)}{Liability} = \frac{E(I)}{t} = \frac{sf(\tilde{t}) - (\overline{X} - t)F(\tilde{t})}{t}$$

since $t = c\overline{X}$, $w = \frac{sf(\tilde{t}) - (\overline{X} - c\overline{X})F(\tilde{t})}{c\overline{X}} = \frac{s}{c\overline{X}} f(\tilde{t}) - \frac{1-c}{c}F(\tilde{t})$

Note that $\frac{s}{\overline{X}}$ is the coefficient of variation. Not only is $\frac{s}{\overline{X}}$ explicit in the above equation, but it is also implicit in $\tilde{t}$. Hence, relative variation is an important component of rates when using the normal distribution.

Pros and Cons of Using the Normal Distribution:

Pros:

1. The normal distribution depends upon two parameters for which we have the most confidence in estimating – *mean* and *variance*.
2. Experience has shown that the normal distribution is widely applicable to many situations. For example, consider the distribution of yields on each hectare of a given field. That distribution can be further separated into yields distributed across each square meter within each hectare. Hence, producing an average yield for a field is a process of aggregating averages upon averages. Thus, the central limit theorem suggests that the yield of a given field should be close to normally distributed.

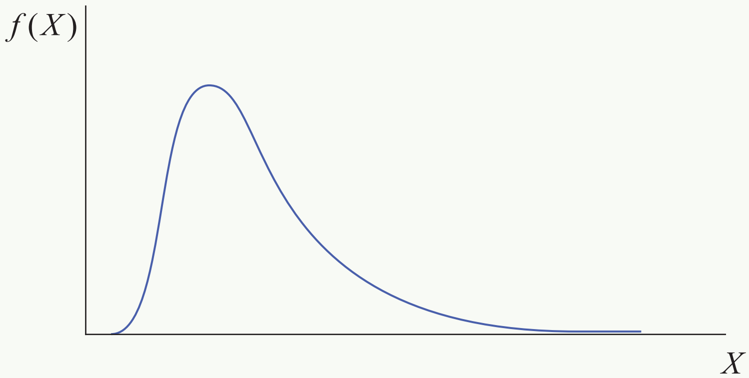
Cons:

1. The normal distribution encompasses unrealistic negative values for yields. Methods that account for truncations at zero yields are available, but usually result in *ad hoc* assumptions and must be used carefully.
2. Depending upon the Central Limit Theorem to justify the normal distribution is problematic because it is premised on individual observations being independently distributed. Weather generally causes yields to be spatially correlated even within relatively expansive regions. Hence, this violates the tenets of the central limit theorem.
3. The normal distribution is symmetric and does not allow for skewness. This restricts the distribution's flexibility for many yield outcomes.
4. The normal distribution's kurtosis is also predetermined so that the distribution may not be sufficiently flexible to be used in certain situations.

Even with these shortcomings, the normal distribution is widely used for rating many insurance products.

4.3.3.4. Log-Normal Distribution. The graph for the log-normal distribution is presented in figure 4.6.

Figure 4.6. The Log-Normal Distribution.



Properties:

(1) The $\log(X)$ is normally distributed;

(2) Distribution is skewed to the right with long tail;

(3) $\overline{X} > median$.

The log-normal distribution is used for specific modeling procedures. Substantial research supports the premise that prices are distributed log-normally. Although yields are not usually well-represented by log normal distributions, revenue distributions ($price \times yield$) may be approximated reasonably well by log-normal distributions.

4.3.3.5. The Beta Distribution is characterized by a highly flexible functional form with a finite range. It is complicated because it contains the Beta function. Although widely used for statistical and research purposes, its use in rating processes has been less frequent. It is flexible to the point of being highly sensitive to outliers. However, efforts to refine its use may result in greater future applications.

4.3.3.6. Extreme Value Functions are a combination of functions rather than a single unique function. Usually, it is a collection of exponential functions which often includes the normal function. After estimating the distributions for each of these functions, the one with the thickest tail to the left is often selected. It has been widely applied in South American areas that are prone to flooding. The Gumbel function is a member of the extreme value function and is probably the most widely used. Extreme value functions are often used when insured events occur infrequently (low frequency) but generate high indemnities (high severity). Rates are usually high relative to "true," but unknown, expected indemnities.

4.3.4. Choice of Distribution. Rates obtained from parametric distributions are often compared to those obtained from empirical rating processes. The choice of distribution function is often a matter of judgment. A variety of tests can be used to provide guidance in this process. However, such tests are weak and inconclusive. Mean and standard deviation estimates calculated from available data are usually used as parameters in selected distribution functions. The standard deviation of skewness and kurtosis are approximately $\frac{6}{\sqrt{n}}$ and $\frac{24}{\sqrt{n}}$, where n is sample size. A simple test for skewness and kurtosisis illustrated using the following estimated statistics.

Skewness = 0.30

Kurtosis = 0.20

n = 64

Recall that the normal distribution has zero skewness and a kurtosis of 3. Therefore, tests for normality are conducted based on the following t-tests:

Skewness: $\frac{0.3-0}{6/\sqrt{64}}$ = 0.40.

Kurtosis: $\frac{0.2-3}{24/\sqrt{64}} = \frac{-2.8}{3}$ = -0.93.

Neither test provides evidence of non-normality. Even so, practitioners should not be lulled into using any distribution without substantial analyses. More sophisticated tests are available. However, these tests often provide little added insight. Analyses should not only include statistical tests, but also graphical approaches, histograms, and other more elementary efforts.

4.4. Regional Level Actuarial Methods

Farm-level data are often less than desirable in terms of quality, quantity, and historical length for rating processes. In such cases, regional data are often used to complement limited farm-level data. However, the use of regional data requires specific actuarial procedures.

4.4.1. Deficiencies of Long Term Farm Data. Sufficiently long time series data at the farm-level are often not available. However, regional data (for example, at the country or district level) may be available for as many as 35-50 years. Regional data are not as variable as farm-level data because the former are developed from averages of the latter. In addition, farm-level data may only be available for a relatively short period of time (say, 6-10 years) across multiple farms. The challenge is to adopt an appropriate yield distribution using both farm-level and regional data that represents the true underlying variability of farm-level data. Both empirical and parametric approaches are used to extract such information. However, the logic used in both methods is similar. In general, regional yield variance is identified. Any remaining variability is allocated to the farm-level. This process is similar to variance decompositions, analysis of variance, and bootstrapping approaches.

4.4.2. Variance Decomposition Approaches. Variance decompositions are used to combine longer term regional data with shorter term farm-level data. This allows for the variability of farm-level data to be projected over a longer time period through its relationship to regional level data.

Farm-level variance is decomposed into regional variance and farm-level residual variance. We first review the statistical notion of the variance of a sum before discussing this decomposition.

4.4.2.1. Variance of the Sum of Variables. Let X and Z be two random variables that are paired so that X_j and Z_j occur simultaneously and $\overline{X} = \overline{Z}$. Then:

$$s^2_{X+Z} = \sum \frac{\left[(X_j + Z_j) - \overline{(X+Z)}\right]^2}{df}$$, where df is the degrees of freedom.

Note that $\overline{X+Z} = \overline{X} + \overline{Z}$ and defining $x = X - \overline{X}$ and $z = Z - \overline{Z}$

$$s^2_{X+Z} = \frac{\sum\left[(X_j - \overline{X}) + (Z_j - \overline{Z})\right]^2}{df} = \frac{\sum (x_j + z_j)^2}{df} = \frac{\sum x_j^2}{df} + \frac{\sum z_j^2}{df} + \frac{2\sum x_j z_j}{df} =$$

$$= s^2_X + s^2_Z + 2 s_X s_Z r_{X,Z}$$

where, $r_{X,Z}$ is the correlation between x and z

If X and Z are linearly independent, then $r_{X,Z} = 0$ and, consequently, $s^2_{X+Z} = s^2_X + s^2_Z$.

4.4.2.2. Variance Decomposition. For simplicity, assume that farm-level mean yields are equal within a region. For those years in which farm-level data are available, regional yields are subtracted from farm-level yields. The difference is referred to as the farm residual or farm deviation. Because in each year the regional yield is the average of farm yields, the farm residuals across farms for each year sum to zero. Because each farm-level average yield is equal to the regional average yield by assumption, the sum of the farm residuals over time for each farm will also equal zero. Mathematically, this process is:

Let R_t = regional yield in year t=1,T, and $X_{i,t}$ = farm i's yield in year t, i=1,n; t=k,T. So, there are T years of regional yields, and n farms with T-K+1 years of yields.

Then, the farm residuals are $v_{i,t} = X_{i,t} - R_t \; \forall \; i=1,n$ and $t = K,T$

As mentioned above, $\sum_{i=1}^{n} v_{i,t} = 0 \quad \forall \; t = K,T$ and $\sum_{t=K}^{T} v_{i,t} = 0 \quad \forall \; i = 1,n$

It follows that: $\sum_{i=1}^{n} \sum_{t=K}^{T} v_{i,t} = 0.$

The $v_{i,t}$'s are farm-level yields and will be used directly in the empirical rating process. For parametric rating processes, the standard deviation of $v_{i,t}$ is used.

An interesting property is that farm residuals and regional yields are linearly independent. To illustrate, let $\overline{R}_{K,T}$ = regional average yield over years K to T and the covariance between the regional yields and farm residuals is:

$$s_{v,R} = \frac{\sum_{t=K}^{T}\sum_{i=1}^{n} v_{i,t}\,(R_t - \overline{R}_{K,T})}{df}$$

$$= \frac{\sum_{t=K}^{T}(R_t - \overline{R}_{K,T})\sum_{i=1}^{n} v_{i,t}}{df} = 0$$

because the sum of the $v_{i,t}$'s equals 0. Therefore, the regional variance plus the farm-level residual variance is equal to total farm-level variance.

For purposes of rating accuracy, it is important to use as much information as possible. In this case, the information from long term regional data is combined with short term farm-level data (residuals). The information about regional variability uses all T years. The most information about the residual farm variability is incorporated by including all of the v's over n farms from T-K+1 years.

At least 5-6 years of observations for at least 35 representative farmers are necessary to apply the above procedure. There should also be a balance between the number of years and the number of farms. For example, it would be acceptable to have 10 years of observations for 20 representative farms rather than 5 years of data for 40 representative farms.

4.4.3. Rating. The previously developed decomposition of farm-level variability can be used either for empirical or parametric rating.

4.4.3.1. Empirical Rating. The challenge is to develop a sufficiently large data set that captures farm-level variability and other characteristics representative of farm-level yields. The data can be generated by randomly choosing an R_t and a $v_{i,t}$ to generate a simulated farm yield or $Y = R_t + v_{i,t}$. This process is repeated to generate as large a data set of Y's as desired.

An alternative approach is to generate a sample of Y's by using all possible combinations of R_t and $v_{i,t}$. If 50 years of regional data and 6 years of farm-level data from 80 farms are available, then all combinations generate (50)(6)(80) = 24,000 values for Y. This sample can then be used for rating as discussed earlier. This approach is similar to bootstrapping methods.

4.4.3.2. Parametric Rating requires that certain parameters of a selected distribution be estimated. For example, the triangular distribution with $a=0$ and the normal distribution each require that two parameters be estimated – the mean and variance. Calculating means and variances directly from empirical data is straightforward. Given certain assumptions, the best estimate available for the regional yield variance is the variance calculated from all T years of regional data, and the best estimate of the farm residual variance is the variance of the v's. Because the v's and R's are linearly independent, the best estimate of the farm-level variance is $s^2 = s_R^2 + s_v^2$.

This approach combines information contained in longer term regional data and shorter term farm-level data. The calculated farm-level variance is then used to calculate rates as discussed above.

4.4.3.3. Correlations between Regional and Farm Yields. The correlation between regional yields and individual producer yields is often useful in rating processes. The correlation between R and X is:

$$r_{R,X} = \frac{s_{R,X}}{s_R s_X};$$

$$s_{R,X} = \frac{\sum_t \sum_i (R_t - \overline{R})(X_{i,t} - \overline{X})}{df};$$

$$= \frac{\sum_t \sum_i (R_t - \overline{R})(R_t + v_{i,t} - \overline{X})}{df}.$$

Recall that by assumption: $\overline{R} = \overline{X}$, so

$$s_{R,X} = \frac{\sum_t \sum_i (R_t - \overline{R})^2 + \sum_t \sum_i (R_t - \overline{R})\, v_{i,t}}{df};$$

$$= \frac{s_R^2 + \sum_t (R_t - \overline{R}) \sum_i v_{i,t}}{df}.$$

Since $\sum_i v_{i,t} = 0$, $S_{R,X} = S_R^2$.

So, $r_{RX} = \frac{S_R^2}{S_R S_X} = \frac{S_R}{S_X}$.

Furthermore, $r_{RX}^2 = \frac{S_R^2}{S_X^2}$.

Research often indicates that $r \approx 0.7$. In such cases, $r^2 \approx 0.5$. So, farm-level yield variance is about twice regional yield variances.

4.4.3.4. Farm Yield Difference. The previous discussion of combining long term regional data with short term farm data can easily be modified to include situations in which farm average yields are not equal. Prior to using the v's for rating, they can be adjusted by $\overline{X}_i - \overline{R}_{T,K}$ for each i farm. After this adjustment, the rating proceeds as before.

Variance is implicitly assumed to be independent of average yields. This assumption often holds in practice, but the data need to be analyzed to determine if the variance is homogenous across farms. If the data are heterogenous, techniques are available to make the appropriate adjustments (discussed below).

Rates can easily be calculated for different expected yields. Again, the variance of yields is often similar across expected yields. Therefore, by changing expected yields and using the same variance, desired rates can be calculated.

4.5. Problems Associated with Rating Processes

An ideal data set would include many producers with farm-level records extending over a long period of time. The data would have been carefully and thoroughly collected and would be absent of anomalies, trends, missing observations, and structural change. Data related to health and life insurance tends to approximate the ideal situation, but this seldom occurs when rating crop insurance. In general, one can expect to obtain only a limited number of years of individual farm-level data. In addition, technological change and other trends occur, and we often find substantial differences in productivity across farms. We need to identify and address these challenges.

4.5.1. Trend. In many cases, yields appear to exhibit a systematic change that does not appear to be random. These changes may be abrupt or gradual. It is important to determine whether yields have been increasing over time (i.e., trend). Yield trends represent a structural change that could be caused by changes in technology, education, internal policies, or external policies.

> ***Example.*** *In the United States, the importation of avocados was prohibited for many years. Because of recent trade liberalization agreements with Mexico, imports of avocados into the United States have increased, lowering avocado prices. Avocado production is input-intensive and a single avocado tree can generate as much as $1,500 of revenue per year. Historically, avocado producers intensely managed avocado trees. For example, they often fertilized, watered, and monitored each tree. Increased imports and lower avocado prices have rendered these practices cost-prohibitive. As managers reduce the use of inputs, yields decline. As a result, avocado insurance programs need to be reviewed because of lower expected yields, given increased foreign competition.*

Some crop yields are relatively price-sensitive, while others are less so. Price sensitive crop yields need to be reviewed frequently. Input availability (especially irrigation water) must be considered during rating processes.

Furthermore, technological changes cause crop yields to trend upward over time. Trends are usually calculated at the regional level and then projected to farm levels. In the case of Ukraine, historical regional yield data across all rayons could be regressed onto time to test whether the time trend slope coefficient is statistically different from zero. A map of Ukraine with color coding relative to the level of those regression coefficients would provide a visual illustration of yield trends. Such information should be compared across regions.

Assume that yield trends have occurred over time. To adjust for those changes, one can fit a regression line such as:

$$Y_t = \alpha + \beta t + e_t,$$

where t is year with $t = 1, T$. To determine if a statistically significant trend exists, we test the hypothesis:

$$H_0 : \beta = 0.$$

We estimate a regression equation to obtain estimates of the parameters:

$$Y_t = \hat{\alpha} + \hat{\beta} t + e_t.$$

Consequently,

$$e_t = Y_t - (\hat{\alpha} + \hat{\beta} t).$$

The next step is to calculate the expected yield for the last year in the sample. If the last year of the sample is 2007, we use the regression equation to predict the expected yield for 2007:

$\hat{Y}_{2007} = \hat{\alpha} + 2007\hat{\beta}$. In general: $\hat{Y}_T = \hat{\alpha} + \hat{\beta} T$, where $T = \textit{the last year of data}$.

It is recommended that one does not use such regression to forecast trends beyond the sample data. Using the last year of data to anchor yield distributions to the insurable year is preferable.

Detrended yields are constructed using $\tilde{Y}_t = \hat{Y}_T + e_t$. For small samples (less than 30 observations), then e_t should be adjusted for the degrees of freedom lost in the estimation process. In small samples, the errors underestimate the true variability because of the estimation of $\hat{\alpha}$ and $\hat{\beta}$. The degrees of freedom are used to adjust the residuals as:

$$e_{t\,adj.} = e_t \left(1 + \frac{1}{T} + \frac{3}{1+T}\right)^{1/2}.$$

The correction factor may also be used to generate the trend adjusted sample as:

$$Y_{t\,adj.} = \hat{Y}_T + e_{t\,adj}.$$

As T becomes large, the term $\left(1 + \frac{1}{T} + \frac{3}{1+T}\right)$ monotonically declines to 1.

A regression of $Y_t = \alpha + \beta t + e_t$ reduces the variability of the error term as much as possible. Hence, using regression equations to detrend yield data should only be done if there is compelling evidence of a trend. If no trend actually exists in the data, this procedure will result in underrating premiums.

It is also possible that non-linear trends exist in the data. Non-linearities may take the form of a polynomial trend or may be caused by structural breaks in the data. The most complicated case occurs when a trend is non-linear in the parameters.

Warning: *Do not over-fit the data! For example, the regression fit of a data set with 20 observations modeled with a 19 degree polynomial will be perfect, but will not allow for any residual variance.*

Trend adjusted regional yields are used in the usual manner for rating.

4.5.2. Heteroskedasticity. Data analysis may reveal the existence of heteroskedasticity, which is the technical term for systematic changes in data variability. Heteroskedasticity is often revealed by graphical analyses. Explicit statistical tests also exist to identify heteroskedasticity.

Two different regressions can be used to test for the existence of heteroskedasticity:

(1) $|e_t| = \alpha + \beta t$, or

(2) $e_t^2 = \alpha + \beta t$,

where t is time (year). Each regression can be used to test the null hypothesis that:

$$H_0\colon \beta = 0.$$

If the estimate of β is statistically different from zero, then one can reject the null hypothesis of no heteroskedasticity. The first method has an intuitive interpretation for the parameter β – namely, that the absolute value of the deviations increases by β in each period.

The presence of heteroskedasticity requires the use of sophisticated rating methodologies. Heteroskedasticity is a complex problem that needs to be addressed by an experienced and trained actuary. If present, several approaches can be used to mitigate its impact (such as Generalized Least Squares). Even small values of β can cause large changes in generated samples (or in standard deviations of yield). Therefore, premium rates are quite sensitive to the presence of heteroskedasticity.

If heteroskedasticity is present in the data, the regression errors, e_t, should be transformed using:

$$\tilde{e}_t = e_t \frac{a + bT}{a + bt},$$ a and b are the estimated values for alpha and beta in (1) above

The $\tilde{e}_t$ are used in the rating process as before.

4.5.3. Autocorrelation occurs when subsequent year yield observations are correlated.

> ***Example.*** *Autocorrelation can occur in wheat yields that are accompanied by the production of heavy straw that is left in a field. Soil nitrogen decomposes straw during a year and reduces soil fertility. Unless supplemental nitrogen is added, the following year's wheat yield may be reduced.*

Autocorrelation is important because farmers may adversely select to purchase crop insurance after a good year because an indemnity payment is more likely in the subsequent year. A possible solution is to increase rates or reduce trigger yields after a high-yielding year. However, both of these adjustments reduce market appeal. Producers do not appear to react strongly to mild autocorrelation. Therefore, much of the problem can be addressed through load factors.

In the absence of trend, autocorrelation can be detected by either the correlation coefficient, $r_{Y_t, Y_{t-1}}$, or by the following regression:

$$Y_t = a + bY_{t-1} + e_t$$

and testing to see if b is statistically different from zero.

If a trend is present, autocorrelation could be detected by the following regression

$$e_t = \psi + \gamma e_{t-1} + u_t,$$

where e_t is the residual from the trend regression. A t-test is used to determine if γ is statistically significant. Although the test is biased, it is generally sufficient for this purpose.

It is possible to have very low levels of autocorrelation that are statistically significant (say an estimate for γ of 0.06-0.07 or 6-7%). In these situations, adverse selection problems will be minimal. However, when autocorrelation reaches the 10-12% range, adverse selection can become problematic.

4.5.4. Moral Hazard and Adverse Selection are issues that exist within all insurance programs. Rating problems, however, are created as moral hazard and adverse selection problems become prevalent.

4.5.4.1. Moral Hazard occurs when the purchase of insurance results in detrimental management practices that increase the likelihood or level of indemnity payments. Moral hazard may be of two types: (1) *ex ante*, when people behave in a more risky manner before loss becomes evident, and (2) *ex post*, when there is a failure to mitigate losses. For example, an *ex ante* moral hazard is a failure to use high-quality seed because of an insurance purchase. An *ex post* example occurs when a producer does a poor job of harvesting a crop because the insurance price is greater than the market price.

Moral hazard often results from **asymmetric information** in which one party (a farmer) has more information about production practices or market prices than a second party (an insurer). A farmer usually possesses more information about his operation and practices than an insurance agent, in part, because it is expensive for insurance providers to monitor farm inputs and management.

4.5.4.2. Adverse Selection is also caused by asymmetric information. Those farmers most likely to receive indemnity payments are also more likely to purchase insurance. For example, inter-temporal yield variations differ substantially by location. If all producers are charged the same premium rates, those with small amounts of yield variation will choose not to insure and those with high variation will insure. If an insurer could identify those with lower yield variability, those producers could be placed in a separate risk pool and rated accordingly.

If insurance rates are developed from a representative pool of producers, those producers who are more likely to receive indemnity payments (riskier producers) will choose to participate in an insurance program. Because only the high-risk producers are participating, the initial premium rates will be too low. Eventually, the rates will be increased consistent with the risks incurred. As rates increase, the least risky of the high-risk group will cease to participate leaving the highest risk producers in the program. This process will continue until participation erodes to an unacceptable level. The process is called **participation erosion**.

Participation erosion can be mitigated in a variety of ways:

1. Accurately pool similar risk producers and charge each pool premiums according to their risk levels;
2. Subsidize premiums so insurance is desirable to all producers (including low-risk producers);
3. Determine if a sufficient number of producers are highly risk-averse;
4. Mandate participation by all producers;
5. Use proxy variables (e.g., precipitation) as indemnity triggers.

4.5.5. Spatial Smoothing. Even with 25 or 30 years of data, we may not have observed every possible extreme event. Failure to observe extreme outcomes will cause premium rates to be too low to provide actuarial soundness. On the other hand, it is also possible that extreme events may have occurred more frequently when using short time series data. In these cases, premium rates will be too high and much more data will be needed to accurately estimate the probability of severe outcomes. However, rating is restricted to available data. Consequently, the quantity of data may be increased through Spatial Smoothing.

In the United States, for example, spatial smoothing is conducted within states because a catastrophic occurrence in one area of a state may reveal the probability of such an occurrence in another area. Spatially smoothing rates can help account for data limitations.

There are two methods of spatial smoothing: Catastrophic Pooling and smoothing based upon Regional Correlations.

4.5.5.1. Catastrophic Pooling places some portions of the worst outcomes into a central pool. Assume that this portion is 20%. Then, a rate is calculated based on the remaining 80% of outcomes for each region. Then, a rate for the 20% of worst outcomes aggregated across regions is established. Finally, premium rates are calculated as a weighted average of 80% of the regional rate and 20% of the aggregate rate. In the United States, this pooling process is called "20-80."

Example. Consider the following Loss/Liability ratios.

Table 4.2. Regional Loss Cost Ratio Example

Region Loss Cost Ratio								
	Region					Region		
Year	A	B	C		Year	A	B	C
1	0.118	0.150	0.164		11	0.113	0.077	0.036
2	0.134	0.059	0.075		12	0.153	0.500	0.263
3	0.057	0.057	0.000		13	0.700	0.242	0.242
4	0.063	0.002	0.065		14	0.120	0.078	0.042
5	0.023	0.000	0.023		15	0.175	0.121	0.150
6	0.213	0.082	0.131		16	0.108	0.080	0.028
7	0.090	0.110	0.123		17	0.146	0.124	0.133
8	0.125	0.086	0.092		18	0.121	0.145	0.151
9	0.084	0.123	0.056		19	0.170	0.195	0.199
10	0.078	0.148	0.117		20	0.029	0.040	0.069

These data are first ranked in descending order in each region.

Table 4.3. Loss Cost Ratios in Descending Order

Region Loss Cost Ratio								
	Region					Region		
Year	A	B	C		Year	A	B	C
1	0.700	0.500	0.263		11	0.118	0.086	0.092
2	0.213	0.242	0.242		12	0.113	0.082	0.075
3	0.175	0.195	0.199		13	0.108	0.080	0.069
4	0.170	0.150	0.164		14	0.090	0.078	0.065
5	0.153	0.148	0.151		15	0.084	0.077	0.056
6	0.146	0.145	0.150		16	0.078	0.059	0.042
7	0.134	0.124	0.133		17	0.063	0.057	0.036
8	0.125	0.123	0.131		18	0.057	0.040	0.028
9	0.121	0.121	0.123		19	0.029	0.002	0.023
10	0.120	0.110	0.117		20	0.023	0.000	0.000

The worst four outcomes in each region (a total of twelve) are allocated to the Catastrophic Pool and rates are established for the average outcome. The remaining sixteen outcomes for each region are likewise averaged and a rate is established.

Table 4.4. Premium Rates

	Regions		
	A	B	C
Regional Rates without pooling	0.141	0.121	0.108
Regional Rate based on 80% Pool	0.098	0.083	0.081
Catastrophic Pool	0.268	0.268	0.268
Regional Pooled Rate	0.132	0.120	0.118

Notice that the Catastrophic Pool has the same rate for each region. The weighted average regional pooled rates are less variable across regions than the unpooled rates.

This method places all of the catastrophic events in one pool. This practice is widely used in crop insurance but is an *ad hoc* approach. In certain areas of the United States, however, this method results in substantial rate increases because the catastrophic component is relatively large. Because the catastrophic component may be disproportionally generated from certain areas, resulting rate levels may cause low participation from low-risk areas.

4.5.5.2. Regional Correlation. This approach smooths rates by considering the correlation of yields between regions. The correlations among regions are used to develop weights to calculate weighted average rates.

Let:

$$w_i = \text{rate based only on a region's data,}$$

$$P_{i,j} = \text{correlation between yield history in regions } i \text{ and } j.$$

Let:

$$\tilde{w}_i = \text{weighted rate,}$$

$$\tilde{w}_i = \frac{\sum_j P_{i,j} w_j}{\sum_j P_{i,j}}$$

We calculate a correlation matrix among the regional yields:

Table 4.5. Correlation Matrix

Region	Region A	Region B	Region C
A	1.000	0.393	0.596
B	0.393	1.000	0.817
C	0.596	0.817	1.000
Individual Rates w_i	0.141	0.121	0.108
Smoothed Rate $\tilde{w}_i$	0.127	0.120	0.120

Note that the smoothed rate for Region A is

$$\frac{(1.0)(0.141)+(0.393)(0.121)+(0.596)(0.108)}{1+0.393+0.596}=0.127$$

The rate for Region A is lower than its individual rate, while the rate for Region B is almost identical. Region C's smoothed rate is higher than its individual rate. Hence, the differences in rates across regions have been reduced. This process is generally consistent with the spatial statistics literature. A similar process can also be used to smooth rates across crops.

4.5.6. Pooling. The process of pooling groups or individuals who face similar probabilities of losses for rating purposes is called pooling. Pooling occurs for several reasons. If a group of farmers face substantially different risks, then participation will depend on risks relative to rates. As discussed earlier, failure to account for risk differences will result in participation erosion. Therefore, grouping farmers into similar risk pools is important for long term program viability. Furthermore, farmers will perceive programs that do not account for risk differences to be unfair.

Pools can be based on a variety of criteria. For example, risk pools may be established by region, crop, expected yield, production practice (irrigated versus dry land), crop type (spring wheat versus winter wheat versus durum wheat), or geography. Another important pooling decision involves coverage levels. Mature crop insurance programs offer producers the opportunity to choose coverage levels, usually between 50% and 80%. Premium rates are differentiated by coverage level, allowing producers to choose from a variety of insurance options.

The difference between rates across coverage levels is called **Rate Spreading**. Some countries have used common rate spreads across regions and crops based on coverage levels. These efforts have generally been unsuccessful in developing actuarially sound rates.

A common F-test can be used to compare the yield variability across pools. Such tests are used to determine if pools can be combined for rating purposes.

4.6. Rate Loading and Judgment

Pure risk rates are usually based on expected indemnity payments. In addition, other costs are incurred in the provision of crop insurance. These costs are collectively referred to as loads. Loads differ based on each insurance situation.

Loads are commonly used to offset costs associated with:

1. Confidence in the rating method – loads are larger for those products that have more tenuous rating procedures.
2. Insurance services (e.g., operations, loss adjustment, agents, returns on investment, etc.).
3. Political, judicial, or other institutional risks.

4. Data quality and quantity – loads are larger for rates that are based on poor data quality and quantity.
5. Market size – small insurance markets are more costly to service and are loaded more heavily.

Loads are sometimes determined by cost accounting/budgeting or historical performance of similar insurance products. In addition, experience with similar products across regions or countries are used to determine loads.

The following is a hypothetical loading example for a multiple peril insurance product.

Rating	10%
Service	25%
Data Volume	10%
Other	5%
Total	50%

Given these loads and a pure risk rate of 0.087, a total rate premium rate would be calculated as: (Pure Risk Rate) x (1 + Load) = (0.087) x (1+0.50) = 0.1305. This is an example of a proportional loading process.

If the pure risk rate for a product is relatively low, then loads may be attached as "add ons" rather than proportionally. Suppose the pure risk premium is only 0.03. Then, loads may be set in the following manner:

Rating	0.5%
Service	2.5%
Data Volume	0.8%
Other	1.0%
Total	4.8%

The total rate would then be calculated as: (Pure Risk Rate) + (Load) = 0.03 + 0.048 = 0.078.

Loading for data volume and other concerns often becomes a matter of judgment. Loads are seldom less than 25% of the pure risk premium. In situations in which the pure risk premium is low and data volume and quality is poor, loading may be 100% or more.

4.7. Updating Premium Rates

Rating reviews are not only important for updating purposes, but can also uncover trends in underwriting procedures. Updating involves reviewing and evaluating rates, underwriting, data management, and fraud/illicit activity at all levels (farmer, agent, adjuster, issuing agency, regulator, and auditors). Premium rates should be updated when possible with data from actual insurance histories. Program operational anomalies that are included in actual insurance histories may not be reflected in the data used to initially rate the programs. Thus, actual loss data from the operational insurance program need to be included in updating processes and combined with recent insurance experience. All aspects of initial insurance programs should be designed to facilitate future rating updates. Rate reviews should be conducted annually and intensive updates at least every three years.

Rates can be updated in several ways:

1. Use the original rating process and include recent yield data to recalculate rates.
2. Use actual loss ratios incurred since the last rating process. Eventually, rates should be based on actual loss histories. In practice, the usual approach is to use a weighted average of recent loss ratios and those used in the initial rating process. For example, one might base 5% of the rate on the most recent year's losses and 95% on the last rating procedure. However, it can be shown that this process places disproportionate influence on single year outcomes.

The following process avoids placing disproportionate weights on specific years. The loss experience of an operating insurance program is expressed either as a **Loss Cost Ratio *(LCR)*** or **Loss Ratio *(LR)*** and defined by the following formulas:

$$Loss\ Cost\ Ratio = \frac{Indemnity\ Paid}{Liability};$$

$$Loss\ Ratio = \frac{Indemnity\ Paid}{\Pr emium}.$$

Assume that three years of actual insurance experience are available that generate a loss cost ratio of 8% and the original rating process generated a pure risk premium of 10%. Suppose that you want to eventually use only actual experiences for rating processes once 40 years of data are accumulated. Consequently, every actual year of experience represent 2.5% of the 40 year horizon. Updating rates after three years of experience would result in (0.075)(0.08) + (0.925)(0.10) = 0.0985. If after another four years the rates are again updated and the loss cost ratio for the first seven years is 9%, a new rate is calculated as (0.175)(0.09) + (0.825)(0.10) = 0.09825.

After 40 years, the original rating will be phased out so that only actual actuarial experience is used. It is, of course, important to consider yield trends when using lengthy data sets.

5.0. Proxy Index Insurance

Two of the major issues associated with yield insurance programs are monitoring/servicing costs and moral hazard. These problems can be severe enough to cause crop insurance programs to fail. Proxy index insurance products may reduce monitoring/servicing costs and problems created by moral hazard. Such products use metrics to trigger indemnity payments that obviate some of the problems associated with the use of individual yield triggers such as individual field loss adjusting.

Proxy insurance is also useful for providing risk management for specialty crop production. Developing individual insurance programs for small volume, specialty crop is often prohibitively expensive because of small crop volumes, yield data deficiencies, and harvest windows. For example, strawberries are continuously harvested over many months. Consequently, loss adjusting is prohibitively expensive. Furthermore, proxy index insurance can be used to insure multiple crops simultaneously. However, the acquisition and management of weather data for proxy index insurance products is often more expensive than anticipated. In addition, weather stations do not exist in many regions and the development and maintenance of such stations is expensive.

Proxy index triggers may be based on weather, regional yields, satellite imagery, or other factors. This section focuses on weather-based proxies.

5.1. Introduction

Insurance based on weather outcomes is similar to insurance obtained through the use of weather derivatives. Weather derivatives are becoming more common as a means for managing risk. Weather derivatives are often used in non-agricultural applications. For example, weather derivatives are used to insure against:

- Unusual cold temperatures resulting in increased demands for and prices of heating oil or natural gas;
- Rain events resulting in the cancellation of outdoor concerts or events;
- Lack of snow at ski areas;
- Cold temperatures that reduce consumer activity during prime shopping periods.

Weather derivatives could be used to insure against many agricultural weather perils including:

- Excess precipitation during harvest or planting;
- Freezing during apple blossom;
- Excess precipitation and cold temperatures on orchard crops;
- Excess heat effects on canola;
- Lack of heating degree-days during the growing season;
- Drought.

5.2. Proxy Weather Indices

Various forms of weather proxies are used in crop insurance programs. Weather index products base indemnities on "All or Nothing" or Prorated approaches. In addition, dual triggers are frequently used, as are methods for partitioning liability and the use of survival coverage.

5.2.1. Indemnity Payments may be calculated in a variety of ways when using proxy insurance products. Two common methods are "All or nothing" or Prorated approaches.

5.2.1.1. "All or Nothing" products refer to the payment of the entire insured liability as an indemnity when a proxy trigger occurs. For example, assume that a temperature trigger of 0°C has been established. If temperatures are colder than the trigger during the insured period, the entire insured liability is paid as an indemnity.

5.2.1.2. Prorated Products use dual triggers to calculate indemnity payments. For example, assume a producer insures a crop against excessive precipitation at two trigger points, 3cm and 6cm. If the precipitation outcome is greater than 6 cm, the entire liability is paid. If the precipitation is less than 3 cm, no liability is paid. If the precipitation outcome is between 3cm and 6 cm, the indemnity is prorated based on a proportion of the liability. For example, if the precipitation outcome is 4cm (one-third of the distance between 3cm and 6cm), then one-third of the liability is paid as an indemnity.

5.2.2. Product Types. Single or multiple variable proxy indices can be used as indemnity triggers.

5.2.2.1. Single Variable proxy index products use a single weather variable such as precipitation or temperature for triggering indemnities. The indemnity may be of an "All or Nothing" or Prorated form.

5.2.2.2. Multiple Variable proxy index products are also used in crop insurance programs. These products may insure against more than one weather factor simultaneously. For example, multiple variable products have been developed that insure against both temperature and precipitation outcomes. Alternatively, a multiple variable product could insure against extreme temperature or precipitation events during different periods of the year. Multiple variable products may calculate indemnities using "All or Nothing" or Prorated approaches. Indemnities for Prorated products can be calculated using Liability Partitioning or Survival methods.

Example 1: Liability Partitioning with Prorated Indemnity

Assume that a farmer insures an apple orchard against both cold temperatures and excess precipitation (multiple perils). The dual temperature triggers are
-1°C and -5ºC, and the dual precipitation triggers are 3cm and 6cm. The farmer chooses to split his total liability of 1,000UAH between excess precipitation and freezing temperatures on a 50-50 basis.

If the temperature outcome during the insured period is -2°C, then the indemnity (I_t) is:

$$I_t = \left[\frac{-1-(-2)}{(-1-(-5))} \times 0.5 \right] \times 1,000 = (0.125) \times 1,000 = 125UAH$$

If the temperature outcome is greater than -1°C, then I_t is zero. If the temperature outcome is less than -5°C, then I_t equals 0.5 x 1,000 = 500UAH.

In addition, if the precipitation outcome is 5 cm, then the indemnity (I_p) is:

$$I_p = \left[\frac{(5-3)}{(6-3)} \times 0.5 \right] \times 1,000 = (0.333) \times 1,000 = 333UAH$$

If precipitation is less than 3cm, then I_p is zero. If precipitation is greater than 6cm, then the indemnity equals 0.5 x 1,000 = 500UAH.

The total indemnity is $I = I_t + I_p$.

Example 2: Survival Products and Prorated Indemnities

Assume the above apple production example in which indemnities are calculated on a Survival basis rather than with Liability Partitioning. A farmer insures an apple orchard against both cold temperatures and excessive precipitation using dual triggers. The temperature triggers are -1ºC and -5ºC, and the precipitation triggers are 3cm and 6cm. The total liability is 1,000UAH.

Assume the actual temperature is T and actual precipitation is P. If T and P are between their trigger levels, then the indemnity is calculated as:

$$I = \left[1 - \left(\left(1 - \frac{-1-T}{(-1-(-5))}\right)\left(1 - \frac{P-3}{(6-3)}\right)\right)\right] \times 1{,}000$$

where the term $\left(1 - \frac{-1-T}{(-1-(-5))}\right)$ is a proxy for the proportion of the apple crop that survives the temperature peril. The term, $\left(1 - \frac{P-3}{(6-3)}\right)$, is a proxy for the proportion of the apple crop that survives the precipitation peril.

Therefore, the entire term in square brackets represents a proxy for the proportion of the apple crop that survives both perils.

Assume that $T = -2$ and $P = 5$, such that:

$$I = \left[1 - \left(\left(1 - \frac{1}{4}\right)\left(1 - \frac{2}{3}\right)\right)\right] \times 1{,}000 = 0.75 \times 1{,}000 = 750UAH$$

If the temperature is less than -5°C or precipitation is greater than 6cm, the entire liability is paid as an indemnity. If the temperature is greater than -1°C, then T is set equal to -1 in the above equation. If precipitation is less than 3cm, then P is set equal to 3. Table 5.1 presents indemnities for various combinations of temperature (T) and precipitation (P).

Table 5.1. Relationship between Temperature, Precipitation, and Indemnities for Survival Products with Prorated Indemnities

	Precipitation (cm)			
Temperature	**3**	**4**	**5**	**6**
-1	0.0	333.3	666.7	1,000.0
-2	250.0	500.0	750.0	1,000.0
-3	500.0	666.7	833.3	1,000.0
-4	750.0	833.3	916.7	1,000.0
-5	1,000.0	1,000.0	1,000.0	1,000.0

5.3. Effectiveness of Weather Proxies for Yields

The effectiveness of proxy index products must be evaluated based upon correlations between proxy index metrics and yields. Because proxy index measures and yields are not perfectly correlated, indemnities will not always match the frequency and severity of low yields. Typical correlations between precipitation and yield range from 0.45 to 0.65.

5.3.1. Simulations of Indemnity Probabilities. The importance of correlations between yields and weather variables are illustrated with the following simulation. Assume that yields and precipitation are linearly related and each has a mean of 1 and a standard deviation of 0.4. In addition, an indemnity is paid if precipitation is below 0.7cm. Define yields (y) as very low if $y<0.5$, low if $0.5<y<0.7$, and moderate or above average if $y>0.7$. Such outcomes are possible regardless of precipitation levels unless yields and precipitation are perfectly correlated. Indemnities are triggered by precipitation rather than yield. For this example, indemnities are defined as "large" if precipitation (p) is below 0.5cm, "small" if precipitation is between 0.5cm and 0.7cm, and "zero" if precipitation is above 0.7cm.

If we assume that yields are very low (less than 0.50), then table 5.2 illustrates the probability of indemnity payments based on the various correlations between precipitation and yield.

Table 5.2. Yields are Very Low (y<0.5)

Precipitation-Yield Correlation	Probability of Indemnity Payment		
	Zero (p>0.7cm)	Small (0.5<p<0.7cm)	Large (p<0.5cm)
0.000	0.777	0.123	0.100
0.200	0.671	0.155	0.174
0.400	0.526	0.206	0.268
0.600	0.362	0.243	0.395
0.800	0.168	0.251	0.580
0.900	0.067	0.247	0.686
0.950	0.024	0.199	0.777
1.000	0.000	0.000	1.000

If the correlation between precipitation and yield is only 0.20, the probability of receiving a zero indemnity payment is 0.67 for very low yields. However, the probability of receiving a large indemnity payment is 0.17.

Table 5.3. Yields are Low (0.5<y<0.7)

Precipitation-Yield Correlation	Probability of Indemnity Payment		
	None (p>0.7cm)	Small (0.5<p<0.7cm)	Large (p<0.5cm)
0.000	0.784	0.114	0.103
0.200	0.723	0.145	0.132
0.400	0.661	0.179	0.159
0.600	0.590	0.219	0.192
0.800	0.479	0.316	0.205
0.900	0.377	0.429	0.194
0.950	0.278	0.559	0.163
1.000	0.000	1.000	0.000

If the correlation between precipitation and yield is only 0.20, the probability of receiving a zero indemnity payment is 0.72 for low yields. However, the probability of receiving a large indemnity payment is 0.13.

If we assume that yields are moderate or above average (greater than 0.7), then table 5.4 illustrates the probability of indemnity payments based on correlations between precipitation and yield.

Table 5.4. Yields are Moderate or Above Average (y>0.7)

Precipitation-Yield Correlation	Probability of Indemnity Payment		
	None (p>0.7cm)	Small (0.5<p<0.7cm)	Large (p<0.5cm)
0.000	0.774	0.119	0.107
0.200	0.820	0.103	0.077
0.400	0.867	0.081	0.051
0.600	0.921	0.057	0.023
0.800	0.972	0.025	0.003
0.900	0.993	0.006	0.001
0.950	0.999	0.001	0.000
1.000	1.000	0.000	0.000

If the correlation between precipitation and yield is only 0.20, the probability of receiving a zero indemnity payment is 0.82 for moderate or above average yields. However, the probability of receiving a large indemnity payment is 0.077.

5.3.2. "Misses" occur when indemnities generated by the proxy trigger do not match indemnities based on actual crop yields. In each of the above tables, the situation in which a perfect match occurs between the proxy and yield triggers only occur in the last row because the correlations between precipitation and yield is 1.0. Misses can be reduced by increasing the correlation between proxy triggers and actual yields. This is usually accomplished by combining various proxies. For example, combining temperature and precipitation proxies into a single index may increase the correlation between yield and the composite index from 0.60 to 0.75. In some case, larger increases in the correlation can be obtained by combining a weather index with satellite imagery.

5.3.3. Timing of Precipitation. Various studies have evaluated the relationships between yield and the timing of precipitation (as opposed to precipitation levels). Approximately 60% of these studies do not show statistically significant improvement by including monthly specific precipitation rather than total growing season or annual precipitation levels.

5.3.4. Nonlinearities. Economic theory suggests that the relationship between weather and yields must be non-linear. Incorporating non-linear relationships between weather and yields can often increase correlations between weather conditions and actual yields and result in fewer "misses".

5.4. Data

Most actuarial rating processes require substantial high-quality historical data. For rating weather proxy index products, densities of weather stations are an important factor in determining correlations between weather metrics and actual crop yields. In addition, the number of years for which data are available and the accuracy of recording processes are critical for establishing premium rates and determining indemnity payments.

5.4.1. Data Sources. Two weather data repositories are currently being maintained: (1) Global Summary of Day (GSOD), and (2) Global History Climate Network – Daily (GHCND). Terms of access to these data sources are country-specific. Of course, some regions/countries have larger numbers of weather stations and weather station densities than others.

The GSOD has the following features:

- Data are free of charge to all non-commercial users;
- Data are derived from global integrated surface data;
- Over 9,000 stations are typically available for recent years;
- Data for 1929 to present are available for a limited number of stations;
- Many airport and some additional city locations are available;
- Various daily summary elements such as temperature (mean, maximums, minimums), dew point, wind speed (mean, maximum, peak gust), air pressure, visibility, precipitation, snow depth, etc. are available;
- Data are usually updated daily within 1-2 days after the date of each observation;
- A substantial number of observations are often missing for many weather stations.

The GHCND has the following features:

- Data are developed from numerous sources and subjected to quality assurance reviews;
- Archives include daily maximum temperatures, daily minimum temperatures, precipitation (i.e., rainfall and snow water equivalents), snowfall, and snow depth;
- A substantial number of observations are often missing for many weather stations.

By its nature, weather data represent large numbers of observations. For example, 50 years of weather data translates into 18,000 observations for a single weather station. Therefore, electronic tools and processes are particularly important for rating and quality control.

5.4.2. Data Uses. Weather data are needed for two purposes in developing and implementing weather proxy products: (1) rating, and (2) indemnity calculations.

5.4.2.1. Rating. Ideally, 50 years of high-quality historical data would be available for rating. The data would include all daily observations during the portion of the year to be indemnified. If less than 30 years of data are available, then more sophisticated statistical analyses are likely required and resulting premium rates are less reliable. Missing observations may occur because of recording or mechanical failures, human error, coding errors, or other reasons. Unfortunately, missing observations are usually systematic. That is, it is more likely that an observation is missing during unusually cold periods than when temperatures are more moderate. Therefore, missing observations cannot be ignored and must be replaced with realistic values. Filling techniques are technically complex and often use information gleaned from other, similar weather stations.

Updating weather proxy index rates requires the continuous maintenance of and access to weather station data. Given the volume of data, automated electronic collection and storage system must be developed and maintained. It is imperative that appropriate quality control systems be continuously employed.

5.4.2.2. Indemnities. Weather data are necessary to calculate indemnities for all weather proxy index products. The data pipeline must be timely, dependable, and accurate, because contracts specify timeframes for the payments of indemnities. The data pipeline must be able to support indemnity timeliness requirements. All data pipelines eventually fail either in terms of unreported observations or timeliness. Hence, backup provisions must be developed and maintained to support data pipelines and obviate problems created by failures. Furthermore, insurance contracts and other legal documents must specifically address the potential for data pipeline failures and specify appropriate alternatives for calculating indemnities if a failure occurs.

5.5. Rating

Weather insurance products pose substantial rating challenges. These challenges are difficult to overcome if large quantities of weather data are unavailable.

5.5.1. Autocorrelation and Truncation. Most statistical rating analyses assume that observations are independently and identically distributed. This is not the case for weather data. Weather outcomes at any point in time are likely highly correlated with nearby time periods (i.e., autocorrelated). Overcoming these autocorrelation problems is complex and heavily dependent on data volume and statistical expertise. Furthermore, precipitation is necessarily truncated at zero which creates additional statistical problems. These problems are particular troublesome for data filling and rating processes. Failure to address autocorrelation and truncation issues often results in biased premium rates and failed programs. In particular, biased rates invite numerous problems associated with adverse selection.

5.5.2. Seasonality. Clearly, mean weather variables change with seasons. However, both the distributional and distributional forms can vary across seasons. Therefore, the development of weather insurance contracts often must consider changes in mean temperatures, variability, and distributional forms. Such problems are particularly prevalent in products that insure precipitation or growing temperature days over a lengthy period.

5.5.3. Trend. The influence of climate change can be modelled as trends in weather variables. Yield trends were discussed earlier in this manual. To the extent that weather trends exist, they can be treated in a similar manner to yield trends. Statistical techniques need to be used to maintain actuarial soundness if such trends are present.

5.5.4. Timing of Attachment. Timing of attachment refers to the time between when a contract is purchased and when the insured period begins. As the interval between purchase and attachment declines, purchasers are able to incorporate more accurate weather information into their purchase decisions because of the autocorrelated nature of weather. If the rating process is based on historical weather conditions, shorter attachment intervals can result in adverse selection. Therefore, the purchase/attachment interval must be sufficiently long to minimize a purchaser's ability to forecast future weather outcomes through observations of current weather conditions and the predictions of others. An alternative is to include publically-available weather forecasts into the rating process. However, these approaches have had only limited success. If forecasts are not included in the rating process, then the purchase attachment interval should be no shorter than 30 days.

5.5.5. Rating Illustration. Assume that we wish to create an "All or Nothing" insurance product triggered by temperature below 0°C on September 15 and we have 50 years of data. During those 50 years, the temperature had been below 0°C only four times on September 15.

The probability of an indemnity trigger based on this information is 8%, which represents the pure risk premium without any loads. Suppose, however, that the historical temperature information surrounding September 15 results in the rates illustrated in Table 5.5.

Table 5.5. Example Rates by Date

Date	Years <0°C	Probability (Rates)
Sept 10	6	0.12
Sept 11	7	0.14
Sept 12	6	0.12
Sept 13	5	0.10
Sept 14	8	0.16
Sept 15	4	0.08
Sept 16	5	0.10
Sept 17	7	0.14
Sept 18	4	0.08
Sept 19	9	0.18
Sept 20	10	0.20

Notice that the 8% probability on September 15 is lower than that occurring within five days on either side of the insurance date. Therefore, the 8% rate is not a reasonable expectation of the future probability of temperatures being below 0°C because of the expectation that temperature is autocorrelated. Insurance market speculators are well aware of these types of rating errors and arbitrage these problems to their own advantage. Such activity can cause insurance programs to fail.

5.6. Multiple Weather Stations

Sometimes indemnities are triggered by a combination of weather station outcomes. For example, a producer may have fields that spatially separated. Thus, indemnities could be triggered by more than one weather station or a combination of weather stations. Multiple weather stations can be used to develop an index that is used as an indemnity trigger. Such indexes are developed as a weighted average of relevant weather station measurements.

Purchasers of insurance contracts are often allowed to select specific weather stations for developing weather triggers. Producers are generally allowed to choose weather outcomes as measured by up to six or seven weather stations for determining a weather index. Historical data from selected stations are used rating purposes. The weights for the selected stations are used both for rating purposes and for determining indemnity payments.

In some cases, producers are allowed to select weights that are then applied to measurements from the relevant weather stations. In other cases, distance weighted smoothing is used to calculate weights for each weather station. Distance weighted smoothing requires the identification of a geographic center, such as a field, for the insured production. Weather stations close to the center position are used to create a weather index. The weights for each weather station are based upon the distance of each weather station from the center position. The distance from the center position to weather station i (i=1,..,n) is defined as d_i . For example, if i =1 and n=3, then the weight for weather station 1 is:

$$w_1 = \frac{d_2 d_3}{d_1 d_2 + d_1 d_3 + d_2 d_3}.$$

The weights for weather station 2 and 3 are calculated in a like fashion. Note that if weather station 1 is zero distance away from the center position (i.e., the weather station itself is at the center position) then $d_1 = 0$ and $w_1 = 1$, w_2=w_3=0. Therefore, all of the weight is placed on weather station 1. Alternatively, if $d_1 = d_2 = d_3$, then $w_1 = w_2 = w_3 = 1/3$.

Note that the weights and distances are inversely proportional, i.e., $\frac{w_i}{w_j} = \frac{d_j}{d_i} \;\forall\; i, j$.

Finally, a variety of alternative weighting methods exist, but those presented above are the most common.

6.0. Reinsurance

Primary insurance companies or issuing agencies usually do not have sufficient capital to absorb losses caused by extreme outcomes. Large indemnities are more likely to occur in crop insurance than many other types of insurance because crop insurance losses are often correlated across farms. Adverse weather events are the primary cause of low yields. In particular, drought is often a regional event that impacts the yields of many neighboring farmers within an area. In addition, crop insurance sales for any given company tend to occur within limited regions, and few crop insurance companies operate on a global scale. Furthermore, crop insurance issuing agencies are often not diversified into other forms of insurance such as health, life, or property insurance. And, crop insurance agreements are usually short-term and limited to annual contracts. Therefore, primary crop insurance companies are not generally diversified by space, time, or sector and often incur risks that are correlated within their insurance portfolio. For these reasons, primary insurance companies transfer risk to other companies who are better able to manage these risks. The transfer involves paying fees to such companies.

Reinsurance refers to the ceding of risk from one insurance company (such as an issuing agency or primary insurance company) to another (a reinsurer) or to the government. Reinsurers are usually large companies that are well-diversified across space, sectors, and insurance types. Governments often provide reinsurance in terms of stop-loss agreements.

Crop insurance risk absorption in the United States (as measured by indemnity payouts) has been approximately:

Table 6.1. Indemnity Payment Responsibilities by Risk Bearer

Risk Bearers	Percentage
Issuing Agency	5-10
Federal Government (Special Pool)	5-15
Reinsurers	80-90

The United States has a special pool for high-risk farmers in which the federal government assumes all the risk. Most of the insurance policies placed in the Special Pool have associated risks that are not palatable to the private insurance industry. However, private insurance companies are limited in the number of policies that can be placed in the Special Pool, about 15%. In addition, the government provides a stop-loss for indemnities that exceed about 350% of premiums.

The private sector has absorbed most indemnity payments in recent years because losses have been relatively low. Although government indemnity outlays have not been triggered under the stop-loss agreement in recent years, some outlays have occurred within the Special Pool.

6.1. Forms of Reinsurance

Reinsurance agreements specify the method in which losses or indemnity payments will be shared between risk bearers (e.g., between a primary insurance company and a reinsurer). These reinsurance relationships are often complex and not standardized across reinsurance companies. The forms of reinsurance specify how losses across a primary insurance companies' entire portfolio (i.e., not on a policy-by-policy basis) are to be shared.

6.1.1. Dollar One refers to an arrangement in which the reinsurer is responsible for all indemnity payments. The primary insurance company is responsible for servicing the insurance program but bears none of the indemnity risk.

6.1.2. Co-pays (Co-insurance, Percentage Participation) refer to the sharing of losses in mutually agreeable proportions between a primary insurance provider (or issuing agency) and reinsurer(s). A 40-60 co-pay would mean that a primary insurance company would pay 40% of indemnity costs and a reinsurer would pay 60%.

6.1.3. Reinsurance Deductible is the amount of loss the primary insurance provider or issuing agency sustains prior to a reinsurer indemnifying losses. The deductible may be specified as a percentage of liability or as a premium multiple. If the premium rate is 8% and the multiple is 1.5, then the primary insurer must absorb 12% of the liability before the reinsurer begins absorbing remaining indemnity payments.

6.1.4. Stop-Loss is the amount (or premium multiple) at which all remaining loss responsibility is transferred to another entity. Stop-losses are similar to deductibles except that they are usually intended to protect against catastrophic losses. Often, stop-losses are triggered less than 2-5% of the time. Stop-losses are often provided by governments.

6.1.5. Tranched (Stacked, Layered) Approaches divide liability among primary issuing agencies, reinsurers, and/or governments. Indemnity responsibilities are hierarchical in that those "lower" in the tranche make indemnity payments only after those higher in the tranche have exceeded pre-specified limits. Consider the following layered or tranched example for a $1 billion liability:

Table 6.2. Layered or Tranched Reinsurance Responsibility Example

	Loss Responsibility	
	Percentage	**Amount**
Primary Insurance Provider	5%	$50,000,000
Lead (or Junior) Reinsurer	15%	$150,000,000
Secondary (or Senior) Reinsurer	20%	$200,000,000
Government Stop Loss	60%	$600,000,000

The Government stop-loss liability is the lowest level in the tranche and is only triggered after $400,000,000 of liability is paid by the primary insurance provider ($50,000,000), the lead reinsurer ($150,000,000), and the secondary reinsurer ($200,000,000).

To illustrate payment responsibilities of different levels of total losses for a $1 billion liability, consider the following table:

Table 6.3. Layered or Tranched Reinsurance Indemnity Example

		Loss Payment			
Example	**Total Loss**	**Primary Insurance Provider**	**Lead Reinsurer**	**Secondary Reinsurer**	**Government**
A	$40,000,000	$40,000,000	0	0	0
B	125,000,000	50,000,000	$75,000,000	0	0
C	300,000,000	50,000,000	150,000,000	$100,000,000	0
D	500,000,000	50,000,000	150,000,000	200,000,000	$100,000,000

For example C, a total loss of $300,000,000 has occurred. This loss will be absorbed (or indemnified) by the primary insurer ($50,000,000), the lead reinsurer ($150,000,000), and the secondary reinsurer ($100,000,000). The government's stop-loss is not triggered in this case. Therefore, the Government would not be responsible for any indemnity payments. However, in example D, the government's stop-loss responsibility is triggered by the $500,000,000 total loss and their indemnity payout would be $100,000,000.

6.2. Rating or Costs of Reinsurance

As in the case of primary insurance, reinsurance premium rates are comprised of both a pure risk premium and a load. The following discussion is intended to introduce reinsurance rating concepts. Actual reinsurance rating is complex and the following discussion is only intended to provide an introduction to reinsurance rating.

6.2.1. Pure Risk Premiums are usually estimated from historical data. Ideally, historical data on indemnity payments from an identically-designed insurance program are used. However, such data are often not available and one must depend upon historical yield data. The rate estimation process depends on the specific reinsurance agreement, data availability, and a myriad of other factors. The process may involve parametric distributions, empirical distributions, or a combination of the two.

For illustration purposes, the following simplified example is provided. Assume there are four farms in a portfolio. The farm data is sufficiently short that it cannot be used directly for reinsurance rating. However the summary yield statistics of farm means, standard deviations, and between-farm correlations can be estimated. The underlying individual yield distribution is assumed to be normally distributed. Furthermore, assume that the yield average, standard deviations, and correlations between farms are the same across all four farms. The objective is to develop a long term individual farm yield data set that can be used for rating purposes. The developed yield data set will have the same average, standard deviation, and correlations as estimated from actual farm data. However, the developed data will be extend over a much longer period and will provide a better approximation of events that are likely to result in indemnity payments. For illustration purposes, the developed data set will have 40 observations (40 years of data) for the 4 farms.

The process begins by generating a series of numbers (z_{jt}) for the jth farm in year t that are identically, independently, and normally distributed with a mean of 0 and a standard deviation of 1. In this example j=1,4 and t=1,40. Random number generators (which are widely available in software programs such as Excel) are used to generate the z_{jt}'s. Table 6.4 presents the z_{jt}'s for this example.

Table 6.4. Generated z_{jt}'s

Year/j	1	2	3	4
1	-1.8406	1.2993	-0.8629	0.1445
2	0.4117	0.1775	-0.5174	1.0973
3	1.7318	0.4913	1.3368	-1.9142
4	0.7219	-0.9426	1.2908	-0.3431
5	2.5949	0.3164	1.5141	0.4159
6	-0.6683	-0.1645	0.6198	0.4421
7	0.6913	-1.1765	0.5799	0.5368
8	-0.4778	0.0594	-1.9046	1.0907
9	0.3378	-1.5548	-0.2350	-0.1389
10	0.4796	0.3599	-1.6275	0.1638
11	0.0540	-0.1248	1.3376	-0.3184
12	-0.8056	-0.7555	-0.0723	0.8713
13	-0.7380	-0.8942	0.2321	-0.6883
14	-1.2805	-0.5580	-0.3428	-2.2118
15	0.0313	1.6004	0.2323	-1.2977
16	0.7029	1.1291	1.0583	-1.2328
17	0.6105	-0.7984	0.7099	0.0712
18	-0.6355	0.3096	-0.3816	0.7267
19	-1.1084	-0.2061	-0.7363	0.4491
20	-0.1091	0.0229	-0.1508	0.7239
21	0.7792	-0.7969	0.2623	0.0235
22	0.4625	0.3868	-0.2984	-0.0458
23	-1.4991	1.0685	0.7370	-1.4002
24	-1.2592	-0.6143	-0.6867	-0.0006
25	1.1723	0.6617	-0.3905	1.3793
26	0.5917	0.7673	0.4027	-0.3815
27	-0.2708	-1.6572	0.6885	-1.4351
28	1.5040	-2.1869	-0.9240	-0.3071
29	-0.0625	-0.0447	1.2280	1.9711
30	0.7297	-0.1855	-0.4780	0.5982
31	0.5822	-0.8311	-0.0981	0.6367
32	0.6272	0.9206	-1.0397	-0.1446
33	-1.2240	0.9468	1.3809	-0.8026
34	-0.5241	0.4059	0.1981	0.6305
35	-0.5134	0.1996	-2.1044	0.8452
36	2.0105	1.2584	-0.8073	-0.4590
37	0.1007	0.1300	-1.9018	-1.7221
38	-1.5969	0.8522	1.0642	1.6368
39	-0.9223	-1.2253	0.0799	-0.7562
40	-1.5579	1.3298	0.4349	1.1543

Next, the z_{jt}'s are transformed into x_{jt}'s that have the desired correlations, a mean of 0, and standard deviation of 1 using the following process. (The t subscripts have been suppressed below for brevity.)

$x_1 = z_1$

$x_2 = a_{21}z_1 + a_{22}z_2$ where a_{kj} are parameters found as below

$r = \sum_t x_1x_2 / df$ where r is the between farm correlation, df is the degrees of freedom, with the standard deviation of x equal 1

$$= \sum_t z_1 \left(a_{21}z_1 + a_{22}z_2\right) / df$$

$$= \sum_t a_{21}z_1^2 / df$$

$$= a_{21}$$

To constrain the variance of x to 1

$1 = a_{21}^2 + a_{22}^2$ so

$$a_{22} = \sqrt{1 - a_{21}^2}$$

$$a_{22} = \sqrt{1 - r^2}$$

$x_3 = a_{31}z_1 + a_{32}z_2 + a_{33}z_3$

Using a process similar as above

$$a_{31} = r$$

$$a_{32} = \frac{r(1-r)}{\sqrt{1-r^2}}$$

$$a_{33} = \sqrt{\frac{1 + r - 2r^2}{1 + r}}$$

and similarly for x_4

$x_4 = a_{41}z_1 + a_{42}z_2 + a_{43}z_3 + a_{44}z_4$ where

$$a_{41} = r$$

$$a_{42} = \frac{r(1-r)}{\sqrt{1-r^2}}$$

$$a_{43} = \frac{r\sqrt{\frac{1 + r - 2r^2}{1 + r}}}{1 + 2r}$$

$$a_{44} = \sqrt{\frac{1 + 2r - 3r^2}{1 + 2r}}$$

Notice that the x_j's have a mean of 0, a standard deviation of 1, and a correlation of r , except for sampling error introduced when the z_j's were randomly generated.

Table 6.5 presents the transformed data.

Table 6.5. Transformed x_{jt}'s

	Farm Number			
Year/j	**1**	**2**	**3**	**4**
1	-1.8406	0.2049	-1.2498	-0.6071
2	0.4117	0.3596	-0.1654	1.0189
3	1.7318	1.2913	2.0992	-0.2327
4	0.7219	-0.4553	1.1428	0.0811
5	2.5949	1.5715	2.6250	2.0266
6	-0.6683	-0.4766	0.1244	0.0944
7	0.6913	-0.6733	0.4795	0.5488
8	-0.4778	-0.1874	-1.7769	0.2517
9	0.3378	-1.1776	-0.4718	-0.4377
10	0.4796	0.5515	-0.9851	0.1410
11	0.0540	-0.0811	1.0831	0.0122
12	-0.8056	-1.0571	-0.6799	0.0531
13	-0.7380	-1.1434	-0.4376	-1.1239
14	-1.2805	-1.1234	-1.0812	-2.6199
15	0.0313	1.4017	0.6673	-0.5009
16	0.7029	1.3293	1.5415	-0.0812
17	0.6105	-0.3862	0.6544	0.2760
18	-0.6355	-0.0496	-0.5399	0.2682
19	-1.1084	-0.7327	-1.2148	-0.4090
20	-0.1091	-0.0347	-0.1710	0.4936
21	0.7792	-0.3006	0.3737	0.2317
22	0.4625	0.5662	0.0993	0.2458
23	-1.4991	0.1758	0.1607	-1.3976
24	-1.2592	-1.1616	-1.3676	-0.9476
25	1.1723	1.1592	0.4583	1.7879
26	0.5917	0.9604	0.8462	0.2980
27	-0.2708	-1.5706	-0.0517	-1.6079
28	1.5040	-1.1419	-0.6337	-0.3107
29	-0.0625	-0.0700	0.9585	1.7648
30	0.7297	0.2042	-0.0790	0.6867
31	0.5822	-0.4287	-0.0289	0.5345
32	0.6272	1.1109	-0.2696	0.2528
33	-1.2240	0.2079	0.7888	-0.6913
34	-0.5241	0.0895	0.0169	0.3940
35	-0.5134	-0.0838	-1.9173	0.0395
36	2.0105	2.0950	0.7094	0.8408
37	0.1007	0.1629	-1.4650	-1.6618
38	-1.5969	-0.0604	0.3165	0.9588
39	-0.9223	-1.5223	-0.7496	-1.3963
40	-1.5579	0.3727	-0.0400	0.6063

The x_j's are then transformed into developed yields using:

$Y_j = \overline{Y} + s x_j$ where $\overline{Y}$ is the mean and s is the standard deviation of the developed yields, Y_j

The Y_j's have the desired means, standard deviations, and correlations. Table 6.6 presents the Y_j's with a mean of 1, a standard deviation of 0.4, and a correlation of 0.5 except for sampling errors.

Table 6.6. Developed Yields Y_{jt}

	Y			
Year/j	**1**	**2**	**3**	**4**
1	0.2637	1.0820	0.5001	0.7571
2	1.1647	1.1438	0.9339	1.4076
3	1.6927	1.5165	1.8397	0.9069
4	1.2888	0.8179	1.4571	1.0324
5	2.0380	1.6286	2.0500	1.8107
6	0.7327	0.8094	1.0498	1.0378
7	1.2765	0.7307	1.1918	1.2195
8	0.8089	0.9250	0.2892	1.1007
9	1.1351	0.5290	0.8113	0.8249
10	1.1919	1.2206	0.6060	1.0564
11	1.0216	0.9676	1.4332	1.0049
12	0.6778	0.5772	0.7280	1.0213
13	0.7048	0.5427	0.8249	0.5505
14	0.4878	0.5506	0.5675	-0.0480
15	1.0125	1.5607	1.2669	0.7996
16	1.2812	1.5317	1.6166	0.9675
17	1.2442	0.8455	1.2618	1.1104
18	0.7458	0.9802	0.7840	1.1073
19	0.5567	0.7069	0.5141	0.8364
20	0.9564	0.9861	0.9316	1.1974
21	1.3117	0.8798	1.1495	1.0927
22	1.1850	1.2265	1.0397	1.0983
23	0.4004	1.0703	1.0643	0.4409
24	0.4963	0.5354	0.4529	0.6210
25	1.4689	1.4637	1.1833	1.7152
26	1.2367	1.3842	1.3385	1.1192
27	0.8917	0.3718	0.9793	0.3569
28	1.6016	0.5432	0.7465	0.8757
29	0.9750	0.9720	1.3834	1.7059
30	1.2919	1.0817	0.9684	1.2747
31	1.2329	0.8285	0.9884	1.2138
32	1.2509	1.4443	0.8922	1.1011
33	0.5104	1.0832	1.3155	0.7235
34	0.7904	1.0358	1.0067	1.1576
35	0.7946	0.9665	0.2331	1.0158
36	1.8042	1.8380	1.2838	1.3363
37	1.0403	1.0652	0.4140	0.3353
38	0.3613	0.9758	1.1266	1.3835
39	0.6311	0.3911	0.7002	0.4415
40	0.3769	1.1491	0.9840	1.2425

Table 6.7 presents the indemnity payments calculated from the Y_j's for each farm for a primary insurance coverage level of 0.75. The average indemnity is also calculated across all farms for each year. It is assumed that a reinsurance deductible is used and that the reinsurer absorbs all indemnities exceeding 12% of liability (or, 9% of expected yield). The table presents the indemnity payments from the primary insurance company and the reinsurer.

Table 6.7. Average, Primary, and Reinsurer Indemnities

	Indemnity						
Year/j	1	2	3	4	Average	Primary	Reinsurer
1	0.486	0.000	0.250	0.000	0.184	0.090	0.094
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.017	0.000	0.000	0.000	0.004	0.004	0.000
7	0.000	0.019	0.000	0.000	0.005	0.005	0.000
8	0.000	0.000	0.461	0.000	0.115	0.090	0.025
9	0.000	0.221	0.000	0.000	0.055	0.055	0.000
10	0.000	0.000	0.144	0.000	0.036	0.036	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000
12	0.072	0.173	0.022	0.000	0.067	0.067	0.000
13	0.045	0.207	0.000	0.200	0.113	0.090	0.023
14	0.262	0.199	0.182	0.798	0.361	0.090	0.271
15	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	0.000	0.000	0.000	0.000	0.000	0.000	0.000
17	0.000	0.000	0.000	0.000	0.000	0.000	0.000
18	0.004	0.000	0.000	0.000	0.001	0.001	0.000
19	0.193	0.043	0.236	0.000	0.118	0.090	0.028
20	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	0.350	0.000	0.000	0.309	0.165	0.090	0.075
24	0.254	0.215	0.297	0.129	0.224	0.090	0.134
25	0.000	0.000	0.000	0.000	0.000	0.000	0.000
26	0.000	0.000	0.000	0.000	0.000	0.000	0.000
27	0.000	0.378	0.000	0.393	0.193	0.090	0.103
28	0.000	0.207	0.003	0.000	0.053	0.053	0.000
29	0.000	0.000	0.000	0.000	0.000	0.000	0.000
30	0.000	0.000	0.000	0.000	0.000	0.000	0.000
31	0.000	0.000	0.000	0.000	0.000	0.000	0.000
32	0.000	0.000	0.000	0.000	0.000	0.000	0.000
33	0.240	0.000	0.000	0.027	0.067	0.067	0.000
34	0.000	0.000	0.000	0.000	0.000	0.000	0.000
35	0.000	0.000	0.517	0.000	0.129	0.090	0.039
36	0.000	0.000	0.000	0.000	0.000	0.000	0.000
37	0.000	0.000	0.336	0.415	0.188	0.090	0.098
38	0.389	0.000	0.000	0.000	0.097	0.090	0.007
39	0.119	0.359	0.050	0.309	0.209	0.090	0.119
40	0.373	0.000	0.000	0.000	0.093	0.090	0.003

Table 6.8 presents the expected indemnity payments (the mean of the annual average indemnity column in Table 6.7) and the pure risk premium rate which is calculated as the quotient of the expected indemnity and the liability.

Table 6.8. Expected Indemnities and Pure Risk Rates

Indemnity Payer	Expected Indemnity	Rate as % of Liability
Total or Primary Insurance Rate	0.062	0.083
Primary Insurer	0.036	0.049
Reinsurer or Reinsurance Rate (Pure Risk)	0.025	0.034

The previous example is greatly simplified from actual reinsurance rating. Usually, the generated data set contains several hundred annual observations and the number of farms is often much larger than four. However, if the several thousand farms are considered, then rates may be calculated with sufficient accuracy by using a subsample of farms

6.2.2. Loads. Pure risk reinsurance rates are increased by a load that compensates the reinsurer for a variety of costs that are not easily quantified. Loads are usually determined more by judgment than quantitative analyses. Considerations for developing loads include:

- Reinsurer's servicing costs;
- Confidence in the reinsurance rating including data quality and quantity;
- Program design;
- Program integrity or "tightness" of underwriting and loss adjusting;
- Reinsurance form and level of risk (e.g., is there a government stop loss?);
- Personnel competency;
- Reputation of involved parties;
- Cost of capital;
- Transparency, monitoring, and auditing of insurance activity;
- Political, judicial, and legal risk;
- Regulations;
- Taxes or government fees.

Furthermore, loads may differ within different tranche levels. Often, the shallowest tranche (the tranche that makes the first indemnity payments) and the tranche that has the lowest payment probability (i.e., those that are only triggered by severe losses) are loaded more heavily than middle tranches. The shallowest tranches are loaded more heavily because of higher per contract servicing costs; the deepest tranches are loaded more heavily because of higher capital requirements, even though frequency may be quite low.

7.0. Summary

Actuarial processes include activities that establish insurance premium rates and related quantitative analyses. The major stakeholder groups associated with agriculture insurance are agricultural producers, insurance firms, and governments. Each has different, and sometimes opposing, interests.

Crop insurance shares many elements associated with other forms of insurance. The purchaser of an insurance contract pays a premium to an issuing agency to transfer undesirable outcome risks. Actuarial sound premiums are established such that the expected indemnities (payouts in the case of insured events) and costs of providing insurance are offset by premium collections. An issuing agency often pays a reinsurer to accept much of the risk that has been acquired through the sale of insurance contracts.

All crop insurance contracts require that insured parties absorb a deductible in the event of an insured outcome. The deductible helps reduce moral hazard behavior. For crop insurance, coverage levels refer to the difference between 100 percent coverage (zero deductible) and the stated deductible. Indemnity payments represent a transfer of funds from an insurer to an insured party to partially or fully compensate for insured losses. Such payments are triggered by yield (or in many cases, revenue) levels which are below expected yields less the deductible.

One of the major actuarial activities is developing premium rates that are actuarially sound. That is, premium collections must be sufficient to offset indemnity payments and the costs of providing insurance. The rating process depends on the quality of data.

As with all insurance products, issues related to moral hazard and adverse selection must be vigilantly examined. The use of index insurance as an indemnity trigger mechanism reduces moral hazard and monitoring costs. Index products have been developed around weather variables, area yields, and satellite imagery. Such products are also highly data intensive. Adverse selection issues are often managed by pooling producers into appropriate risk categories and then rating each group according to their level of risk.

Rating processes must not ignore issues important to reinsurers. Private reinsurance companies have the diversification and financial reserves to manage risks taken by issuing agencies. In general, issuing agencies pay fees to reinsurance companies as a means for transferring risk to entities that have the capacity to absorb it. In many cases, governments serve as reinsurers to varying degrees. Actuarially sound rating processes help reduce the costs of reinsurance.

Accuracy, transparency, careful data management, statistical expertise, detailed written contracts, strong legal institutions, well-developed contract law and property rights, and clearly identified government roles are all important elements for developing actuarially sound rating processes. Crop insurance programs will not be successful if any of these elements are ignored or poorly realized.

Index

Contact information:
International Finance Corporation
2121 Pennsylvania avenue, NY
Washington, DC 20433, USA
www.ifc.org

WORLD BANK GROUP