Methodology Analysis for Weighting of Historical Experience

Implementation Report

September 5, 2011

By

Keith H. Coble
Mary Frances Miller
Roderick M. Rejesus
Ryan Boyles
Thomas O. Knight
Barry K. Goodwin

Contact:

George Duffield
Sumaria Systems
618-632-8450
gduffield@sumaria.net

Keith H. Coble
662-352-9491
khcoble11@aol.com

Table of Contents

Executive Summary

In March 2010, Sumaria Systems Inc. (Coble, et al. 2010) completed a comprehensive review of the methodology and procedures used to determine APH target rates and COMBO rates under the Federal Crop Insurance Program.[1] The study provided several recommendations for modifying the current APH and Combo methodologies and suggested further evaluation of several other issues. One of those issues involved the current RMA practice of using equally-weighted, adjusted, historical, loss cost experience for a county/crop program as the cornerstone of the current rating procedures. Sumaria Systems was subsequently contracted by the USDA/ Risk Management Agency (RMA) to conduct additional analysis of this issue. The statement of work for this project directs Sumaria to perform a detailed investigation and to develop an improved methodology for weighting, or otherwise adjusting, RMA's historical loss cost data in order to maximize its statistical validity for developing premium rates.

The project commenced in September 2010. A draft technical report examining various options and approaches to the issues was submitted to RMA in March 2011. The agency evaluated those options and determined which were preferred. The agency also commissioned further analysis and detailed explanation and documentation of the procedures selected for implementation. This report provides that additional analysis, explanation and documentation.

Our team, including experienced crop insurance analysts, a leading professional actuary, and a professional climatologist, has reviewed the materials provided by RMA and additional materials that we collected independently. The credentials of our team are discussed in greater detail in the Appendix of this report.

This report examines a number of conceptual considerations related to the issues we were tasked to address. Our team evaluated the alternative weather data available and issues associated with using those data to characterize weather probabilities. We have conducted analysis nationally for nine crops (apples, barley, corn, cotton, potatoes, rice, sorghum, soybeans, and wheat). Based on this analysis we make several recommendations.

Weather Probabilities

Recommendation 1. – We recommend that RMA use Climate Division Data for calculating crop-specific weather indexes. We believe the weather data collection that best meets the weather-data criteria outlined in Section 4 of this report is the National Climatic Data Center's Time Bias Corrected Divisional Temperature-Precipitation-Drought Index data, also called the climate division data. The climate division data provide several drought indexes and other weather variables that are time-aggregated to the monthly level and spatially-aggregated to the climate division level for the years back to 1895. Thus, the data allow RMA to compare the weather experience incurred by the modern program to weather extending 80 years prior to the 1975 cut-off of available loss-cost data.

[1] This report is available at http://www.rma.usda.gov/pubs/2009/comprehensivereview.pdf.

Recommendation 2. . – We recommend that RMA use fractional logit models estimated at the climate division level to relate loss cost experience to the Palmer Drought Severity Index (PDSI) and Cooling Degree Days (CDD). Time period variants of both weather indicators should be used for different crops and locations. An out-of-sample forecasting competition is suggested to select the time-period/variables for a crop/climate division, and if the models are not found statistically significant we recommend no weather weighting. This process creates a weather index from 1895-present which characterizes the growing conditions experienced in each year.

Recommendation 3. – Given recommendation 2 we propose that RMA categorize the loss cost experience observed over the period chosen into weather 'probability bins' or categories. These bins would be chosen according to an incremental procedure which would select a parsimonious number of bins for the crop/climate division. Once observed loss costs are categorized within bins, all historical loss costs within a bin are given equal weather probability. The bins recommended would have variable width but equal probability. The variable width binning process we propose ensures that at least one year during the rating period is classified in each bin, thereby providing proper weights that reflect all of the historical weather data.

Recommendation 4. – While not a directive in the statement of work, a conclusion reached during our analysis is that RMA should use all years available to calculate the catastrophic load and that extreme loss costs within the catastrophic load should be weighted using the weather index probabilities. Further, we recommend changing the catastrophic load cap to the 90th percentile and reducing the aggregation region for catastrophic load from the state level to a climate division, which is consistent with the weather weighting procedure. We also recommend dampening of the weight given to the most extreme weather years. Specifically, if the weather index for a particular year is above the 97th percentile, we recommend that the weight given to that year's input to the catastrophe load be adjusted to reflect the percentile of the weather index. That is, if the data span 30 years of experience, a year with a weather index at the 98th percentile should be given 2% (1-in-50) weight rather than 3.33% (1-in-30) weight. The weight taken from the adjusted year should then be spread evenly among the remaining years.

Changing Severity of Loss Costs

We were also directed to consider changing severity of loss costs over time due to technological advances and changing agronomic conditions. Finally we were asked to address how to incorporate program participation changes over time in a way that represents the current program. In response to these tasks, we added an additional recommendation:

Recommendation 5. – A variety of factors suggests non-stationarity in some RMA loss cost data. Such factors include an expanding participant pool, evolving production systems, the advent of biotechnology, and changing program underwriting rules. In many cases it is difficult, if not impossible, to disentangle these effects. We recommend that RMA use adjustments to remove non-stationarity from the loss cost history when statistical analysis supports the adjustment. We recommend estimating these adjustments at the national level for a crop and that weather should be taken into account when these models are estimated. Further, symmetric caps on the magnitude of the adjustments should be imposed to avoid excessive modification of the loss history in any particular location.

We first recommend application of a discrete adjustment for data prior to 1995 to the adjusted loss cost data. Specifically, we recommend estimating the effect at the national level and calculating a percentage difference by state using the effect relative to the post-1995 average loss cost. However, we stress that where analysis indicates that non-stationarity in the loss cost history is not statistically significant, no adjustment should be made

Second, we recommend shortening the loss history for base rates to 20 years while using a longer series of years for catastrophic loading. This recommendation reflects the recognition that a longer time series is needed to capture extreme events than for measuring the risk quantified by the base rate. Finally, we recommend using net acreage weighting within probability categories or 'bins', which recognizes the additional credibility of experience that is based on more exposed acres.

Report Organization

The primary focus of this report is to describe in detail the proposed method and illustrate how it would be made operational by RMA. First, a background of the issues investigated is provided and then the conceptual basis for the proposed method is discussed. In chapter 4 details of the components of the rating system are described and illustrated. In chapter 5 a summary of the aggregate effects for corn and soybean base rates is provided.

1. Study Background and Motivation

The Federal Crop Insurance Program provides insurance products to agricultural producers in the U.S. In 2010, the program insured 256 million crop acres with a total liability of $78 billion. This public-private partnership involves private delivery of products designed and rated by the USDA. Private firms sell and service the products and are compensated for delivery and offered reinsurance. Producers are offered subsidized rates for the various insurance products. These rates are predicated upon RMA being able to quantify the actuarially fair insurance rate. Specifically, the Federal Crop Insurance Act was amended by the Agricultural Risk Protection Act of 2000 (PL106-224) to state the following regarding rate making:

1) Sec. 508(i) (2) states "Review of rating methodologies. To maximize participation in the Federal crop insurance program and to ensure equity for producers, the Corporation shall periodically review the methodologies employed for rating plans of insurance under this subtitle consistent with section 507(c)(2)."

2) Sec. 508(i) (3) states "Analysis of rating and loss history. The Corporation shall analyze the rating and loss history of approved policies and plans of insurance for agricultural commodities by area."

3) Sec. 508(d) (2) states "the amount of the premium shall be sufficient to cover anticipated losses and a reasonable reserve."

These three statements can be interpreted through standard actuarial definitions. The *Statement of Principles Regarding Property and Casualty Insurance Ratemaking* identifies a fundamental principle of insurance ratemaking as: "A rate is an estimate of the expected value of future costs." Typically, the largest component of the rate is the provision for losses. While there are other important considerations in rate development, most of the actuarial foundations of ratemaking are intended to provide a framework for estimating the expected loss component of the rate.

The current RMA COMBO programs are composed of a mix of individual-level revenue and yield insurance. The COMBO product rates are constructed on a foundation of yield insurance rates with revenue rates being overlaid on the APH yield rating system. In this analysis the focus is on proper weighting of historical experience to derive actuarially fair yield insurance rates. However, these results would then carry through to the related revenue insurance rates. Because different crops are subject to different perils and, therefore, varying loss costs, the APH procedure establishes rates for each crop separately. It is rare that a single insured, for any insurance coverage, will have a sufficiently large insurance history to allow expected losses to be derived solely from the insured's own loss history. Thus, it is common and appropriate to consider the aggregate experience of a group of similar risks in developing rates. For APH, the aggregation is primarily done geographically. Rates are developed by geographic area, usually the county. Thus, for each crop, the APH ratemaking process typically derives Loss Cost Ratios (LCRs), and consequently rates, by county.

In March 2010, Sumaria Systems Inc. (Coble, et al. 2010) provided a comprehensive review of the methodology and procedures used to determine APH target rates and COMBO rating in the Federal Crop Insurance Program.[2] The study provided several recommendations for modifying the current APH and Combo methodologies and suggested further evaluation of several other issues. One of those issues involved the current RMA practice of using equally-weighted, adjusted, historical, loss cost experience for a county/crop program as the backbone of the current rating procedures. The current system uses a fairly lengthy data series of observed loss costs and gives each year's experience equal weight.

More specifically, RMA currently utilizes insurance experience back to 1975, where available. An earlier report by Josephson, et al. (2000) summarizes the history of how RMA has evaluated the length of experience period.[3] According to this document, in a study in 1983 performed for FCIC, Milliman and Robertson (M&R) evaluated the length of the experience period. That study concluded "…. the FCIC should continue to use all available past history in the ratemaking process with possibly greater weight given to the more recent years." (Josephson, et al. 2000, p. 17). At the time of the 1983 study, each year was given equal weight in the determination of the county average. The suggestion to give greater weight to more recent years was made because of concerns about the impact of amendments to the FCIC Act of 1980, and the possibility that the pre-1980 experience might not be relevant. The issue was addressed again by M&R in 1995 and in 1996. In the latter report, M&R again recommended no changes to the practice of equal weighting of all years.

The review of the APH Rating Methodology by Sumaria (Coble, et al. 2010) recommended that RMA continue to use loss experience as the foundation of the rating system. However, the study recommended that RMA evaluate alternative loss cost experience weighting procedures that incorporate additional information such as weather data, historical yields, or the amount of participation. The study recommended that RMA consider altering the weight given to its historical loss costs. The weights could potentially be based on a longer time series of weather variables. Another possibility, not necessarily mutually exclusive with the previous approach, is to adjust the weights according the level of participation (potentially measured by liability or the proportion of total acres insured). The study also suggested that changes in technology or in the composition of the pool of insured producers over time may suggest that the loss costs observed from a particular historical event would be different had it occurred in today's crop insurance program (see Section 6.11 of Coble, et al. 2010).

Sumaria Systems was subsequently contracted by the RMA to conduct additional analysis of this issue. The project commenced in September 2010. This implementation report describes a

[2] This report is available at http://www.rma.usda.gov/pubs/2009/comprehensivereview.pdf.
[3] This report is available at http://www.rma.usda.gov/pubs/2000/mpci_ratemaking.pdf .

proposed procedure for weighting the historical experience used to develop rates for the APH product. Sumaria performed a detailed investigation to develop an optimal methodology for weighting, or otherwise adjusting, RMA's historical loss cost data in order to maximize its statistical validity for developing premium rates. That report was provided to RMA which then selected the approach that best fit the needs of the program. This is a second report developed by the Sumaria team, which includes experienced crop insurance analysts, a leading professional actuary, and a professional climatologist (the credentials of our team are discussed in greater detail in the Appendix of this report.). In this report we present an implementation plan or model that RMA can incorporate into its current methodology.

2. Background Summary of the Current RMA Loss Cost Rating System

The RMA rating procedures use historical loss cost experience for a crop in a county in developing county base rates. These county base rates are then adjusted for factors such as coverage level, unit format, crop type, and crop practice to obtain a rate for an insured unit. In this chapter we describe current procedures followed in developing county base rates. The summary provided here draws heavily from detailed descriptions contained in an RMA internal document entitled "Rate Methodology Handbook: Actual Production History" which is applicable for 2011 and subsequent years. We also draw upon the aforementioned 2010 Sumaria review (Coble et al. 2010).

The Statplan database forms the foundation for the APH rating process. The result of these procedures is the construction of a set of data tables. Two of these tables, the production ratio table and the county summary table, contain the essential data that support the actual production history rating process. The production ratio table contains the data used in computing production ratios, which are discussed in detail in the 2010 Sumaria review (Coble et al. 2010),and the county summary table, which contains information summarized at the county level and is used in evaluating specific risks such as prevented planting. The following are several specific issues addressed in the development of the Statplan database.

- Adjusting for *Winter Kill Experience* in winter wheat and barley.
- *High Risk Experience* -- Because high-risk experience is not considered to be consistent with experience from other land in a county, this insurance experience is excluded from the production tables upon which base rates are determined.
- *Whole Farm Units*-- The Revenue Assurance product offered whole farm units which combine the coverage for two or more crops in a county. Experience for this combined coverage cannot be segregated by crop and is therefore excluded from all Statplan data tables.
- *Prevented Planting*-- Prevented planting is not considered to be a production loss and so prevented planting indemnities and associated liability are excluded from the production ratio tables. These indemnities and liabilities are captured in other Statplan databases for use in prevented planting reviews.
- *Written Agreements*-- Insurance experience established under a written agreement is excluded from the standard Statplan rating data.
- *Late Planted/Planting Adjustments*-- Late planting insurance experience is first adjusted to reflect the correct liability/coverage (if it were not late planted) and is then included in the Statplan database.
- *Replants*-- Indemnities paid to insured producers to cover the cost of replanting are not included in the base rate calculations and thus are not stored in the yield ratio or county summary tables. However, the liability and any indemnities paid on replanted acreage are included in the Statplan tables and in base rate development because the acreage is planted under conditions that are expected to produce at least the guaranteed yield.

- *Revenue Adjustments*-- Three revenue insurance products were introduced by the RMA in the mid 1990s--Revenue Assurance (RA), Crop Revenue Coverage (CRC), and Income Protection (IP). All of these products insure producers against shortfalls of gross revenue below a guaranteed level and in all three the yield risk component of the coverage is based on APH procedures. *RMA* transforms indemnities for CRC and RA to be equal to what they would have been had the coverage been based on the fixed APH Price Election rather than the revenue plan base price and harvest price. The result is a calculated indemnity, *for insured units that are indemnified*, that is equal to what the indemnity would have been under APH yield insurance. This achieves consistency within the Statplan data across the APH yield insurance product, CRC and RA, with or without a harvest price feature or option[4].
- *Revenue Adjustments for Replanted Acreage*-- The process described in the previous bullet is used to convert revenue product loss experience to equivalent yield losses. A similar process is followed for replant losses.
- *Coverage Level*--The common coverage level used as the base for APH rating is the 65% coverage level. Therefore, loss experience for units insured at levels above 65% must be adjusted downward to reflect what it would have been at the 65% coverage level and loss experience for coverage levels below 65% must be adjusted upward to what it would have been at the 65% coverage level.
- Once RMA has adjusted existing loss experience in the Statplan data development process, the actuarial branch begins a multi-step process to develop a target rate for each county/crop program. In effect, the target rate is the rate that should serve as the base upon which rates in a county are anchored.

RMA uses a catastrophic loading procedure to reduce the influence of outliers in the experience of a county/crop program. Because crop losses are often characterized by infrequent but severe losses, even several decades of county loss experience may be subject to sampling error. Catastrophic loading is an actuarial technique used to mitigate the effect of sampling error when the true magnitude of the sampling error is not known. Catastrophic loading is intended to remove anomalous experience from the county/crop data while preserving normal loss experience. In general, losses deemed catastrophic are spread across all counties for a crop in a state. Thus, the capping of loss cost experience in a county/crop program is not a load in the sense that it is an additional factor added to rates, but rather it redistributes loss experience within a state/crop program.

The current RMA procedure censors the county loss experience at the 80th percentile of the historical county experience. No distributional assumptions are required for the procedure. To illustrate this, assume 30 years of data are available for the county/crop program. Then the 80th percentile of the loss cost is the 24th highest observed loss cost ratio (note when the percentile does not fall on a discrete observation, a linear interpolation is used). All indemnities above the

[4] IP experience is not included in APH base rate calculations because of differences in product design.

truncation point are aggregated to the state/crop program level. For a county, the catastrophic (CAT) indemnity is calculated as follows:

3. Conceptual Explanation of Weighting and Loss Cost Adjustments

a. Weather Weighting

One issue that should be considered in the weighting of historical loss experience is the representativeness of weather experience reflected in the Statplan data used for calculating county base rates. Statplan is a loss experience data set that utilizes information from 1975 (where available) onward (i.e. 35 years of data in 2010). In many lines of insurance, 35 years of loss history would be considered a very "long" time series of data to use in rate making. However, 35 years may be a relatively short series for accurately reflecting probabilities of the weather events that are a dominant factor in crop losses.

For example, given the current use of simple averaging of loss cost data to calculate county base rates, the severe loss years of 1988 and 1993 are each given 1/35 weight but the long term frequency of the weather events that drove these losses may be greater or less than 1/35. It could be that the 1988 drought was a 1 in 20 year event rather than a 1 in 35 year event. If so, a larger weight than 1/35 would be appropriate for that year. Alternatively, it could be that drought events observed in 1988 only occur 1 in 50 years in a longer weather time series and should be given less weight than 1/35. The intent of weather weighting of loss cost data is to bring additional information from a longer series of weather variables to more properly weight the loss cost data used to calculate average county rates.

In developing a system to weight short loss experience data using longer weather/climate data, one has to consider the following issues: (1) the weather or climate data to use for weighting (e.g., the relevant weather data, the length of the data, the degree of coverage and/or level of aggregation, the relationship of such weather to losses, and the availability of weather variables), and (2) the development of a procedure to properly weight each year in the short loss data (e.g., categorizing each loss data year and creating weights for each year in a manner that is consistent with other parts of the rating process).

Weather/Climate Data

There is an abundance of weather data available in the U.S. that can be used for weather-based weighting of loss experience data. However, there are several issues to consider in choosing the weather data to be used. First, one has to consider the length of the different climate data series that are available. In the context of weighting insurance data, one would like to have the longest series of historical weather data available. This would help ensure that different weather outcomes, especially the rare extreme weather events that cause losses, would be adequately represented in the longer data series. Information about the probabilities of different weather events will be better captured if one has a very long climate data series.

However, the need for a long data series must be balanced with the second issue to consider – the degree of coverage and level of aggregation. For example, there may be weather data that are available for 200 years, but these data sets may only contain data for a particular part of the country and/or aggregated at the national level. Crop insurance covers a large portion of the U.S. and so weather data covering most or all states are needed. In addition, there is significant heterogeneity of the weather events that drive losses at the county level for a particular year. There is value in having data at a lower level of aggregation (i.e., county level or 5 x 5 mile grids) rather than at the national level only. However, in using weather/climate data at lower levels of aggregation, it may be the case that data interpolation methods were involved in the construction of the data, especially at the sub-county level where there frequently are no weather stations in a particular location.

Another factor to consider in choosing the weather or climate data to use in weighting loss experience is the availability of different weather variables that can be used. Longer series of climate data may be available for some basic variables like temperature or precipitation, but variables like drought indexes may not be available for this longer period of time. Climate data at lower levels of aggregation and with wider coverage may only be available for certain weather variables and may be absent for others. Hence, to have flexibility in determining the weather variables that can help to explain losses, the availability of different weather variables in a particular climate data set is also an important consideration.

Finally, in choosing climate data for weather weighting crop insurance loss cost data, the source of the data and the availability of the data in the future are also important considerations. The source of the climate data has to be reliable and must have a good reputation in terms of reporting weather/climate data. Moreover, there should be a reasonable expectation that the weather/climate data will continue to be routinely available in the future to support updating of weather weights and an official rapid assessment for each subsequent year’s weather influence.

Development of Weather Weighting Procedure

Once the weather data have been chosen, the next thing to consider is the development of a weather weighting procedure. The first important issue to evaluate is the choice of weather variables to use in classifying and weighting each loss experience year. A number of weather variables during a specific time period could be related to crop losses and one approach is to simply include all available weather variables (and all time periods) that exist in the chosen climate data. However, this straightforward approach would add complexity to the procedure and might generate a lot of noise, especially if there are a number of weather variables in specific time periods that do not have a statistically significant effect on losses. Further, there are often many different weather variables available such that the capacity to use everything that might exist is limited. Hence, there has to be a balance between simplicity/noise in the data and the number of weather variables (for different time periods) used in the weighting. Consulting the professional literature should provide some guidance as to which weather variables are relevant to yields and what time period to use (i.e., what weather variable at what time periods best explain crop losses). Procedures to evaluate the “best” combination of weather variables to use should also be considered. For example, regressions of losses on different weather variables at different time periods could be conducted. In-sample or out-of-sample model fitting criteria.such as the adjusted r-square or a root mean squared error (MSE) can be used to evaluate the best combination of weather variables to be used in the weighting. Presumably, the weather variables and time periods chosen will be the ones that “best” explain crop losses over time. A weather index can then be created using the chosen weather variables and time periods. One issue to consider here is the level of aggregation to use in constructing the combinations of weather variables and time periods to be used. In other words, will the same weather variables and time periods be used for each county, state, and crop? Alternatively, is one combination appropriate for the entire nation?

Based on the weather index developed, each year in the “shorter” loss experience data set has to be classified relative to the longer term weather index. This will allow for developing the proper weights to assign to each of the actual loss experience years in the shorter data series. There are a number of ways to classify a year and assign a weight. One approach is to generate a histogram with equal bin widths and variable probabilities (or frequencies) (see Coble et al., 2010, p. 85 and Figure 3.1).[5] The bins or groupings with equal widths can then be used to classify each year of the loss experience (i.e. which bin does the loss year belong to given the actual experience) and the probability associated with the bin assigned to the year will serve as the weather weight. An alternative that is recommended is to develop variable bin or grouping widths with equal probabilities associated with each bin (See Figure 3.2). The bins or groupings will again be used to classify each year, but since these are variable width bins with equal probability, there is no need to have differential weights for each actual year of experience. In both of these procedures, one has to evaluate the number of bins to be used and make sure that all bins are represented in the shorter loss data. If not, the weighted average may not fully reflect the available historical experience. In addition, the complexity of the procedure and the ease of implementation should also be a considered in choosing the approach to classify and assign weights to the actual loss years.

Another issue to consider in the development of the weather weighting procedure is its consistency with other rating procedures such as the catastrophic loading (i.e. state excess load). To the extent possible, the proposed weather weighting procedure should allow for the catastrophic loading currently used by RMA, which caps the adjusted loss cost ratio at a fixed percentile for all available years. There should also be some conceptual evaluation of the appropriateness of the catastrophic loading methods, given the introduction of weather weighting in the rating system.

[5] Alternative methods such as generating kernel densities or fitting parametric distributions can also be used instead of histograms. However, one should recognize that these more complex procedures may have implications for implementation. One has to weigh the relative benefits of more complex approaches against the efficiency and ease of more simple approaches (like using a histogram).

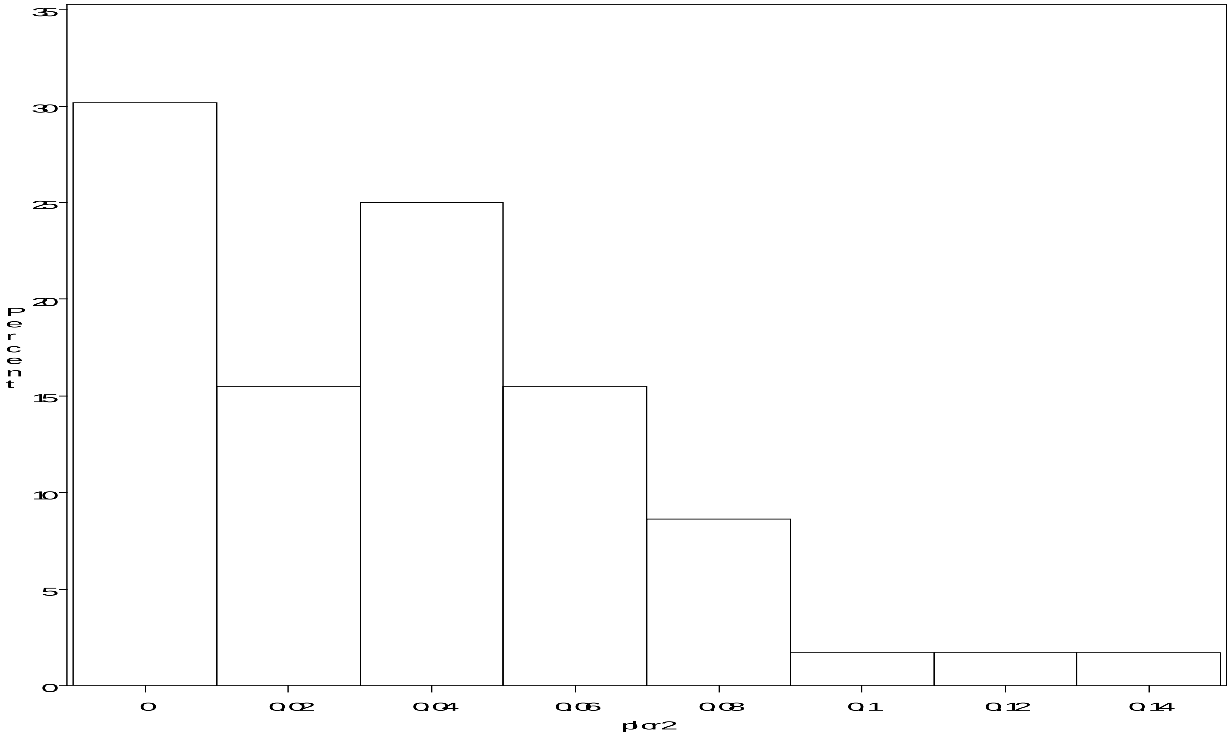


Figure 3.1. Histogram with equal bin widths and variable probabilities for each bin.

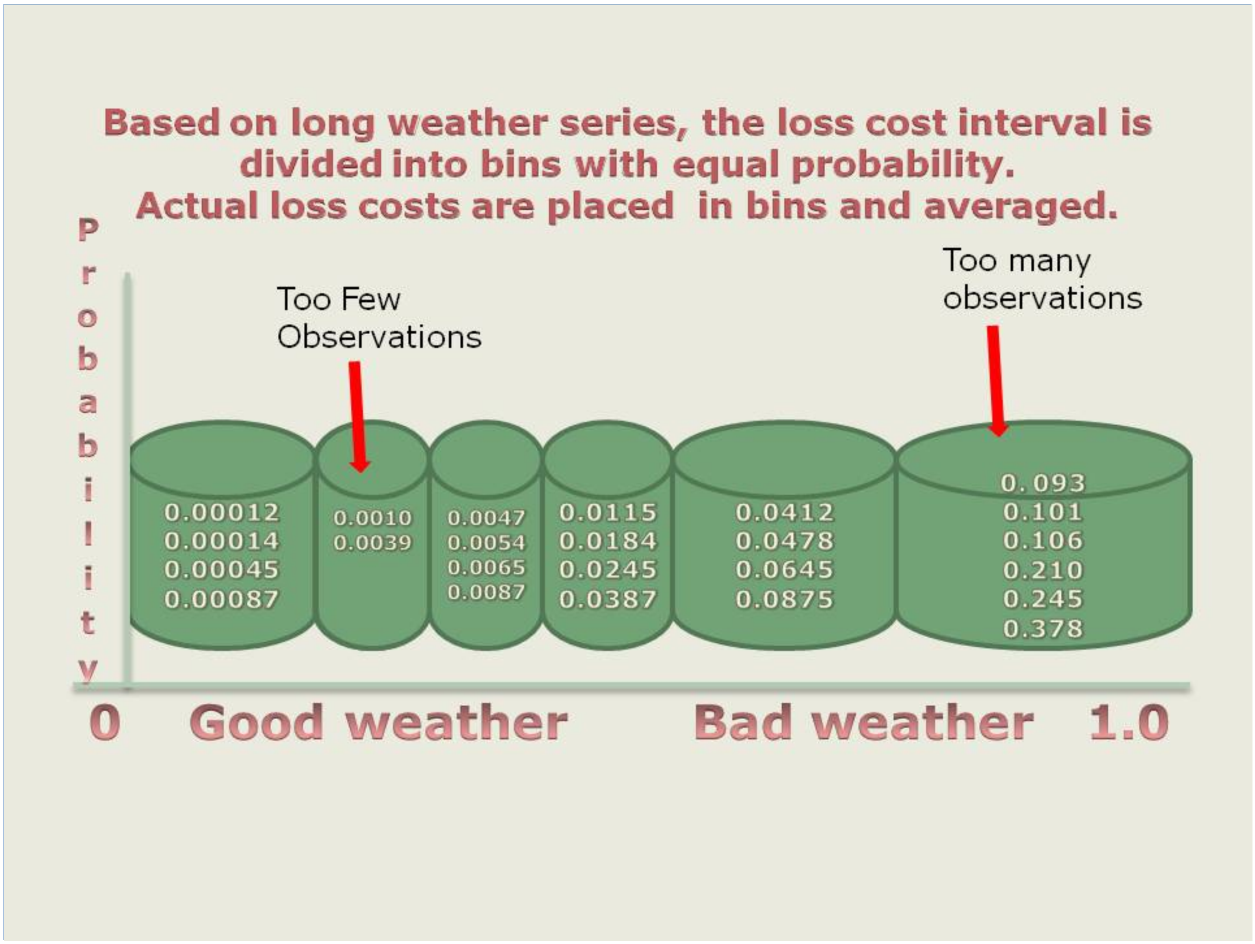


Figure 3.2. Variable bin widths with equal probabilities for each bin.

b. Conceptual Assessment of Non-stationarity in Loss Cost Data

The objective of ratemaking is to provide an estimate of the expected value of future costs. While historical exposure and loss experience provide the starting point for ratemaking, the relevance of the historical experience must always be considered. The *Statement of Principles Regarding Property and Casualty Insurance Ratemaking*[6] (Casualty Actuary Society 1988) notes that ratemaking begins with historical experience, but then goes on to discuss necessary considerations in the ratemaking process that may affect the reliance the actuary can place on the data. Among other considerations, the *Principles* (Casualty Actuary Society 1988) call on the actuary to consider the following factors.

- Homogeneity of the data: including subdividing or combining data so as to minimize the distorting effects of operational or procedural changes.
- Trends: past and prospective changes in claim costs, frequencies, and exposures.
- Policy provisions: past and prospective changes in coinsurance, coverage limits, deductibles and other policy provisions.
- Mix of business: past and prospective changes in the distribution of policies among deductible, coverage selections or type of risk that may affect frequency or severity of claims.
- Operational changes: past and prospective changes in the marketing or underwriting process.

When the effect of such changes can be measured (historically) or projected (prospectively), the actuary adjusts the data accordingly. There is extensive actuarial literature on adjustments such as trending of loss and premium or exposure data, including a standard practice on trending procedures in ratemaking (Actuarial Standards Board 2009).[7]

The property/casualty ratemaking process is a dynamic activity – insured characteristics, the mix of business and the economic environment are constantly shifting, making incorporation of appropriate adjustments for such changes extremely difficult even for relatively recent experience. Precedent and common usage within the actuarial profession steer the actuary to minimize the length of time spanned by the historical data used in the ratemaking process to just enough to be statistically reliable. In the absence of statistically reliable data beyond a relatively short historical time period, actuaries turn to credibility weighting against other contemporary estimates rather than expanding the history.

The reasoning behind using a relatively short time span for an insurer in a competitive market is clear: an insurer's mix of business is bound to shift over time as its market position changes.

---

[6] This document can be found at http://www.casact.org/standards/princip/sppcrate.pdf.
[7] This document can be found at http://www.actuarialstandardsboard.org/pdf/asops/asop013_114.pdf.

However, it is not only competitive pressures that affect the mix within an insurer's experience. Characteristics of the same insureds change over time: policyholders age, their homes become older, they turn over older vehicles for new ones, etc. Commercial exposures also change over time: ownership changes, workplace safety improves, manufacturing processes are upgraded, etc. Insurer procedures also affect the underwriting results: policy provisions and settlement processes evolve over time. Capturing – and appropriately reflecting – all such changes (and their interactions) is virtually impossible.

Thus, it is typical for the actuary to consider *how short* a time period is required for reliable ratemaking rather than *how long* is the period of available data. In general, the larger the size of the exposure, the smaller the time period utilized by the ratemaking actuary. While relatively small commercial carriers may use five to ten years of their own experience (weighted against rating bureau rates) and small personal lines carriers typically use five years of experience for property exposures and three to five years for automobile ratemaking, the National Council on Compensation Insurance (the rating bureau for workers compensation in most states) utilizes only two policy years in its standard ratemaking procedure. The NCCI's database encompasses virtually all of the insured business within a state, so the mix of business itself is not an issue; however, changes within the insureds themselves are still assumed to be present, and the NCCI limits the historical data in its ratemaking process to the minimum needed to produce a stable indication.

In cases where the data over a short time span are not considered to be fully credible, it is also common practice for the actuary to use a somewhat longer time period (such as five years of data rather than three), but then to judgmentally assign less credibility to the older periods through the use of decreasing weights.

The exception to the common practice of using fewer rather than more years of data exists in procedures used to account for very infrequent extreme events. In that case, the actuary is forced to expand the time spanned by the ratemaking data in order to ensure that a reasonable estimate of the frequency and/or severity of large loss events are captured. In order to preserve the desired short-term nature of the historical data used for the non-catastrophic portion of the rate, extreme events are sometimes projected entirely separately from the rest of the rate. This method assumes that large events are independent of the smaller events, and also that the need to capture the extreme events in the rate outweighs shifts in the business that have not been captured by adjustments to the data. Alternatively and more typically, extreme event data over a longer time period are analyzed in terms of their ratio to losses excluding extreme events, and then the projected extreme event ratio is applied to the non-catastrophic portion of the rate (Werner and Modlin 2010).[8] This technique assumes that the extreme event ratio is relatively constant over time, and when it is applied to a non-catastrophic rate based on recent data, any changes in the mix of business will be captured. This assumption tends not to hold in the event of natural catastrophes because (a) the time period of available data is too short to capture the potential range of loss outcomes and (b) the mix of business has shifted dramatically toward

[8] This publication is available at http://www.casact.org/pubs/Werner_Modlin_Ratemaking.pdf see pages 107-111.

higher exposure areas, resulting in understated historical catastrophe to non-catastrophe ratios. The third method for catastrophe ratemaking involves direct modeling of the projected experience using a comprehensive tabulation of the exposures and an extreme event model derived from outside sources. The insurance industry typically relies on sophisticated natural catastrophe models for its hurricane and earthquake exposures and on scenario-based models for extreme events such as terrorism.

When applying these principles to RMA's ratemaking process, we ask the following questions.

- How are current conditions and the mix of business different from those reflected in the historical experience, and can the data be adjusted appropriately?
    - Are there identifiable significant shifts in the program that would be expected to affect all data prior to a particular date?
    - Are there identifiable trends in the experience that can be captured?
- How many years of data are necessary for determination of the base rate?
- How many years of data are necessary for determination of the catastrophe provision?

Explicit Adjustments for Changes

We have observed that there is a significant discontinuity in the data for many crops that occurs around 1995. This corresponds to known changes in the way the program was administered and to a marked increase in program participation. Figure 3.3 provides evidence of the change in RMA's book of business over the past three decades. This graph plots the net acres insured for the six major crops (corn, soybeans, wheat, cotton, rice, barley) from 1981-2009. One can see a distinct change in participation in 1995. Prior to 1995 there had been a strong upward trend in participation but legislative changes in 1994 resulted in an almost doubling of insured acres in 1995. Further, after a slight drop off in 1996 and 1997 net acres have largely remained above 170 million acres. While not shown in this graph, much of the 1995 participation was in catastrophic coverage policies, however much of that acreage has now migrated to buy-up coverage.

This type of program dislocation is appropriately captured in the ratemaking process by measuring the average effect of the change at a macro level and then applying an adjustment to the data prior to the change. Comparable adjustments, for example, can be found in the NCCI's process for accounting for benefit changes adopted by state legislatures. The expected effect of the benefit change is calculated, and all experience prior to the change is adjusted uniformly for the expected effect so long as it remains in the ratemaking data. Once the years prior to the change roll off, no adjustment is necessary.

Selection of the Number of Years for Basic Ratemaking

RMA's program differs from most property/casualty exposures in that the loss experience is highly correlated with weather patterns. Even after capping the experience at the 80th or 90th percentile, it is still very important to capture a representative sample of weather outcomes within the ratemaking process. The need to capture variation in the weather precludes the exclusive use of a very short time series of data as would be used in a more typical exposure. However, once we have identified how "typical" a year's weather is via the weather index, we need only to ensure that we have captured sufficiently many observations within each range of weather outcomes. Although the range of modeled loss costs within the bins at the high end of the weather index will be very large, the loss costs within high end bins will generally be capped by the catastrophe procedure prior to their use in the basic ratemaking procedure.

We examined the number of observations (years of data) necessary to ensure that there is a high probability that no one year will get too much weight in the calculation due to being the sole observation in a very large bin. With only 15 years of observations, there is about a 25% probability that the binning procedure will end up with five or fewer bins and that one of the bins will still only have one observation. That one observation would receive 20% weight in the calculation of the indicated base rate, which is more weight than we would like to place on a single year's experience. With 20 years of data, the probability of a single year getting that much weight drops to about 8%, with 25 years it's around 2.5%, and with 30 years of data it is about 1%.

The high probability of placing 20% weight on a single observation indicates that 15 years of data is probably insufficient. However, at some point actuarial judgment would lobby for dropping data years that are so far removed in time as to be unlikely to be representative of current experience: hence we recommend that RMA consider limiting the number of years of data for *base* ratemaking to 20 years.

Judgmental Credibility Weighting for More Recent Data

Given the long time span required to assure a reasonable weather distribution in the base rate calculation, generally accepted actuarial practice would consider judgmentally assigning less weight to older years in the data. The effect of the necessary adjustments for program changes and trend discussed above tend to be compounded over time, causing the loss cost estimates from older years to become more and more dependent on estimated adjustments over time. The proposed method for grouping the data by weather indexes would allow for a judgmental credibility weighting of observations based on time within the same weather index range.

Selection of the Time Period for the Catastrophe Load

When considering the catastrophe load, however, the maximum amount of relevant data should be used. RMA's current procedure uses all available data, and we recommend that the full data series continue to be used, with the possible exclusion over time of early years if the covered acreage is very low relative to current acreage. On the other hand, if the weather index for a particular year is above the 97th percentile, one may want to adjust the weight given to that year's input to the catastrophe load to reflect the frequency reflected by the percentile of the weather index. That is, if the data span 30 years of experience, a year with a weather index at the 98th

percentile should be given 2% (1-in-50) weight rather than 3.33% (1-in-30) weight. The weight taken from the adjusted year should then be spread evenly among the remaining years.

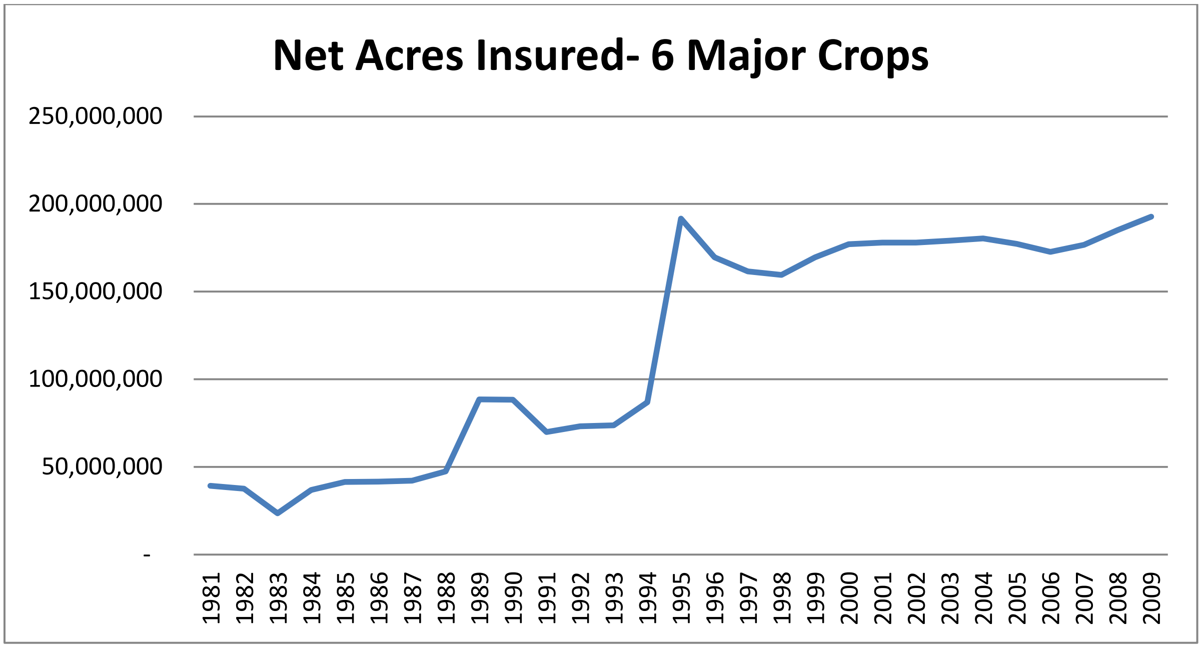


Figure 3.3 Net acre insured change from 1981 to 2009.

4. Proposed Weighting of Historical Experience

a. Weather Weighting

In order to quantify the relative frequency of extreme weather events that may be associated with loss experience, a reference set of climate data is needed that meets the following idealized criteria.

(1) Provides climate information across all geographies where loss experience is observed.
(2) Provides climate information at sufficiently local scales to explain local loss experience.
(3) Provides the longest possible temporal record of climate events to ensure adequate analysis of the frequency of both normal and extreme climates.
(4) Provides specific climate variables that provide meaningful explanation of loss experience.
(5) Is operationally and routinely updated for use in future analysis and weighting.

There are several climate datasets that partially meet these 5 criteria. First, the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) Unified Precipitation Analysis is an interpolation of the available point-based precipitation gauge data collected by both NOAA and USGS. It meets the above criteria (1), (2), and (5), but provides only information on precipitation and has data only since 1948. Important information on temperature and drought are not provided, and these data do not allow for characterization of the relative frequency of known extreme drought events in the 1920s and 1930s nor hurricane or flooding events prior to 1948.

A national analysis of Palmer Drought Severity Index developed by Dai et al. (2004) meets criteria (1), (3), and possibly (4), but is not updated regularly and provides drought severity information only every 250 kilometers, which is insufficient to explain local loss experience.

Another group of data that partially meet the criteria involves atmospheric model simulations, including NCEP Re-analysis and the North American Regional Reanalysis (NARR). These products meet criteria (1), (2), (4), and (5), but NCEP re-analysis (and similar) only contain information since 1948 and NARR only since 1979. Similarly, PRISM data from Oregon State University provide detailed model estimates of temperature and precipitation for a very long record and meet criteria (1), (2), and possibly (3), but PRISM does not provide drought estimates and is not considered operationally available.

The data collection that best meets all 5 criteria is the National Climatic Data Center’s Time Bias Corrected Divisional Temperature-Precipitation-Drought Index data, also called the Climate Division Data. Climate Division data provide monthly, serially complete information on temperature, precipitation, relative severity of dry and wet periods using drought indexes, and degree day metrics of heat and cold accumulation since 1895 for the continental United States, grouped into 344 divisions. Updates are operationally provided each month by NOAA's National Climatic Data Center. A useful description of the history and current status of climate division data is provided by Guttman and Quayle (1996). More technical details on the data and adjustment methods are given in NCDC (1994) and Karl et al. (1996).

Climate Division data are produced using more than 5,000 National Weather Service cooperative observer gauge reports. Climate Division boundaries group stations of similar climates into regions that adhere to state political borders. In most cases, the climate division boundaries also follow county boundaries. However, in regions with more complex geography (including some states with complex topography and/or shorelines), climate division boundaries follow river basins within each state. While climate divisions were originally designed in 1912, boundaries were adjusted in the 1940s to align with crop reporting districts or drainage basins. The Climate Division boundaries are shown in Figure 4.1a. In some instances climate divisions split counties. The assignment of counties used in our study is shown in Figure 4.1b. This allocation is based on relative area, geography and other factors.

There are limitations to using Climate Division data. Climate division boundaries are not always delineated for climate homogeneity. Especially in the mountainous terrain of the western US, the boundaries follow drainage basins and all locations within those boundaries are not likely to have very similar climate characteristics as climate changes quickly with elevation. Another weakness is that the station network used for each division calculations is not constant. Stations move, cease operation, and new ones are introduced throughout the history of the observing network. This introduces some error with any divisional calculations. Another weakness is the accuracy of division level data prior to 1931, when regression equations are used to estimate division-level data from statewide average data that were standard during that period.

Despite these weaknesses, Climate Division data provide the best operationally available climate information for crop loss analysis. They provide serially complete national coverage (with no missing data) at a geographic scale sufficient to characterize local climate extremes with a period of record sufficient to identify the relative frequency of climate events that may be associated with loss experience.

Data Preparation

The development of the weather weighting procedure starts by merging the climate data set (see previous sub-section) with RMA's Statplan loss experience data (See Figure 4.1 for the different climate divisions within states). Note that the climate data are observed at the climate division level as described above, while the RMA Statplan data are reported at the county level.[9] This

[9] The county loss data utilized in this study are typically aggregated for all types/practices (with the exception of wheat, where the data are separated to identify winter and spring wheat). This type of aggregation is consistent with the county level data used in calculating the base county rate (see Coble et al., 2010 p. 38).

difference necessitates the use of an additional data set that assigns counties to particular climate divisions. Most counties are entirely or nearly entirely contained by a climate division. Counties associated with each climate division are provided by NOAA NCDC. However, as some divisions (especially in the mountainous western US) are delineated to follow drainage basins, there are many counties (approximately 300) that are split by climate division boundaries. We developed a data set such that each county is assigned to a specific climate division[10] based on 2 criteria:

(1) Counties that are split by 1 or more climate divisions are assigned to the climate division that covers the greatest amount of area in the county.
(2) For counties that are not easily assigned according to (1), the county is assigned to the larger climate division as the larger climate division should have more weather stations in the aggregated value and therefore should have more confidence in the weather representation.

Based on this data set we are able to generate a merged climate-loss experience data set at the county and at the climate division levels.

All counties within a particular climate division have the same weather data. The loss data also must be aggregated to the climate division level. This is done by summing the adjusted indemnities and liabilities of all counties within a climate division level and then calculating loss cost ratios (LCR) based on these summed amounts. The climate division data are used to generate a weather index that is needed for classifying loss years, while the county data are used in averaging the loss cost data to calculate a base county rate.

Weather Index Development

A critical component in the development of a weather weighting approach is the choice of the weather variables that are used to determine the relative weights assigned to each year of loss data. One can use a single weather variable or a combination of different weather variables. Based on the literature (Wilhemy, Hubbard and Wilhite 2002) and the expert opinion of the climatologist in our team, we chose to examine a parsimonious set of weather variables – the Palmer Drought Severity Index (PDSI) and Cooling Degree Days (CDD). PDSI is a particularly good weather variable to examine because it subsumes effects of both precipitation and temperature and provides a locally relative scale ranging from very wet to very dry conditions. Wilhemy, Hubbard and Wilhite 2002 show that much loss experience is associated with drought

[10] We build on the NOAA data set that assigns particular counties to climate divisions to develop this data set. This data set cannot be used 'as is' because there are a number of counties (approximately 300) that are assigned to multiple climate divisions. The starting point for the assignments is based on the listing provide by NOAA NCDC. The climatologist in our team (Dr. Ryan Boyles) set a criterion to decide which county is uniquely assigned to a particular climate division (see previous section). In addition, there are county codes created by RMA that are unique to the program (and FSA), such as having East (code=155) and West (code=156) Pottawatamie, IA while the NOAA data simply have Pottawatamie, IA (code =155). These occurrences were accommodated in the data set developed.

conditions, but PDSI also allows for very wet (flood) conditions that may also be associated with loss. CDD allows for examining heat units for a particular time period that affects crop growth. CDD is equivalent to Growing Degree Days (GDD) at a base of 65F, and allows exploration of loss experience that may be associated with extended cold or heat that would not be captured in PDSI.

For the PDSI, we created two variables to represent positive PDSI and negative PDSI values. Positive PDSI values represent wet spells (i.e., larger positive numbers indicate more moisture) and negative PDSI values represent drought conditions (i.e., larger negative numbers represent more severe drought conditions). In addition, the positive and negative PDSIs we use are limited to the May-June and July-August periods (i.e., average May-June and average July-August PDSIs are utilized in the study). In summary, four PDSI measures are examined in the development of our weather index – May-June PDSI for positive values (mj_pdsi_p), May-June PDSI for negative values (mj_pdsi_n), July-August PDSI for positive values (ja_pdsi_p), and July-August PDSI for negative values (ja_pdsi_n). The CDD variables used in developing the weather index are total season CDD (from May to September) (total_cdd) and June-July total CDD (jaj_cdd). The June-July periods are periods in which crop growth is frequently adversely affected by heat units.[11]

Based on these six weather variables, an index is created by estimating a fractional logit regression model (at the climate division level) where the dependent variable is the climate division adjusted loss cost ratio and the independent variables are the six weather variables discussed above (see Papke and Wooldridge 1994 for a discussion of fractional logit models). Fractional logit regression is used to account for the proportional nature of the data and censoring of loss costs at zero and one. This approach ensures that predicted values do not fall below zero or above one. [12] Based on our investigation of the degree of censoring of the data at zero, we believe that using the fractional logit is appropriate in this case. The degree of zero censoring in the data ranges from 6-11% for corn and soybeans, to about 30% for barley and potatoes. On the other hand, the degree of censoring at one is significantly lower in the data and it is below 1% for most crops (the exception is apples with censoring at one of about 1.1%).

[11] These six variables apply to all crops except winter and spring wheat. For winter wheat, the following variables are used: Sept./Oct average PDSI (positive and negative), April /May average PDSI (positive and negative), September to May total season CDD, and March to April total CDD. For spring wheat, the following variables are used: April/May average PDS (positive and negative), June/July PDSI (positive and negative), April to August total season CDD, and May to June total CDD. Further note that the durum wheat type has been aggregated with spring wheat.

[12] Note that ordinary least squares (OLS) regression can also be used to estimate the index. The disadvantage of OLS is that predictions are not constrained to lie on the [0,1] interval. Nevertheless, one can argue that the predicted loss costs here are only used as a "tool" to rank the years in terms of having "good" vs. "bad" weather (i.e., one could interpret negative values as indicating good weather years). The magnitudes of the predictions are not used 'per se'. Using the OLS model to estimate the model did not result in significantly different classifications of the loss years (relative to the fractional logit model). However, we recommend using the fractional logit given the degree of censoring in the data and the intuitive concept of limiting predicted loss costs to lie between zero and one.

To have an even more parsimonious model specification, an out-of-sample competition for each state is conducted to determine which combination among the six initial weather variables best predicts losses (i.e., in this case which combination best predicts adjusted loss cost out-of-sample).[13] A minimum mean square error (MSE) criterion is used to evaluate the model with best out-of-sample predictions:

$$MSE = \left(\frac{1}{n} \sum_{i=1}^{n} e_i^2 \right),$$

where e_i is the difference between the actual adjusted loss cost and a predicted adjusted loss cost (out of sample) based on the fractional logit regression model. A lower MSE means that there is a smaller discrepancy between the actual and predicted adjusted loss cost ratios and one would prefer the combinations of weather variables that produce the lowest MSE values. Note that we run independent regressions for each climate division within the state (i.e., climate divisions do not cross state lines), but undertake the out-of-sample competition to find the best combination of weather variables for the entire state. This implies that each regression model is estimated independently but a common specification, in terms of the weather variables included in the regression model, is applied for all climate divisions within a state for each individual crop. In other words, for a crop in a state, the same weather variables are used in the loss-cost regression though parameters on weather variables may differ across climate divisions.

To facilitate the out-of-sample competition for each state, we limit the number of weather variable combinations to be considered to seven: (1) May-June PDSI positive and May-June PDSI negative, (2) July-August PDSI positive and July-August PDSI negative, (3) total season CDD and June-July total CDD, (4) May-June PDSI positive, May-June PDSI negative, July-August PDSI positive, and July-August PDSI negative, (5) May-June PDSI positive, May-June PDSI negative, total season CDD, and June-July total CDD, (6) July-August PDSI positive, July-August PDSI negative, total season CDD, and June-July total CDD, and (7) May-June PDSI positive, May-June PDSI negative, July-August PDSI positive, July-August PDSI negative, total season CDD, and June-July total CDD. Limiting the combinations to these seven choices and estimating the model for each crop, covering all states allows for less of a computational burden (i.e., runs not to exceed six hours for each crop). A hypothetical example of how an out-of-sample competition works can be seen in Table 4.1. In this example, the lowest MSE is for combination 4. This means that, for this state, the best combination of weather variables to use in creating an index is the following: May-June PDSI positive, May-June PDSI negative, July-August PDSI positive, and July-August PDSI negative. This combination best predicts loss costs out-of-sample.

[13] In-sample fit criteria (such as in a stepwise regression using an adjusted R-squared criterion) could also be used to determine the optimal combination of weather variables. However, there are a number of criticisms to this approach (i.e., bias in the tests to iteratively choose the best variables from the sample, as well as over fitting) that makes out-of-sample competition more attractive in this case (See, for example, Rencher and Pun 1980 and Copas 1983).

Once the optimal combination of weather variables is chosen for a particular crop and state, this combination of weather variables is used to produce a weather index for all of the climate divisions within the state producing the crop. Essentially, the predicted values of the "best" regression model specification are used as the weather index for each year of weather data. Using predicted values (i.e., predicted loss costs in this case), makes it possible to "backcast" a weather index for each year in which weather data are available (e.g., from 1895 onwards) even when there are no available loss experience data for the pre-crop insurance years (See Table 4.2). The relative probability of an extreme weather event can therefore be assessed over a 116 year time span (1895-2010) based on the predicted values. For example, the weather index for 1988 can be compared to other years from 1895 onward to determine the relative probability of this weather event occurring in the larger sample.

A concern with using the predicted values is that there may be cases when even the "best" combination of weather variables does not produce a statistically significant model that explains losses over time. For example, in some climate divisions, the Pearson chi-square test of overall model fit for the preferred model specification is not statistically significant and the correlation of the predicted values with the actual loss costs is actually negative. This means that the weather variables we considered do not have enough power to explain the pattern of losses observed over time and that there is no significant positive correlation between the model predictions and the actual loss costs. This situation occurs, for example, in some areas where extreme wet conditions are very rare and widespread irrigation makes drought conditions unimportant. In that case, we can fairly state that the weather has little effect on the crop insurance loss experience. We flag these cases, and the weighting methods based on the weather index developed are not applied (See a hypothetical example in Table 4.3).

Example Results

In Table 4.4, we show a hypothetical example of the estimation results from a fractional logit regression model based on data for corn in Illinois (climate division 5) and soybeans in Indiana (climate division 1). In these examples, the independent variables used are the "best" weather variables chosen based on the out-of-sample forecasting competition. For example, based on the out-of-sample competition results for corn (See Table 4.5) the "best" weather variables to explain losses in Illinois are the CDD variables (total_cdd and jaj_cdd), which are used in the fractional logit regression in Table 4 (top panel).

Once the out-of-sample competitions and fractional logit regression estimations are undertaken, we flag climate divisions where the chosen models do not produce a statistically significant model fit. In Table 4.6, we show hypothetical examples for Indiana, Iowa, and Kansas where we flagged counties that have insignificant fractional logit regression models (in particular see the Iowa (19) climate divisions where Flag=1). Note that we also flag those climate divisions with less than 10 years of loss cost data (See State Proxy flag in Table 4.7). In these cases, we aggregate to the state level and use the fractional regression estimates at the state level to get the predicted values for these "thin" data climate divisions. In rare cases where there is a climate division in a state without at least 10 observations, we do not apply the models and instead recommend that some form of subjective rating continue to be used to establish rates.

A hypothetical example of predicted loss costs for corn in Iowa (climate division 5) is presented in Table 4.2. The "backcasted" loss costs from 1975 to 1979 are presented in order to show that the predicted loss costs can be calculated for years in which there are no actual loss data. This facilitates the classification of years based on the weather index (predicted loss cost) for the 116 years for which the weather variables are available.

Loss Year Classification and Weight Assignment

Using the weather index values from the regression model, each year needs to be classified and assigned a weight that represents its likelihood as indicated by the longer weather series. As mentioned in section 3 above, one approach is to develop variable width bins (or groupings) with equal probabilities or weights. This approach is done by first determining the number of bins or percentiles and assigning the weather indexes to the appropriate bin or percentile cut-off. For example, assuming that we are interested in 10 bins we would like to find the weather indexes in the long history of weather data that correspond to the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th percentiles, in addition to the minimum and maximum values. In this case, we have variable width bins, since the ranges of the loss cost values used to delineate the bins are not equal across bins, but the probability of falling into each bin is always equal to 10% (See Figure 3.2 in previous section). If the predicted values are normally distributed, the tails (at both ends of the distribution) tend to have wider bin ranges since only a few observations fall in these areas, but the middle bins tend to have smaller widths because a lot of observations fall in these middle bins.

Once the variable width bins are delineated, the weather index value for each year (from 1895 onward) can be classified and assigned to the bin in which it falls. Using the above example, if the bin width for the 10th bin (from the 90th percentile to the maximum) is, say, from 0.09 to 0.15 and the year 1988 weather index is 0.13 (i.e. one of the high loss years), then year 1988 is in the 10th bin. Each year is similarly classified using predictions from the fractional logit regression models. Since the probability of each bin is equal in this approach, there is no need to assign a specific differential weight to each bin.

One issue that needs to be addressed is the number of bins needed to ensure that there is at least one year of loss cost data in each bin (to avoid empty bins). As discussed in further detail below, once the years from 1895 onward are classified based on the weather index, the RMA's actual adjusted loss cost data from 1990 through 2009 are utilized to calculate the average loss cost for a county. Hence, it is possible that years from 1990 through 2009 do not contain a dispersion of data such that each bin has one or more loss costs (i.e., not all bins are represented in the 1990-2009 period). For example, it may be that no year in the 1990 through 2009 period is classified as falling into bin 9. This will have adverse implications for the calculation of the average loss costs if not all bins are represented in the 1990-2009 period (i.e., not all bin probabilities are represented). In particular, a range of observed weather history is not being captured in the weighting of loss costs. Therefore, to address the issue of empty bins and, at the same time, determine the appropriate number of bins, the approach we pursue is to first look at 15 bins and then move down one bin at a time (i.e., from 15 to 2 bins) to establish the largest number of bins for which there are no cases of empty bins in the years with loss data (1990-2009). A minimum bin number of 5 is applied so that no county will have less than five bins. If the "optimal" bin number without empty bins is less than five, then no weather weighting is applied (i.e. essentially just using one bin). This binning process is done for each climate division, and so the number of bins may vary for each climate division within a state.

The variable bin width with equal probability approach is a fairly straightforward method compared to the approach of using kernel densities or parametric distributions. This "simplicity" facilitates the practical implementation of this procedure for multiple crops and for nationwide coverage. Moreover, we believe this variable bin width approach may be better than a standard histogram approach (that has equal bin widths and variable probabilities for each bin) because this mitigates the "empty bin" issue described above. That is, the likelihood of having empty bins for the years with loss data (1990-2009) is smaller under this approach as compared to a histogram approach with equal bin widths and variable probabilities. The number of bins in the variable bin width with an equal probability approach tends to be greater than if we used the histogram approach.

A hypothetical example of bin classification results for soybeans in Mississippi (climate division 1) is presented in Table 4.7. In this example, the number of bins is 10 and this number assures that there are no "empty bins" from 1990-2009 at the climate division level. All bin classifications are represented in the 1990-2009 data (i.e., see Bin Classification column in Table 4.7). We also show in this table that the model insignificance flag and state proxy flag are both equal to zero, which means that the model fit results for this climate division is significant and the number of observations used in the estimation is at least 10.

After each year is classified into a particular bin at the climate division level (for all 116 years), the classified data for each year and the insignificance flags (based on regression model) are then merged with the county level loss data. Since the regressions and year classifications based on the weather indexes are done at the climate division level, all counties within a particular climate division will initially have the same year classification and insignificance flags.

At the county level (where the climate division bin classification and actual county loss cost data are merged), it is possible that there are counties without actual loss cost data for the full period from 1990-2009. For example, there may be counties where no crop insurance was sold on a crop in, say, 1990-1995. This means that actual loss cost data from 1990-1995 are missing for this county. Hence, this again will mean that there may be empty bins when calculating the average loss costs. The binning process above is then re-applied to these counties with less than the full number of years of actual loss cost data. Only the years with actual data are considered in the binning process. Although this is at the county level, the weather index values at the climate division level that have corresponding (i.e. existing) actual county loss cost data are still used as the basis for the second binning process. As with the climate division binning process, a minimum of five bins is the minimum threshold in this county level binning process. No weather weighting is applied if the "optimal" bin number is less than five.

Loss Cost Averaging Procedure

The average loss costs are next calculated at the county level using the 1990-2009 data. We first calculate the aggregate loss cost for each county, which is the current procedure used for computing the county base rate. Then we do a "weather weighting" average of loss costs for each county. This weather weighting is done by first taking the average loss cost within each of the defined bins and then taking the "average of the average loss costs" across the bins. For example, if there are 9 bins within a county, we first calculate a simple average of the loss costs within each of these 9 bins (i.e., one average loss cost for each bin that results in 9 "average" observations). Then, we take the average of the 9 average loss costs for the 9 bins (i.e., "average of the average loss costs"). Since the bins are constructed to have equal probabilities, there is no need for taking a "weighted average of the average loss costs". However, given the approach described above, a recency weighting procedure can be applied when taking the average loss cost within a bin. That is, more recent years of data can be given more weight relative to older years within each bin. Net acreage weighting is the chosen recency weighting procedure applied within the bins. This means that a "net acre weighted average" of actual loss costs within bins is calculated. This accounts for more recent years being more "informative" than older years (given that more recent years tend to have more acres insured than older years). The net acre weighting here also means that actual county loss cost calculated based on more acres insured tend to have more weight than county loss cost data computed from smaller acres insured.

To allow for consistency with the current catastrophic loading procedure, we also calculate the unweighted and weather weighted average loss costs where the adjusted loss cost data are censored at the 80th and 90th percentile.

Example and National Summary Results

A hypothetical example where county level loss costs are merged with the bin classification data can be seen in Table 4.3 for corn in Dewitt County, IL. The unweighted and weather weighted average loss costs at the county level can be calculated using the data presented in Table 4.3. The bin classification column allows us to conduct the weather weighting procedure described above. If the insignificance flag for model fit is equal to one in any county, we do not recommend using weather weighting for the county.

Hypothetical examples of unweighted and weather weighted average loss costs for several counties in Iowa are presented in Table 4.8. Note that we calculate six loss costs averages (i.e. six weighting types) per county where: Weighting type = 1 if the average loss cost is calculated with no weather weighting; Weighting type =2 if the average loss cost is calculated with weather weighting; Weighting type = 3 if the average loss cost is calculated with censoring at the 80th percentile and no weather weighting; Weighting type = 4 if the average loss cost is calculated with censoring at the 80th percentile and with weather weighting; Weighting type = 5 if the average loss cost is calculated with censoring at the 90th percentile and no weather weighting; Weighting type = 6 if the average loss cost is calculated with censoring at the 90th percentile and with weather weighting. In the example in Table 4.8, it can be seen that the weather weighted average loss cost tends to be smaller than the unweighted average loss cost. However, this is not a pattern observed in every county-crop combination. There are cases where the weather weighted average loss costs are higher than the unweighted average loss costs.

Table 4.9 presents the national average of the calculated unweighted and weather weighted loss costs for all crops we examined. This is the liability weighted average across counties (i.e., the liability weighted average (not simple average) of the average county level loss costs based on the 2009 liability of each county). For apples, barley, cotton, potatoes, rice, and spring/winter wheat, the weather weighted average loss costs (at the national level) tend to be smaller than the unweighted loss costs. However, for corn, cotton, sorghum, and soybeans the weather weighted average loss costs (at the national level) tend to be larger. A map showing the pattern of the difference between unweighted and weather weighted average loss costs for corn is presented in Figure 4.5. Around 51% of the counties have weather weighted average loss costs lower than the unweighted loss costs.

Table 4.1. Example of a hypothetical out-of-sample competition for choosing the best weather variables to create a weather index for a state.

Combination No.	Weather Variable Combinations	Mean squared error
1	ja_pdsi_n ja_pdsi_p	0.91210
2	ja_pdsi_n ja_pdsi_p total_cdd jaj_cdd	0.96825
3	mj_pdsi_n mj_pdsi_p	1.14213
4	mj_pdsi_n mj_pdsi_p ja_pdsi_n ja_pdsi_p	0.86039
5	mj_pdsi_n mj_pdsi_p ja_pdsi_n ja_pdsi_p total_cdd jaj_cdd	0.98366
6	mj_pdsi_n mj_pdsi_p total_cdd jaj_cdd	1.01876
7	total_cdd jaj_cdd	0.98623

Note: In the example above, Combination No. 4 is the best combination of weather variables based on a Mean Squared Error criteria. These will be the variables used in the fractional logit regression to create the weather index for a particular state and crop.

Table 4.2. Hypothetical example of weather index values, net acres and actual adjusted loss costs for corn in Iowa (State=19), climate division 5 (1975-2009).

State	Climate Division	Year	Net Acres	Actual Adjusted loss costs	Weather indexes
19	5	1975			0.013088
19	5	1976			0.0066332
19	5	1977			0.01381172
19	5	1978			0.0155085
19	5	1979			0.00979918
19	5	1980	386569.9	0.00850007	0.01860698
19	5	1981	682904.5	0.00165572	0.0066969
19	5	1982	399409.3	0.00290903	0.00939713
19	5	1983	190959.8	0.03955581	0.02977137
19	5	1984	446252.2	0.00654651	0.00991062
19	5	1985	502489.2	0.00422874	0.00852455
19	5	1986	542506.4	0.00542233	0.0075103
19	5	1987	510334.5	0.00063739	0.01377865
19	5	1988	599368.5	0.1357396	0.05201126
19	5	1989	1392289.1	0.01159806	0.00765586
19	5	1990	1166061.1	0.00804332	0.0129854
19	5	1991	852311.4	0.00912895	0.01383259
19	5	1992	897023.9	0.00145545	0.00394455
19	5	1993	818194.7	0.1242836	0.00734594
19	5	1994	981496.4	0.00096833	0.00734679
19	5	1995	1035910.2	0.0045309	0.0170454
19	5	1996	599679.6	0.00172944	0.005732
19	5	1997	1033995.6	0.0015911	0.00671987
19	5	1998	1074943.8	0.0094961	0.04710033
19	5	1999	1150101.9	0.00057391	0.00677308
19	5	2000	1243181.5	0.00022049	0.01780193
19	5	2001	1237287.7	0.00437922	0.0106978
19	5	2002	1311398.1	0.00041306	0.00989905
19	5	2003	1334522.2	0.00168785	0.01217966
19	5	2004	1374407.5	0.00262745	0.00778709
19	5	2005	1332961.6	0.00067134	0.01534896
19	5	2006	1284211.9	0.00101743	0.01114424
19	5	2007	1469130.3	0.00063091	0.02610541
19	5	2008	1440665.6	0.0116602	0.00627608
19	5	2009	1567807.9	0.01434467	0.00631721

Note: The weather indexes are available from 1895-2010. In the interest of space, we only present data from 1975-2009. However, this demonstrates that "backcasted" predicted values can be calculated in years without the actual loss cost data.

Table 4.3. Hypothetical example of county-level data used for calculating weather weighted average loss costs for De Witt county (County=39), IL (State=17): corn.

State	County	Climate Division	Year	Actual Adjusted loss costs	Bin Classification	No. of Bins	Flag =1 if insignificant
17	39	4	1980	0.1237103	10	11	0
17	39	4	1981	0.0083081	3	11	0
17	39	4	1982	0.0040853	2	11	0
17	39	4	1983	0.1285333	11	11	0
17	39	4	1984	0.0081736	5	11	0
17	39	4	1985	0	2	11	0
17	39	4	1986	0	5	11	0
17	39	4	1987	0	9	11	0
17	39	4	1988	0.1321881	10	11	0
17	39	4	1989	0.0007658	2	11	0
17	39	4	1990	0.0031037	3	11	0
17	39	4	1991	0.0008012	10	11	0
17	39	4	1992	0.0006445	1	11	0
17	39	4	1993	0.0004054	3	11	0
17	39	4	1994	0	3	11	0
17	39	4	1995	0.0185295	8	11	0
17	39	4	1996	0	2	11	0
17	39	4	1997	4.105E-05	2	11	0
17	39	4	1998	0.0009253	8	11	0
17	39	4	1999	0.0004244	6	11	0
17	39	4	2000	0	4	11	0
17	39	4	2001	0.0007537	4	11	0
17	39	4	2002	0.0125182	9	11	0
17	39	4	2003	9.802E-05	3	11	0
17	39	4	2004	0.0011999	1	11	0
17	39	4	2005	0.0031927	10	11	0
17	39	4	2006	0.0006764	7	11	0
17	39	4	2007	0.0020617	9	11	0
17	39	4	2008	0.0008186	3	11	0
17	39	4	2009	0.0026792	1	11	0

Table 4.4. Hypothetical example of fractional logit regression results using selected "best" weather variables for the state: corn in climate division 5, Illinois (17) and soybeans in climate division 1, Indiana (18).

Corn: Climate Division 5, Illinois

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-17.6357	15.7925	-48.5884	13.3171	1.25	0.2641
total_cdd	1	0.0101	0.0181	-0.0254	0.0456	0.31	0.5774
jaj_cdd	1	0.0055	0.0333	-0.0598	0.0707	0.03	0.8692

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	27	0.5804	0.0215
Scaled Deviance	27	0.5804	0.0215
Pearson Chi-Square	27	0.5873	0.0218
Scaled Pearson X2	27	0.5873	0.0218
Log Likelihood		-2.7963	
Number of Observations Used		30	

Soybeans: Climate Division 1, Indiana

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.9453	3.2812	-11.3764	1.4857	2.27	0.1318
ja_pdsi_n	1	-0.8383	1.4242	-3.6296	1.9531	0.35	0.5561
ja_pdsi_p	1	0.2246	1.3966	-2.5127	2.9619	0.03	0.8722

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	27	0.4373	0.0162
Scaled Deviance	27	0.4373	0.0162
Pearson Chi-Square	27	0.5476	0.0203
Scaled Pearson X2	27	0.5476	0.0203
Log Likelihood		-2.5996	
Number of Observations Used		30	

Note: All fractional logit results for all "state-climate division-crop" combinations are available from the authors upon request.

Table 4.5. Weather variables chosen for each state to calculate the weather index based on the out-of-sample competition: A hypothetical example for corn.

state	Weather Variable Combinations	Mean squared error
1	ja_pdsi_n ja_pdsi_p	5.1859665
4	ja_pdsi_n ja_pdsi_p	0.1061328
5	total_cdd jaj_cdd	9.7063898
6	total_cdd jaj_cdd	1.8864413
8	total_cdd jaj_cdd	0.5599298
9	ja_pdsi_n ja_pdsi_p	1.8823582
10	ja_pdsi_n ja_pdsi_p total_cdd jaj_cdd	0.3794463
12	total_cdd jaj_cdd	0.9065109
13	total_cdd jaj_cdd	7.304132
16	total_cdd jaj_cdd	1.4533495
17	total_cdd jaj_cdd	0.9245507
18	total_cdd jaj_cdd	0.9172543
19	total_cdd jaj_cdd	1.3256818
20	mj_pdsi_n mj_pdsi_p ja_pdsi_n ja_pdsi_p	3.0718882
21	ja_pdsi_n ja_pdsi_p	1.1527433
22	mj_pdsi_n mj_pdsi_p	7.3063892
23	ja_pdsi_n ja_pdsi_p	2.1138152
24	ja_pdsi_n ja_pdsi_p	3.0374947
25	total_cdd jaj_cdd	0.6621126
26	total_cdd jaj_cdd	6.7038752
27	total_cdd jaj_cdd	3.8878839
28	total_cdd jaj_cdd	10.240115
29	mj_pdsi_n mj_pdsi_p ja_pdsi_n ja_pdsi_p	2.141111
30	mj_pdsi_n mj_pdsi_p total_cdd jaj_cdd	2.4403504
31	ja_pdsi_n ja_pdsi_p	0.4422113
33	ja_pdsi_n ja_pdsi_p	0.1377493
34	ja_pdsi_n ja_pdsi_p	1.5424982
35	total_cdd jaj_cdd	4.1921117
36	total_cdd jaj_cdd	4.940591
37	mj_pdsi_n mj_pdsi_p ja_pdsi_n ja_pdsi_p total_cdd jaj_cdd	3.6997904
38	mj_pdsi_n mj_pdsi_p ja_pdsi_n ja_pdsi_p total_cdd jaj_cdd	9.0963143
39	mj_pdsi_n mj_pdsi_p ja_pdsi_n ja_pdsi_p	0.9783726
40	total_cdd jaj_cdd	8.4841414
41	ja_pdsi_n ja_pdsi_p	0.461521

42	ja_pdsi_n ja_pdsi_p	4.7989437
44	mj_pdsi_n mj_pdsi_p ja_pdsi_n ja_pdsi_p	0.3507331
45	total_cdd jaj_cdd	5.3265052
46	mj_pdsi_n mj_pdsi_p	6.1671228
47	mj_pdsi_n mj_pdsi_p ja_pdsi_n ja_pdsi_p	0.9630748
48	mj_pdsi_n mj_pdsi_p	6.7365194
49	ja_pdsi_n ja_pdsi_p	0.4809822
50	ja_pdsi_n ja_pdsi_p	0.9504381
51	mj_pdsi_n mj_pdsi_p ja_pdsi_n ja_pdsi_p	1.8095605
53	ja_pdsi_n ja_pdsi_p	0.1583591
54	total_cdd jaj_cdd	6.3146905
55	total_cdd jaj_cdd	3.8595231
56	mj_pdsi_n mj_pdsi_p	2.7716175

Table 4.6. Hypothetical example of climate divisions flagged as statistically insignificant in Indiana (State=18), Iowa (State=19), and Kansas (State=20) for corn.

State	Climate division	Correlation	P value	Flag =1 if insignificant
18	1	0.697348	1.849E-05	0
18	2	0.8206735	2.804E-08	0
18	3	0.703589	1.442E-05	0
18	4	0.6154699	0.0002946	0
18	5	0.6592917	7.421E-05	0
18	6	0.7147064	9.123E-06	0
18	7	0.4857597	0.0065023	0
18	8	0.5676294	0.0010696	0
18	9	0.4039534	0.0268396	0
19	1	0.1176057	0.5359587	1
19	2	0.087774	0.6446394	1
19	3	0.4513596	0.0122938	0
19	4	0.2842945	0.1278601	1
19	5	0.4576954	0.0109846	0
19	6	0.8277632	1.67E-08	0
19	7	0.2724787	0.1451881	1
19	8	0.5400418	0.0020673	0
19	9	0.7837669	3.015E-07	0
20	1	0.8007809	1.072E-07	0
20	2	0.8111501	5.434E-08	0
20	3	0.7218704	6.715E-06	0
20	4	0.732416	4.203E-06	0
20	5	0.8057017	1.341E-07	0
20	6	0.8578067	1.388E-09	0
20	7	0.6950983	2.019E-05	0
20	8	0.4734226	0.0082312	0
20	9	0.9378357	2.15E-14	0

Note: If the Flag (last column) is equal to one then the fractional logit regression model is deemed to be insignificant (i.e. the correlation between actual and weather indexes has a p-value > 0.1) or the correlation is negative.

Table 4.7. Hypothetical example of bin classification for soybeans in Mississippi (State=28) climate division 1 (1980-2009).

State	Climate Division	Year	State proxy flag=1 if used state predicted values	Bin Classification	No of Bins for the Climate Division	Flag =1 if insignificant
28	1	1980	0	4	10	0
28	1	1981	0	8	10	0
28	1	1982	0	2	10	0
28	1	1983	0	5	10	0
28	1	1984	0	4	10	0
28	1	1985	0	8	10	0
28	1	1986	0	9	10	0
28	1	1987	0	1	10	0
28	1	1988	0	10	10	0
28	1	1989	0	8	10	0
28	1	1990	0	4	10	0
28	1	1991	0	6	10	0
28	1	1992	0	5	10	0
28	1	1993	0	2	10	0
28	1	1994	0	4	10	0
28	1	1995	0	1	10	0
28	1	1996	0	1	10	0
28	1	1997	0	5	10	0
28	1	1998	0	10	10	0
28	1	1999	0	5	10	0
28	1	2000	0	8	10	0
28	1	2001	0	5	10	0
28	1	2002	0	4	10	0
28	1	2003	0	5	10	0
28	1	2004	0	3	10	0
28	1	2005	0	7	10	0
28	1	2006	0	9	10	0
28	1	2007	0	8	10	0
28	1	2008	0	6	10	0
28	1	2009	0	4	10	0
28	1	2010	0	10	10	0

Note: The state proxy flag is equal to 1 if there are not enough observations (n>10) in the climate divisions to run a credible fractional regression model and calculate a weather index.

Table 4.8. Hypothetical example of unweighted and weather weighted loss costs at the county-level for Boone County (county=15), Dallas County (county=49), and Grundy County (county=75), IA (State=19).

State	Climate Division	County	County Average loss costs	Flag =1 if insignificant	Weighting Type
19	5	15	0.0096378	0	1
19	5	15	0.0076921	0	2
19	5	15	0.0028386	0	3
19	5	15	0.0027737	0	4
19	5	15	0.0035587	0	5
19	5	15	0.0033862	0	6
19	5	49	0.0100697	0	1
19	5	49	0.0097928	0	2
19	5	49	0.0058953	0	3
19	5	49	0.0058029	0	4
19	5	49	0.007514	0	5
19	5	49	0.0075715	0	6
19	5	75	0.0091694	0	1
19	5	75	0.0051299	0	2
19	5	75	0.001323	0	3
19	5	75	0.0010593	0	4
19	5	75	0.0044935	0	5
19	5	75	0.0032308	0	6

Note: Weighting type = 1 if the average loss cost is calculated with no weather weighting and no censoring; Weighting type =2 if the average loss cost is calculated with weather weighting but no censoring; Weighting type = 3 if the average loss cost is calculated with censoring at the 80th percentile and no weather weighting; Weighting type = 4 if the average loss cost is calculated with censoring at the 80th percentile and with weather weighting; Weighting type = 5 if the average loss cost is calculated with censoring at the 90th percentile and no weather weighting; Weighting type = 6 if the average loss cost is calculated with censoring at the 90th percentile and with weather weighting.

Table 4.9. Liability weighted national average (across counties) of unweighted and weather weighted average loss costs for apples, barley, corn, cotton, potatoes, rice, sorghum, soybeans, spring wheat and winter wheat.

Crop	No. of Counties	Unweighted loss costs (no censoring)	Weather weighted loss costs (no censoring)	Unweighted loss costs (censoring at 80th)	Weather weighted loss costs (censoring at 80th)	Unweighted loss costs (censoring at 90th)	Weather weighted loss costs (censoring at 90th)
apples	140	0.1839529	0.1756118	0.1509251	0.1458255	0.1722479	0.1649113
barley	646	0.1033683	0.0952631	0.071994	0.0677116	0.088203	0.0820236
Corn	1930	0.0505333	0.0525652	0.028726	0.0293841	0.0394102	0.0409063
cotton	437	0.143511	0.1459077	0.1103868	0.1110684	0.1292813	0.1305584
potatoes	128	0.083174	0.0807186	0.0659818	0.0646853	0.0752233	0.0730846
Rice	84	0.0263574	0.0251909	0.015527	0.0148564	0.0203618	0.0193536
sorghum	750	0.1208383	0.1317581	0.0887164	0.09226	0.1079448	0.1140653
soybeans	1523	0.0542112	0.0538458	0.0384229	0.0379807	0.0467105	0.0460899
spring wheat	244	0.1218715	0.1171909	0.0887732	0.0872793	0.1094074	0.1063092
winter wheat	951	0.0982152	0.0852073	0.0719574	0.065563	0.0851164	0.0759965

Note: These are the liability-weighted, national average loss costs across all counties (i.e., liability weighted average)

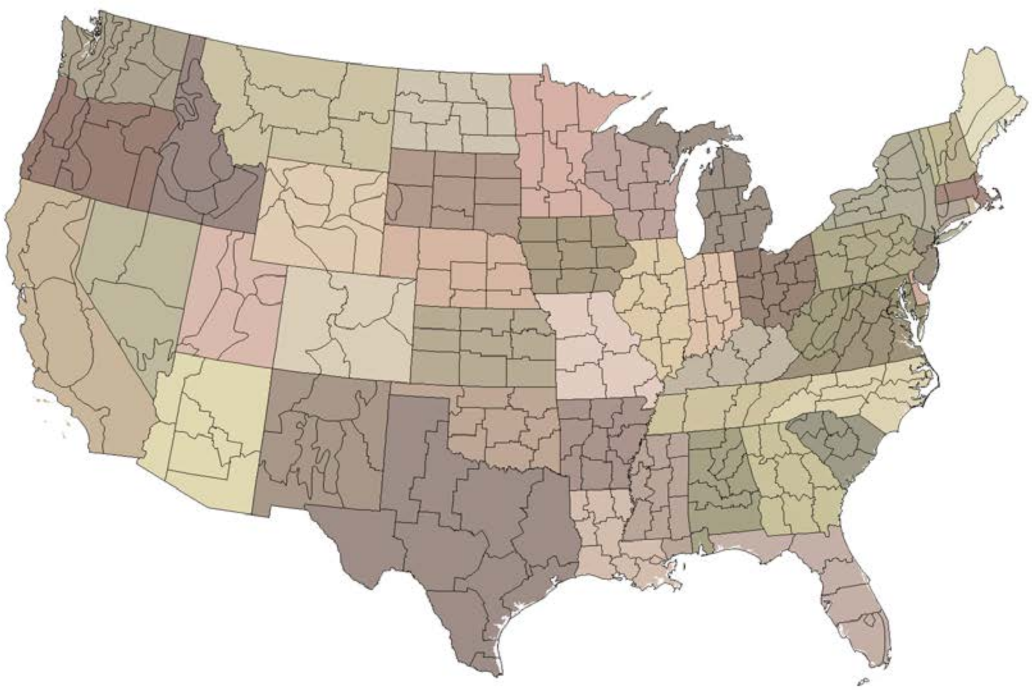


Figure 4.1a. Map of U.S. climate divisions.
(Established by the National Climate Data Center of NOAA)

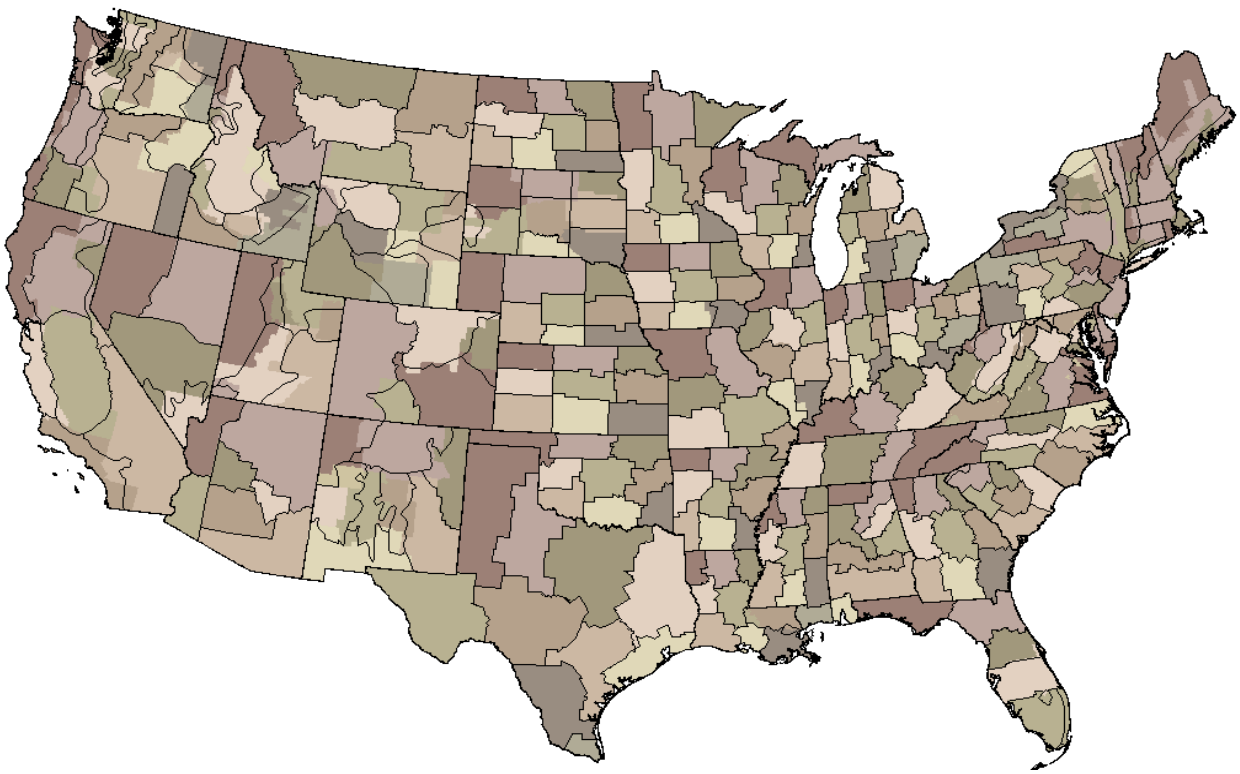


Figure 4.1b. County assignment to climate divisions delineated within states.

b. Proposed Non-Stationarity Loss Cost Adjustments

Based on the discussion of non-stationarity of loss costs in section three, we conducted several empirical analyses to quantify these effects where possible. Ultimately, a combination of regression analysis and a credibility weighting scheme was selected. Based on regression analysis, the most consistent and robust difference in loss costs is represented by a fixed effect for the pre-1995 period. We also recommend an approach that gives more weight to observations with more insured acres.

Regression Adjustment for the Change in Participation

Because we are attempting to measure broad effects such as technological change or significant program changes, our analysis is conducted by aggregating crop/climate division data to the crop/state level. Also note that because of the issues discussed in our analysis of weather effects we include the weather variables aggregated to the state level. This allows us to evaluate program non-stationarity while controlling for unique weather events that may drive the observed loss costs in the 1990-2009 data. The model applied is of the form:

$$\text{adj\_yr\_lcr} = f(\text{pre-1995, weather variables}).$$

We suggest estimation with the fractional logit procedure used for developing the weather index. The Pre-1995 variable takes a value of 1 if crop year is prior to 1995. This variable is posited to capture differences in expected loss costs before and after the fundamental program changes that took place in 1995. No state/crop result is reported if there were not at least 15 years of loss cost data for the state. We also required at least 20,000 acres insured in the year except for apples where the limit was lowered to 5000 acres. Our previous analysis evaluated state, regional and national level estimates of the pre-1995 effect. In our opinion there is a trade-off between allowing regional variation of the pre-1995 effect and robustness of national estimates. Therefore we do not make a specific recommendation as to what level of aggregation should be used.

Weighting Approaches

In addition to the estimated pre-95 effect, we recognize program experience based on more data is more credible. Thus, we recommended net acre weighting. In this approach, we use the county level net acres insured as weights to account for credibility and recency. The county level net acre variable is a “proxy” for recency weights given that it has generally been increasing over time (from 1980 till the 2000s) (See Figure 3.3).

Summary

For the major commodity crops (e.g., corn, soybeans, spring wheat, and cotton), accounting for recency using all the weighting approaches above generally reduces the average loss costs compared to when recency is not accounted for. But for other crops like apples and sorghum, recency weighting generally increases the liability weighted average loss cost at the national level.

Liability average loss costs for corn, soybeans, spring wheat and cotton tend to be lowest under the linear weighting or net acre weighting procedure. It should be noted, however, that the resulting liability-weighted average loss costs for all the recency weighting schemes described above are similar to the approach of simply using 20 years of the most recent data in the calculation (See Table 4.11). Hence, there is appeal to using the "most recent 20 years" approach to account for recency because of its simplicity in implementation compared to the other approaches described above.

Using Shortened Loss Cost Series for Base Rates

Table 4.11 reports the national level averages from an analysis that reflects another alternative means to address non-stationarity in program loss cost expectations. These results largely follow the same approach as reported in Table 4.9. However, in this analysis data from earlier years are omitted from the base rate calculation if it more than 20 years old. The approach assumes that a longer time series would be used to quantify the catastrophic load. The results in Table 4.11 are derived by conducting the weather weighting procedure described earlier, but then any data older than 20 years is dropped from the binning step of the process.

The table reflects three scenarios relative to the catastrophic load. First, we present the no censoring scenarios which use the full loss cost record and ignore catastrophic loading, then we report estimates assuming the loss cost is censored at the 80th or the 90th percentile. In each scenario we report both weather –weighted and unweighted results. All results are report as a percentage of the Table 4.9 results. Values greater than 100% indicate that the shorter series would increase rates relative to using 30 years of data, while values of less than 100% indicate that current rates would be lowered by shortening the base rate series. Within a crop the results are largely consistent across censoring scenarios. For example, all values for apples are above 100% while all values for barley are less than 100%.

The results for apples, cotton and winter wheat indicate that limiting loss cost histories to 20 years would lead to substantially higher rates. Conversely, barley, corn, soybeans and spring wheat all are observed to have substantially lower rates. Note that significant variation is observed within a crop.

Table 4.11. Aggregate implications of shortening loss cost history to twenty years.

20 year loss cost as a percentage of 30 year loss cost							
Crop	No. of Counties	Unweighted loss costs (no censoring)	Weather weighted loss costs (no censoring)	Unweighted loss costs (censoring at 80th)	Weather weighted loss costs (censoring at 80th)	Unweighted loss costs (censoring at 90th)	Weather weighted loss costs (censoring at 90th)
Apples	138	106%	106%	107%	107%	107%	107%
Barley	629	80%	85%	89%	92%	84%	88%
Corn	1914	82%	88%	88%	90%	86%	89%
Cotton	431	106%	97%	109%	103%	109%	101%
Potatoes	127	97%	98%	100%	100%	99%	100%
Rice	83	82%	90%	97%	98%	92%	96%
Sorghum	727	101%	102%	101%	102%	102%	104%
Soybeans	1512	84%	87%	84%	87%	84%	87%
Spring wheat	242	86%	96%	90%	95%	87%	94%
Winter wheat	937	104%	105%	109%	107%	108%	107%

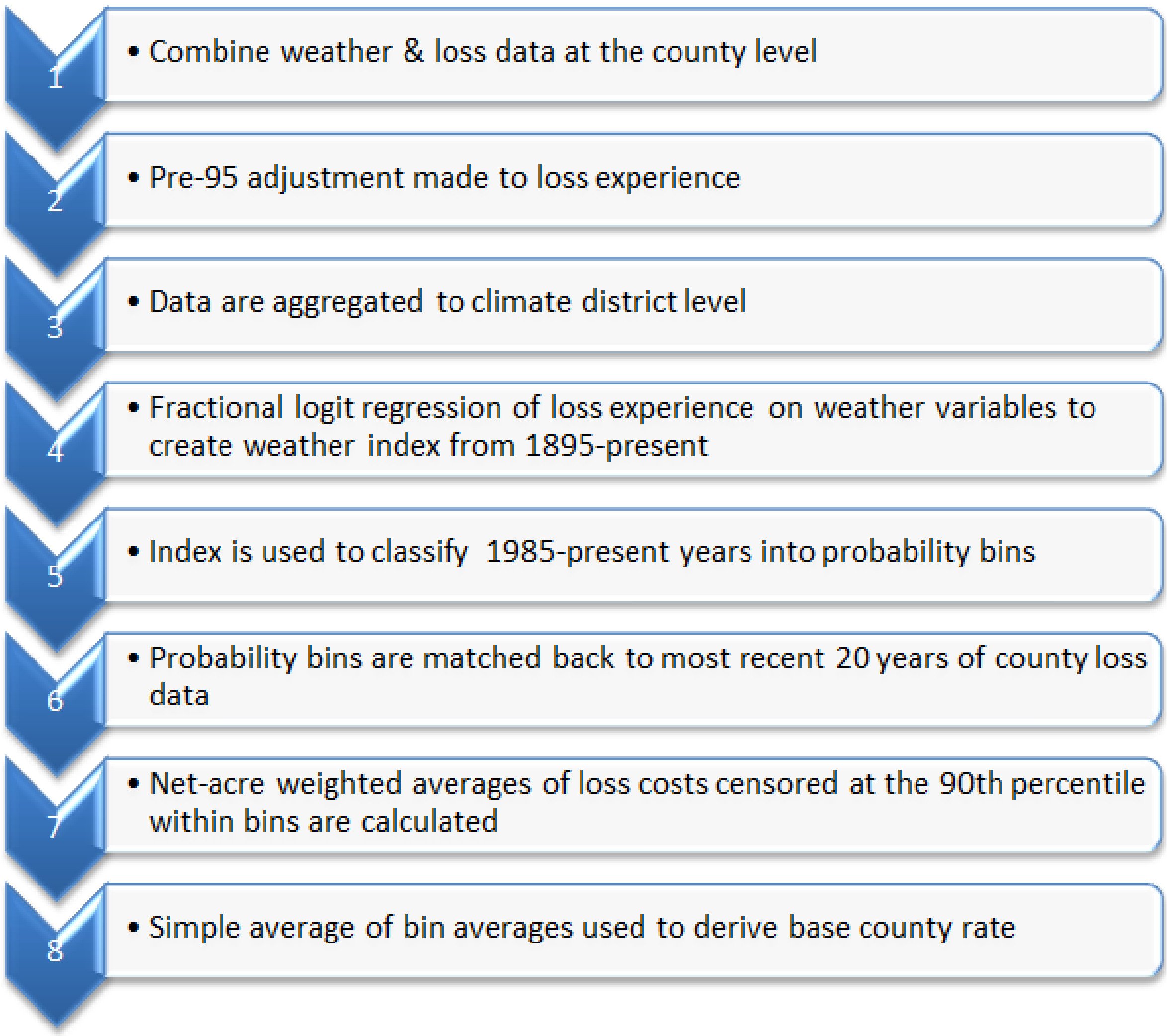
c. Illustration of Combined Base Rating Procedures

The proposed procedure involves four primary changes to the current simple averaging of historical loss cost:

1. A pre-1995 adjustment,
2. Weather weighting,
3. Net acre weighting within probability bins, and
4. The use of a 20 year moving average of loss data.

To illustrate how these adjustments will be integrated with each other and into the RMA base rating system we provide a detailed illustration for corn in climate division 5 of Illinois. First a flow chart is provided in Figure 4.3 which provides an overview of the eight step process.

Figure 4.3 Flow chart of the loss experience reweighting system for base rating.



Having given a general overview, a more detail explanation is provided next. We illustrate the process of calculating the expected loss cost when the following adjustments are accounted for: pre95 adjustment, weather weighting, net acre weighting, and use of 20 years of loss data. The example here utilizes data from the state of Illinois (17) and uses all the counties within climate division 5 (eastern climate division). Climate division 5 in Illinois is composed of the following counties: Champaign (19), Ford (53), Iroquois (75), Kankakee (91), Livingston (105), Piatt (147), and Vermillion (183).

The first step in the reweighting process is to combine the weather and loss data at the county level. In this case, all counties within a climate division have the same weather data. After this step, the pre-95 adjustment factors are applied to all data prior to 1995. Table 4.12 shows an example of how average loss costs are different with and without the pre95 adjustment factors applied to county level data (for corn).

After applying the pre95 adjustments, the data are aggregated up to the climate division level. These climate division data are then used in the fractional regression model to determine the "optimal" combinations of weather variables used in each state and to create predicted values that serve as the "weather index" to classify each year (from 1895- present) based on the long-term weather behavior in the climate division. An example of the weather indexes is seen in Table 4.13 for Illinois climate division 5.

The weather indexes at the climate division level are then used to classify each year into bins (i.e. years within the bins are years with similar weather). In this binning process, the number of bins is determined by looking at the 20-year period from 1991-2010 and making sure that there are no "empty bins" (i.e. each bin category is represented in this 20-year period). Fifteen (15) years is the initial number of bins investigated and if not all fifteen bins are represented in the latest 20-year period then 14 bins are examined (and so on). This process is continued until we find the largest number of bins where all bin categories are represented in the 20-year period (i.e. this process goes from 15 bins to as few as 5 bins). An example of the result of this binning process can be seen in Table 4.14 for Illinois climate division 5. In this case, there are 11 bins and all bins (from 1 to 11) are represented in the 1991-2010 period. Note that some bins have more years in them than others. As mentioned in p. 28-29 above, a county-level binning process is also implemented if there are counties with less than 20 years of data since these missing observations could also result in empty bins. But in this example, all counties have the full 20 years of actual data and, therefore, the climate division level bin classification is the only one used here. In addition, the "optimal" number of bins is 11 and so the minimum threshold of five bins is not reached.

The next step is merging the binning results (at the climate division level) to the county level actual loss cost data censored at the 90th percentile (See Table 4.15). This is the data set used for calculating the averages within bins and the overall average (average of the average within bins).

The average of the "average within bins" is the final result we are interested in (i.e., the final base premium estimate where all the adjustments in the first paragraph above are applied).

In Table 4.16, we show the average within bins for two counties in climate division 5 of Illinois. Note that the averages within bins are "net acre weighted" averages. This net acre weighting generally gives more weight to more recent years within bins (since more recent years typically have more insured acres).

After calculating the net-acre-weighted averages within bins, the average across all the bins (11 bins in Table 4.16) is calculated. This is the base premium rate. The averages across bins for counties in Illinois climate division 5 is reported in Table 4.17.

Table 4.12. Average loss cost ratios (LCR) for counties in Illinois climate division 5 (Corn; 1980-2010): With and without pre95 fixed adjustment.

County	Average LCRs without pre95 adjustment	Average LCRs with pre95 adjustment
Champaign (19)	0.0193	0.0124
Ford (53)	0.0374	0.0229
Iroquois (75)	0.0299	0.0191
Kankakee (91)	0.0307	0.0197
Livingston (105)	0.0355	0.0222
Piatt (147)	0.0132	0.0085
Vermillion (183)	0.0275	0.0176

Table 4.13. Predicted LCRs from the fractional logit regression for Illinois climate division 5 Corn; 1980-2010).

Crop Year	Net Acres	Actual Loss cost ratio	Weather index
1980	46666	0.068	0.023
1981	99413	0.001	0.002
1982	72679	0.001	0.001
1983	35962	0.092	0.045
1984	147325	0.015	0.006
1985	171342	0.001	0.000
1986	196450	0.005	0.008
1987	189592	0.003	0.008
1988	203502	0.156	0.145
1989	890345	0.001	0.001
1990	667228	0.001	0.001
1991	590842	0.114	0.128
1992	865613	0.001	0.000
1993	798127	0.005	0.004
1994	858636	0.000	0.002
1995	888169	0.034	0.033
1996	834373	0.004	0.001
1997	838033	0.004	0.001
1998	832875	0.003	0.015
1999	890781	0.001	0.009
2000	964957	0.000	0.001
2001	946922	0.002	0.001
2002	970740	0.007	0.019
2003	1015135	0.001	0.001
2004	1049753	0.001	0.001
2005	1068085	0.004	0.025
2006	970250	0.004	0.004
2007	1214946	0.000	0.011
2008	1246866	0.001	0.002
2009	1357670	0.001	0.001
2010	1320642	0.007	0.040

Note: The predicted LCRs above serve as the “weather index” that allows for classification of all years based on the long-term weather pattern. Predicted values are available from 1895 to the present but the values reported above are only from 1980-2009.

Table 4.14. Bin Classification for Illinois climate division 5 (Corn; 1991-2010).

Crop Year	Weather index	Bin Classification	Number of Bins in Climate Division
1991	0.128	11	11
1992	0.000	1	11
1993	0.004	6	11
1994	0.002	4	11
1995	0.033	10	11
1996	0.001	2	11
1997	0.001	1	11
1998	0.015	9	11
1999	0.009	7	11
2000	0.001	3	11
2001	0.001	3	11
2002	0.019	9	11
2003	0.001	2	11
2004	0.001	2	11
2005	0.025	10	11
2006	0.004	5	11
2007	0.011	8	11
2008	0.002	3	11
2009	0.001	2	11
2010	0.040	10	11

Table 4.15. Bin classification data at the climate division 5 level merged with the county level actual loss cost data for Champaign (19), Illinois (Corn, 1991-2010).

Crop Year	Net Acres	County Actual Loss cost ratio	Climate division Weather index	Bin Classification	Number of Bins in Climate Division
1991	61179	0.086	0.128	11	11
1992	96491	0.000	0.000	1	11
1993	94245	0.001	0.004	6	11
1994	99534	0.000	0.002	4	11
1995	114330	0.049	0.033	10	11
1996	116142	0.000	0.001	2	11
1997	121611	0.001	0.001	1	11
1998	123625	0.002	0.015	9	11
1999	148622	0.000	0.009	7	11
2000	154932	0.000	0.001	3	11
2001	160025	0.001	0.001	3	11
2002	171460	0.005	0.019	9	11
2003	168850	0.000	0.001	2	11
2004	171604	0.001	0.001	2	11
2005	169562	0.000	0.025	10	11
2006	161666	0.001	0.004	5	11
2007	195712	0.000	0.011	8	11
2008	195980	0.001	0.002	3	11
2009	225837	0.000	0.001	2	11
2010	211735	0.003	0.040	10	11

Table 4.16. Net Acre weighted Average loss costs within bins for Champaign and Ford counties, Illinois (Corn, 1991-2010).

County	Bin Classification	No. of years averaged	Average loss costs within bins
Champaign (19)	1	2	0.00040
Champaign (19)	2	4	0.00032
Champaign (19)	3	3	0.00053
Champaign (19)	4	1	0.00009
Champaign (19)	5	1	0.00115
Champaign (19)	6	1	0.00089
Champaign (19)	7	1	0.00028
Champaign (19)	8	1	0.00004
Champaign (19)	9	2	0.00363
Champaign (19)	10	3	0.00741
Champaign (19)	11	1	0.02689
Ford (53)	1	2	0.00076
Ford (53)	2	4	0.00062
Ford (53)	3	3	0.00040
Ford (53)	4	1	0.00012
Ford (53)	5	1	0.00260
Ford (53)	6	1	0.00191
Ford (53)	7	1	0.00030
Ford (53)	8	1	0.00037
Ford (53)	9	2	0.00129
Ford (53)	10	3	0.00661
Ford (53)	11	1	0.01409

Table 4.17. Average loss costs across the “average within bins” for Illinois counties in climate division 5 (Corn, 1991-2010).

County	Average loss costs (across bins)
Champaign (19)	0.00378
Ford (53)	0.00264
Iroquois (75)	0.00700
Kankakee (91)	0.00612
Livingston (105)	0.00392
Piatt (147)	0.00183
Vermillion (183)	0.00491

Note: The average loss costs above are the base premium rate estimates for each county where the following adjustments/procedures are applied: pre95 adjustment, weather weighting censoring at the 90th percentile, net acre weighting, and use of 20 years of loss data.

d. Illustration of Proposed Catastrophic Loading Procedures

The first step in the recommended cat loading procedure is the same as in the current process: determine which years in the loss history are above the 90th percentile. (We note that the current procedure caps losses at the 80th percentile. Also note the change in catastrophic loading region from state to climate division.) We assume that the effect of changes in the program reflected in the pre-1995 adjustment apply uniformly to all losses; hence the catastrophe capping calculation is applied after the pre-1995 adjustment to the loss cost data. Table 4.18 continues our example by showing the determination of the 90th percentile for Champaign County, Illinois, with data from 1980 through 2010.

The next step is to determine whether any of the excess losses should be adjusted for highly unusual weather. We now rank each loss cost year's weather index relative to full set of weather indexes and determine if any year in our data had weather above the 97th percentile. Figure 4.5 illustrates this procedure. There are 116 years from 1895 through 2010, so the 4 years with the highest weather indexes make up the years above the 97th percentile. For corn in Illinois climate district 5, the 4 years with the highest indexes were 1936 (.5006), 1933 (.4408), 1988 (.1449) and 1991 (.1284). The long term probability of weather as bad as that observed in 1988 is estimated at about 3/116 = 2.6%, and we would expect to see weather that bad only once in 39 years, not once in 31 years (the number of years in our data). We would expect to see weather as bad as 1991 about once in 29 years, so its appearance once in the 31 year history is about as expected.

We now adjust the excess indemnity for 1988 to reflect the lower long-term probability of the weather for that year. Table 4.19 shows this calculation. Because we have 31 years of data (1980 through 2010), a simple average of the excess losses would give each year 1/31 (3.2%) weight, but we want to limit the weight given to 1988 to 2.6%. We therefore adjust the excess losses for 1988 by 2.6%/3.2% = .802 before adding them into the county's contribution to the excess load.

The final step in the process combines the adjusted excess losses for all counties in the climate division to determine the cat load for the climate division. The cat load is the sum of the adjusted excess losses divided by the sum of the exposures (liability). An illustration is provided in Table 4.20.

Figure 4.4 Flow chart of the proposed catastrophic loading procedure

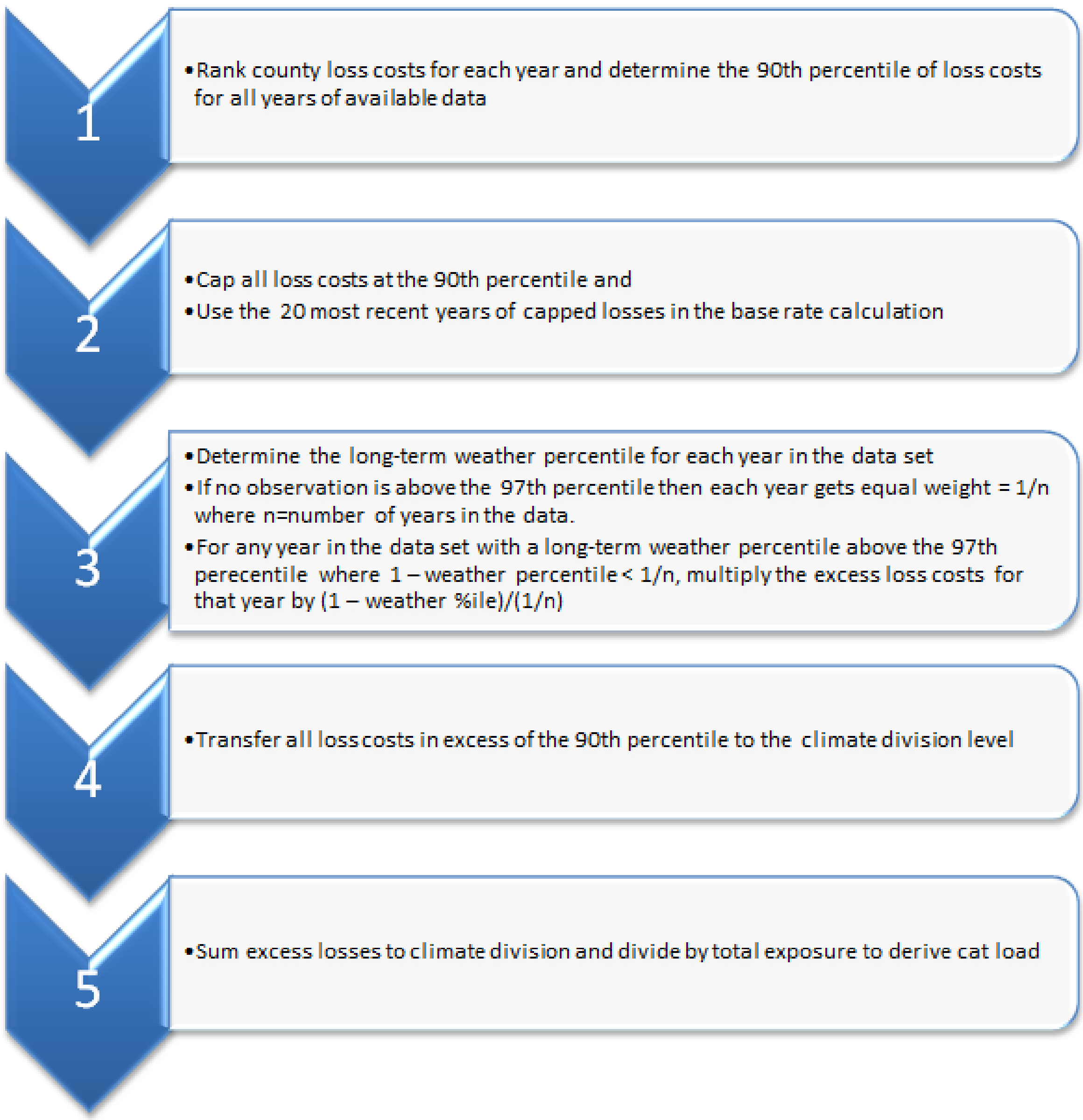


Table 4.18. Determination of excess loss costs after pre-1995 adjustment, actual loss cost data for Champaign (19), Illinois (Corn, 1980-2010).

	Loss	**pre-1995**	**After**	**Capped**	**Excess**	
Year	Cost	Adjustment	Pre95 Adj	Loss Cost	Loss Cost	Liability
1980	0.08319	0.599730	0.04989	0.04989	0.00000	1,198,550
1981	0.00120	0.599730	0.00072	0.00072	0.00000	2,924,223
1982	0.00239	0.599730	0.00143	0.00143	0.00000	2,171,021
1983	0.15678	0.599730	0.09403	0.04989	0.04414	1,215,445
1984	0.01126	0.599730	0.00675	0.00675	0.00000	5,089,301
1985	0.00070	0.599730	0.00042	0.00042	0.00000	5,496,762
1986	0.00136	0.599730	0.00081	0.00081	0.00000	4,864,574
1987	0.00039	0.599730	0.00023	0.00023	0.00000	4,464,462
1988	0.13052	0.599730	0.07827	0.04989	0.02839	4,791,357
1989	0.00073	0.599730	0.00044	0.00044	0.00000	19,197,580
1990	0.00121	0.599730	0.00073	0.00073	0.00000	13,592,310
1991	0.14373	0.599730	0.08620	0.04989	0.03631	12,083,506
1992	0.00026	0.599730	0.00015	0.00015	0.00000	18,539,735
1993	0.00148	0.599730	0.00089	0.00089	0.00000	18,596,087
1994	0.00015	0.599730	0.00009	0.00009	0.00000	21,004,237
1995	0.04905	1.000000	0.04905	0.04905	0.00000	23,880,386
1996	0.00046	1.000000	0.00046	0.00046	0.00000	27,971,377
1997	0.00060	1.000000	0.00060	0.00060	0.00000	26,916,018
1998	0.00210	1.000000	0.00210	0.00210	0.00000	29,561,993
1999	0.00028	1.000000	0.00028	0.00028	0.00000	28,968,497
2000	0.00000	1.000000	0.00000	0.00000	0.00000	28,362,667
2001	0.00077	1.000000	0.00077	0.00077	0.00000	31,574,979
2002	0.00473	1.000000	0.00473	0.00473	0.00000	33,632,325
2003	0.00009	1.000000	0.00009	0.00009	0.00000	35,879,189
2004	0.00052	1.000000	0.00052	0.00052	0.00000	41,574,931
2005	0.00021	1.000000	0.00021	0.00021	0.00000	37,705,030
2006	0.00115	1.000000	0.00115	0.00115	0.00000	33,645,797
2007	0.00004	1.000000	0.00004	0.00004	0.00000	72,988,367
2008	0.00075	1.000000	0.00075	0.00075	0.00000	102,526,794
2009	0.00026	1.000000	0.00026	0.00026	0.00000	102,148,611
2010	0.00265	1.000000	0.00265	0.00265	0.00000	94,752,540
		90th %ile	0.04989			

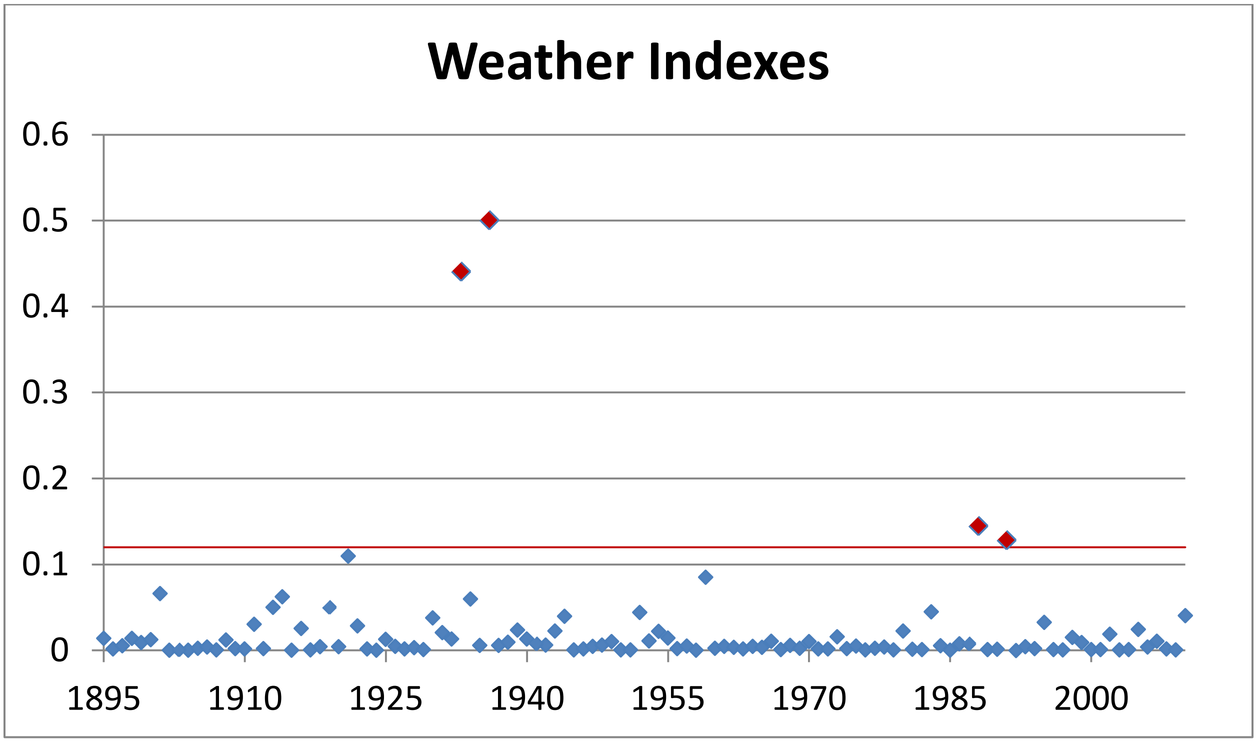


Figure 4.5. Weather indexes for Illinois Corn, Climate Division 5 (1895-2010).

Table 4.19. Weather adjusted excess losses, actual loss cost data for Champaign (19), Illinois (Corn, 1980-2010).

	Excess		**Excess**	**Weather**		**Adjusted**
Year	Loss Cost	Liability	Losses	Probability	Adjustment	Excess Losses
1980	0.00000	1,198,550	-	19.8%	1.000	-
1981	0.00000	2,924,223	-	76.7%	1.000	-
1982	0.00000	2,171,021	-	78.4%	1.000	-
1983	0.04414	1,215,445	53,647	10.3%	1.000	53,647
1984	0.00000	5,089,301	-	46.6%	1.000	-
1985	0.00000	5,496,762	-	98.3%	1.000	-
1986	0.00000	4,864,574	-	41.4%	1.000	-
1987	0.00000	4,464,462	-	40.5%	1.000	-
1988	0.02839	4,791,357	136,007	2.6%	0.802	109,040
1989	0.00000	19,197,580	-	81.9%	1.000	-
1990	0.00000	13,592,310	-	77.6%	1.000	-
1991	0.03631	12,083,506	438,766	3.4%	1.000	438,766
1992	0.00000	18,539,735	-	100.0%	1.000	-
1993	0.00000	18,596,087	-	55.2%	1.000	-
1994	0.00000	21,004,237	-	68.1%	1.000	-
1995	0.00000	23,880,386	-	14.7%	1.000	-
1996	0.00000	27,971,377	-	82.8%	1.000	-
1997	0.00000	26,916,018	-	93.1%	1.000	-
1998	0.00000	29,561,993	-	25.0%	1.000	-
1999	0.00000	28,968,497	-	37.9%	1.000	-
2000	0.00000	28,362,667	-	79.3%	1.000	-
2001	0.00000	31,574,979	-	81.0%	1.000	-
2002	0.00000	33,632,325	-	23.3%	1.000	-
2003	0.00000	35,879,189	-	89.7%	1.000	-
2004	0.00000	41,574,931	-	84.5%	1.000	-
2005	0.00000	37,705,030	-	18.1%	1.000	-
2006	0.00000	33,645,797	-	57.8%	1.000	-
2007	0.00000	72,988,367	-	33.6%	1.000	-
2008	0.00000	102,526,794	-	75.9%	1.000	-
2009	0.00000	102,148,611	-	91.4%	1.000	-
2010	0.00000	94,752,540	-	12.1%	1.000	-
Sum			628,420			601,453

Table 4.20. Determination of climate division cat load, Illinois Corn Climate Division 5.

County	Excess Losses	Liability
Champaign	601,453	887,318,651
Ford	274,103	541,515,564
Iroquois	1,638,295	1,172,295,057
Kankakee	1,072,824	666,762,452
Livingston	4,373,358	1,240,403,411
Piatt	243,080	436,890,436
Vermilion	1,248,597	640,813,173
Climate Division	9,451,710	5,585,998,744
Division Cat Load		0.001692

5. Implications of the Proposed Methods

As directed by the statement of work for this project, Sumaria has performed a detailed investigation of the proposed methodology for weighting, and otherwise adjusting, RMA's historical loss cost data in order to maximize its statistical validity for developing premium rates. Our team has also provided analysis of the implications of the proposed approach. This section of the report summarizes the effect the proposed approach will have on RMA rates. Because corn and soybeans are a priority for implementation our results analysis focuses on those crops.

Table 5.1 reports national average estimated changes in corn and soybeans base premium rates. These results are liability weighted averages of county level data. They are derived by assuming catastrophic loading will occur at the 90th percentile in the future rather than at the 80th as has been used in the past. The estimated base rate change is calculated by comparing the base rate derived using current procedures versus proposed procedures. Current procedures are modeled using 30 years of adjusted loss cost data and using a simple average of the adjusted loss costs after the catastrophic loading procedure is applied. The proposed procedure includes four modifications of the current base rating procedure:

1. A pre-1995 adjustment,
2. Weather weighting,
3. Net acre weighting within probability bins, and
4. The use of a 20 year moving average of loss data.

The results in Table 5.1 reflect the combined effect of all four modifications. Note that these results do not impose restrictions on the annual magnitude of adjustment and do not include the catastrophic load portion of the rate. Further, these estimated changes impact only the yield portion of a rate and would not alter the price risk portion of a revenue insurance rate.

The national average change in corn base premium rates is 19.1 percent and 25.2 percent for soybeans. However, while the percentage change for soybeans is larger than for corn, the national average soybean base rates are also higher. The table also reports a break out for four states (Illinois, Indiana, Iowa, and Minnesota). For corn, the percentage rate reduction in all four of these states is well above the national average. For soybeans, the rate reduction in Illinois is over 43.6 percent, but in the other three states the rate reduction is on par with the national average.

Further disaggregation of the results can be seen in Figure 5.1 which shows county-by-county comparisons in a map. These results show even greater heterogeneity across locations. In general the greatest percentage rate reduction for corn occurs in major production regions and some outlying irrigated counties. While the national average base rate declines 19 percent, there are regions with substantial rate increases such as portions of western Kansas and portions of New England.

Figure 5.2 reports the county-level results for soybeans. The variation across counties is somewhat less dramatic than for corn. In general, the Corn Belt is observed to have rate reductions which are largest in Illinois. Some other regions have similar reduction such as the Mid-South. Rate increases are suggested in some Western Plains states and portions of the Eastern Seaboard.

Table 5.1 Estimated effects on base rates.

		National Average	Illinois	Indiana	Iowa	Minnesota
Corn	Current Procedure	3.49%	1.66%	2.37%	1.45%	2.33%
	Proposed Modification	2.83%	1.04%	1.60%	1.01%	1.31%
	Percent change	-19.1%	-37.7%	-32.6%	-30.7%	-43.8%
Soybeans	Current Procedure	4.41%	1.82%	2.31%	1.38%	3.13%
	Proposed Modification	3.29%	1.02%	1.77%	1.02%	2.32%
	Percent change	-25.2%	-43.6%	-23.3%	-25.7%	-25.7%

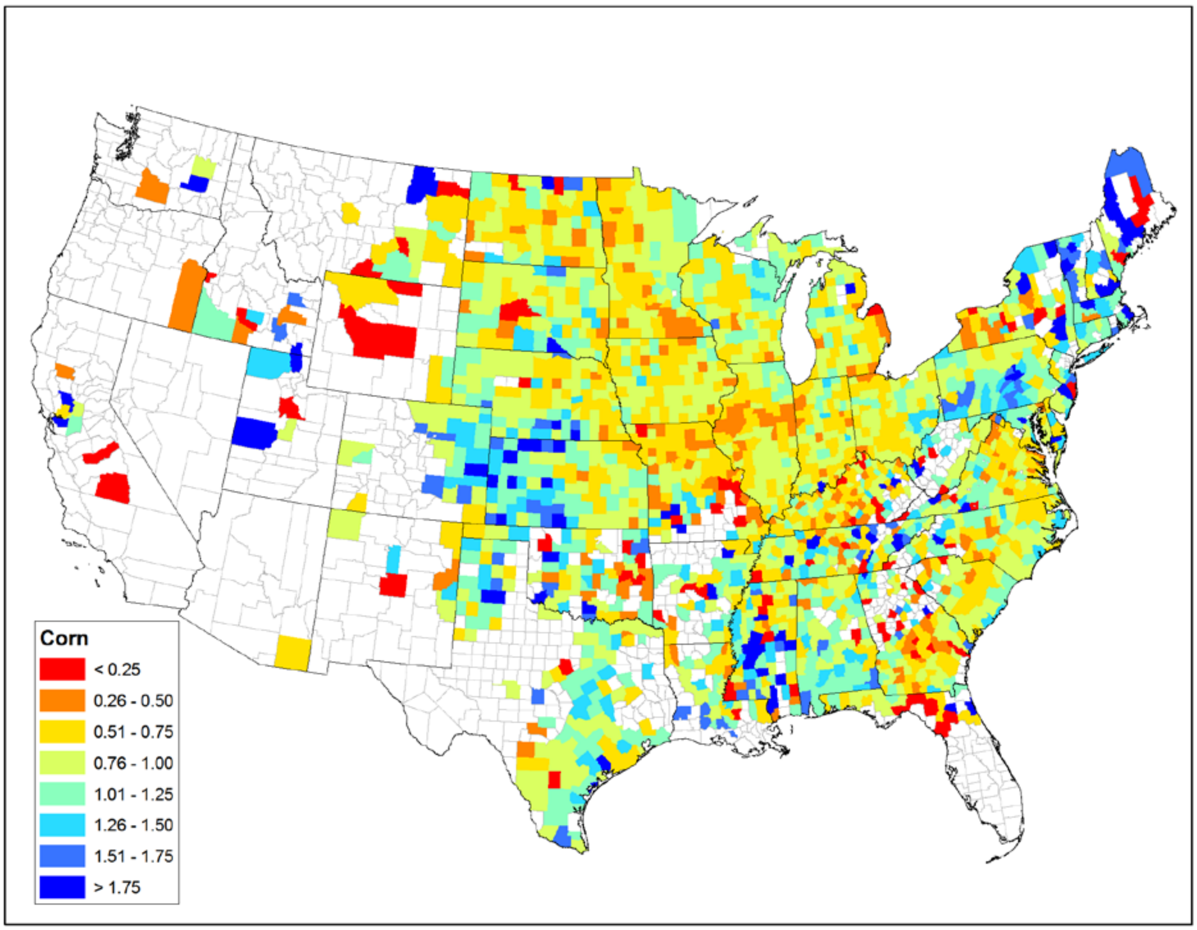


Figure 5.1. County level changes in estimated base rate for corn.

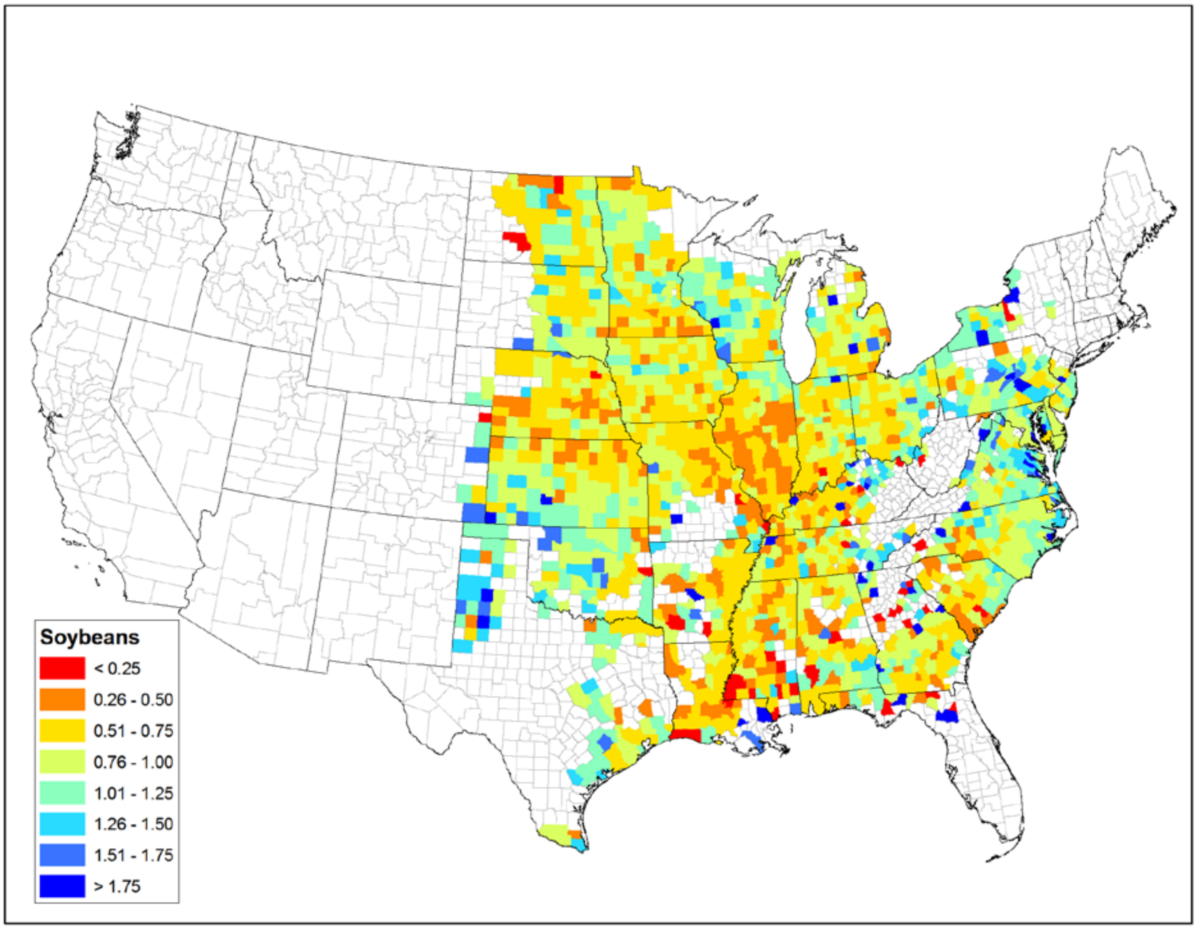


Figure 5.2. County level changes in estimated base rate for soybeans.

References

Actuarial Standards Board. 2009. “Actuarial Standard of Practice No. 13: Trending Procedures in Property/Casualty Insurance—Revised Edition.” Developed by the Subcommittee on Ratemaking of the Casualty Committee of the Actuarial Standards Board.

Casualty Actuary Society. 1988. “Statement of Principles Regarding Property and Casualty Insurance Ratemaking.” Guidelines adopted by the Board of Directors of the Casualty Actuary Society.

Coble, K.H., T.O. Knight, B.K. Goodwin, M.F. Miller and R.M. Rejesus. 2010. “A Comprehensive Review of the RMA APH and COMBO Rating Methodology: Final Report.” Report prepared for the USDA – Risk Management Agency-USDA by Sumaria Systems, Inc.

Copas, J.B. 1983. Regression, prediction and shrinkage. Journal of the Royal Statistical Society, Series B, vol 45, pp 311-354.

Dai, A., K. E. Trenberth, and T. Qian. 2004. A global data set of Palmer Drought Severity Index for 1870-2002: Relationship with soil moisture and effects of surface warming. J. Hydrometeorology, vol. 5, 1117-1130, American Meteorological Society, Boston, MA.

Guttman, N. and R. Quayle. 1996. A historical perspective of U.S. climate divisions. Bulletin of the American Meteorological Society, vol. 77(2), pp. 293-303, American Meteorological Society, Boston, MA.

Josephson, G.R., R.B. Lord, and C.W. Mitchell. 2000. “Actuarial Documentation of Multiple Peril Crop Insurance Ratemaking Procedures.” Report prepared for the USDA – Risk Management Agency by Milliman & Robertson, Inc., Brookfield, WI.

Karl, T.R., C. N. Williams, P. J. Young, and W. Wendland. 1986. A model to estimate the time of observation bias associated with monthly mean maximum, minimum, and mean temperatures for the United States. Journal of Climate and Applied Meteorology, January 1986, vol. 25, pp. 145-160, American Meteorological Society, Boston, MA.

NCDC. 1994. Time Bias Corrected Divisional Temperature-Precipitation-Drought Index. Documentation for dataset TD-9640. Available from DBMB, NCDC, NOAA, Federal Building, 37 Battery Park Ave. Asheville, NC 28801-2733. 12 pp.

Papke, L. E. and J. Wooldridge. 1996. “Econometric methods for fractional response variables with an application to 401(k) plan participation rates.” *Journal of Applied Econometrics* 11: 619–632.

Rencher, A.C. and F.C. Pun, 1980. Inflation of R^2 in Best Subset Regression. Technometrics, vol 22, pp 49-54.

Appendix . Research Team Credentials

Our research team includes: Dr. Keith H. Coble, Dr. Barry K. Goodwin, Dr. Rod Rejesus, Ms. Marry Frances Miller, Dr. Thomas O. Knight, and Dr. Ryan Boyles. Four of these team members (Coble, Goodwin, Rejesus, and Knight) are agricultural economists who have extensive experience in conducting risk management research. All of these individuals have active ongoing research programs focusing on crop insurance and risk management. In addition, Drs. Coble, Goodwin, Rejesus, and Knight have experience working with the RMA and other federal agencies on various research projects, as well as established records of successful project completion and timely performance. Mary Frances Miller has evaluated various crop insurance projects: Experience-based insurance discounts, biotechnology discounts, and aquaculture insurance. She also played a central role in our recent review of the APH and COMBO rating methodologies. Dr. Boyles is an applied climatologist with experience in using and evaluating climate observations and models for applications to crop models and decision systems. He is the North Carolina State Climatologist and is Director of the North Carolina State Climate Office at NC State University.

Dr. Keith H. Coble

Dr. Keith H. Coble is a W.L. Giles Distinguished Professor at Mississippi State University. He earned his Ph.D. in Agricultural Economics at Texas A&M University in 1993. Dr. Coble came to Mississippi State in 1997 after serving as leader of the Crop Risk Management Team at the Economic Research Service of the U.S. Department of Agriculture. Dr. Coble has testified three times before the Congressional Agriculture Committees regarding government risk policy. His research has covered a broad range of risk, agricultural policy, and crop insurance issues, and has been frequently published in scientific research journals. He currently serves as the founding Chair of the Applied Risk Analysis Section of the Agricultural and Applied Economics Association. Dr. Coble has performed numerous analyses for the Federal Crop Insurance Corporation and Risk Management Agency including serving as an underwriting reviewer and technical expert for the Board of Directors of the Federal Crop Insurance Corporation. Among the other issues he has examined for the RMA are: (a) actuarially fair premium rate adjustments for optional versus basic units, (b) crop insurance demand, (c) analysis of rate relativities, (d) reviews of the Dollar Revenue Plans, (e) review of GRP, (f) review of AGR, (g) experienced-based discounts, (h) reference yield updating, and (i) APH and Combo Rate Review.

Dr. Barry K. Goodwin

Dr. Barry Goodwin is William Neal Reynolds Professor in the Departments of Economics and Agricultural and Resource Economics at North Carolina State University. He holds a Ph.D. in economics with a minor in statistics from North Carolina State University. He has written over 100 publications, including two books. He has held faculty positions at Kansas State University, Ohio State University, and North Carolina State University. He has worked on crop insurance issues for the past 18 years. He has authored a book on crop insurance that is widely cited. He has participated in many insurance plan reviews, including a recent comprehensive review of revenue insurance policies. He is a frequent consultant to the Risk Management Agency, and has also worked in consultation on actuarial matters with the insurance industry.

Dr. Roderick M. Rejesus

Dr. Roderick M. Rejesus is Associate Professor in the Department of Agricultural and Resource Economics at North Carolina State University. Prior to this position Dr. Rejesus was an Assistant Professor at Texas Tech University. Dr. Rejesus received his M.S. degree at Clemson University and his Ph.D. degree at the University of Illinois at Urbana-Champaign. He has an active research program that focuses on risk management and crop insurance issues, precision agriculture, and other issues related to agricultural economics. Dr. Rejesus has published research findings on ex-post moral hazard in crop insurance, determinants of anomalous prevented planting claims, the added land and new producer provisions in crop insurance, and many other crop insurance related issues. He has been involved in several RMA funded projects, namely: Unit Division Structure Review, Premium Rate Discount Project, the Reference Yield Update Methodology Project, the Climate Change Impacts on Crop Insurance, and (most recently) the Comprehensive Review of the APH Rating Methodology.

Mary Frances Miller
FCAS, MAAA, FCA, Hon FIA, CPCU, ARe, AIM

Mrs. Miller is a founder and the senior consulting actuary with Select Actuarial Services. With more than 20 years of property and casualty actuarial experience, she provides actuarial consulting services on several major accounts and is additionally responsible for the professional development of the other members. Her expertise is frequently called upon to assist clients in making decisions regarding the maintenance and design of their risk management programs. Prior to the formation of Select Actuarial Services, Mrs. Miller was the Senior Vice President and Chief Actuary for five years at Sedgwick Actuarial Services. As Chief Actuary she performed a wide spectrum of actuarial studies and also managed the actuarial staff. Before joining Sedgwick in 1993, Mrs. Miller was reinsurance actuary with American States Insurance Companies, where her duties included pricing within the Reinsurance Division, as well as the design and development of specialized software targeting property catastrophe exposures, case reserving for automobile and workers' compensation long-term disability claims, and evaluating treaty commutation proposals.

Mrs. Miller graduated with highest honor from the Honors College at Michigan State University with Bachelor of Arts degrees in Mathematics and Linguistics. She is a Fellow of the Casualty Actuarial Society, a member of the American Academy of Actuaries, a Fellow of the Conference of Consulting Actuaries, and a Chartered Property and Casualty Underwriter. She was elected an Honorary Fellow of the Institute of Actuaries (UK) in 2005. She has been an active contributor to the actuarial profession since achieving fellowship in 1988, and has chaired the CAS Professionalism Education Committee, the Education Policy Committee, and task forces on mutual recognition and future education planning. She was Vice-President for Admissions of the Casualty Actuarial Society from 2000 to 2002, President-Elect in 2003, President in 2004, and she chaired the CAS Board in 2005. She has been a member of the Board of Directors of the American Academy of Actuaries and is a current member of the Board of Directors of the Conference of Consulting Actuaries. As a member of the Actuarial Standards Board subcommittee on reserves, Mrs. Miller was a drafter of ASOP #36, the standard for reserve opinions in the United States.

Career Experience

- Founding Member of Select Actuarial Services in 1999
- Senior VP and Chief Actuary, Sedgwick Actuarial Services, 1993-1999
- American States Insurance Company, 1984-1993

Professional Activities

- Vice President of Admissions of the CAS, 1999-present
- CAS Representative, International Actuarial Association Education Committee, 2001-present
- American Academy of Actuaries Committee on Qualifications, 1999-present
- Actuarial Standards Board Subcommittee on Reserving Standard, 1997-1999
- Frequent speaker at PRIMA, RIMS, CAS Loss Reserve Seminar and several state-sponsored captive and self-insured organizations

Education

- Bachelor of Arts Degrees in Mathematics and Linguistics, Michigan State University
- Obtained FCAS designation in 1988, CPCU in 1995, Associate in Reinsurance in 1993, Associate in Management in 1994

Dr. Thomas O. Knight

Thomas O. Knight is Emabeth Thompson Professor of Agricultural Risk Management in the Department of Agricultural and Applied Economics at Texas Tech University. He earned a Ph.D. in Agricultural Economics in 1984. Knight's research program has focused on agricultural risk analysis. Since 1990, the primary focus of Knight's research has been a wide range of issues relating to Federal crop insurance programs. Among the issues he has examined are: (a) crop insurance demand, (b) actuarial effects of moral hazard in crop insurance, (c) premium rate adjustments for optional versus basic units, (d) proper reference yields, and (e) experience-based premium rate discounts. Knight has (a) conducted a cost benefit analysis in support of the RMA's implementation of prevented planting and double insurance provisions of the Agricultural Risk Protection Act of 2000 (ARPA), (b) completed a review of Dollar Revenue Plans of insurance applicable for 13 specialty crops, (c) collaborated in a team that examined the effects of crop insurance programs on cotton acreage changes, (d) collaborated in a team that examined premium rate relativities for alternative coverage levels, (e) and collaborated in a team that conducted a comprehensive review of rating methods used for the APH-based yield and revenue insurance programs. Knight also serves as an underwriting reviewer and technical expert for the Board of Directors of the Federal Crop Insurance Corporation.

Dr. Ryan Boyles

Dr. Boyles is the State Climatologist and Director, State Climate Office of North Carolina and Extension Assistant Professor, Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University.

His educational background includes a M.S. and Ph.D. in Marine, Earth, and Atmospheric Sciences, from NC State University. He has served as the State Climatologist for North Carolina since 2006 and was Associate State Climatologist from 2001-2006. In that role his duties include oversight of climate services, outreach and extension, research, and systems development. He has served as a Member of the North Carolina Legislative Commission on Global Climate Change, has been a member of the American Meteorological Society since 1995 and serves on the AMS Committee on Applied Climatology and Committee on Climate Services since 2009. Dr. Boyles has also served on the NC Drought Management Advisory Council for over 10 years and has received awards for his work on drought analysis and monitoring.

Dr. Boyles has a proven track record as a practicing climatologist who also conducts rigorous research. He has 11 refereed journal articles relating to climatology or meteorology. Several of these articles have direct application to agricultural applications. His work has appeared in *Applied Engineering for Agriculture, Bulletin of the American Meteorological Society, Pure and Applied Geophysics*, and *Natural Hazards*. He is quite familiar with the collection and uses of historical weather data, the limitations of such data and models derived from these data.