Methodology Analysis for

Weighting of Historical Experience

Revised Technical Report

July 12, 2011

By

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

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) provided a comprehensive review of the methodology and procedures used to determine APH target rates and COMBO rating 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 project commenced in September 2010. This report evaluates alternative methodologies for weighting and adjusting historical experience used to develop rates for the APH product.

The statement of work for this project directs Sumaria to perform a detailed investigation and 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.

We were directed to consider the Palmer Drought Index, other weather variables, changing severity of loss costs over time, and program participation changes over time. The statement of work also directed us to deliver a report that offers multiple approaches that compare and contrast the varying combinations of the factors based on statistical validity, feasibility, sustainability, and a balance of improvement versus complexity. That is the purpose of this report. It is expected that the Government will select the approach that best fits its needs and that is optimal for the program. We will also provide an implementation plan and models that RMA can incorporate into its rating methodology.

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 Appendix C 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

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

Corrected Divisional Temperature-Precipitation-Drought Index data, also call the Climate Division Data. The climate division data provide several drought indexes and other weather variables time-aggregated to the monthly level and spatially-aggregated to the climate division level for the years back to 1895. Thus, it allows RMA to compare the weather experience incurred by the modern program to weather extending 80 years past the 1975 cut-off of loss-cost data.

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 ranks 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 we add:

Recommendation 5. – A variety of factors suggests non-stationarity in some RMA loss cost data. Primary factors we perceive in RMA data are 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 effect. 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 state or more aggregate 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 a particular location. We examined several approaches including:

1. A discrete adjustment to data prior to 1995
2. A discrete adjustment to data prior to 1995 plus a trend adjustment since 1995
3. Adjusting loss cost based as a function of net acres insured
4. Shortening the loss history for base rates (not catastrophic loads) to twenty years
5. Decadal weights comparing median loss cost bins
6. A linear recency effect
7. Net acre weights within probability bins

All of these approaches have instances where they appear to perform well. The first three procedures require model estimation while the fourth is a procedure that would only slightly alter current RMA practices. We believe all could be made compatible with other RMA procedures and with weather weighting. However, we stress that where statistical analysis indicates non-stationarity in the loss cost history, making no adjustment results in a rate that is not actuarially sound. Ultimately we recommend a combination of option 1, 4 and 7. The discrete adjustment for data prior to 1995 would be applied to the adjusted loss cost data first. Specifically we would estimate the effect at a national level and calculate a percentage difference by state using the effect relative to the post-1995 average loss cost. Shortening the loss history for base rates to 20 years while using more years for catastrophic loading reflects the recognition that a longer time series is needed to capture extreme events than for the risk quantified in the base rate. Finally, using net acre weighting within probability categories “bins’ recognizes the additional credibility of experience that is based on more exposed acres.

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 USDA. Private firms sell the product 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.

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 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 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 under 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 of 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 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 should 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 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 report evaluates alternative methodologies for weighting historical experience used to develop rates for the APH product. The statement of work for this project directs Sumaria to perform a detailed investigation and 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. We were directed to consider the Palmer Drought Index and other weather variables to control for both

[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 .

good and bad growing conditions. 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.

The statement of work also directed us to deliver a report that offers multiple approaches that compare and contrast the varying combinations of the factors based on statistical validity, feasibility, sustainability, and a balance of improvement versus complexity. That is the purpose of this report. It is expected that the Government will select the approach that best fits its needs and that is optimal for the program. We will then produce a second report presenting an implementation plan or model that RMA can incorporate into its current methodology.

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 Appendix C of this report.

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 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 subsequent sections, and the county summary table contains information summarized at the county level and 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 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 that are 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 section 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 down 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 RMA believes 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 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 Assessment of Potential 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 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 US 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 only at the national level. Crop insurance covers a large portion of the US 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 available in the future to support updating of weather weights as more data become available.

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 and 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 to this approach 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 the 80th 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).

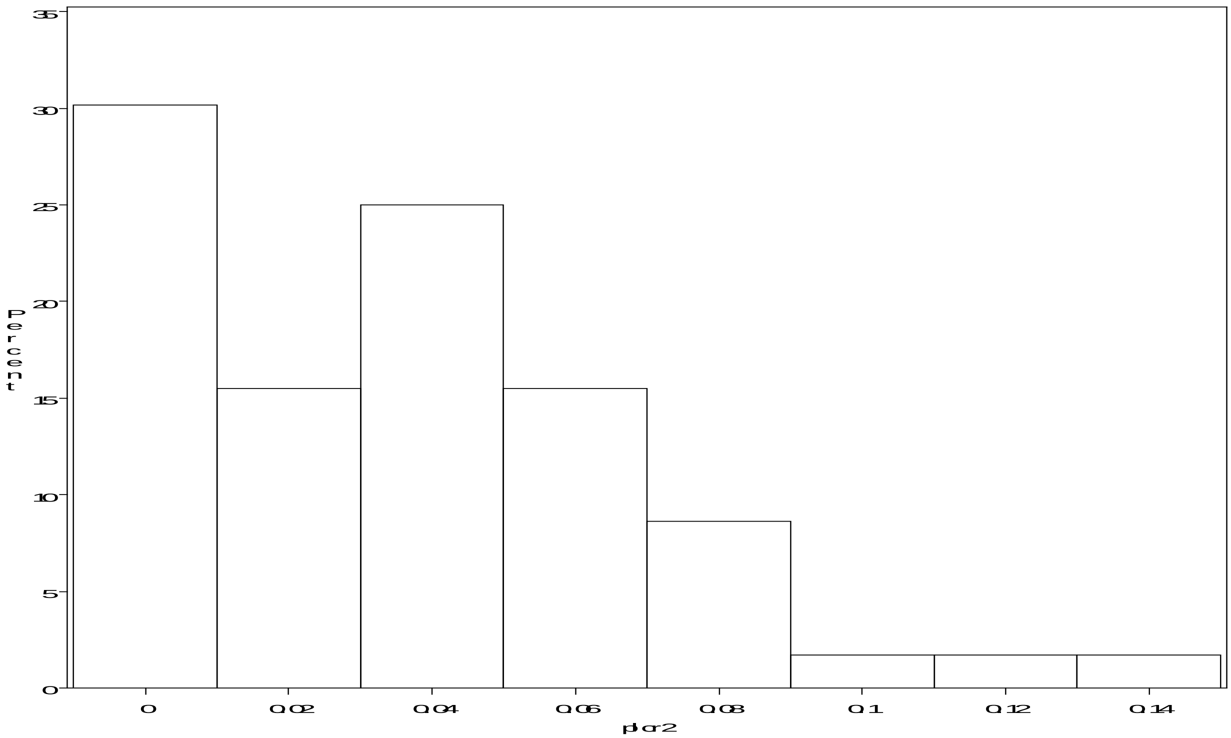


Figure 3.1. Histogram with equal bin widths and variable probabilities for each bin (508 compliant data is in Appendix D-1).

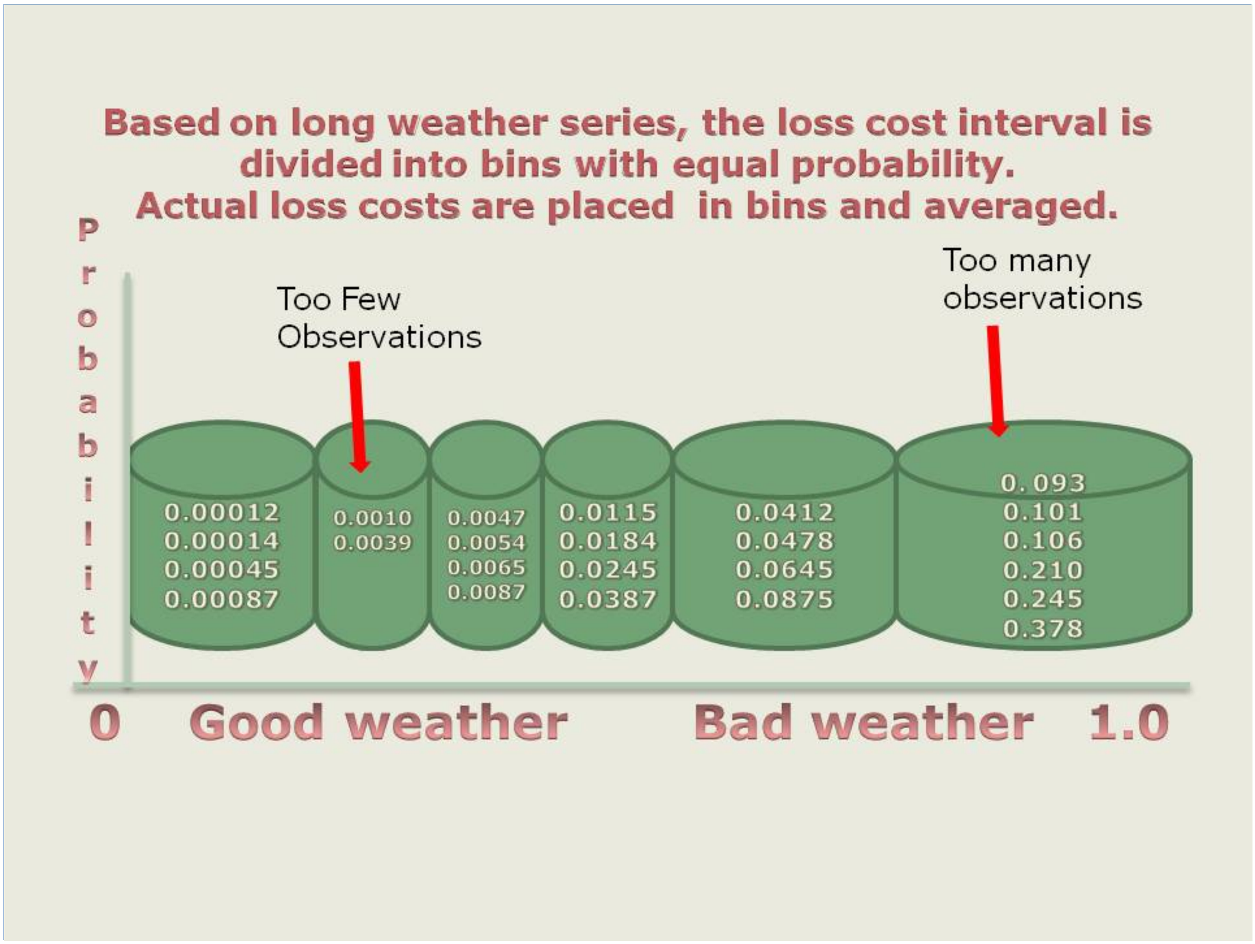


Figure 3.2. Variable bin widths with equal probabilities for each bin.

b. Conceptual Assessment of Non-stationarity in Loss Cost

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.

Where 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 market penetration. 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 170,000 acres. While not shown in this graph, much of the 1995 participation was in catastrophic coverage policies, 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.

Even after adjusting for the 1995 change and after accounting for weather, there is a discernable downward trend in loss costs for some crops in some regions. This may reflect changes in technology, including the increasing prominence of biotech crops, or other factors such as program changes or shifts in participation. Again, the loss experience should be adjusted at least to current trend levels, and consideration should be given to whether the trend should be extrapolated into the future.

As discussed in our prior study (Coble et al. 2010), if there have been shifts in the mix of business by type or practice – including shifts in technology that may not be directly reflected in the rates – the history should also be adjusted as much as possible to reflect current exposures. Such changes may be sufficiently captured in a trend analysis, but different changes can affect the experience differently, making it difficult to capture their effects using trend.

These adjustments should be made to the uncapped adjusted loss costs.

Selection of the Number of Years for Basic Ratemaking

RMA's program differs from most property/casualty exposures in that the loss experience is very 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, the probability that the number of bins required to ensure at least one observation per bin will drop to five and that there will still be only one observation in one or more of the bins is about 25%. With 20 years of data, that probability 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 are 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 or at most 25.

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 compound 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 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.

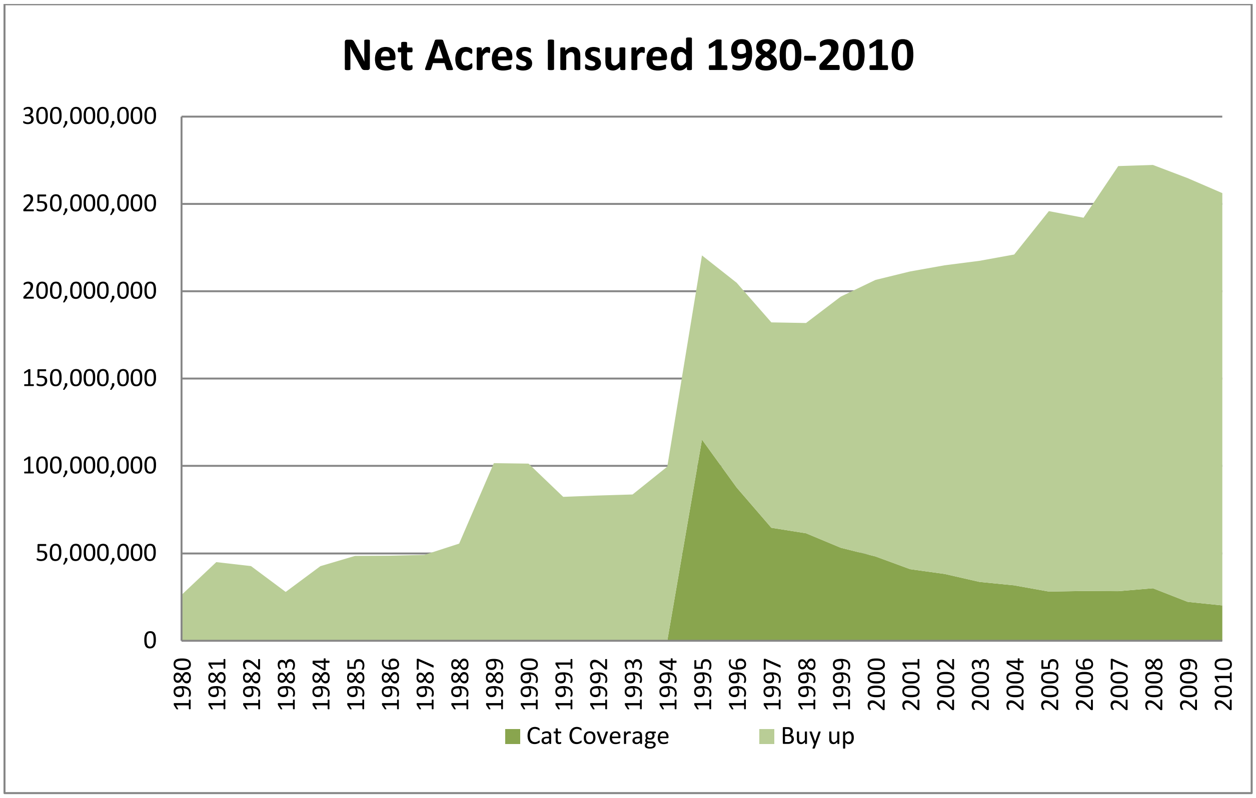


Figure 3.3 Net acre insured change from 1981 to 2009 (508 compliant data is in Appendix D-2).

4. Analysis of Weighting Issues

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 are 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 provide information since 1948 and NARR only since 1979.

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 National Climatic Data Center. A nice 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 provided 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 climate into regions that follow 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 cross 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 changes in 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 dichotomy 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

---

[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).

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 get assigned 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 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 base 65F, and allows exploration of loss

[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 (~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.

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). 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 (See Figure 4.2 for zero censoring in the corn data and Figure 4.3 for zero censoring in the barley data). 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% (See Figure 4.4)).

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

---

[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 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 between zero and one.

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 (or an extreme loss 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. We flag these cases, and the weighting methods based on the weather index developed are not applied (See Table 4.3).

Example Results

In Table 4.4, we show an 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 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 no climate division in a state with at least 10 observations, we do not apply the models and instead recommend that some form of subjective rating be used to establish rates.

An example of predicted loss costs for corn in Iowa (climate division 5) is presented in Table 4.8. 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 predicted loss cost 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 predicted loss costs to the appropriate bin or percentile cut-off. For example, assuming that we are interested in 10 bins we would like to find the predicted loss costs in the long history of weather data that correspond to the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th percentile, 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 predicted loss cost 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 predicted loss cost 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 to assume and the possible existence of empty bins during the years with loss cost data (from 1980 till 2009). 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 1980 till 2009 are utilized to calculate the average loss cost for a county. Hence, it is possible that years from 1980 to 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 1980-2009 period). For example, it may be that no year in the 1980 to 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 1980-2009 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 till 2 bins) to establish the largest number of bins for which there are no cases of empty bins in the years with loss data (1980-2009). This 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 (1980-2009) 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.

An example of the 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 1980-2009. All bin classifications are represented in the 1980-2009 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.

Loss Cost Averaging Procedure

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 have the same year classification and insignificance flags.

The average loss costs are next calculated using the 1980-2009 data where there are available actual adjusted loss cost values in the RMA Statplan 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, the recency weighting (discussed in more detail below) 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.

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 percentile. A similar calculation is done where the censoring is done at the 90th percentile (since there was a recent recommendation to increase the censoring for catastrophic loading to this level).

Example and National Summary Results

An example case 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 (i.e., we do not report a weather weighted average in this case).

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 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 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 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. Predicted loss cost 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	Predicted loss costs
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 predicted loss costs 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. Example 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. 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: corn example.

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

Note: Combinations of weather variables used for other crops can be seen in Appendix A.

Table 4.6. 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 predicted loss costs has a p-value > 0.1) or the correlation is negative.

Table 4.7. 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 predicted loss cost (weather index).

Table 4.8. 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 national average loss costs across all counties (i.e., liability weighted average) where the insignificance flags and state proxy flags are not equal to one. All weighted and unweighted loss costs for each county can be seen in the Appendix B.

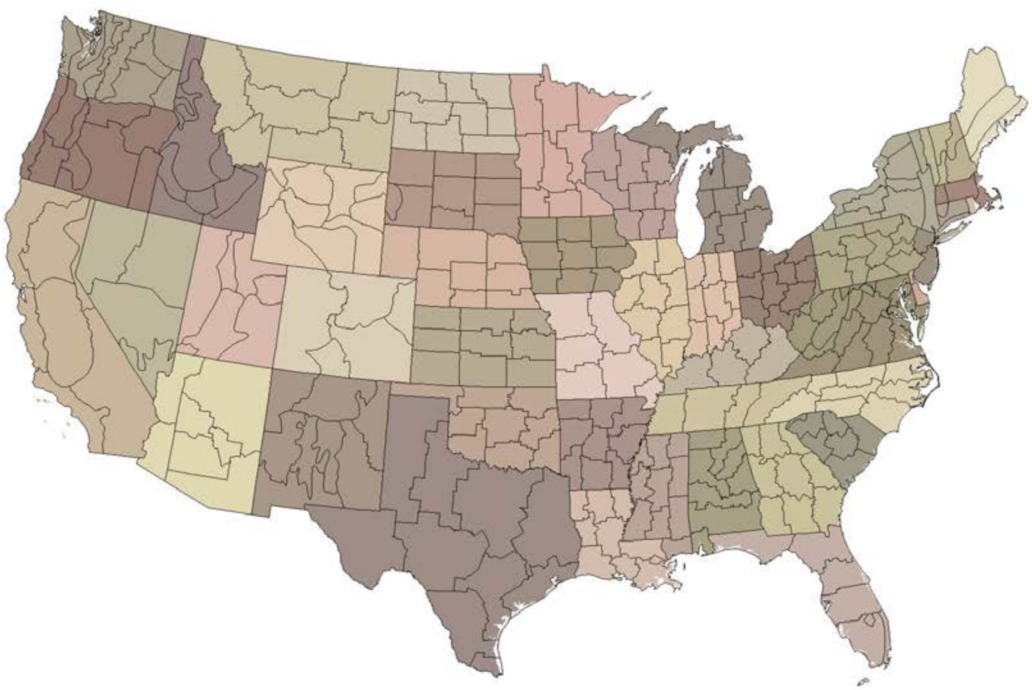


Figure 4.1a. Map of U.S. climate divisions.
(Established by the National Climate Data Center of NOAA)

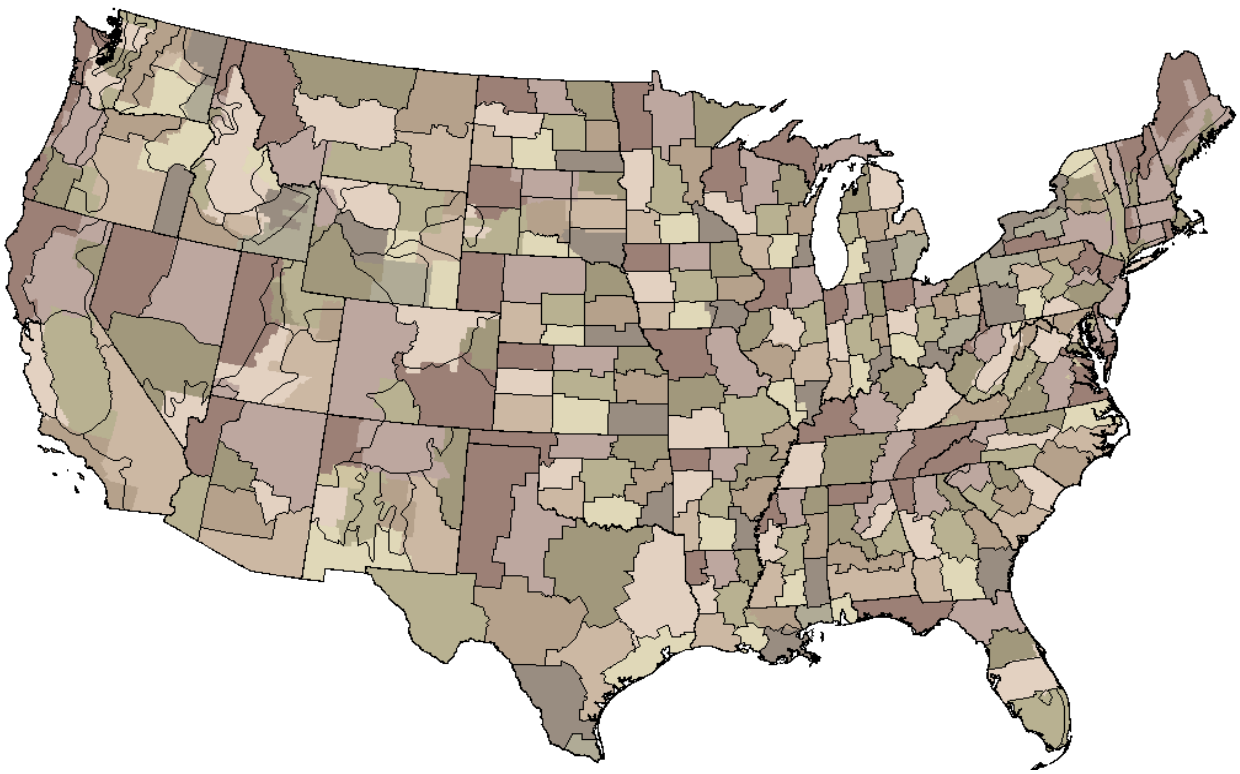


Figure 4.1b. County assignment to climate divisions delineated within states.

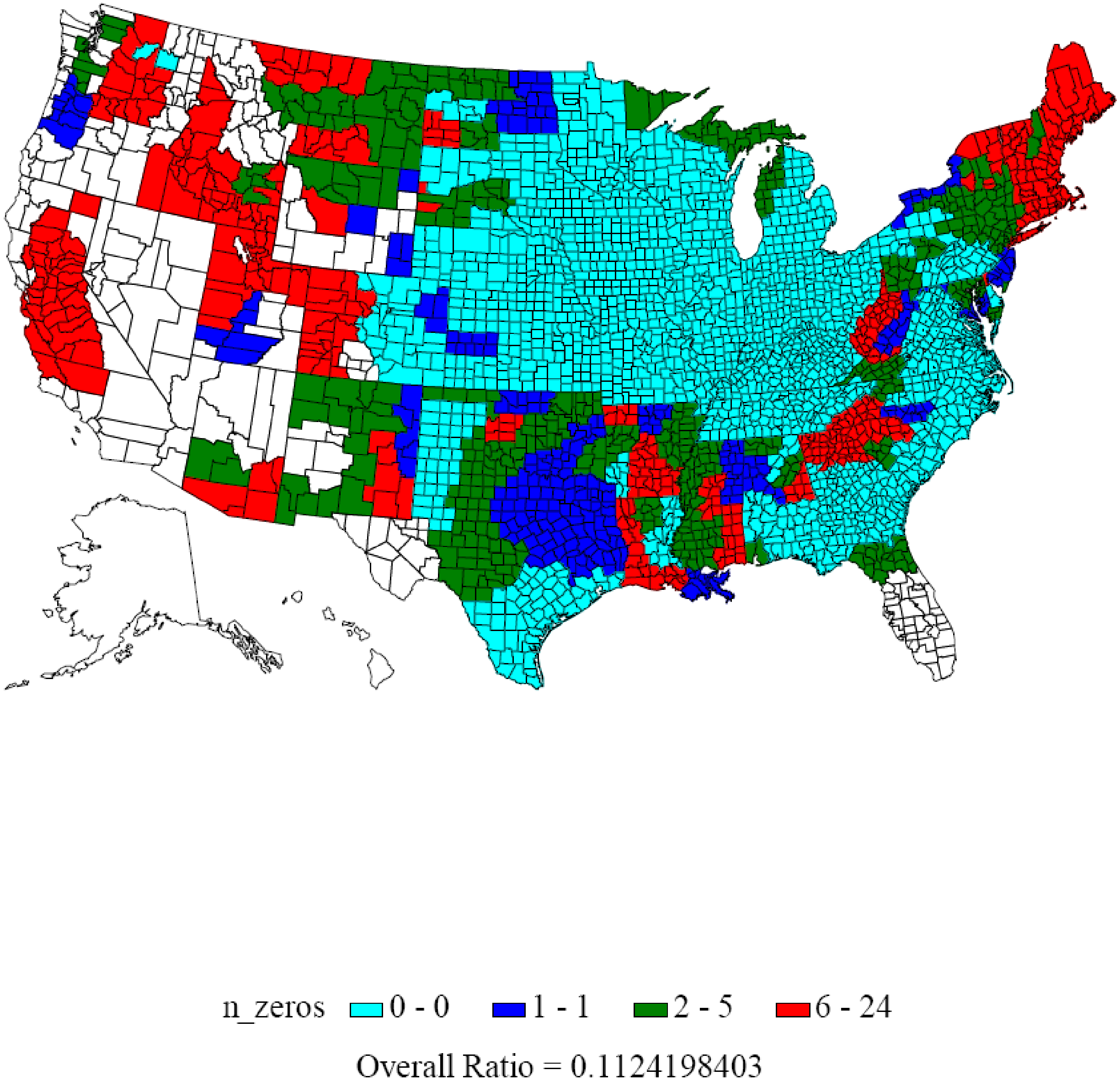


Figure 4.2. No. of years with zero loss costs in the corn climate division level data set (508 compliant data is in Appendix D-3).

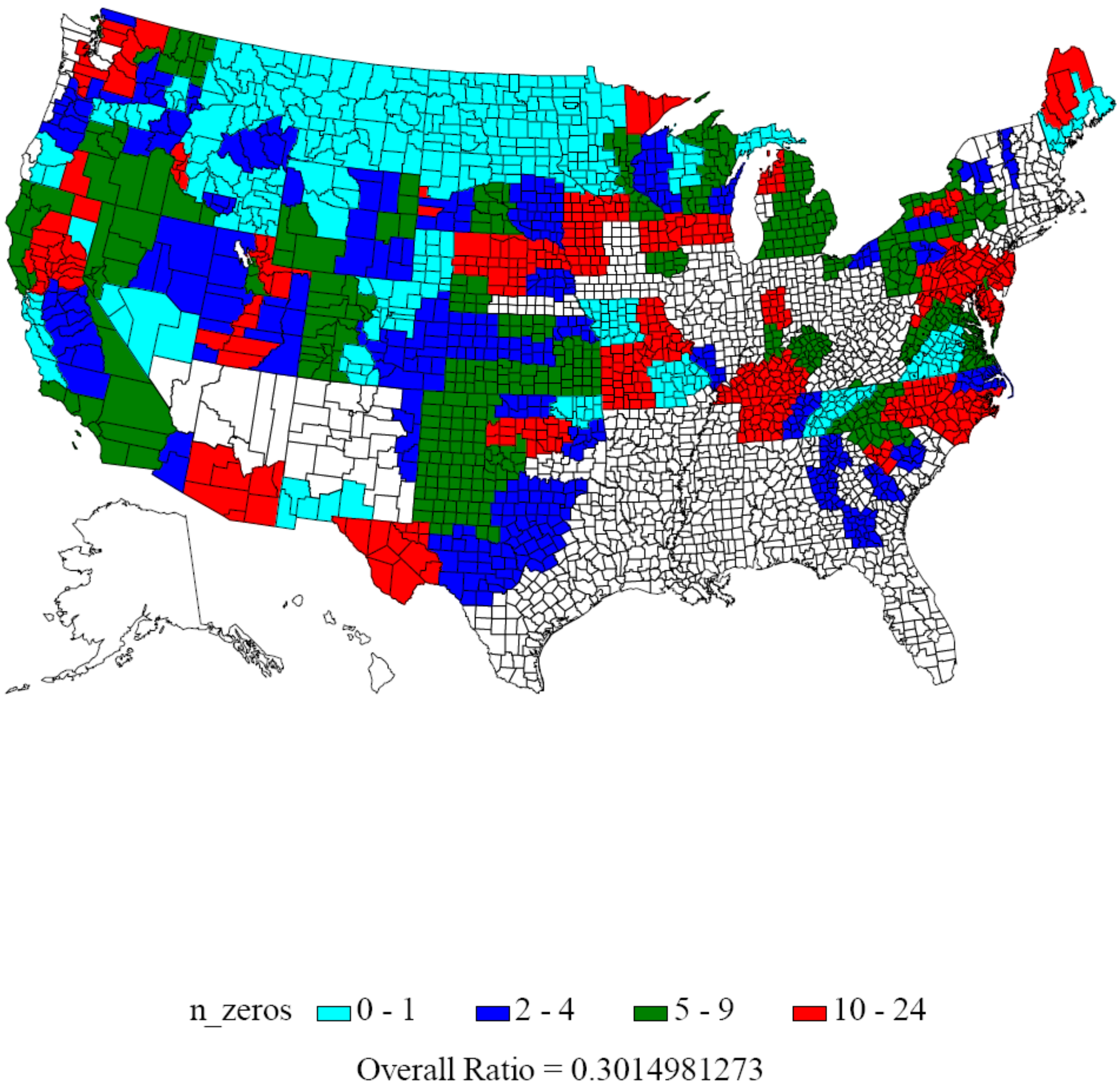


Figure 4.3. No. of years with zero loss costs in the barley climate division level data set (508 compliant data is in Appendix D-4).

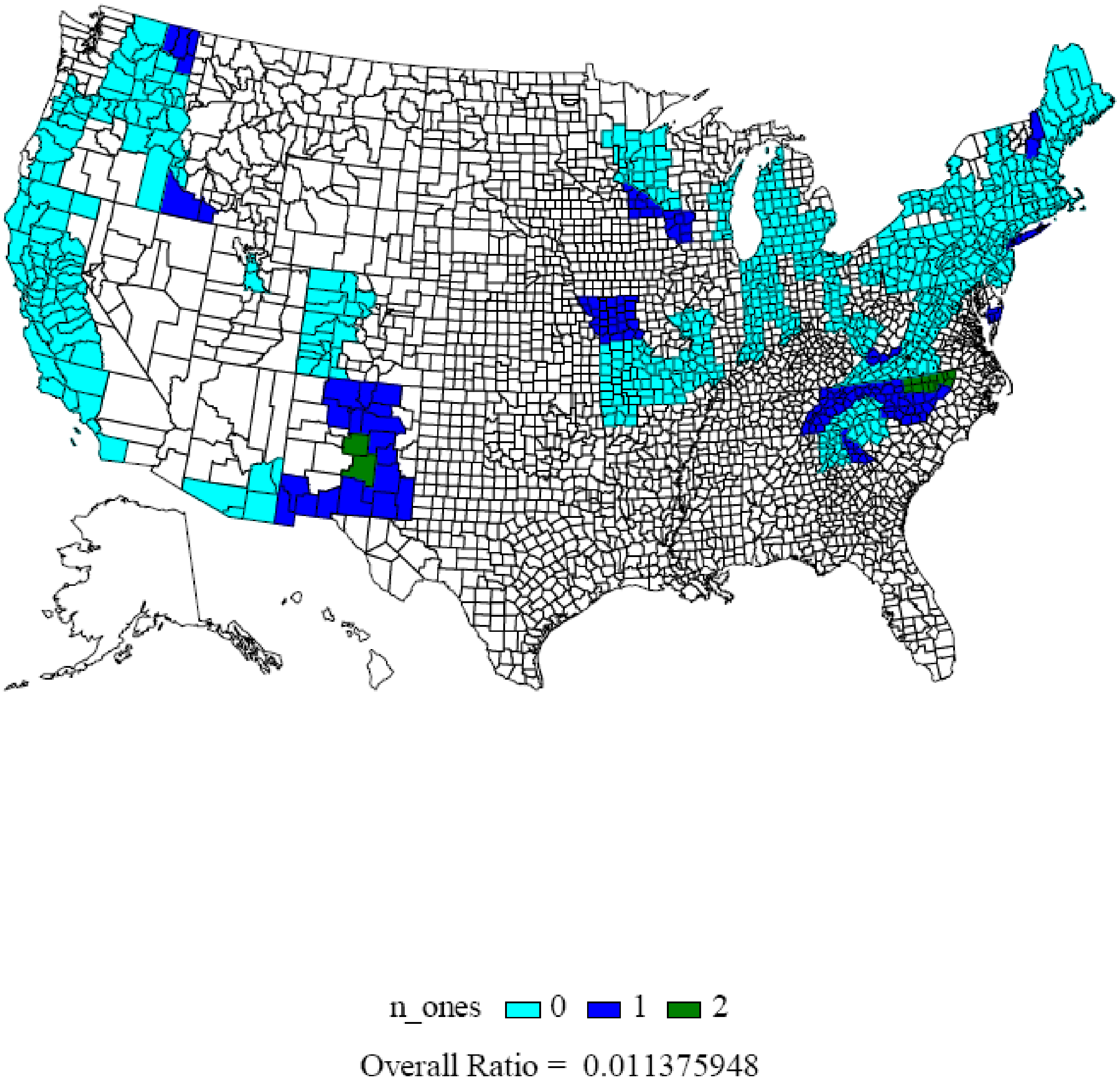


Figure 4.4. No. of years with loss cost equal to one in the apple climate division level data set (508 compliant data is in Appendix D-5).

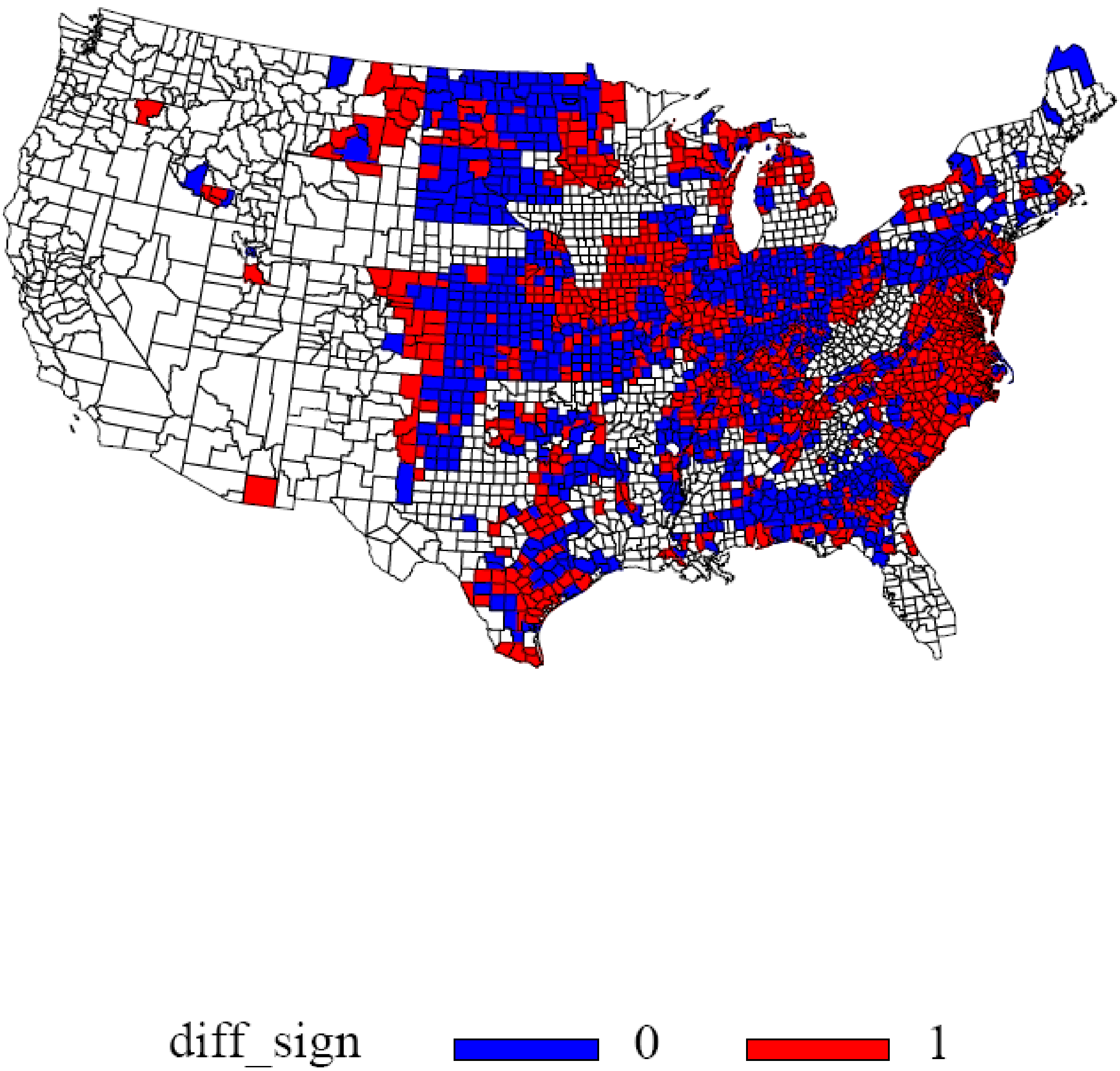


Figure 4.5. Map of the difference between the unweighted average loss cost and the weather weighted loss costs for corn (508 compliant data is in Appendix D-6).

Note: negative difference (weather weighted < unweighted) is in blue (0) and positive difference (weather weighted > unweighted) is in red (1).

b. Analysis of Alternative Loss Cost Adjustments

Based on the discussion of non-stationarity in the loss cost in section three, we conducted several empirical analyses to quantify these effects where possible. Because we are attempting to measure effects such as technological change or program changes that are applied broadly, 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 cost in the 1980-2010 data. Because of the increased aggregation, censoring of the dependant variable was examined and found to no longer be an issue, so for ease of interpretation ordinary least squares regression was used. Three alternative models were estimated:

model 1 adj_yr_lcr = f(pre-1995, weather variables)

model 2 adj_yr_lcr = f(pre-1995, post-1994 trend, weather variables)

model 3 adj_yr_lcr = f(net acres insured, weather variables)

The Pre-1995 variable takes a value of 1 if crop year is less than 1995. This variable is posited to capture differences in expected loss costs before and after the fundamental program changes that took place in 1995. The trend variable takes a value of zero if crop year is less than 1995 and is the difference between crop year and 1994 (crop year – 1994) from 1995 on. This variable is estimated with the pre-1995 variable to see if there are discernable trends in the loss cost experience since the 1995 program changes. Finally, net acres insured is the sum of net acres insured for the crop/state. Net acres reflect participation in insurance programs and crop acres. Further, all else equal, more acres provide greater credibility underlying the program. 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.

In the following tables we report results by crop. The parameter estimates for each state are reported. When the parameter is statistically significant at the ten percent level it is noted. We also shade in green any value that indicates more recent loss costs are lower than older experience. This is indicated by 1) a positive pre-1995 loss cost effect in model 1; 2) a positive pre-1995 loss cost effect in model 2 and/or the post-1994 loss cost trend; 3) a negative effect for net acres in model 3 when net acres insured are trending up through time. Conversely, if these parameters take the opposite signs and are statistically significant we highlight in red indicating higher loss cost patterns in recent years. In some instances the statistical significance for the pre-1995 effect is different between model 1 and model 2. In general we regard the effect to be significant if either model suggests so.

Table 4.10 reports the results for apples: a crop with a single state containing sufficient data to analyze and even then data are missing for the pre-1995 period. The lone significant loss cost adjustment is the net acre variable, which is negative and significant suggesting years with more insured acres have a lower loss cost. Because there has been an upward trend in net insured apple acres, this suggests more recent experience has been better in Washington apples.

Table 4.11 shows the loss cost adjustment results for barley. The results for the pre-1995 variables are decidedly mixed. In two states (CO, ID) more recent experience appears more actuarially sound. However, in three other states (MN, OR, SD) this result is contradicted. In no barley states is there a statistically significant trend effect, but in three states the net acres variable is significant (CO, ID, OR). However, the Colorado and Idaho results appear to conflict with the pre-1995 results.

The analysis for corn is reported in table 4.12. Note first that when results are statistically significant, they suggest more recent experience is better. Seven states are found to have a significantly higher loss cost prior to 1995. This includes major producing states of Illinois, Indiana, and Missouri. Downward trends in post-1995 lost costs are observed in Iowa, North Dakota, and Tennessee. The net acres variable is negative in every state where it is significant. In total, 13 states fall in this group which includes nearly every Corn Belt state.

Table 4.13 reports the results for grain sorghum. The pre-1995 variable is positive and significant in three states and negative and significant in Oklahoma. The post-1995 loss cost trend is not significant in any state. The net acres model contradicts the pre-1995 dummy model in Illinois. Further, in the major producing states of Kansas, Oklahoma, and Texas the net acre model suggests loss costs are increasing as net acres increase.

The results for potatoes in two states are shown in Table 4.14. Only one effect is significant. The pre-1995 effect in Idaho suggests higher loss costs prior to 1995.

Rice results are shown in Table 4.15. Only four states had sufficient data for analysis, but statistical significance is found in all four. The pre-1995 effect is significant in Arkansas, Louisiana, and Mississippi, but not Texas. The post-1994 trend is not significant in any state. However, the net acres model suggests that loss costs improve with increased insured acres in Arkansas, Louisiana, and Texas.

Soybean results are reported in Table 4.16. The pre-1995 effect is positive and significant in several states suggesting more recent experience is better. Most of the states with significant effects are Southern states such as Texas, Alabama, Arkansas, Mississippi, and Louisiana. However, Illinois and Nebraska also have the same sign. Virginia is the lone exception with a negative pre-1995 effect. The post-1995 trend model is only significant in two states – North Carolina and Virginia. In both cases it suggests recent loss costs are lower. When significant, the net acres insured model always suggests that loss costs improve with increased insured acres in the state. It is significant in seven regionally diverse states, but notable Illinois and Missouri from the Corn Belt.

Table 4.17 reports the results for spring wheat. Idaho and Montana are found to have a positive pre-1995 effect while South Dakota has an opposite sign. Only Idaho has a positive and significant post-1994 effect. The net acres insured effect is only significant in Montana.

The final table in this section is table 4.18 for winter wheat. While data are available from several states, the results for winter wheat are mixed. The pre-1995 dummy variable is positive and significant in several southeastern states, but takes the opposite sign in Texas, Oklahoma, Oregon, and Colorado. The post-1994 effect is only significant in three states. In California and Kentucky, it suggests upward trends in loss costs while in Colorado the trend is negative. The net acres variable is positive and significant in five states and negative and significant in South Carolina.

State Models

Table 4.10. Estimation of alternative state-level loss cost adjustments -apples.

Apples	State	Model 1 Pre-1995 Loss Cost Effect	Significant (10% level)	Model 2 Post 1994 Loss Cost Trend	Significant (10% level)	Model 3 Net Acres (10,000 ac)	Significant (10% level)
WA	53	N.A.					
WA	53	N.A.	.	-0.002515	.		
WA	53		.		.	-0.01898	*

Table 4.11. Estimation of alternative state-level loss cost adjustments -barley.

Barley	State	Model 1 Pre-1995 Loss Cost Effect	Significant (10% level)	Model 2 Post 1994 Loss Cost Trend	Significant (10% level)	Model 3 Net Acres (10,000 ac)	Significant (10% level)
CA	6	0.0003		.			
CA	6	0.08014		0.007645			
CA	6	.		.		0.02404	
CO	8	0.13256	*	.			
CO	8	0.0996		-0.002845			
CO	8	.		.		-0.00582	*
ID	16	0.04299	*	.			
ID	16	0.02962		-0.001751			
ID	16	.		.		-0.00086	*
MN	27	-0.08543	*	.			
MN	27	-0.0708		0.002076			
MN	27	.		.		-0.00325	
MT	30	0.06871		.			
MT	30	-0.00777		-0.009119			
MT	30	.		.		-0.00052	
ND	38	-0.02891		.			
ND	38	-0.05598		-0.003077			
ND	38	.		.		0.00006	
OR	41	-0.12441	*	.			
OR	41	-0.02063		0.01273			
OR	41	.		.		0.02837	*Acres stable
SD	46	-0.07658		.			
SD	46	-0.19942	*	-0.013524			
SD	46	.		.		0.00462	
WA	53	0.00874		.			
WA	53	-0.01023		-0.002346			
WA	53	.		.		0.00064	
WY	56	0.00477		.			
WY	56	-0.02075		-0.002717			
WY	56	.		.		-0.00017	

Table 4.12. Estimation of alternative state-level loss cost adjustments -corn.

Corn	State	Model 1 Pre-1995 Loss Cost Effect	Significant (10% level)	Model 2 Post 1994 Loss Cost Trend	Significant (10% level)	Model 3 Net Acres (10,000 ac)	Significant (10% level)
AL	1	0.05397		.			
AL	1	0.01514		-0.005237			
AL	1	.		.		-0.01103	*
CO	8	-0.01082		.			
CO	8	-0.02857		-0.00201			
CO	8	.		.		0.00012	
DE	10	-0.03129		.			
DE	10	-0.03946		-0.000921			
DE	10	.		.		0.00321	
GA	13	-0.00341		.			
GA	13	-0.07951		-0.010552			
GA	13	.		.		-0.00606	
IL	17	0.03233	*	.			
IL	17	0.01847		-0.001487			
IL	17	.		.		-0.00008	*
IN	18	0.03072		.			*
IN	18	0.01706		-0.001713			
IN	18	.		.		-0.00021	*
IA	19	0.01765		.			
IA	19	-0.01356		-0.003442	*		
IA	19	.		.		-0.00005	*
KS	20	0.00013		.			
KS	20	0.00014		0.000001			
KS	20	.		.		-0.00002	
KY	21	0.09438	*	.			
KY	21	0.09772	*	0.000401			
KY	21	.		.		-0.00199	*
LA	22	0.06448		.			
LA	22	0.03784		-0.003211			
LA	22	.		.		-0.00091	
MD	24	-0.0179		.			
MD	24	-0.02377		-0.00075			
MD	24	.		.		0.00084	
MI	26	0.04805	*	.			
MI	26	0.01403		-0.004335			

MI	26	.		.		-0.00121	*
MN	27	-0.00199		.			
MN	27	0.03671		0.005493			
MN	27	.		.		0.00002	
MS	28	0.12261		.			*
MS	28	0.07581		-0.005209			
MS	28	.		.		-0.00392	
MO	29	0.05629	*	.			
MO	29	-0.01254		-0.008116			
MO	29	.		.		-0.00071	*
NE	31	0.00704		.			
NE	31	-0.00793		-0.001761			
NE	31	.		.		-0.00003	
NC	37	-0.00556		.			
NC	37	-0.00788		-0.000317			
NC	37	.		.		-0.00024	
ND	38	0.02687		.			
ND	38	-0.10847		-0.013724	*		
ND	38	.		.		-0.00062	*
OH	39	0.02491		.			
OH	39	0.00948		-0.001992			
OH	39	.		.		-0.00024	*
OK	40	-0.03678		.			
OK	40	-0.04379		-0.000801			
OK	40	.		.		0.00519	*
PA	42	-0.01854		.			
PA	42	-0.01005		0.001093			
PA	42	.		.		0.00068	
SC	45	0.00518		.			
SC	45	-0.03907		-0.006141			
SC	45	.		.		-0.00406	
SD	46	-0.00299		.			
SD	46	-0.0701		-0.007388			
SD	46	.		.		0.00002	
TN	47	0.03634	*	.			
TN	47	0.00366		-0.003805	*		
TN	47	.		.		-0.00184	*
TX	48	0.08919	*	.			
TX	48	0.0564		-0.003803			
TX	48	.		.		-0.00076	*

VA	51	0.02188		.		
VA	51	0.00828		-0.001749		
VA	51	.		.	-0.0023	*
WV	54	0		.		
WV	54	0		0.004302		
WV	54	.		.	0.08797	
WI	55	0.04375	*	.		
WI	55	-0.00753		-0.006336		
WI	55	.		.	-0.00048	*
WY	56	0.04215		.		
WY	56	0.05978		0.002224		
WY	56	.		.	-0.00007	

Table 4.13. Estimation of alternative state-level loss cost adjustments –grain sorghum.

Grain Sorghum	State	Model 1 Pre-1995 Loss Cost Effect	Significant (10% level)	Model 2 Post 1994 Loss Cost Trend	Significant (10% level)	Model 3 Net Acres (10,000 ac)	Significant (10% level)
AR	5	0.07993	*	.		.	
AR	5	0.06457		-0.001825		.	
AR	5	.		.		-0.00851	
CO	8	-0.03422		.		.	
CO	8	0.12454		0.017516	*	.	
CO	8	.		.		0.00519	
IL	17	0.07109	*	.		.	
IL	17	0.10496		0.0035		.	
IL	17	.		.		0.02795	* Acres stable
KS	20	-0.03363		.		.	
KS	20	-0.02108		0.001516		.	
KS	20	.		.		0.00032	*
LA	22	0.09289		.		.	
LA	22	0.11792		0.003167		.	
LA	22	.		.		0.00399	
MO	29	0.07442	*	.		.	
MO	29	0.04461		-0.003915		.	
MO	29	.		.		0.00732	* Acres declining
NE	31	-0.01168		.		.	
NE	31	-0.0239		-0.001438		.	
NE	31	.		.		-0.0006	
NM	35	-0.01621		.		.	
NM	35	-0.11169		-0.010528		.	
NM	35	.		.		0.00064	
OK	40	-0.11504	*	.		.	
OK	40	-0.12353	*	-0.00102		.	
OK	40	.		.		0.00912	*
SD	46	-0.14014		.		.	
SD	46	-0.31675		-0.018606		.	
SD	46	.		.		0.01553	
TX	48	-0.01622		.		.	
TX	48	-0.08615		-0.008175		.	
TX	48	.		.		0.00041	*

Table 4.14. Estimation of alternative state-level loss cost adjustments -potato.

Potato	State	Model 1 Pre-1995 Loss Cost Effect	Significant (10% level)	Model 2 Post 1994 Loss Cost Trend	Significant (10% level)	Model 3 Net Acres (10,000 ac)	Significant (10% level)
ID	16	0.02631	*	.			
ID	16	0.02479		-0.000183			
ID	16	.		.		-0.00149	
ND	38	0.00769		.			
ND	38	0.03158		0.002529			
ND	38	.		.		0.01461	

Table 4.15. Estimation of alternative state-level loss cost adjustments -rice.

Rice	state	Model 1 Pre-1995 Loss Cost Effect	Significant (10% level)	Model 2 Post 1994 Loss Cost Trend	Significant (10% level)	Model 3 Net Acres (10,000 ac)	Significant (10% level)
AR	5	0.05267	*	.		.	
AR	5	0.05573	*	0.000506964		.	
AR	5	.		.		-0.00133	*
LA	22	0.07592	*	.		.	
LA	22	0.08937	*	0.001595231		.	
LA	22	.		.		-0.00417	*
MS	28	0.01978	*	.		.	
MS	28	0.02262	*	0.000333516		.	
MS	28	.		.		-0.00157	
TX	48	0.01066		.		.	
TX	48	0.01103		0.000041966		.	
TX	48	.		.		-0.00117	*

Table 4.16. Estimation of alternative state-level loss cost adjustments -soybeans.

Soybeans	state	Model 1 Pre-1995 Loss Cost Effect	Significant (10% level)	Model 2 Post 1994 Loss Cost Trend	Significant (10% level)	Model 3 Net Acres (10,000 ac)	Significant (10% level)
AL	1	0.12153	*	.			
AL	1	0.08483		-0.00495			
AL	1	.		.		0.00121	
AR	5	0.11899	*	.			
AR	5	0.06552		-0.00724			
AR	5	.		.		-0.00135	*
DE	10	-0.06008		.			
DE	10	-0.04477		0.0016			
DE	10	.		.		0.00432	
GA	13	0.04816		.			
GA	13	-0.03076		-0.01094			
GA	13	.		.		-0.00097	
IL	17	0.03186	*	.			
IL	17	0.04357	*	0.001257			
IL	17	.		.		-0.00011	*
IN	18	0.0129		.			
IN	18	0.00284		-0.00126			
IN	18	.		.		-0.0001	*
IA	19	0.00848		.			
IA	19	0.00149		-0.00077			
IA	19	.		.		-0.00001	
KS	20	0.04147		.			
KS	20	0.02152		-0.00241			
KS	20	.		.		-0.00032	
KY	21	0.05654	*	.			
KY	21	0.05204		-0.00055			
KY	21	.		.		-0.00086	
LA	22	0.17882	*	.			
LA	22	0.15714	*	-0.0026			
LA	22	.		.		-0.00055	
MA	24	-0.00688		.			
MA	24	-0.02017		-0.0017			
MA	24	.		.		-0.00052	
MI	26	0.05011		.			
MI	26	0.03285		-0.00216			

MI	26	.		.		-0.00111	*
MN	27	-0.00566		.			
MN	27	0.01279		0.002618			
MN	27	.		.		0.00003	
MS	28	0.11254	*	.			
MS	28	0.09199	*	-0.00247			
MS	28	.		.		-0.00127	*
MO	29	0.06373	*	.			
MO	29	0.07277		0.001067			
MO	29	.		.		-0.00029	*
NE	31	0.02344	*	.			
NE	31	0.01064		-0.00151			
NE	31	.		.		-0.00007	*
NC	37	0.01511		.			
NC	37	-0.05389		-0.00941	*		
NC	37	.		.		-0.00084	*
ND	38	-0.01142		.			
ND	38	-0.02464		-0.0015			
ND	38	.		.		0.00002	
OH	39	0.01033		.			
OH	39	0.00355		-0.00088			
OH	39	.		.		-0.00007	
OK	40	0.0238		.			
OK	40	0.02165		-0.00027			
OK	40	.		.		0.0008	
SC	45	0.13705	*	.			
SC	45	-0.04658		-0.01734			
SC	45	.		.		-0.00638	*
SD	46	0.01005		.			
SD	46	-0.02739		-0.00412			
SD	46	.		.		-0.00004	
TN	47	0.03078		.			
TN	47	-0.00486		-0.00438			
TN	47	.		.		-0.00106	
TX	48	0.17472	*	.			
TX	48	0.28287	*	0.012657			
TX	48	.		.		-0.00376	
VA	51	-0.01865		.			
VA	51	-0.09645	*	-0.01001	*		
VA	51	.		.		-0.00113	

WI	55	0.00697	.	
WI	55	0.02185	0.001985	
WI	55	.	.	0.00018

Table 4.17. Estimation of alternative state-level loss cost adjustments –spring wheat.

Spring wheat	State	Model 1 Pre-1995 Loss Cost Effect	Significant (10% level)	Model 2 Post 1994 Loss Cost Trend	Significant (10% level)	Model 3 Net Acres (10,000 ac)	Significant (10% level)
CA	6	-0.52676		.		.	
CA	6	-0.57888		-0.01298		.	
CA	6	.		.		-0.01356	
ID	16	0.07706	*	.		.	
ID	16	-0.02428		-0.02241	*	.	
ID	16	.		.		-0.0032	
MT	30	0.07593	*	.		.	
MT	30	0.04504		-0.00369		.	
MT	30	.		.		-0.00036	*
ND	38	-0.00807		.		.	
ND	38	-0.04989		-0.0047		.	
ND	38	.		.		-9.8E-06	
SD	46	-0.06522	*	.		.	
SD	46	-0.12937	*	-0.00737		.	
SD	46	.		.		0.00026	
WA	53	0.01824		.		.	
WA	53	-0.00158		-0.00433		.	
WA	53	.		.		-0.00014	

Table 4.18. Estimation of alternative state-level loss cost adjustments –winter wheat.

Winter Wheat	State	Model 1 Pre-1995 Loss Cost Effect	Significant (10% level)	Model 2 Post 1994 Loss Cost Trend	Significant (10% level)	Model 3 Net Acres (10,000 ac)	Significant (10% level)
AR	5	0.06946		.		.	
AR	5	0.07271		0.00043		.	
AR	5	.		.		-0.00349	
CA	6	-0.1172		.		.	
CA	6	0.18789		0.0839	*	.	
CA	6	.		.		0.02449	*
CO	8	-0.01127		.		.	
CO	8	-0.0767	*	-0.00778	*	.	
CO	8	.		.		0.00047	*Declining Net Acres
GA	13	0.03441		.		.	
GA	13	0.07069	*	0.00489		.	
GA	13	.		.		0.00103	
ID	16	0.00886		.		.	
ID	16	0.01471		0.00078		.	
ID	16	.		.		-0.00008	
IL	17	0.07741		.		.	
IL	17	0.05071		-0.00331		.	
IL	17	.		.		-0.00357	
IN	18	0.05584		.		.	
IN	18	0.00431		-0.00962		.	
IN	18	.		.		-0.00499	
KS	20	-0.01623		.		.	
KS	20	-0.0109		0.00063		.	
KS	20	.		.		0.00004	
KY	21	0.03965		.		.	
KY	21	0.10339	*	0.00909	*	.	
KY	21	.		.		0.00003	
LA	22	0.16253	*	.		.	
LA	22	0.17153		0.00108		.	
LA	22	.		.		-0.01089	
MI	26	-0.0074		.		.	
MI	26	-0.03854		-0.00429		.	
MI	26	.		.		-0.00102	
MN	27	0.01866		.		.	

MN	27	0.00315		-0.00205	.	
MN	27	.		.	-0.00013	
MS	28	0.17655	*	.	.	
MS	28	0.14653		-0.00428	.	
MS	28	.		.	-0.01082	
MO	29	0.07196		.	.	
MO	29	0.04356		-0.00352	.	
MO	29	.		.	-0.00523	
MT	30	-0.0611		.	.	
MT	30	-0.0762		-0.00407	.	
MT	30	.		.	0.00057	
NE	31	0.01788		.	.	
NE	31	0.02936		0.001353	.	
NE	31	.		.	-0.00002	
NM	35	-0.0217		.	.	
NM	35	-0.15042		-0.0147	.	
NM	35	.		.	-0.00505	
NC	37	-0.00501		.	.	
NC	37	-0.00536		-4.5E-05	.	
NC	37	.		.	0.00031	
OH	39	0.01238		.	.	
OH	39	-0.01313		-0.00354	.	
OH	39	.		.	-0.00111	
OK	40	-0.07085	*	.	.	
OK	40	-0.05364		0.002009	.	
OK	40	.		.	0.00026	*
OR	41	-0.06108	*	.	.	
OR	41	-0.08991	*	-0.00384	.	
OR	41	.		.	0.00209	*
SC	45	0.05753		.	.	
SC	45	-0.03191		-0.01049	.	
SC	45	.		.	-0.01346	*
SD	46	-0.01197		.	.	
SD	46	-0.03212		-0.00484	.	
SD	46	.		.	-0.00038	
TX	48	-0.11145	*	.	.	
TX	48	-0.03244		0.009376	.	
TX	48	.		.	0.0004	*
UT	49	0.03639		.	.	
UT	49	0.03853		0.000183	.	

UT	49	.	.	0.00341
WA	53	-0.00358	.	.
WA	53	0.00343	0.000891	.
WA	53	.	.	-0.00004
WY	56	-0.02224	.	.
WY	56	-0.03068	-0.00099	.
WY	56	.	.	0.00415

Regional and National Models

We also replicated our non-stationarity analysis at regional and national levels. The regions were defined to reflect similar production areas as shown in Table 4.19. Weather effects were maintained. By combining states, the non-stationarity effects are estimated with more data. This approach provides some smoothing of any effects, but also increases the chance that a general adjustment may not be appropriate in some specific location.

Table 4.19. Region definitions.

Region	States
Southeast Region	1, 12, 13, 37, 45
Southern Plains	20, 31, 8, 40, 48
West	16, 30, 41, 53, 56, 4, 6, 32, 35, 49
East	9, 10, 23, 24 , 25, 33, 34, 36, 42, 44, 50, 51, 54
Midwest	17, 18, 19, 26, 27 29, 38, 39, 46, 55, 21
Delta	5, 28, 47, 22
Southwest	4, 6, 32, 35, 49
Pacific Northwest	2, 16, 41, 53
Northern Plains	30, 38, 46
Great Plains	8, 20, 31, 46, 56

Table 4.20 reports the estimates for spring and winter wheat regions. No effects are significant for spring wheat. However, the results are mixed for winter wheat. In the Delta, Southeast and Midwest there is evidence of significant differences in loss costs over time. Conversely, in the various groupings of the Plains states and the Pacific Northwest there is evidence in at least one of the three model specifications that loss experience is worse in recent years.

Corn and soybean results are reported in table 4.21. In this table any effects that are significant suggest that loss experience is better in more recent years. For corn only the east and west regions show no significance. The clearest result is for the Midwest where all three models suggest improving loss costs over time. For soybeans the pre-1995 effect is significant in the Midwest but has the largest magnitude in the Delta and Southeast.

Table 4.22 reports the regional results for sorghum and cotton. In sorghum the net acres model is never significant, but the pre-1995 effect is significant in the Delta, Midwest, and southeast. Southeast also demonstrates a downward post 1995 trend. Conversely, in the West experience prior to 1995 was better than more recent experience. In cotton, Delta loss cost experience was higher prior to 1995. However, results across models are inconsistent for the Southeast, Southwest, and West regions.

The final regional model is reported in Table 4.23. It reports the results for barley. Only the result for the Midwest region is statistically significant and it suggests that experience prior to 1995 had a lower expected loss cost ratio.

The final table in this set is for national level models (Table 4.24). All crops are reported in a single table. In many instances the national level results mimic the regional results for the major production regions. However, this does not always hold. First, none of the non-stationarity estimates are significant for spring wheat or for barley – similar to most of the regional results.

There is evidence of higher national-aggregate loss costs prior to 1995 for winter wheat, corn, rice and soybeans. When we examine post-1994 loss cost trends at the national level, evidence is found for declining loss costs in corn, soybeans, cotton and sorghum. No evidence is found of increasing trends in loss cost. Our third model uses net acres insured as a variable to explain non-stationarity of loss costs. This model suggests improving loss costs for corn, soybeans, and cotton.

Ultimately, these results are quite similar to the limited sample we examined in our previous rate review report. Statistically significant trends in loss cost while controlling for random weather poses a serious issue for rating. We have suggested several alternative ways to quantify these effects and analyzed them at state, regional, and national levels of aggregation. The other alternative is to shorten the time series used for base rates so that the effects we observe are excluded or dampened. One might also combine approaches. For example one might limit the years used for base rates to 20 years but currently five of those years would be prior to 1995 and one might apply a pre-1995 adjustment to those years until they drop from the 20 year rating period. In general, we believe it is desirable to make these non-stationarity adjustments multiplicative and impose constraints on the magnitude of the effect these adjustments would have on loss costs. Once the non-stationarity adjustments are made then weather probabilities would be applied.

Table 4.20. Region non-stationarity tests for spring and winter wheat.

Region	Pre-1995 Loss Cost Effect		Post 1994 Loss Cost Trend		Net Acres (10,000 ac)	
			Spring wheat			
Northern Plains	0.0185		.			
Northern Plains	-0.01318		-0.003639			
Northern Plains	.		.		-0.00008	
Pacific Northwest	0.01188		.			
Pacific Northwest	0.00661		-0.001259			
Pacific Northwest	.		.		-0.00084	
			Winter Wheat			
Delta	0.13971	*	.		.	
Delta	0.11266	*	-0.0033819		.	
Delta	.		.		-0.00373	
East	0.07386		.		.	
East	0.0304		-0.0048586	*	.	
East	.		.		0.00073	
Great Plains	-0.00941		.		.	
Great Plains	-0.01839		-0.001152		.	
Great Plains	.		.		0.0001	*
Midwest	0.04792	*	.		.	
Midwest	0.00487		-0.0055326		.	
Midwest	.		.		0.00003	
Northern Plains	-0.0611		.		.	
Northern Plains	-0.0762		-0.0040727		.	
Northern Plains	.		.		0.00057	*
Pacific Northwest	-0.02197	*	.		.	
Pacific Northwest	-0.01697		0.0006348		.	
Pacific Northwest	.		.		-0.00006	
Southeast	0.0484	*	.		.	
Southeast	0.04179		-0.0007686		.	
Southeast	.		.		-0.00072	
Southern Plains	-0.08826	*	.		.	
Southern Plains	-0.02858		0.00713888		.	
Southern Plains	.		.		0.00037	*
Southwest	-0.03004		.		.	
Southwest	-0.08856		-0.0074822		.	
Southwest	.		.		0.00651	*

Table 4.21. Region non-stationarity tests for corn and soybeans.

Region	Pre-1995 Loss Cost Effect		Post 1994 Loss Cost Trend		Net Acres (10,000 ac)	
			Corn			
Delta	0.034925	*	.		.	
Delta	0.008874		-0.0029882		.	
Delta	.		.		-0.00154	*
East	0.006703		.		.	
East	-0.00962		-0.0018861		.	
East	.		.		0.00044	
Midwest	0.043685	*	.		.	
Midwest	0.020337		-0.0028144	*	.	
Midwest	.		.		-0.00009	*
Southeast	0.014344		.		.	
Southeast	-0.02243		-0.0047605		.	
Southeast	.		.		-0.00001	*
Southern Plains	0.0087		.		.	
Southern Plains	-0.00459		-0.0016893	*	.	
Southern Plains	.		.		-0.00007	
West	0.035159		.		.	
West	0.0547		0.00222292		.	
West	.		.		0.00191	
			Soybeans			
Delta	0.10693	*	.		.	
Delta	0.07751	*	-0.0037164		.	
Delta	.		.		-0.00067	
East	-0.00347		.		.	
East	-0.02656		-0.0026502	*	.	
East	.		.		0.00010262	
Midwest	0.03513	*	.		.	
Midwest	0.03005	*	-0.0006099		.	
Midwest	.		.		-0.000082	
Southeast	0.07068	*	.		.	
Southeast	-0.00497		-0.0090618	*	.	
Southeast	.		.		-0.0003843	
Upper Midwest	0.01123		.		.	
Upper Midwest	0.00244		-0.0010989		.	
Upper Midwest	.		.		-0000475	
West	0.04458		.		.	
West	0.05379		0.00120945		.	
West	.		.		-0.0003471	

Table 4.22. Region non-stationarity tests for sorghum and cotton.

Region	Pre-1995 Loss Cost Effect		Post 1994 Loss Cost Trend		Net Acres (10,000 ac)	
			Sorghum			
Delta	0.06341	*	.		.	
Delta	0.09199	*	0.003598		.	
Delta	.		.		0.00146	
East	0.04206		.		.	
East	0.02496		-0.002065		.	
East	.		.		0.13723	
Midwest	0.03544	*	.		.	
Midwest	0.00381		-0.003726		.	
Midwest	.		.		0.0001	
Southeast	0.10649	*	.		.	
Southeast	0.02314		-0.010552	*	.	
Southeast	.		.		-0.07146	
Upper Midwest	-0.02973		.		.	
Upper Midwest	-0.10657		-0.008961		.	
Upper Midwest	.		.		0.00277	
West	-0.04838	*	.		.	
West	-0.07605	*	-0.003563		.	
West	.		.		0.00008	
			Cotton			
Delta	0.03686	*	.		.	
Delta	0.00329		-0.003946		.	
Delta	.		.		0.00008	
Southeast	-0.01487		.		.	
Southeast	-0.08212	*	-0.009031	*	.	
Southeast	.		.		0.00008	
Southwest	0.00098		.		.	
Southwest	-0.12678	*	-0.015504	*	.	
Southwest	.		.		0.00008	
West	-0.00996		.		.	
West	-0.04922		-0.005002	*	.	
West	.		.		0.00008	*

Table 4.23. Region non-stationarity tests for barley.

Region	Pre-1995 Loss Cost Effect		Post 1994 Loss Cost Trend	Net Acres (10,000 ac)
			Barley	
Great Plains	0.00648		.	
Great Plains	-0.03612		-0.004959	
Great Plains	.		.	0.00258
Midwest	-0.08571	*	.	
Midwest	-0.0708		0.002078	
Midwest	.		.	-0.00245
Northern Plains	0.01479		.	
Northern Plains	-0.03695		-0.006152	
Northern Plains	.		.	-0.00014
Pacific Northwest	-0.01843		.	
Pacific Northwest	0.00037		0.002385	
Pacific Northwest	.		.	-0.00069
Southwest	0.00694		.	
Southwest	0.12566		0.011005	
Southwest	.		.	0.02428

Table 4.24 National-level non-stationarity estimates.

Pre-1995 Loss Cost Effect		Post 1994 Loss Cost Trend		Net Acres (10,000 ac)	
		Spring wheat			
-0.00503		.		.	
-0.01618		-0.00132		.	
.		.		1.26E-09	
		Winter wheat			
0.024777	*	.		.	
0.011013		-0.0017		.	
.		.		0.000021593	
		Corn			
0.030991	*	.		.	
0.006862		-0.00289	*	.	
.		.		-0.000091562	*
		Soybeans			
0.046959	*	.		.	
0.027417	*	-0.00236	*	.	
.		.		-0.00012155	*
		Sorghum			
0.037107		.		.	
0.004925		-0.00393	*	.	
.		.		-0.000176957	
		Cotton			
0.002815		.		.	
-0.05395	*	-0.00687	*	.	
.		.		0.00019965	*
		Rice			
0.03342	*	.		.	
0.034434	*	0.000128		.	
.		.		-0.000317513	
		Barley			
0.020735		.		.	
0.002854		-0.0021		.	
.		.		0.000020118	

Using Shortened Loss Cost Series for Base Rates

Table 4.25 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 older 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 4.25 are derived by conducting the weather weighting procedure described earlier, but then any data older than 20 years are dropped from the binning step of the process.

The table reflects three scenarios relative to the catastrophic load. First, we show 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 an unweighted result. All results are report as a percentage of the table 9 results. Values greater than 100% suggest 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 three crops suggest that limiting loss cost histories to 20 years would result in substantially higher rates for apples, cotton and winter wheat. Conversely, barley, corn, soybeans and spring wheat all are observed to have substantially lower rates. Note that significant variation is observed within a crop.

Weighting Approaches

In this section, we explore three other “weighting” approaches: net acre weighting, decile weighting, and linear weighting. Given the weather weighting approach discussed above, all the recency weighting here is done in the step where the average loss cost is calculated within each bin (i.e., within-bin averaging). For example, if there are 3 years in bin 5 that includes 1986, 1994, and 2008, then the year 2008 is given more weight than 1994 and 1986, and 1994 has more weight compared to 1986 when we take the within-bin average loss cost. After taking the within-bin average loss costs that accounts for recency, then the average across bins is calculated without anymore weighting (consistent with the weather weighting approach described in the previous section).

Net Acre Weighting

In this approach, we use the county level net acres insured as weights to account for recency. The county level net acre variable is effectively a “proxy” for recency weights given that it has been increasing over time (from 1980 till the 2000s) (See Figure 3.3). In Table 4.26, we present the national summary results for all crops when net acre weighting is applied. Comparing these results with Table 4.9 (no recency weighting), we find that the average loss costs for barley, corn, potatoes, rice, soybeans, and spring wheat tend to be lower with net acre weighting than without. In contrast, the average loss costs for apples, cotton, sorghum, and winter wheat tend to be higher under the net acre weighting scheme as compared to when there is no net acre weighting.

Decile or Decade Weighting

The main idea for the decile or decade weighting approach is to produce three separate "step-down" weights for the following three decades 1980s, 1990s, and 2000s. In this case, we would like to give more weight to the years in the more recent decades such that: weight for 2000s > weight for 1990s > weight for 1980s. Since the data from 2000s is the most recent, the years in the 2000s will always get a weight of one.

The challenge is to be able to calculate the "declining" weights for the 1990s and 1980s. The construction of the decile weights is done at the state level and it depends on the median loss cost of the middle weather bins for each decade. Given the state level construction, all counties within a particular state have the same decile weights and these decile weights only vary across states. In this case, the weather index development and the loss year classification (i.e. assigning years to bins), as described in the weather weighting approach, has to be conducted at the state level as well.

Using the data from the state-level binning results, the following relationship for the middle bins of each state is examined first: median loss cost for 1980s > median loss cost for 1990s > median loss cost for 2000s. If this relationship holds for a particular state, then all counties in the state will have the following decile weights: weight for 2000s =1, weight for 1990s = median loss cost in 2000s/median loss cost in 1990s, weight for 1980s = median loss cost for 2000s/median loss cost in 1980s. Again, the medians here are calculated for the middle weather bins only (to be able to account for/normalize the weather effect on losses). This would result in declining decadal weights, given that the relationship above holds.

One issue is that there are cases where the relationship described above does not hold. In cases like this, we then explore whether at least the following two relationships hold: (a) median loss cost for 1980s > median loss cost for 2000s or (b) median loss cost for 1990s > median loss cost for 2000s. If (a) holds, then the decile weights are set as follows: weights 2000s = 1 and weights for 1980s and 1990s = median loss cost for 2000s/median los cost 1980s. If (b) holds, then the decile weights are set as follows: weights 2000s = 1 and weights for 1980s and 1990s = median loss cost in 2000s/median loss cost in 1990s. If any of the conditions (the original in the previous paragraph, (a), and (b)) above does not hold, then all years are weighted equally in these states (i.e., weights = 1 for all decades; equal weighting).

Given the approach described above, a "declining" decile weighting scheme is applied when the median loss cost for the most recent decade (2000s) is lower than the earlier decades. If the loss cost for a particular state is trending upward (i.e. loss costs in the most recent decade is higher than previous decades), then equal weighting is applied and the higher loss costs in the most recent years are not given more weight. This is an "asymmetric" weighting scheme as it stands, but a "symmetric" scheme that is applied in cases where loss costs are trending up or trending down can still be implemented.

Application of the decile weighting approach described above for all crops is presented in Table 4.27. Comparing this table with Table 4.9 (no recency weighting), we find that the average loss costs for barley, corn, cotton, potatoes, rice, soybeans, spring wheat, and winter wheat tend to be lower with asymmetric decile weighting than without any recency weighting. In contrast, the average loss costs for apples and sorghum tend to be higher under the asymmetric decile weighting method as compared to when there is no recency weighting. Note that these results may change if a symmetric weighting approach is used (i.e., it is likely that there will be more cases/crops where loss costs under this scheme would be higher than the approach without any recency weighting).

Linear Weighting

In Coble et al. (2010; p. 81), it was suggested that a scheme based on a linear weighting system may be one approach for giving more weight to more recent years of loss cost data. In particular, using a declining weight function of the form: $w_{t-j} = \lambda^{j}$, where w is the weight, and λ is a weighting parameter where $0< \lambda<1$ (note that in our case: time t = 2009 to 1980 and j=0 to 29). This approach provides a "smoother" system for assigning recency weights. This is in contrast to the decile weighting described above where there are abrupt "step-down" weights for each decade.

In Table 4.28, we present the results of a linear weighting scheme where lambda is assumed to be equal 0.8 for all counties. Comparing this table with Table 4.9 (no recency weighting), we find that the average loss costs for barley, corn, cotton, potatoes, rice, soybeans, and spring wheat tend to be lower with the linear weighting scheme than without any recency weighting. In contrast, the average loss costs for apples, sorghum, winter wheat tend to be higher under the linear weighting system as compared to when there is no recency weighting.

One limitation of the linear weighting scheme implemented above is the ad hoc assumption of using 0.8 as the value for lambda. In Table 4.28, we simply assume this lambda value. Hence, the question is whether there is a more empirical approach to assigning lambda. One possible approach is to do a grid-search over the domain of lambda (for example, from 0.99 to 0.01) and picking the value where the out-of-sample prediction error is smallest. This exercise can be done at the state level so that the value of lambda will only vary across states. Note that the linear weighting applied above is a "symmetric" weighting scheme where the declining weights are always applied even in cases where loss costs are trending upwards.

Summary

For the major commodity crops (e.g., corn, soybeans, spring wheat, and cotton), accounting for recency using all the weighting schemes 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 of corn, soybeans, spring wheat and cotton tend to be lowest under the linear weighting or net acre weighting scheme. 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 4.29 below). 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.

Table 4.25. 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%

Table 4.26. 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 where recency is accounted for based on weighting with net acres insured.

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	0.2136085	0.18699	0.1771056	0.1576884	0.1976167	0.1758409
barley	646	0.0992693	0.0956556	0.0748457	0.0689429	0.0891087	0.0830908
corn	1930	0.0392705	0.0469135	0.0261702	0.0281878	0.0335283	0.0382481
cotton	437	0.1605034	0.1459869	0.1203881	0.1129118	0.1428675	0.1320719
potatoes	128	0.0822251	0.0799372	0.0689341	0.0653896	0.0775737	0.0734819
rice	84	0.0197633	0.0226357	0.014755	0.0147192	0.0179126	0.0188639
sorghum	750	0.1407187	0.1351842	0.0969112	0.0937687	0.1218699	0.1171366
soybeans	1523	0.0446129	0.0495869	0.0347107	0.0362827	0.0404842	0.0433171
spring wheat	244	0.1215366	0.1150827	0.0919328	0.0874509	0.1112503	0.1055005
winter wheat	951	0.1073631	0.0886324	0.0783354	0.0681608	0.0931634	0.0791437

Note: These are the national average loss costs across all counties (i.e., liability weighted average) where the insignificance flags and state proxy flags are not equal to one.

Table 4.27. 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 where recency is accounted for using asymmetric decile weighting.

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	0.1845765	0.1767554	0.1516272	0.1469343	0.1729096	0.1660752
barley	646	0.1035774	0.0947588	0.0728447	0.0677464	0.0888289	0.0818292
corn	1930	0.0455134	0.0504807	0.0272444	0.0287956	0.0364903	0.0397437
cotton	437	0.1423415	0.1456929	0.1098623	0.1110354	0.1283831	0.130419
potatoes	128	0.081111	0.0806767	0.065913	0.0649548	0.0746961	0.0733913
rice	84	0.0231644	0.0236532	0.0143203	0.0142576	0.0183361	0.0184296
sorghum	750	0.1210101	0.1318028	0.0888684	0.0923029	0.1081026	0.1141039
soybeans	1523	0.0502049	0.0517504	0.0363647	0.0368126	0.0438882	0.0445343
spring wheat	244	0.1195638	0.1164355	0.0881649	0.0871797	0.1080548	0.1059476
winter wheat	951	0.0979213	0.0852977	0.0720352	0.0656954	0.0850161	0.0760926

Note: These are the national average loss costs across all counties (i.e., liability weighted average) where the insignificance flags and state proxy flags are not equal to one.

Table 4.28. 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 where recency is accounted for using linear weighting with λ = 0.8.

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	0.2717638	0.206555	0.2139327	0.1684872	0.2461658	0.1915181
barley	646	0.0764412	0.0889495	0.0634834	0.0666461	0.0724698	0.0793477
corn	1930	0.0303293	0.0439634	0.0228138	0.0272648	0.0275455	0.0365273
cotton	437	0.152981	0.1380243	0.1059132	0.1066816	0.1284575	0.1246372
potatoes	128	0.0682266	0.0792112	0.0604934	0.065126	0.065762	0.0731629
rice	84	0.0135211	0.0204599	0.0111938	0.0134719	0.0126181	0.0170293
sorghum	750	0.1343515	0.1346023	0.0922405	0.0940785	0.1159699	0.1169317
soybeans	1523	0.0348925	0.0463969	0.0287758	0.0339744	0.0326577	0.0404802
spring wheat	244	0.0920505	0.1073243	0.0749443	0.0839555	0.0870996	0.0995594
winter wheat	951	0.1194645	0.0906294	0.0833433	0.0688864	0.1006851	0.0802436

Note: These are the national average loss costs across all counties (i.e., liability weighted average) where the insignificance flags and state proxy flags are not equal to one.

Table 4.29. 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 where recency is accounted for by using the most recent 20 years of data only.

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	0.1954522	0.1855349	0.1619009	0.1557007	0.184652	0.1759565
barley	629	0.0826886	0.0812677	0.064345	0.0621968	0.0744748	0.0724301
corn	1914	0.0415752	0.0461098	0.0253727	0.0263999	0.0337864	0.0364014
cotton	431	0.1519396	0.1419295	0.1202358	0.1141178	0.1406828	0.1317698
potatoes	127	0.0805161	0.0788176	0.0661329	0.0646112	0.074741	0.0727677
rice	83	0.0215826	0.0227762	0.0150129	0.0145862	0.0186913	0.0186174
sorghum	727	0.1221199	0.1338205	0.0893262	0.0941685	0.1104833	0.119158
soybeans	1512	0.0453196	0.0469253	0.0322739	0.0329237	0.0392736	0.0403107
spring wheat	242	0.1048348	0.1119701	0.0801261	0.0825618	0.0949669	0.1000168
winter wheat	937	0.1020952	0.0891515	0.0786298	0.0700948	0.0923004	0.0816391

Note: These are the national average loss costs across all counties (i.e., liability weighted average) where the insignificance flags and state proxy flags are not equal to one.

5. Recommendations and Assessment of Alternative Approaches

As directed by the statement of work for this project Sumaria has 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. The statement of work also directed us to deliver a report that offers multiple approaches that compare and contrast the varying combinations of the factors based on statistical validity, feasibility, sustainability, and a balance of improvement versus complexity. In our assessment and recommendations we will discuss separately the two primary thrusts of this study -- weather weighting and changing severity of loss cost over time.

5.1 Recommendations

Weather Weighting

We were directed to consider the Palmer Drought Index and other weather variables to control for differences in crop growing conditions. We have conducted analysis nationally for nine crops (apples, barley, corn, cotton, potatoes, rice, sorghum, soybeans, and wheat). These crops represent major crops insured by RMA and specialty crops which have unique weather risks. We believe this provides a robust assessment of weather weighting.

Recommendation 1. – After evaluating various alternative sources of data based on several criteria we believe the weather data collection that best meets these criteria is the National Climatic Data Center's Time Bias Corrected Divisional Temperature-Precipitation-Drought Index data, also call the Climate Division Data.

While more detailed data may be available in some cases it is often for shorter time periods or limited locations. The climate division data provide several drought indexes and other weather variables, time-aggregated to the monthly level and spatially-aggregated to the climate division level for the years back to 1895. Thus, these data will allow RMA to compare the weather experienced under the modern program to weather extending 80 years past the 1975 cut-off of loss cost data.

Recommendation 2. – After evaluating various methods to capture the relationship of RMA loss experience to weather we recommend 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 should be used for different crops and locations. Out-of-sample forecasting competition would be used to select the specific model time-period/variables for a crop/climate division and if the models are not found statistically significant we recommend no weather weighting be performed. This process creates a weather index from 1895-present which ranks the growing conditions experienced in each year.

Fractional logit models account for the frequent occurrence of zeros in the regression dependent variable. We recommend out-of-sample competition for model selection to avoid over-fitting the model. PDSI is recommended because it subsumes effects of both precipitation and temperature and provides a locally relative scale ranging from very wet to very dry conditions. The CDD allows measurement of both excess heat in critical periods and sufficient heat for plant growth over a full season.

Recommendation 3. – Given recommendation 2 we propose that RMA categorize the loss cost experience observed over the period chosen into weather 'probability bins'. These bins would be chosen according to a step-wise procedure which would choose a parsimonious number of bins for the crop/climate division. Once observed loss costs are categorized in bins, all historical loss costs within a bin are given equal probability weight. The bins recommended would have variable width but equal probability.

Essentially a form of creating a non-parametric histogram, this procedure was chosen to avoid estimation of parametric densities for each crop/climate division. Loss costs are bounded between zero and one. They are also frequently right skewed with infrequent but severe upward spikes in some years. The weather index provides a measure of weather probability over 115 years, but RMA experience extends for no more than 35 years. Thus, the longer time period will provide probabilities that may not be observed in RMA experience. The variable width binning process we propose ensures actual observations in all bins.

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 cutoff to 90% and reducing the aggregation region for catastrophic load from the state level to a climate division which is consistent with the weather weighting procedure.

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. There are a variety of factors that suggest non-stationarity in loss cost data. Primary factors we find in RMA data are an expanding participant pool, evolving production systems, and changing program underwriting rules. RMA is confronted with the inherent conundrum of significant weather risk which suggests use of a long data series for rating while accounting for the fact that recent experience is often more representative of the current insurance contract and pool of insureds than older experience.

We examined several approaches including:

1. A discrete adjustment to data prior to 1995
2. A discrete adjustment to data prior to 1995 plus a trend adjustment since 1995
3. Adjusting loss cost based as a function of net acres insured
4. Shortening the loss history for base rates (not catastrophic loads) to twenty years
5. Decadal weights comparing median loss cost bins
6. A linear recency effect
7. Net acre weights within probability bins

All of these approaches have instances where they appear to perform well. The first three procedures require model estimation while the fourth is a procedure that would only slightly alter current RMA practices. We believe all could be made compatible with other RMA procedures and with weather weighting. However, we stress that where statistical analysis indicates non-stationarity in the loss cost history, making no adjustment results in a rate that is not actuarially sound. Ultimately we recommend a combination of option 1, 4 and 7. The discrete adjustment for data prior to 1995 would be applied to the adjusted loss cost data first. Specifically we would estimate the effect at a regional level and calculate a percentage difference by state using the effect relative to the post-1995 average loss cost. Shortening the loss history for base rates to 20 years while using more years for catastrophic loading reflects the recognition that a longer time series is needed to capture extreme events than for the risks quantified by the base rate. Finally, using net acre weighting within probability categories “bins’ recognizes the additional credibility of experience that is based on more exposed acres.

5.2 Implications

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 breakout 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-by-county 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 centered 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%

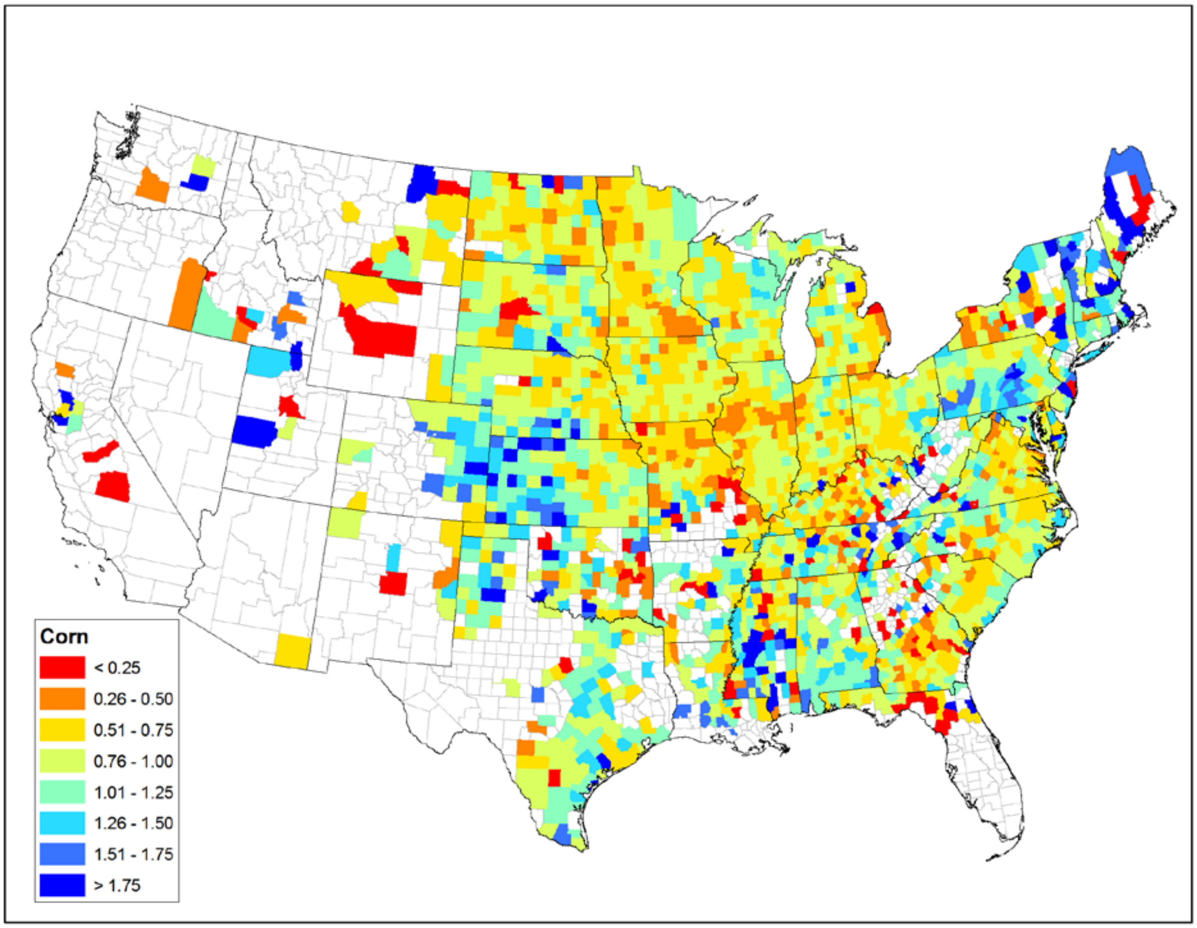
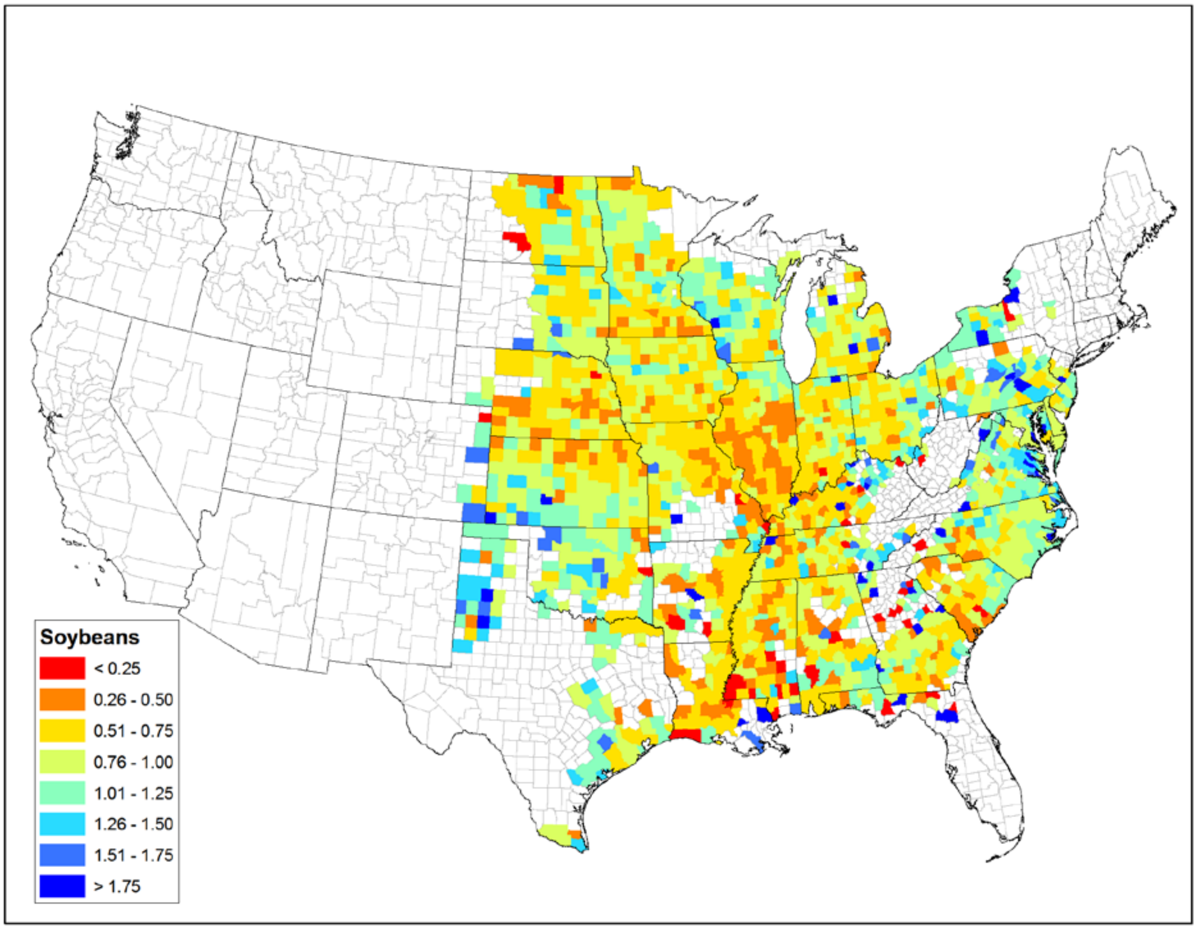


Figure 5.1 county level changes in estimated base rate for corn (508 compliant data is in Appendix D-7).


Soybeans
< 0.25
0.26 - 0.50
0.51 - 0.75
0.76 - 1.00
1.01 - 1.25
1.26 - 1.50
1.51 - 1.75
> 1.75

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.

Wilhelmy, O., K. Hubbard, and D. Wilhite, D. 2002. Spatial representation of agroclimatology in a study of agricultural drought, International Journal of Climatology, vol 22(11), pp. 1399–1414.