

# Modeling Crop Insurance Claims

Nikhil Bhandari  
Rock Creek Analytics, LLC

[www.RockCreekAnalytics.com](http://www.RockCreekAnalytics.com)

[nikhil@rockcreekanalytics.com](mailto:nikhil@rockcreekanalytics.com)

October 8, 2020

## **Abstract**

This article provides an overview of the crop insurance program in the US provided by the Federal Crop Insurance Corporation via a network of private firms. We review the evolution of the program during the past two decades and suggest an approach to model the loss to liability ratio. We show that the Weibull distribution provides a reasonable option to model the loss payments between 0% and 100% of the liability levels, a finding consistent with prior research on insurance claims modeling.

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Data Sources</b>	<b>3</b>
<b>3</b>	<b>General Overview</b>	<b>3</b>
3.1	Premium and Subsidies . . . . .	4
3.2	Losses . . . . .	6
<b>4</b>	<b>Loss Ratio Modeling</b>	<b>8</b>
<b>5</b>	<b>Closure</b>	<b>12</b>
	<b>Appendices</b>	<b>14</b>
<b>A</b>	<b>Inflation Adjusted Premiums and Losses</b>	<b>15</b>
<b>B</b>	<b>Subsidy - Federal, State and Other Programs</b>	<b>17</b>
<b>C</b>	<b>Annual Profiles</b>	<b>18</b>

DRAFT

## 1 Introduction

Crop insurance in the United States is provided by the Federal Crop Insurance Corporation (FCIC) via a network of private firms. FCIC is managed by the Risk Management Agency (RMA) of the US Department of Agriculture [1]. This article provides an introduction to the crop insurance program in the US and provides an overview of policies, liability and claims during the period 2000 - 2020. We also look at the insurance losses and suggest an approach to model these losses.

While the examples and website references provided in this article are US-centric, the ideas presented herein are general and can be applied to all locations. In other regions and countries, the analyst will need to substitute the appropriate regional data sources for crop insurance claims data.

## 2 Data Sources

RMA provides excellent summary level data sets on their website [2]. The analysis presented in this article is primarily based on state / county / crop / coverage level data downloaded from this site by the author in October 2020. The data files are organized for each year and are in *text* file formats that can easily be processed by standard statistical software.

Note that the data from the RMA website is not dis-aggregate data. It is summarized at the level of state, county, crop and coverage level. So we are not able to undertake analysis at an individual policy or an individual claim level. We did request USDA for the policy level data through a Freedom of Information Act request but this request by denied because of prevailing laws and regulations that govern the data.

## 3 General Overview

The US crop insurance program covers over 100 crops and is available to the farmers across the country. FCIC uses several private firms (*Approved Insurance Providers*) to sell and service individual policies. Over the past two decades, the program has had between 1.1 and 1.3 million premium paying policies (see Figure 1.)

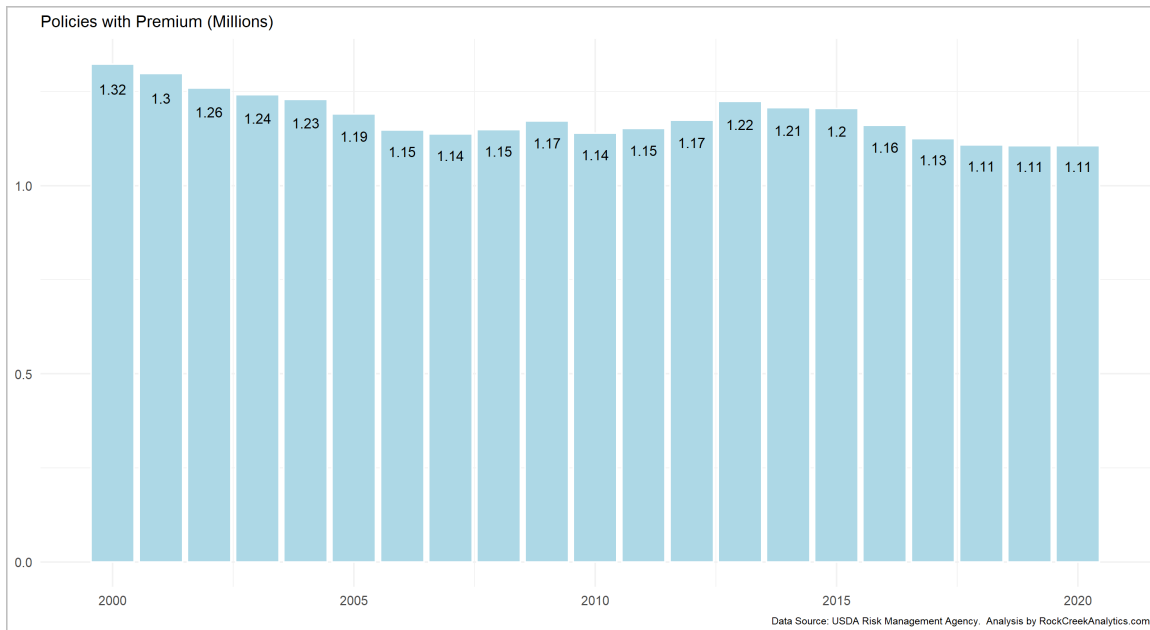


Figure 1: Crop insurance policies with premiums.

### 3.1 Premium and Subsidies

The premiums associated with these policies has been approx. \$10 billion over the past few years (see upper panel of Figure 2). Inflation adjusted premiums are provided in the Appendix.

The program receives a significant subsidy as shown in the lower panel of Figure 2. The subsidy accounts for over 60% of the premiums over the past few years. Note that Figure 2 shows the federal subsidy. Additional subsidies are provided by states, private and other programs but these subsidies are quite low compared to the federal subsidies (details of the subsidy levels are provided in the Appendix). We will disregard these state, private and other subsidies in our analysis. The difference in premiums and federal subsidy can be considered as the “net premium” that the program receives by policy holders.

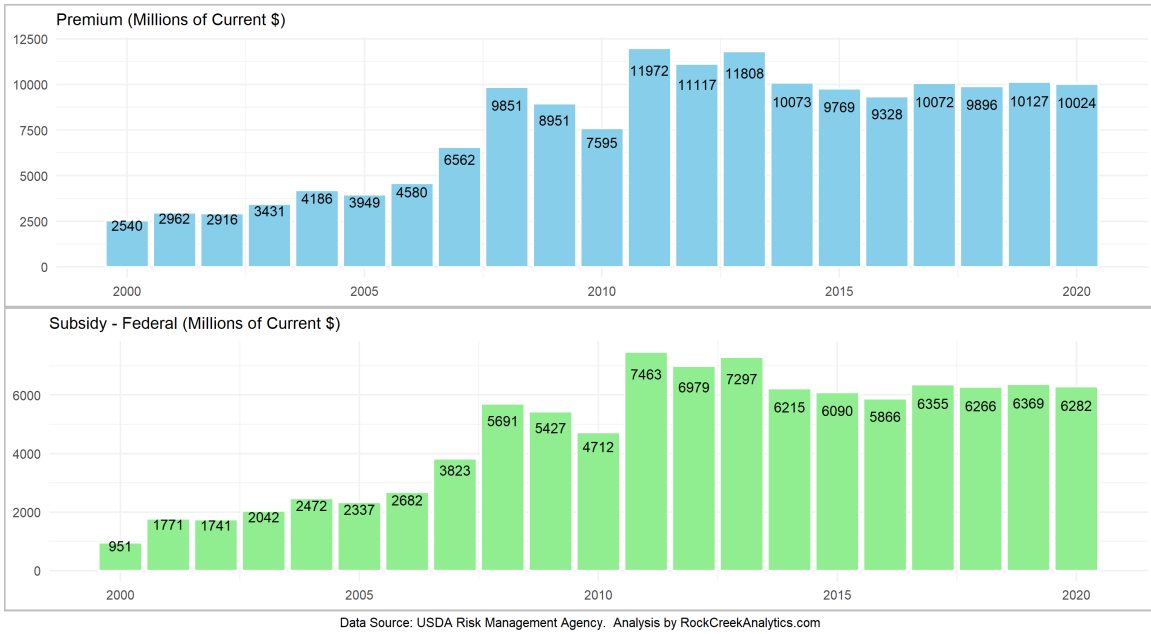


Figure 2: Crop insurance premiums and subsidy.

Four of the over 100 crops that FCIC insures - corn, soybean, cotton and wheat - account for over 70% of the premiums received by the program. The share of the other crops has been steadily growing over the past two decades (see Figure 3.)

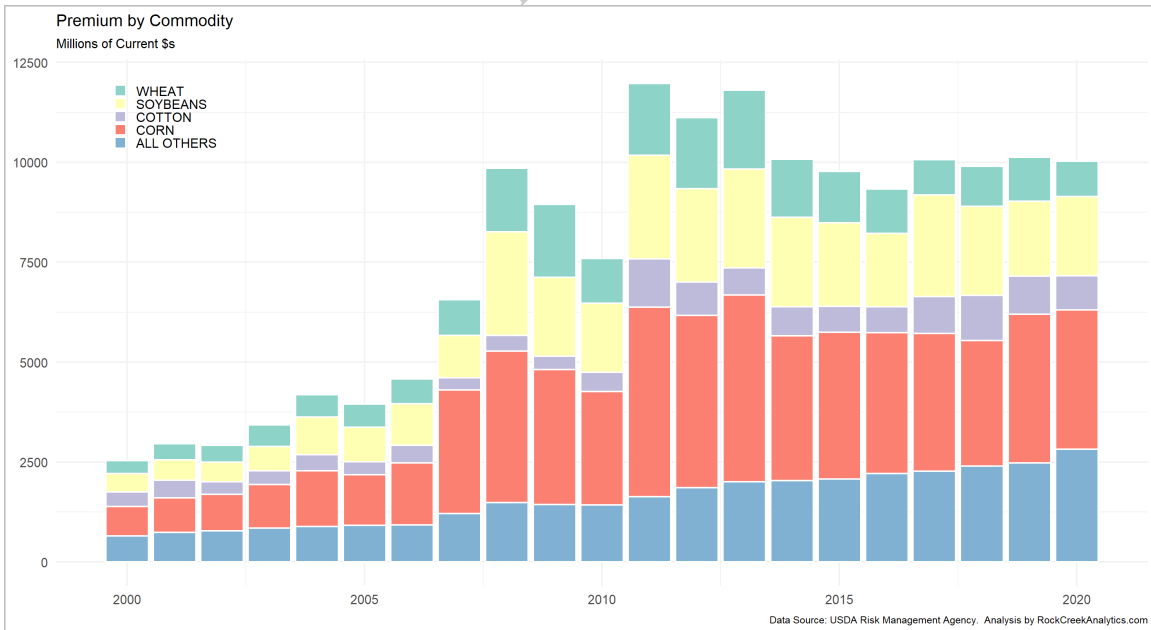


Figure 3: Crop insurance premiums by commodity.

The program is quite widely distributed across the country; Figure 4 shows the premium distribution by county for the year 2020 (similar plots for prior years are provided in the Appendix).

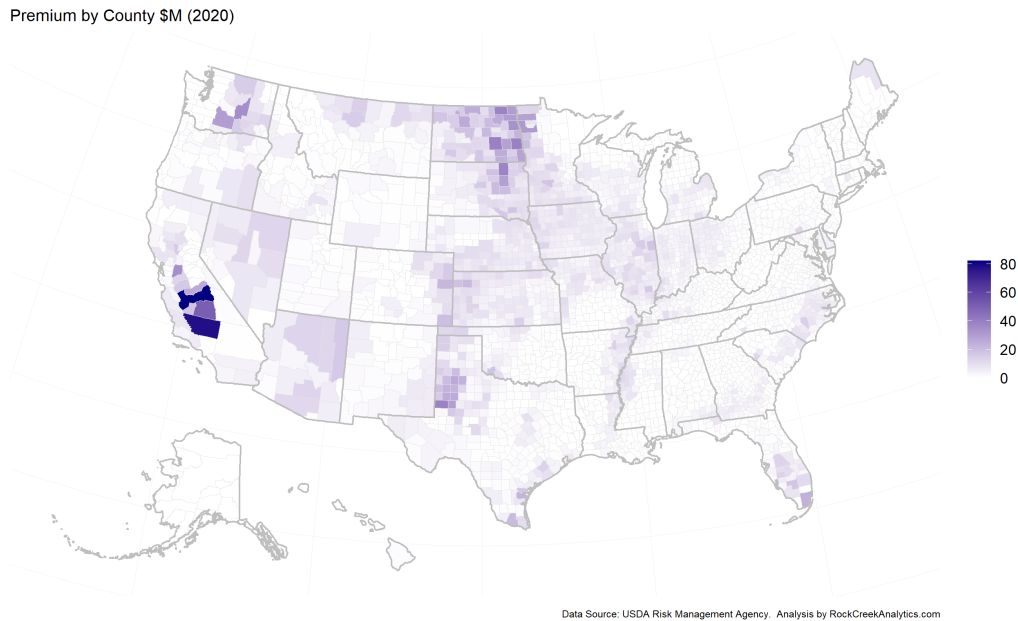


Figure 4: Crop insurance premiums by county (2020).

### 3.2 Losses

As is typical with insurance policies, every year some policy holders file a claim to get indemnified for their losses during the year. These loss amounts (referred to as “Indemnity Amount” by RMA) are shown in Figure 5 by the principal crops and the geographic distribution of the losses for the year 2020 are shown in Figure 6 (prior year plots are provided in the Appendix).

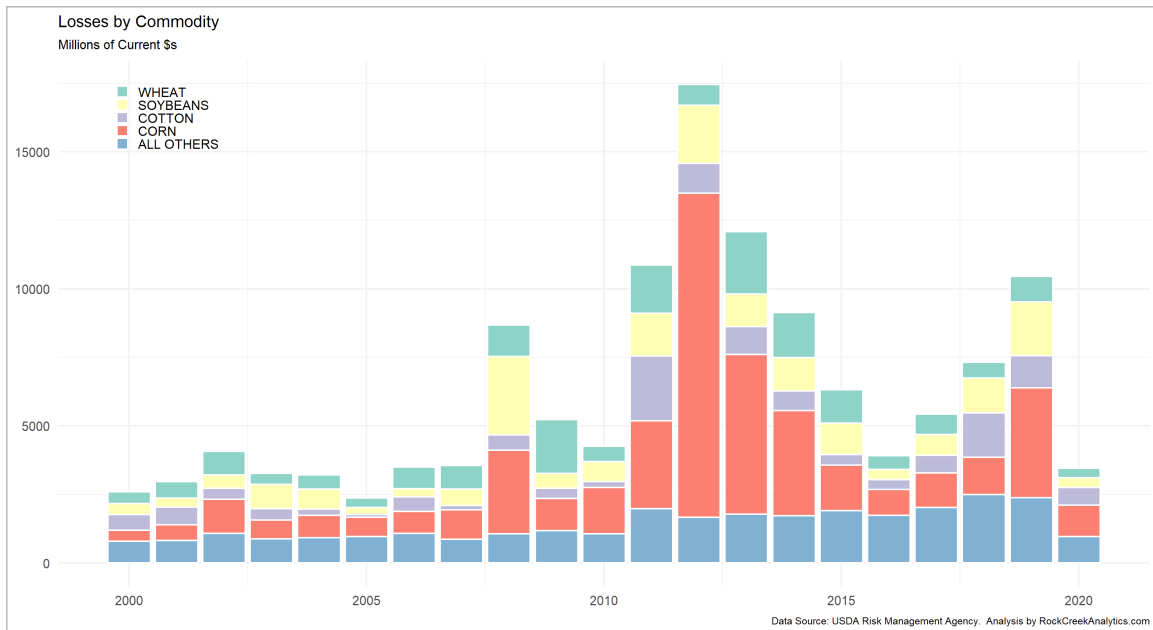


Figure 5: Crop insurance losses by commodity.

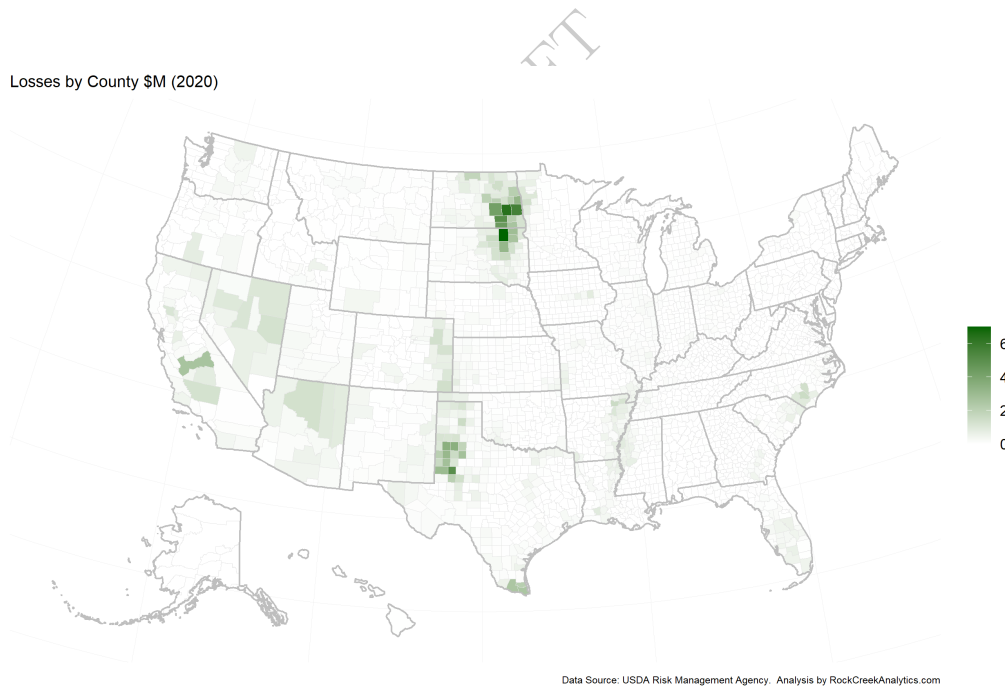


Figure 6: Crop insurance losses by county (2020).

Figure 7 show the losses along side the net premium for the year. The figure clearly shows that in most years the program pays more in losses than what it receives as the net premium.

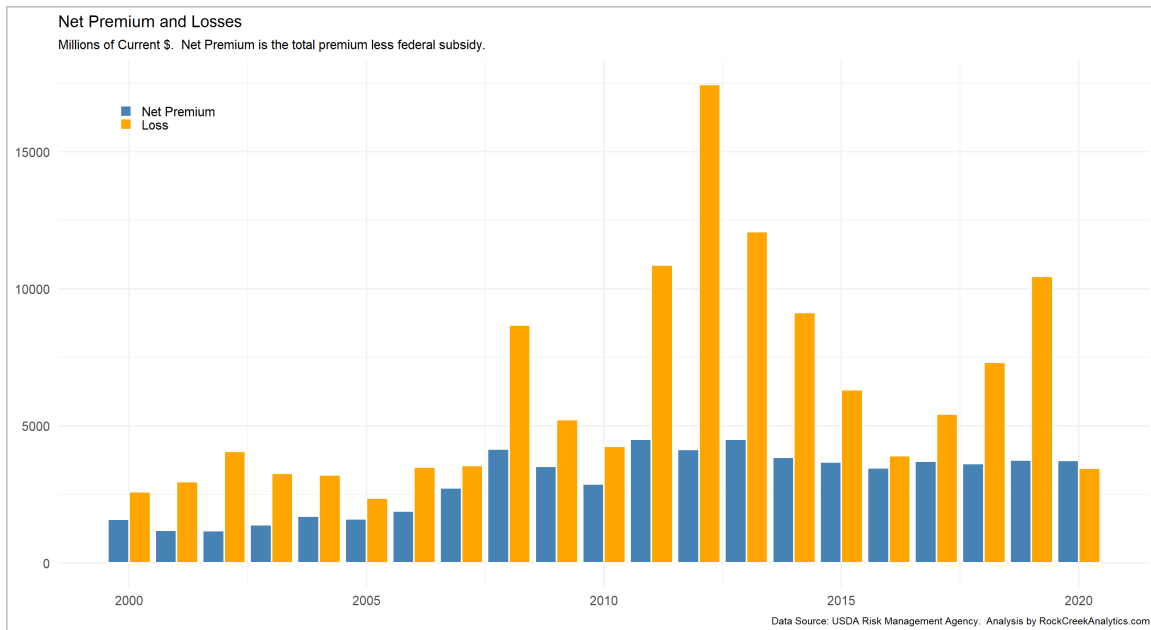


Figure 7: Crop insurance net premium and losses.

## 4 Loss Ratio Modeling

For forecasting losses (claims payments), the ratio of the total loss and total coverage or the total premium is a key component as it defines the amount a claimant is likely to get on their claim. Using a ratio instead of absolute amount allows the analyst to combine data from multiple geographies and from different time periods as it normalizes the coverage levels and effects of inflation.

Figure 8 shows the evolution of the losses as a share of premium (top panel) and liability (bottom panel) and Figure 9 shows the geographic distribution of these ratios for the year 2020. The data for the past two decades indicates that on average losses account for 82% of premium paid or 7.5% of the liability amount.



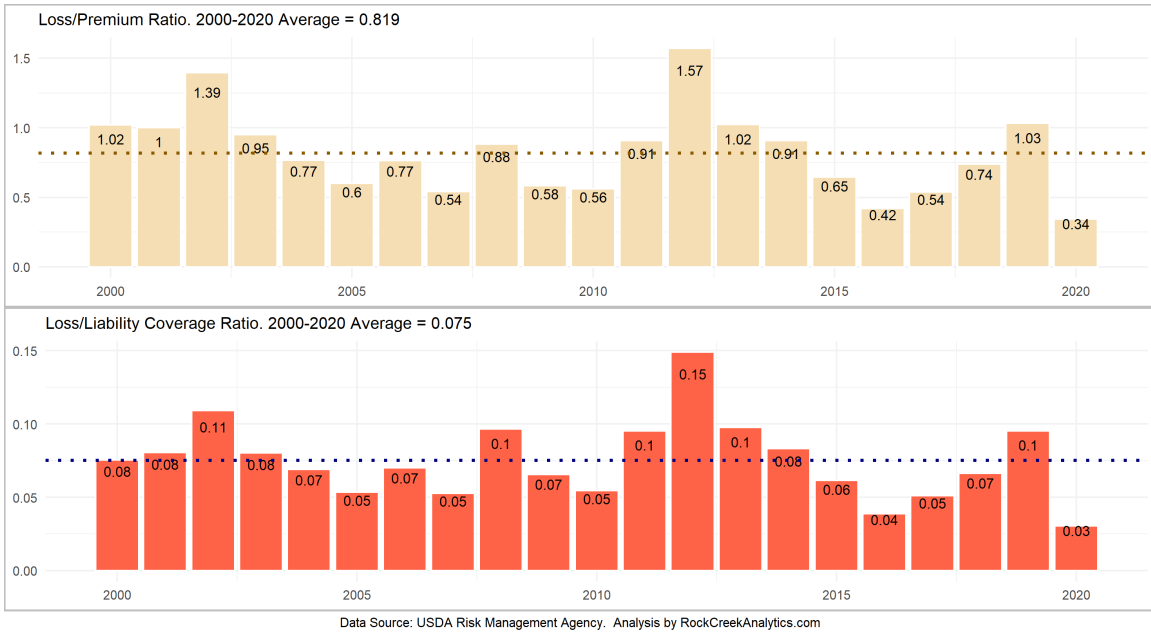


Figure 8: Loss ratio evolution.

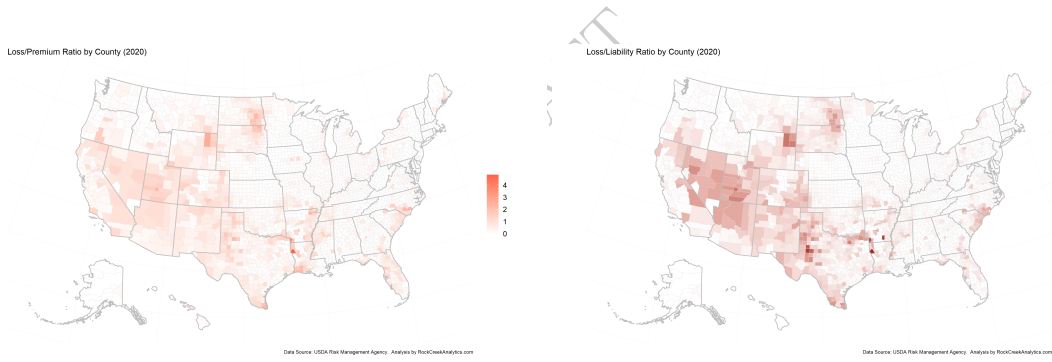


Figure 9: Loss ratios by county (2020).

For modeling the loss ratio, we believe that the loss to liability ratio is a better measure than loss to premium because theoretically the losses can be up to the liability coverage amount that the policy is insured for, and indeed there are individual cases where the policy holder receives 100% of the liability coverage in claims.

As indicated earlier, we did not have access to individual policy or claims data for the modeling task. Instead, we use an aggregated dataset where each data point reflects a combination of state, county, crop, insurance plan name, coverage type and delivery type. For example, for year 2020, there are approx. 140,000 data points reflecting approx. 1.1 million premium paying policies and approx. 187,000 indemnified policies (i.e., losses).

Figure 10 shows the loss to liability ratio for the period 2000 - 2020. There are a large amount of data points with a claim to liability ratio of 0.0 reflecting policies

that did not file claims or claims that were denied. There are several data points with a claim to liability ratio of 1.0 reflecting policies that got 100% of their coverage amount.



Figure 10: Loss to liability ratio distribution (2000 - 2020).

The figure reflects data aggregated by state, county, crop, insurance plan, coverage type and delivery type. It does not reflect individual policies.

Given the particular situation with the aggregate data where the loss to liability of 0% combines both policies that did not file claims and those who got denied the claim, and a *local spike* at the 100% levels, it will not be appropriate to fit one of the standard statistical distributions to the data. One approach to address this is to consider only the data points that have the loss to liability ratio between 0 and 1. The resulting data as shown in Figure 11 follows a more recognizable statistical distribution pattern akin to a log-normal, Weibull or gamma distribution.

After testing some different options, we find that the Weibull distribution fits the data reasonably well (see Figures 11 and 12). Weibull distributions are well studied statistical distributions and used extensively in general insurance analyses for claims modeling, reliability analyses, component lifetime analyses, weather forecasting, hydrology (amount of rainfall, river discharge), etc.

This finding is consistent with the research on insurance claim modeling. Hewitt and Lefkowitz [3] have described the use of five different distributions (gamma, log-gamma, log-normal, gamma+log-gamma, and gamma+log-normal) to fit insurance loss data. Zuanetti *et al.* [4] describe the statistical details of a log-normal model for insurance claims data. Tiwari [5] provides an overview of modeling the claim

frequency using generalized linear models. David and Jemna [6] show how Poisson and negative binomial distributions can be used to model auto insurance claims. Chang *et al.* [7] have suggested the use of Poisson distribution to model the occurrence of individual typhoon/flood events.

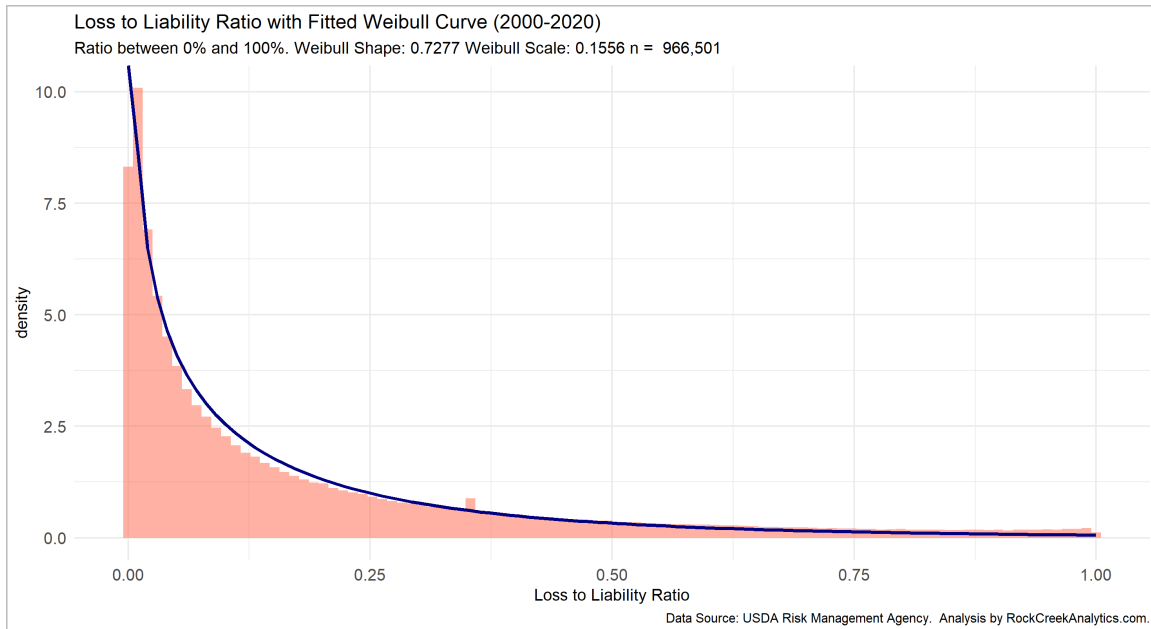


Figure 11: Loss to liability ratio model using aggregate data.

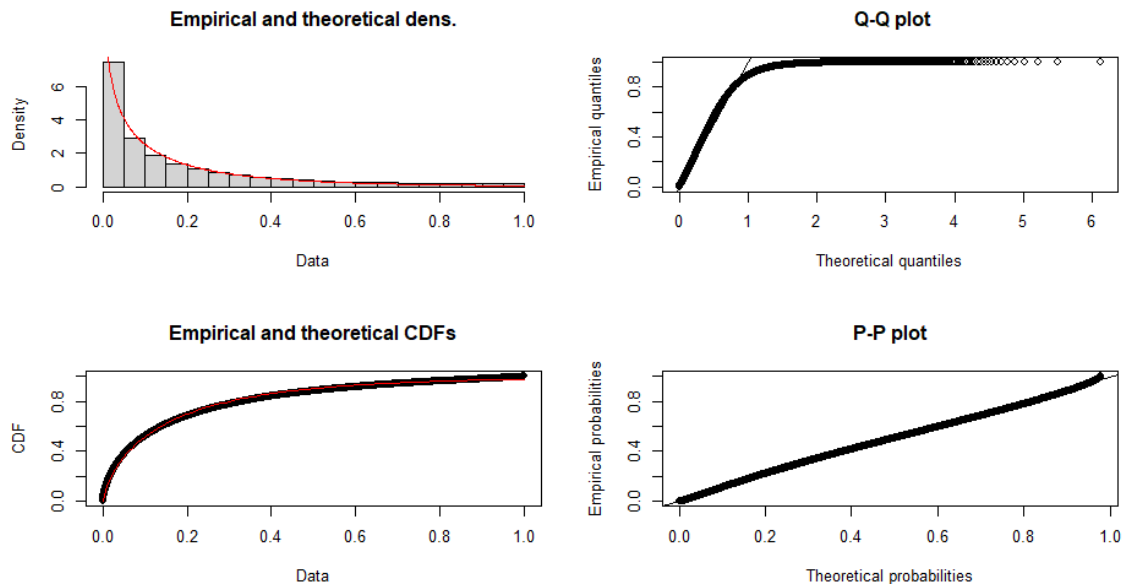


Figure 12: Weibull distribution fit results for loss to liability model.

## 5 Closure

This article has provided an overview of the crop insurance program in the US. We have suggested an approach to model the claims coverage ratio and shown that the Weibull distribution provides a reasonable option to model the claims payments between 0% and 100% of the policy liability levels. While the analysis is based on aggregate dataset, we believe that the results should be applicable to individual policy and claim level data as well.

## References

- [1] *Federal Crop Insurance Corporation*. United States Department of Agriculture - Risk Management Agency. <https://www.rma.usda.gov/FCIC/>.
- [2] *State/County/Crop Summary of Business*. United States Department of Agriculture - Risk Management Agency. <https://www.rma.usda.gov/en/Information-Tools/Summary-of-Business/State-County-Crop-Summary-of-Business>. Accessed October 2020.
- [3] Charles C. Hewitt, Jr. and Benjamin Lefkowitz. *Methods for Fitting Distributions to Insurance Loss Data*. Paper Presented at the November 1979

- Meeting of the Casualty Actuarial Society.  
<https://www.casact.org/pubs/proceed/proceed79/79139.pdf>.
- [4] D.A. Zuanetti, C.A.R. Diniz, and J.G. Leite. *A Lognormal Model for Insurance Claims Data*. REVSTAT – Statistical Journal. Volume 4, Number 2, June 2006.  
<https://www.ine.pt/revstat/pdf/rs060203.pdf>.
- [5] Ajay Tiwari. *Modeling Insurance Claim Frequency*.  
<https://medium.com/swlh/modeling-insurance-claim-frequency-a776f3bf41dc>.  
Accessed September 2020.
- [6] M. David and D. Jemna. *Modeling the Frequency of Auto Insurance Claims by Means of Poisson and Negative Binomial Models*. Scientific Annals of Economics and Business 62(2):151-168. July 2015.  
<https://content.sciendo.com/view/journals/aicue/62/2/article-p151.xml>.
- [7] Ching-Cheng Chang, Wenko Hsu, and Ming-Daw Su. *Modeling Flood Perils and Flood Insurance Program in Taiwan*. 2008 Annual Meeting of the Agricultural and Applied Economics Association.  
<https://ideas.repec.org/p/ags/aaea08/6141.html>.

DRAFT

# Appendices

DRAFT

## A Inflation Adjusted Premiums and Losses

The inflation adjusted premiums (Figure 13), liability (Figure 14) and losses (Figure 15) are shown below. The inflation values used are based on US Consumer Price Index (Series: All items in U.S. city average, all urban consumers, not seasonally adjusted) as shown in Figure 16 from the US Bureau of Labor Statistics.

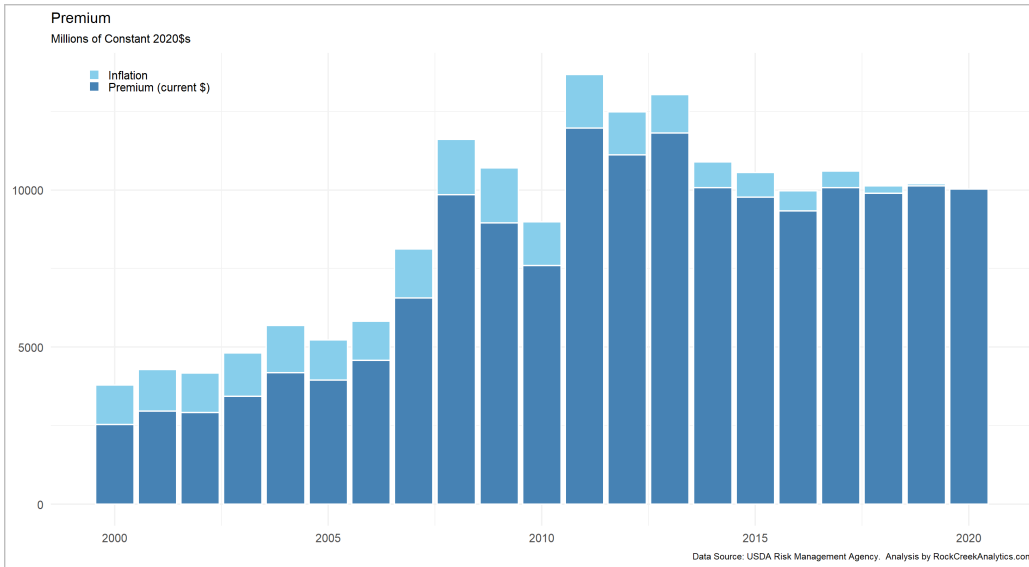


Figure 13: Inflation adjusted premiums.

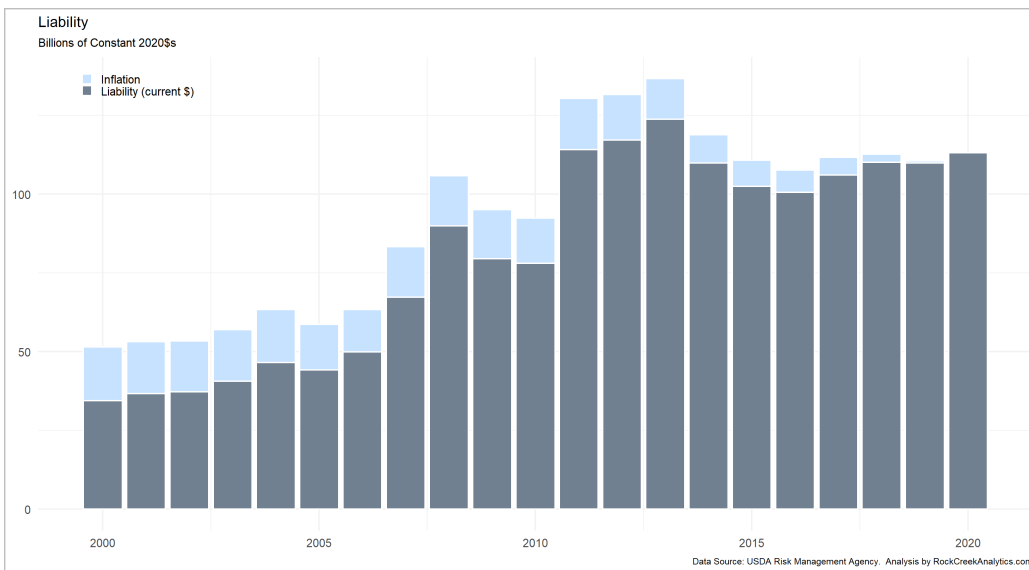


Figure 14: Inflation adjusted liability amount.

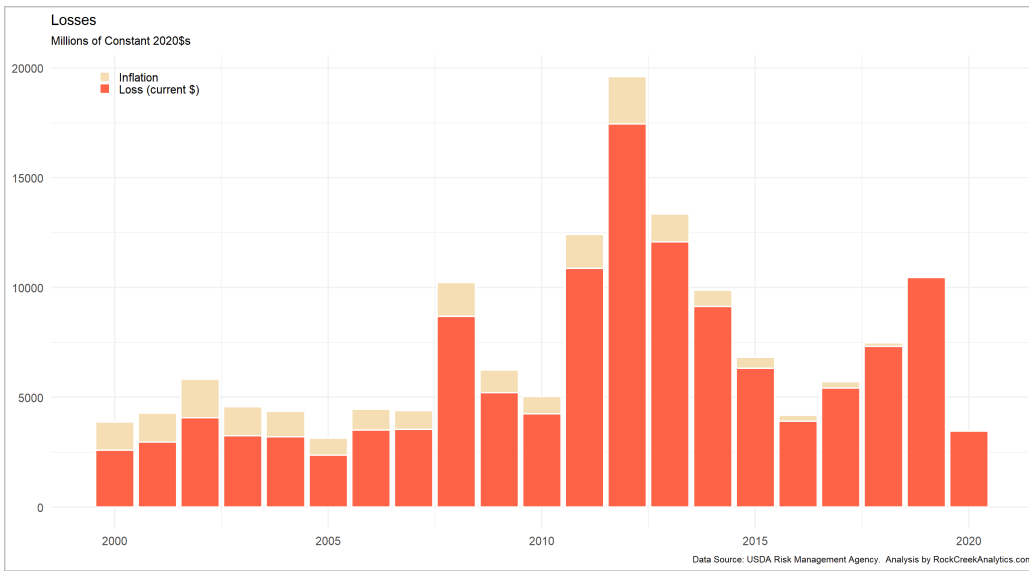


Figure 15: Inflation adjusted losses.

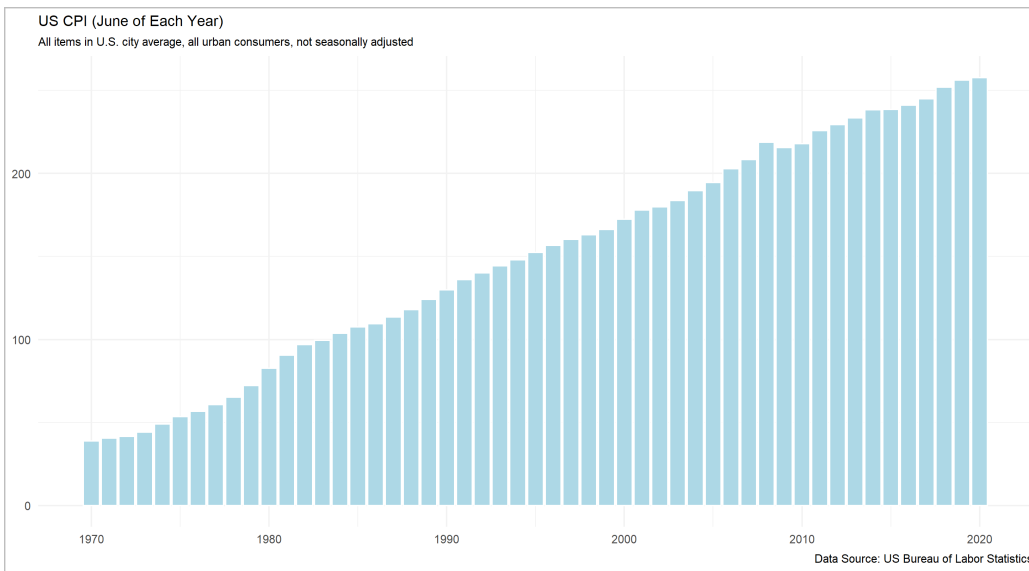


Figure 16: US Consumer Price Index.



## B Subsidy - Federal, State and Other Programs

Subsidy in Millions of Current \$s

Year	Federal	State or Private	Additional
2000	951.19	0.00	0.00
2001	1771.32	0.00	0.42
2002	1741.03	0.00	0.38
2003	2041.66	0.00	0.37
2004	2472.28	0.92	4.22
2005	2337.10	2.62	4.11
2006	2682.01	0.00	0.00
2007	3823.35	0.00	0.00
2008	5690.91	0.00	0.00
2009	5427.41	0.00	0.00
2010	4711.92	0.00	0.00
2011	7463.38	0.00	0.00
2012	6979.36	0.00	0.00
2013	7296.88	0.00	0.00
2014	6215.18	0.00	0.00
2015	6089.77	0.00	0.00
2016	5866.22	0.00	0.00
2017	6355.40	0.00	0.00
2018	6265.93	0.77	0.00
2019	6368.75	0.57	0.00
2020	6282.46	0.87	0.00

Data Source: USDA Risk Management Agency [2].

## C Annual Profiles

DRAFT

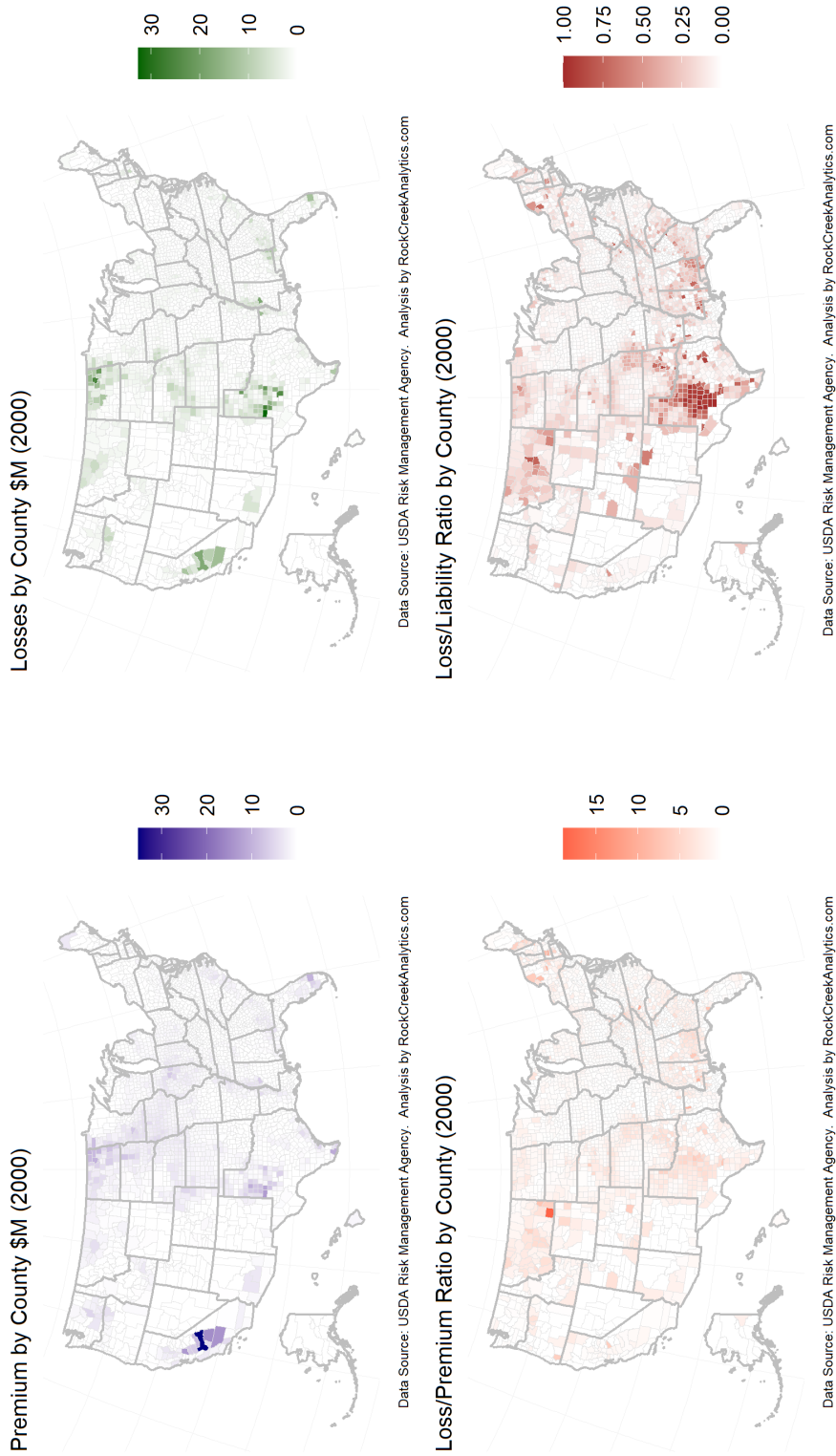


Figure 17: County level premiums, losses and loss coverage ratios (year 2000)

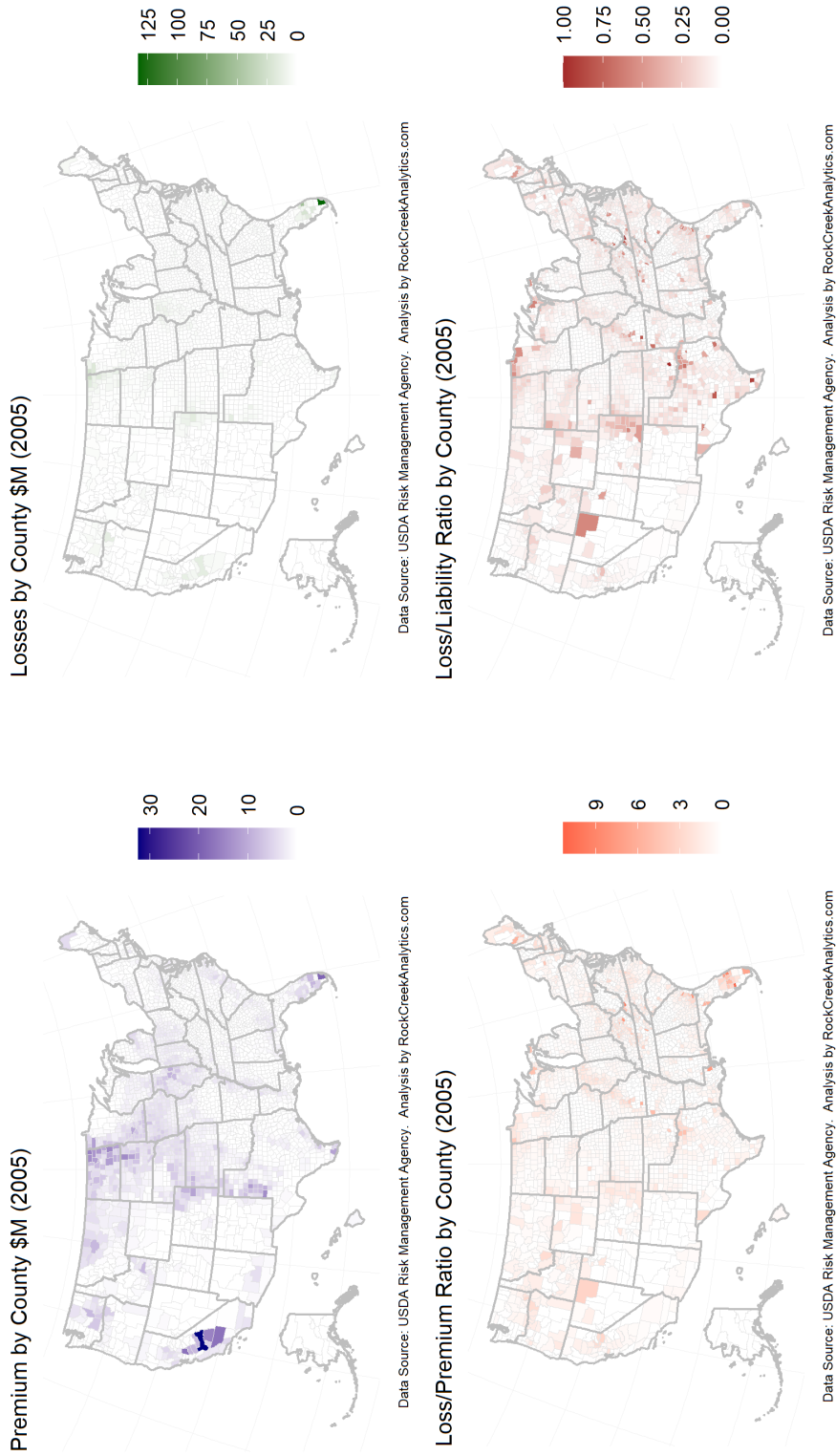


Figure 18: County level premiums, losses and loss coverage ratios (year 2005)

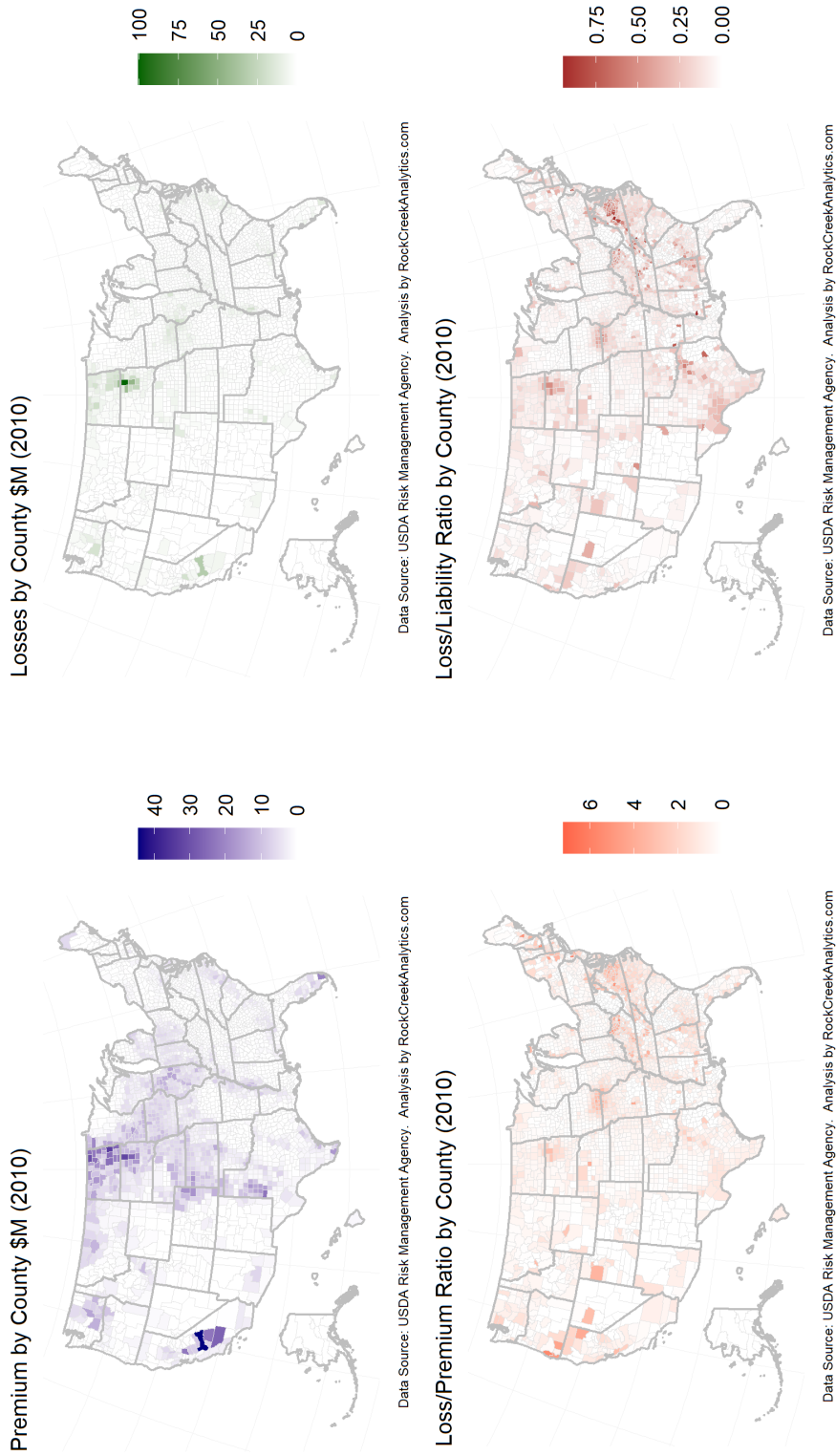


Figure 19: County level premiums, losses and loss coverage ratios (year 2010)

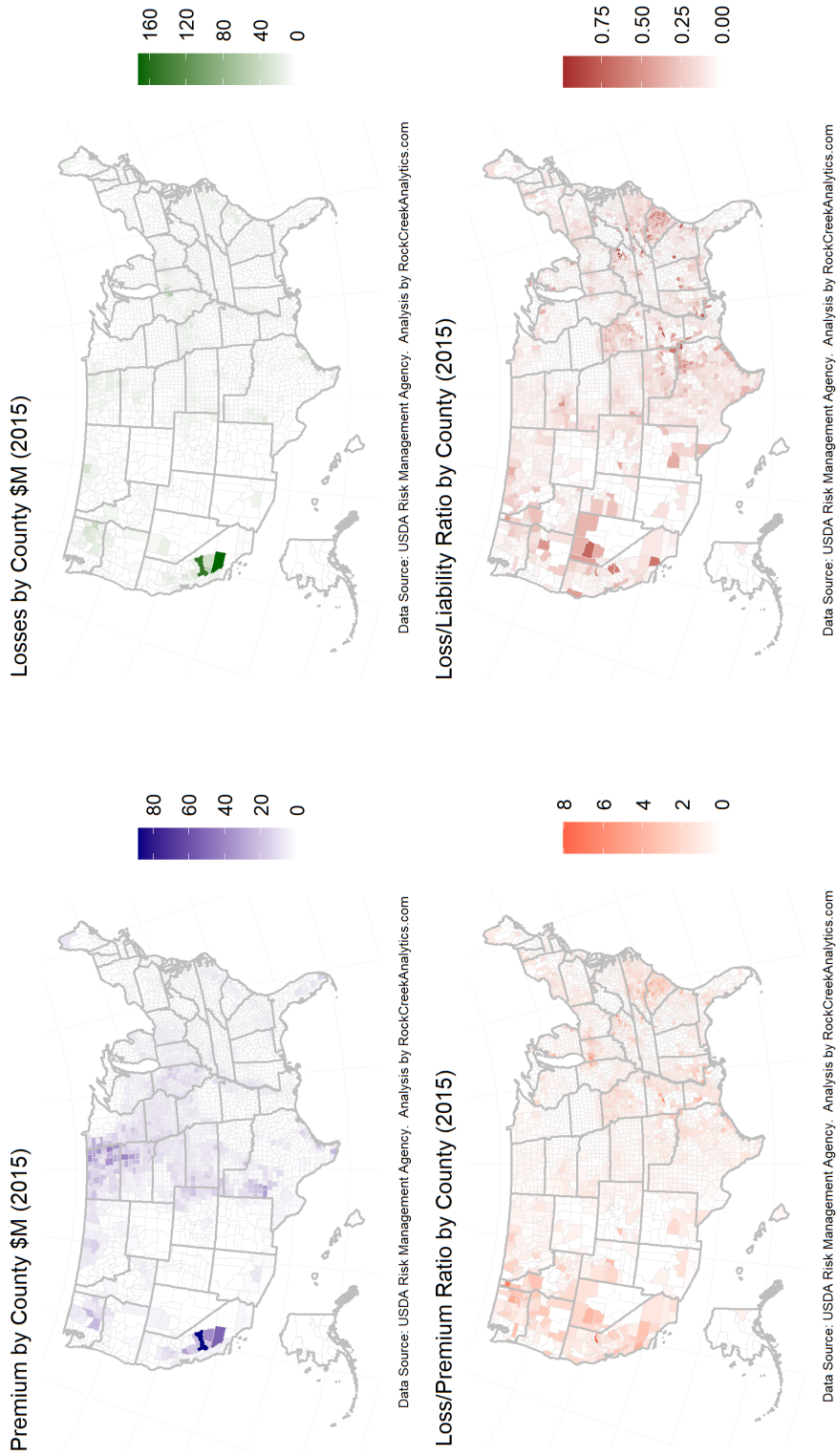


Figure 20: County level premiums, losses and loss coverage ratios (year 2015)

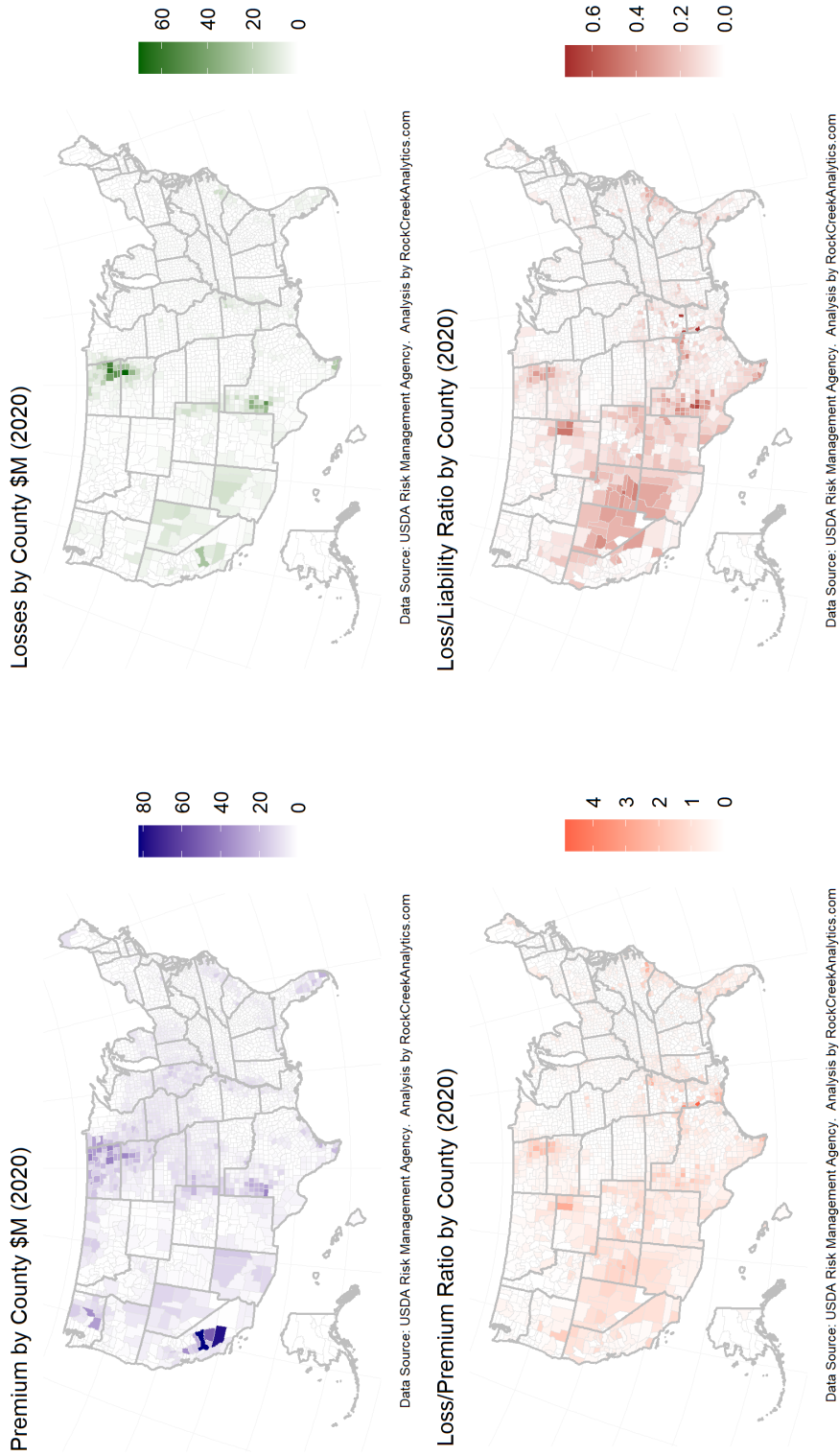


Figure 21: County level premiums, losses and loss coverage ratios (year 2020)