FINAL REPORT: PREMIUM CALCULATIONS

Written By:
Tori White, Conor MacRitchie, Kade Aldrich, Landon Sielaff

Financial Engineering Design Capstone **2023**

Bozeman, Montana
Montana State University

Project Sponsor:
John Kuhling
Watts & Associates

MONTANA STATE UNIVERSITY

MANAGEMENT SUMMARY

The sustainability of the agricultural industry is essential for ensuring food security and economic stability. One of the key components of ensuring the sustainability of this industry is accurately pricing crop insurance premiums for producers. The Financial Engineering capstone team set out to work with Watts & Associates (W&A) to improve the accuracy of Risk Management Agency's (RMA) premium estimates for crop insurance policies, specifically for corn and soybean producers, by utilizing previously unexplored policy attributes.

The project involved analyzing 20 years' worth of data, including Policy Number, Year, State, County, Crop, Type, Practice, Attributes, Premium, and Indemnity, to develop an approach that utilized additional attributes to improve premium estimates. W&A normalized the data to ensure consistency and accuracy. The objective of the project was to accurately bracket producers based on their expected loss ratios, which include <0.5, 0.8, 1.1, and >1.4.

The **approach to this project** involved **leveraging historical producer data from Illinois and Indiana to develop a series of algorithms that could predict premiums for producers in different states**, given the same types of data. The success of the project was measured by comparing the algorithms' in and out of sample performance.

From an ethical perspective, the team was committed to ensuring that the use of additional attributes in determining crop insurance premiums was fair and just. The team understood that any policy changes must take into account the needs and interests of all stakeholders, including small and large agricultural producers, and aim to minimize any discriminatory practices. Additionally, measures were taken to ensure that any data used for this project was anonymized and secure.

The **results** indicated that utilizing additional policy attributes to improve premium estimates and developing algorithms to accurately predict premiums for producers across different states would **lead to improved accuracy in premium estimates**, benefiting both producers and insurers. Moreover, the use of more comprehensive policy attributes could also help to reduce the likelihood of fraudulent claims and limit moral hazard. Accurate premium estimates could **provide incentives for producers to adopt best management practices and reduce risk.** This, in turn, could lead to increased productivity, increased revenue, and reduced financial risks for both producers and insurers.

Based on the findings of this study, it is recommended **that W&A proceed with the development of the algorithms, which could accurately predict premiums for producers across different states.** The business case for this project was to improve the accuracy of RMA premium estimates for crop insurance policies, which could ultimately benefit the entire agricultural industry.

This project had significant implications for the sustainability of the agricultural industry, and it was hoped that the findings would inform policymakers, insurers, and producers alike in their efforts to ensure a secure and sustainable agricultural sector.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

PROJECT OVERVIEW

Background

Watts and Associates, Inc. (W&A) is a prominent economic consulting firm with its headquarters in Billings, Montana. The firm specializes in crop insurance development, agricultural finance, and econometric consulting. W&A has a remarkable track record of working with an array of clients, including the United States Department of Agriculture (USDA), Risk Management Agency (RMA), Agri and AgriFood Canada, World Bank, and World Bank International Finance Corporation (IFC). The firm's primary expertise revolves around providing recommendations for limiting fraud, waste, and abuse within crop insurance programs, with a particular focus on research and development of satellite and weather crop insurance schemes. The senior employees of the firm possess extensive experience in economics, econometrics, database management, and crop insurance.

The project dealt with the RMA and its approach to determining premiums for crop insurance. The RMA is a federal agency that was established by the USDA in 1996 to increase the economic stability of the agriculture industry by providing market-based risk management tools. The RMA is responsible for setting policy premiums, ensuring compliance, and underwriting crop insurance policies. Moreover, the agency collaborates with Approved Insurance Providers (AIPs) and insurance agencies to provide crop insurance policies to agricultural producers (USDA, 2021). They do this with the aim of averaging a net 1:1 ratio of insurance payouts (indemnities) and insurance premiums by both crop and county. This is also known as maintaining a loss ratio equal to 1 (State, 2023).

The objective of the project was to use the expertise of W&A to develop an approach that leveraged historical producer data to improve the accuracy of crop insurance premium estimates, particularly for corn and soybean producers. W&A provided 20 years' worth of data, which would be used to train algorithms that, subsequently, would be capable of accurately bracketing producers by their expected loss ratios, thereby making premium estimates that were more actuarially fair.

W&A is committed to innovation in crop risk management and stabilization of agricultural economies. The firm's interest in working with foreign governments and non-governmental organizations to transfer technical expertise in crop insurance development and agricultural finance underlines its mission of promoting economic stability in the agriculture industry (Watts, 2023).

Project Objectives

The primary objective of this project was to improve the accuracy of crop insurance premium estimates by utilizing additional data attributes and developing algorithms to accurately predict premiums for producers across different states.

An analogy for how the RMA priced crop insurance is that if they sold car insurance, they would estimate premiums solely based on what state and county a person lived in; everyone in a similar geographic area would get the same rate. For instance, a married 44 year old who drove a minivan and had a sterling driving record would pay the same premium as a 19 year old with a sports car and 2 reckless driving charges in the previous 6 months for the same coverage. The goal of this project was to put the '44 year old' in a low risk bucket and the '19 year old' in a high risk bucket. This was because the less risky entity was being charged too much and the more risky entity charged too little for their coverage respective to being actuarially fair. By leveraging historical producer data, our project team intended to enhance the accuracy of crop insurance premiums, creating risk differentiation and avoiding similar situations in the agriculture industry.

The Scrum agile methodology provided the framework for the project. This ensured the consistent timeliness of deliverables and allowed for rapid changes to the project's schedule in case of unforeseen occurrences (Scrum.org, 2023). This framework consisted of seven major "sprints" each with unique goals and deliverables. The first step was to establish team norms, create relationships with the sponsor and advisor, and define the problem to be solved. This was achieved through both in-person meetings and virtual communication with the sponsor and the advisor. The next step was to complete the confidentiality agreement and begin an exploratory analysis of the raw data that was provided by W&A to gain a general understanding of the variables of interest and their relationships. An assumption was made that the legal teams of Montana State and W&A would be able to agree on the terms of the NDA in a timely manner. This allowed the team to gain access to the necessary data for this project that was being provided by W&A.

The subsequent step was to experiment with appropriate mathematical models for each state and crop combination to find which best captured the relationships between the data allocated to the group and the premiums. To develop these models, the team combined skills gained in econometrics, statistics, data structures, algorithms, data mining, and financial engineering courses. These models were compared using both in-sample and out-of-sample performance. Once the best models were selected, they predicted producer premiums and bracketed them based on their perceived risk. Then the observed loss ratios of each group were compared with the predicted loss ratios to determine if the model accurately predicted which observations had higher risk profiles using the attributes of interest. The analysis results indicated that utilizing

additional historical producer data to improve premium estimates and developing algorithms to accurately predict premiums for producers across different states led to improved accuracy in premium estimates, benefiting both producers and insurers.

Stakeholder Needs Assessment

In conducting a stakeholder needs assessment, the student design team for this project has recognized the significance of engaging key stakeholders for better comprehension. An effective solution for this project would involve meeting all of their individual needs. The following stakeholders were identified and categorized into groups depending on if they were affected directly or indirectly by the project.

Direct

- John Kuhling (W&A)
- Montana State University (MSU)
- Dr. Justin Gallager (Primary Advisor, MSU)
- Dr. Sage Kittelman (Director, MSU)
- Dr. Joseph Atwood (Mentor, MSU)
- Student Design Team

Indirect

- Crop Producers
- Risk Management Agency (RMA)
- Agents/Agencies
- Approved Insurance Providers

The primary stakeholder for Watts & Associates (W&A) was John Kuhling, practicing actuary/research analyst, who served as the team's sponsor for this project. Kuhling's main interest was in the final results of the project, which he can utilize to improve the accuracy of crop insurance premium levels.

Other stakeholders in this project included select staff from Montana State University, namely Dr. Justin Gallagher, Dr. Sage Kittelman, and Dr. Joseph Atwood. Dr. Gallagher served as the advisor for this project, while Dr. Kittelman was the professor teaching the Capstone class. Dr. Atwood was a mentor for this project, and their interest was in fully comprehending the project's goal, which enabled them to give relevant and significant advice to the group members.

While Montana State University was a secondary stakeholder for this project, its role was passive as it aimed to uphold the university's reputation. The student design team was also a stakeholder and was responsible for producing a high-quality deliverable for W&A while providing services to John Kuhling throughout the project duration. The student design team was interested in gaining valuable experience through a real-world application of academic material provided throughout the financial engineering undergraduate degree.

Furthermore, agricultural producers who purchase and rely on crop insurance to ensure their living were another stakeholder group in this project. They were interested in making sure that they can provide for themselves and their families through the use of crop insurance. The results of this project could have a direct impact on the premiums paid by the producers. As a result, they may feel that using the attributes to determine their premiums is discriminatory, or they may see it as making crop insurance more efficient.

DATA

The group aimed to improve the accuracy of Risk Management Agency's (RMA) premium estimates for crop insurance policies, utilizing data from two Midwestern states, Illinois and Indiana. Watts & Associates (W&A) was committed to providing a comprehensive dataset that covered 20 years of information to develop an approach that utilized additional attributes to improve premium estimates. The data set included information on a range of variables, including the state and county FIPS codes (represented as stfips, ctyfips and fulfips), the crop year (ranging from 1998 to 2017), the crop type (corn or soybeans), rowID, policy numbers (randomly assigned), insurance plan (with Plan 2 being dominant in the US), coverage level, premium, liability, and indemnity.

Additionally, the group looked at variables, such as the coverage type code, map area dummy, the number of optional units (a subdivision of fields for insurance purposes), the average actual yields reported by the producers, and the average Tflag count (measure of the number of "transitional yields", mock yield values assigned to brand new farmers). They also considered variables such as the average bias (how much they thought yield guarantee was set above or below their true yield), yield bump (increase in yield guarantee allowed by RMA), and enterprise unit dummy (two units far away are insured together). Additionally, the data set included information on the original premium and original indemnity.

Variables of Interest

Stfips: This variable referred to the State Federal Information Processing Standards code for the state where the crop insurance policy was issued. Specifically, stfips 17 represents Illinois, and stfips 18 represents Indiana.

Ctyfips: This variable represented the County FIPS codes for each county within each of the specific states.

Fulfips: This variable represented the State and County Federal Information Processing Standards codes concatenated together. It was used to identify the specific location of the insured crop production.

Crop Year: This variable represented the year in which the crop was produced and insured. The dataset contains crop years ranging from 1998 to 2017.

Crop: This variable represented the type of crop being produced and insured. Specifically, it contained the code for corn (41) or soybeans (81).

RowID: This variable represented a unique identifier assigned to each row or observation in the dataset. This variable was ignored in the analysis as the order was randomly assigned.

Policy Num: This variable was a unique identifier for each crop insurance policy, effectively serving as a random number that was not consistent across years. This variable was ignored in the analysis.

Insurance Plan: This variable represents the type of crop insurance plan that was purchased by the producer. Plan 2 was the dominant insurance plan in the US.

Coverage Level: This variable represented the level of insurance coverage purchased by the producer. The researchers believed that RMA undercharges at an 85% coverage level.

Premium: This variable represented the amount of money an insured party paid to an insurance company in exchange for insurance coverage. In the case of crop insurance, farmers paid a premium to the insurance company to protect their crops against losses due to natural disasters or other unforeseen circumstances. Over the long run, premium collected was expected to equal the indemnities (losses) paid plus provisions for profit and costs associated with administering policies. This variable was normalized across all years to account for inflation and price differentiations.

Liability: This variable referred to the maximum amount of loss that the insurance company was responsible for covering in case of crop damage or loss. This variable was normalized across all years to account for inflation and price differentiations.

Indemnity: This variable represented the amount of money paid to a farmer when the yield or revenue fell below the guaranteed level specified in the insurance policy. It was meant to help farmers recover their losses and mitigate the financial risks associated with crop production. This variable was normalized across all years to account for inflation and price differentiations.

Coverage Type Code: This was a dummy variable that represented the type of crop insurance coverage purchased by the producer. Specifically, it indicated whether the producer elected to insure at a 50% yield election and a 55% price election, or whether the producer opted for a flat fee (1 or 0).

Map Area: This was a dummy variable that represented the risk level of the insured crop production area. Specifically, it identified areas that the RMA considered to be high risk agricultural production areas relative to nearby areas (1 or 0).

Number of Optional Units: This variable represented the number of optional units that the producer had chosen to insure. Optional units are subdivisions of fields for insurance purposes. The researchers had found that producers with higher numbers of optional units tended to be charged higher premiums.

Avg Actual Yields: This variable represented the producer's reported yields on the ten most recent years of crop production at that unit when they signed up for insurance. It is important to note that producers did not always report their previous average actual yields.

Avg Tflag Count: This variable represented the number of transitional yields that the producer was eligible for based on their level of experience. Specifically, new farmers received four transitional yields, while those with one year of experience received three. W&A indicated a positive relationship between the number of transitional yields and the expected loss ratio.

Avg Bias: This variable represented the difference between the yield guarantee and the producer's true yield. A positive bias in the average production history (APH) increased the probability of an indemnity payment.

Yield Bump: This variable represented underwriting rules that allowed producers to increase their yield guarantee. W&A indicated a positive relationship between yield bump and the expected loss ratio.

Enterprise Unit Dummy: This was a dummy variable that indicated whether the insurance policy covered an enterprise unit. An enterprise unit was created when two units are far away but insured together. The researchers had found that enterprise units tended to decrease expected indemnity payment and had a negative relationship with the expected loss ratio.

Original Premium and Original Indemnity: These variables represented the original cost of the insurance policy and the original payment received by the producer in the event of a crop loss, respectively. These variables were ignored in the analysis.

Exploratory Data Analysis

Soybeans and corn were two of the most important crops in the United States, particularly in the Midwest region where Illinois and Indiana are located (USDA, 2022). Together, they accounted for a significant portion of agricultural production in these states, and were major contributors to the economy. Corn was used in a variety of products, including animal feed, ethanol fuel, and food products, while soybeans are primarily used for animal feed and oil production (USDA, 2019; United Soybean Board, 2022). These crops were also significant exports, with the majority of U.S. soybean and corn exports coming from the Midwest (USDA, 2022). Given their importance, analyzing the insurance policies and risk management strategies associated with these crops could provide valuable insights for both producers and policymakers in the region.

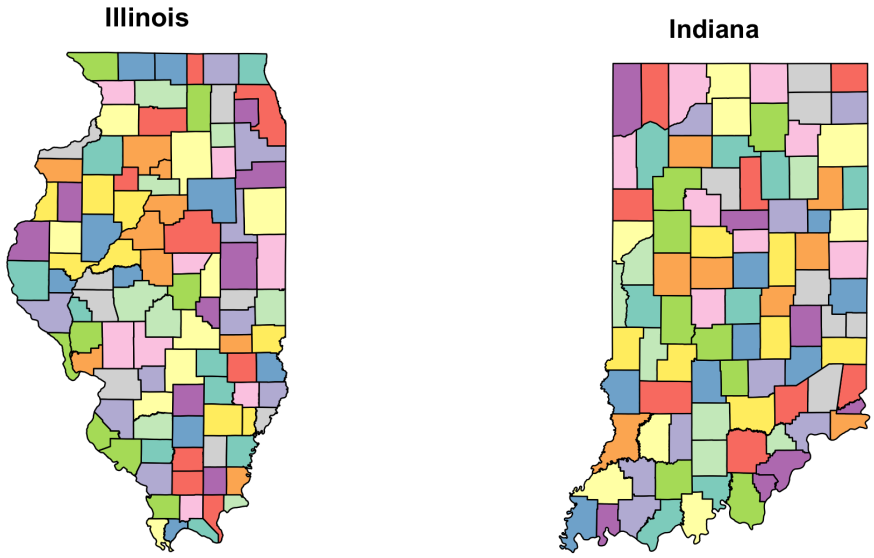


Figure 1: Counties in Illinois and Indiana

Before beginning the project, the team completed a comprehensive exploratory data analysis in order to better understand the data. This data set contained over 2.6 million observations (rows) and 24 variables (columns), resulting in over 63 million data points. Illinois accounted for 74% of the data set with 1,945,325 observations, while Indiana accounted for only 683,568. The split between corn and soy observations was much more balanced with corn accounting for 51% and soy accounting for 49%. In Indiana, the county with the most corn producers was Jasper County, with a total of 11,911 producers. The county with the most soy producers in Illinois was also Jasper County, with a total of 10,766. In Illinois, the county with the most corn producers was Livingston County with a total of 44,513 corn producers. The county with the most soy producers in Illinois was also Livingston County, with a total of 43,380.

To further explore the data, descriptive statistics were computed for each crop. These statistics included the number of observations, mean, standard deviation, minimum, and maximum values for variables such as coverage level, premium, liability, and indemnity. The results showed that for both states and crops, the mean coverage level was around 75% with a standard deviation of approximately 10.4%. The mean premium for corn was around $8125, with a standard deviation of $16,432, while the mean liability was $123,834 with a standard deviation of $208,847. The mean premium for soy was $3,688 with a standard deviation of $7,568 while the mean liability was around $71,520 with a standard deviation of $107,889. The mean indemnity for both crops was calculated to be $5,388 and $1,527, for corn and soy respectively.

Throughout this project, the most important variable of interest was loss ratio, which is indemnity divided by premium. In each of the states, an analysis of average loss ratio and count of producers insured in each state was conducted. The mean loss ratio for corn across both states was around 0.538, with a standard deviation of 2.26, while the mean loss ratio for soy across both states was around 0.423, with a standard deviation of 2.38. Additionally, heat maps were created to visualize the differences in indemnities between each of the counties. The heat maps showed that there was a positive relationship between more southern counties and higher loss ratios. These heat maps provided a quick and easy way to identify potential trends and relationships in the data, which could then be further explored and analyzed.

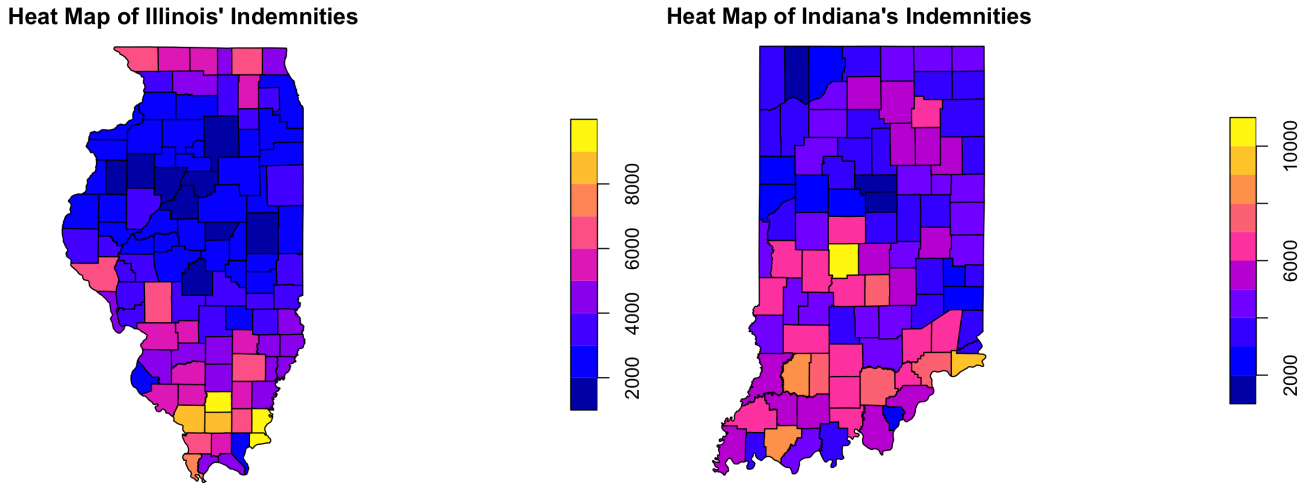


Figure 2: Heat maps of indemnities by county

Overall, this exploratory analysis provided insights into the characteristics of the data, and set the foundation for further analysis and modeling. By understanding the state and crop specific characteristics of the data, the team could better understand the factors that contribute to insurance premiums, liabilities, and indemnities, and make more informed decisions about risk management and crop insurance policies.

METHODS

Approach

Watts and Associates requested a model that would be capable of predicting crop insurance premiums, and thus, loss ratios given certain farm-level data attributes. The loss ratios were very heavily distributed at zero because most crops insurance plans did not pay out any indemnity. The distribution was right-skewed because when there was an indemnity it was often quite large compared to the premium which made some loss ratios extremely large. Premiums had a slightly more normal distribution, however, it was also very right skewed as a majority of premiums were between 0 and 2000 but there were a few that were greater than 500,000. Due to these differences in distributions, it was decided that crop insurance premium would be the dependent value that the models attempt to predict using the other variables. The log-transformed premium had the most normal distribution so models were made using both premium and log-transformed premium as the dependent variable.

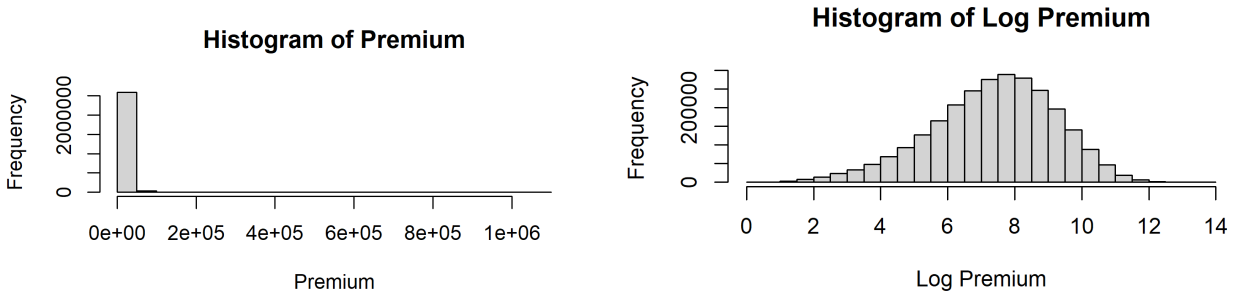


Figure 3: Distribution of premium and log-tranformed premium values

A preliminary analysis of the data was conducted to gain an understanding of the relationships between the potential predictor variables and premium. These relationships were used to determine which variables might be appropriate to include in the models.

Model Building

A series of statistical multiple linear regression models were created using R and trained with a subset of the historical data. The original data set was split into a training subset that included all years but one and a testing subset that included the remaining year. All of the categorical variables were mutated into factors in R so the function that was used to train the models would recognize them as categorical variables and treat them as such. The models were trained and tested with every year combination to compare model performance across years. The goal of these models was to identify relationships between the variables in the model and crop insurance premium. The models differed in which variables were included and whether data transformations were used.

The first model was an overall model that included state fips, county fips, crop, insurance plan, coverage level, liability, a coverage type code dummy variable, a map area dummy variable, number of optional units, average actual yields, average T flag count, average bias, yield bump, and an enterprise unit dummy variable as predictor variables and premium as the response variable. This model was considered the full model and all other models that were created were simplifications of this model.

The second model was a state specific model that included all the variables in the overall model except for the state variable. This model was trained using subsets of the data set that only included observations in the same state. Therefore, for each year's training data set, there was an Indiana version and an Illinois version of this model that was created.

The third model was a crop specific model that included all the variables in the overall model except for crop. This model was trained using subsets of the data set that only included observations that were producing the same crop. There was a corn version and a soy version of the model trained for each year combination.

The final model was a state and crop specific model that included all the variables in the overall model except for state and crop. This model was trained using subsets of the data set that only included observations that were in the same state and were producing the same crop. There was an Illinois corn version, an Illinois soy version, an Indiana corn version, and an Indiana soy version of this model created for each year combination.

All of these models were also created using the same variables except liability and premium were log-transformed. This transformation was performed due to the shape of the distributions of both variables and due to the results from the models built on the original scale. The distributions of premium and liability were quite similar so both were log-transformed.

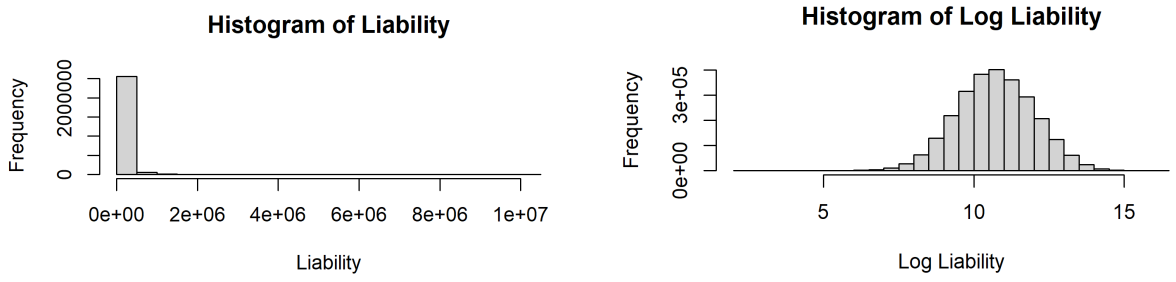


Figure 4: Distributions of liability and log-tranformed liability values

Model Testing

The models were put through an in-sample and out-of-sample testing process to compare the performance of each model and determine which model was most effective. The in-sample testing process consisted of two parts. The first part was using adjusted R-squared as a goodness-of-fit measure to determine how well each model fit the training data set after adjusting for the complexity of the model. The second part was a graphical analysis conducted with the following steps:

1. Use the model to predict premiums for each observation in the training subset of the data based on their attributes
2. Calculate the loss ratio for each of these observations by dividing the observed indemnity by the predicted premium
3. Separate these observations into risk groups based on their predicted loss ratio with the following cut off points <0.5, 0.8, 1.1, >1.4
4. Create a boxplot of the predicted loss ratios separated by risk group
5. Graph a line plot of the mean actual loss ratios for each group
6. Compare the boxplot to the line to determine if the mode accurately grouped the observations into the correct risk group

The out-of-sample testing process was also made up of two parts. The first part involved calculating the Mean Squared Error of the model with the following steps:

1. Use the model to predict premiums for each observation in the testing subset of the data based on the observation's attributes
2. Calculate the loss ratio for each of these observations by dividing the observed indemnity by the predicted premium
3. Calculate the Mean Squared Error by comparing the predicted loss ratio and the actual loss ratio for each of these observations

The second part was a graphical analysis conducted with the following steps:

1. Use the model to predict premiums for each observation in the testing subset of the data based on their attributes
2. Calculate the loss ratio for each of these observations by dividing the observed indemnity by the predicted premium
3. Separate these observations into risk groups based on their predicted loss ratio with the following cut off points <0.5, 0.8, 1.1, >1.4
4. Create a boxplot of the predicted loss ratios separated by risk group

5. Graph a line plot of the mean actual loss ratios for each group
6. Compare the boxplot to the line to determine if the model accurately grouped the observations into the correct risk group

Model Selection

When testing the models, priority was given to out-of-sample performance to ensure the final model could accurately estimate the risk level of agricultural producers that were not in the training data set. This meant that avoiding overfitting the model on the training dataset was paramount. Therefore, the average MSE was the metric primarily used when comparing the models and eventually selecting the final model. The model with the best distribution of MSE was selected as the final model. In the event of two models having similar MSE distributions, the principle of parsimony was used and the simpler model was selected.

Model Validation

After the final model was selected, a model validation process was completed to ensure the final model was valid across all states, crops, and years. This model validation process was completed using the following steps:

1. Train the model using a subset of the original data set for each possible state, crop, and year combination (ex Illinois corn trained using 1998 - 2016)
2. Calculate adjusted R-squared for each model to ensure all models fit the training data set well
3. Use the model to predict premiums for each observation in the testing subset of the data based on their attributes
4. Calculate the loss ratio for each of these observations by dividing the observed indemnity by the predicted premium
5. Calculate the in-sample and out-of-sample Mean Squared Error of the loss ratios using the same process as above for each model to ensure they all perform well out-of-sample
6. Complete graphical analysis both in-sample and out-of-sample for each model to ensure the model is accurately grouping observations into the correct risk group for all state, crop, and year combinations

RESULTS

Results of Models using Untransformed data

The first models built and trained were the models that used the untransformed data. These models performed decently well in-sample with adjusted R-squared values between 0.7 and 0.9. However, these models performed very poorly out-of-sample, predicting both extremely large positive and negative loss ratios. This was because the output of linear regression models that use an untransformed variable are not bounded, allowing for that output to be negative or positive. The extremely large values manifested because the predicted premium was being used as the denominator of loss ratio. The models were predicting some premiums to be extremely small (0.01) and when an indemnity was divided by these very small premiums, it resulted in an extremely large loss ratio (500,000). Using a log-transform on the dependent variable could mitigate both of these issues so the models were rebuilt using the log-transformed premium.

Results of Models using Log-transformed data

The models that were built and trained using the log-transformed premium and liability performed better in-sample than the untransformed models. The models had adjusted R-squared values between 0.9445 and 0.9850 which were much higher than those found for the untransformed models. They also exhibited much better out-of-sample performance as they no longer predicted negative or extremely large values for premiums. When testing, the models were first used to predict log-premiums which were converted back to the scale of US Dollars before calculating the loss ratios.

On average, the state crop model performed the best in regards to both in-sample and out-of-sample performance measures. This model fit the training data subset the best after accounting for the complexity of the model. It was also able to predict loss ratios out-of-sample better than the other model types. This is evidenced by the table below:

Model Type	Mean Adjusted R-Squared	Out-of-Sample Mean MSE
Overall	0.9490	1.4844
State	0.9719	0.6027
Crop	0.9508	1.4053
State Crop	0.9747	0.4713

Table 1: Model performance measures by model type

This state crop model was able to accurately predict which risk group an agricultural producer was in which can be seen in the graph below. The boxplot shows the distribution of predicted loss ratios in each of the five risk groups. The line plot shows the actual mean loss ratio for each risk group. The line plot has an upward trend from group 1 to group 5 which shows that the risk groups that were made using the final model accurately classified the producers' risk level using their attributes. The graphs for every year, crop, and state combination can be found in appendix A. All of these graphs show similar results which shows that the final model is valid across years, states, and crops.

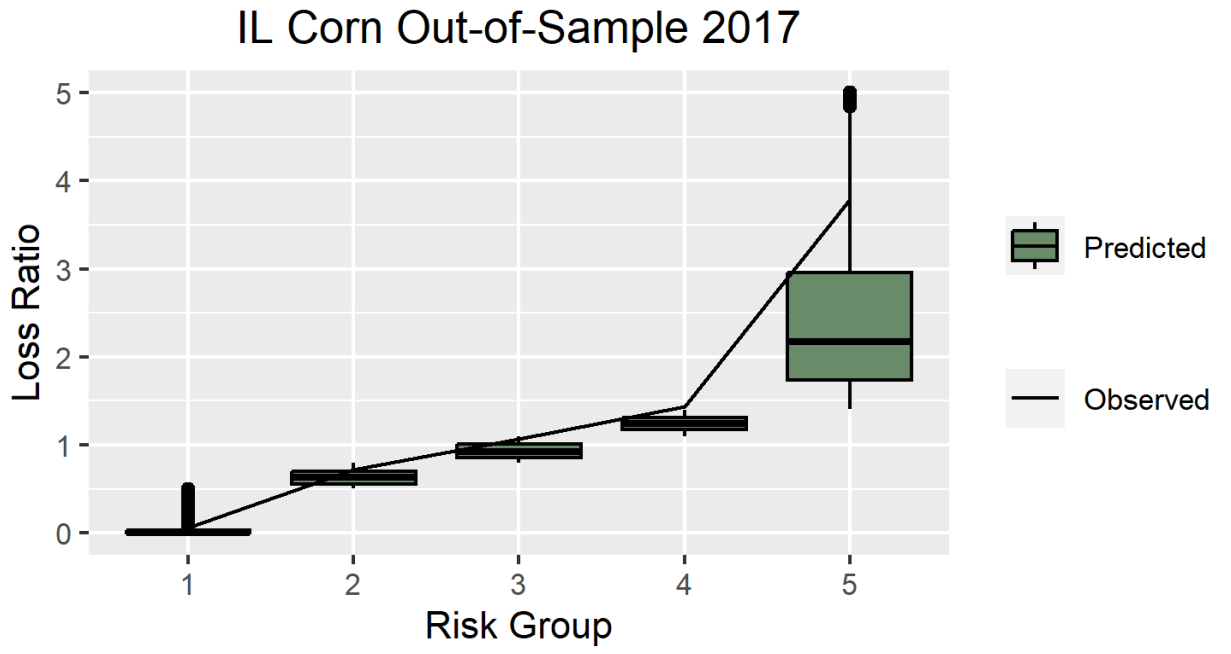


Figure 5: Loss ratio vs risk group for the final model trained on 199 - 2016 IL corn data and tested using 2017 IL corn data

The model validation found that the final model was slightly more accurate in predicting loss ratios for Indiana than Illinois and slightly more accurate in predicting loss ratios for corn than soy. The mean in-sample and out-of-sample MSEs are greater for soy and Illinois producers as can be seen in the table below. However the final model was found to be able to accurately place producers in the appropriate risk group for every state, crop, and year combination so the model is still valid for soy producers in Illinois.

State and Crop	Mean Adjusted R-squared	Mean In-Sample MSE	Mean Out-of-Sample MSE
IL Corn	0.9718	0.4669	0.4807
IL Soy	0.9640	0.6814	0.7560
IN Corn	0.9834	0.2475	0.2808
IN Soy	0.9796	0.2979	0.3675

Table 2: Final model performance measures for each state and crop combination

Final Recommendation

It is recommended that Watts & Associates utilize the state crop specific log-transformed model to predict agricultural producers' actuarially fair crop insurance premiums. This is because of the relative prowess it showed in out-of-sample testing and thus prediction. In order to accurately predict those premiums, the model should be trained only using data stemming from one State and one crop.

The final model was a multiple linear regression model that used the log-transformed *Premium* as the independent variable and used county fips (*Ctyfips*), *Crop Year, Insurance Plan*, *Coverage Level*, log-transformed liability (*ln(Liability)*), a coverage type code dummy variable (*Coverage Type Code)*, a map area dummy variable (*Map Area*), average actual yields (*Avg Yield*), average T flag count (*Avg Tflag Count*), average bias (*Yield Bias*), yield bump (*Yield Bump*), and the enterprise unit dummy variable (*Enterprise Unit Dummy)*. It is depicted below:

$$\ln(Premium) = \beta_0 + \beta_1 Ctyfips + \beta_2 Crop\ Year + \beta_3 Insurance\ Plan + \beta_4 Coverage\ Level + \beta_5 \ln(Liability) + \beta_6 Coverage\ Type\ Code + \beta_7 Map\ Area + \beta_8 Avg\ Yield + \beta_9 Avg\ Tflag\ Count + \beta_{10} Yield\ Bias + \beta_{11} Yield\ Bump + \beta_{12} Enterprise\ Unit\ Dummy$$

REFERENCES

Scrum.org. “What Is Scrum?” Scrum.Org, 2023, www.scrum.org/resources/what-scrum-module.

State of Michigan. “What Is a Loss Ratio?” SOM - State of Michigan, 2023, www.michigan.gov/difs/news-and-outreach/faq/insurance/health-coverage-rate/questions/what-is-a-loss-ratio.

United States Department of Agriculture . “About the Risk Management Agency.” RMA, Aug. 2021, www.rma.usda.gov/en/Fact-Sheets/National-Fact-Sheets/About-the-Risk-Management-Agency.

Watts and Associates. “Agricultural Risk Management.” Watts and Associates, 2 Mar. 2023, wattsandassociates.com/.

U.S. Department of Agriculture, National Agricultural Statistics Service. “Corn and Soybean Production up in 2021, USDA Reports.” *USDA*, 12 Jan. 2022, www.nass.usda.gov/Newsroom/2022/01-12-2022.

United States Department of Agriculture. “Corn is America’s Largest Crop in 2019.” *USDA*, 29 July 2019, www.usda.gov/media/blog/2019/07/29/corn-americas-largest-crop-2019.

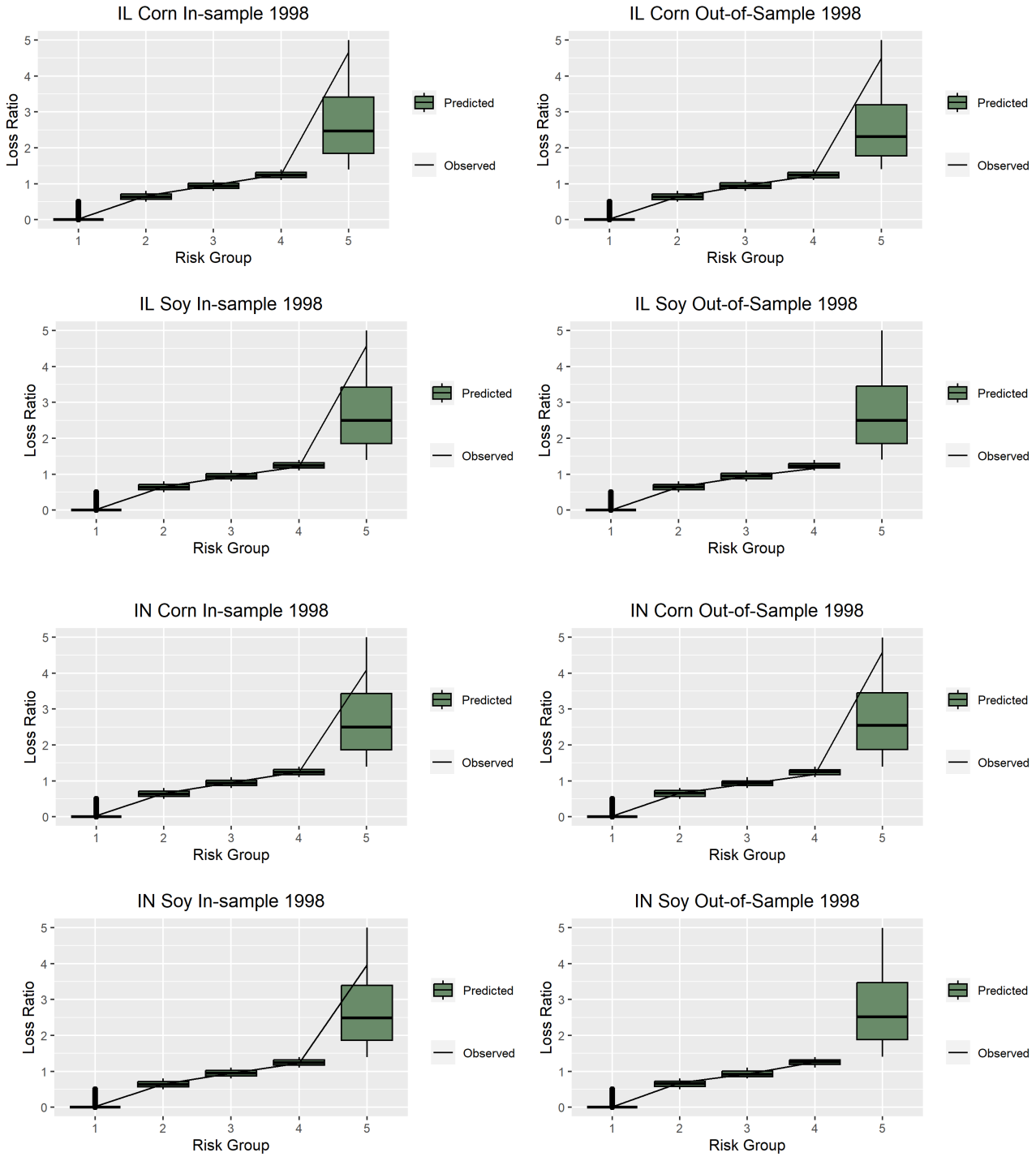
United Soybean Board. “What Are Soybeans Used For?” *United Soybean Board*, 10 June 2022, www.unitedsoybean.org/hopper/what-are-soybeans-used-for/.

APPENDIX

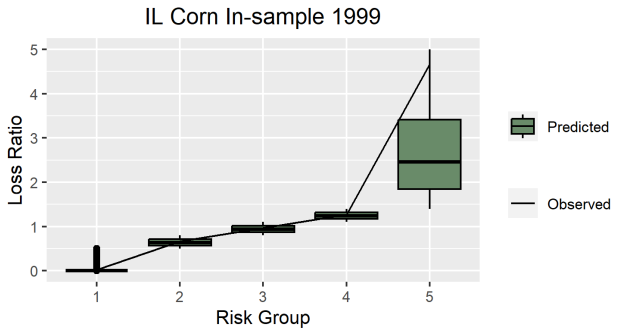
Course	Tools Applied
ECNS 461	Regression Modeling; Prediction; Understanding Potential Biases of Model
ECNS 560	Model/Data Visualization; Exploratory Analysis
EIND 457	Multiple Linear Regression; Model Building and Validation; Measuring Significance of Predictors
EFIN 301	Modeling and Understanding Risk; Understanding and Implementing Key Concepts Surrounding Insurance
EFIN 401	Builds off of Concepts Learned in EFIN 301; Risk Bracketing; Implementing Actuarial Concepts within R
STAT 217	Linear Regression Model Selection; Graphing with R

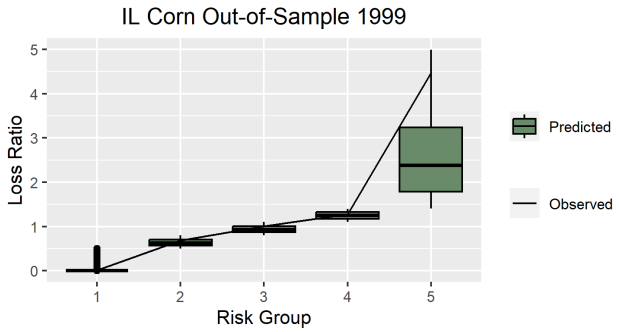
Appendix A: Final model validation graphs

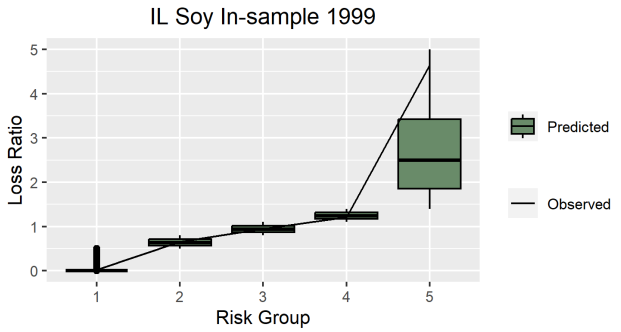
1998

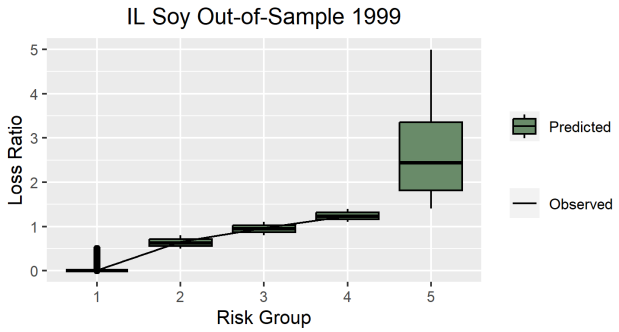


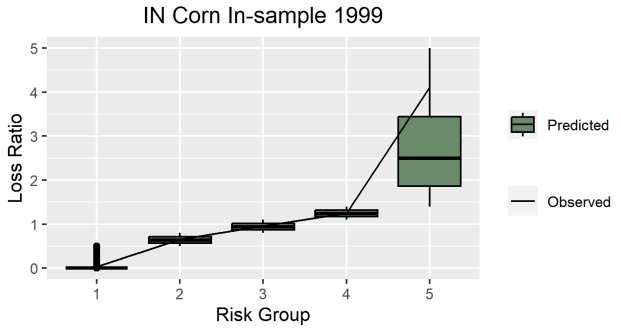
1999

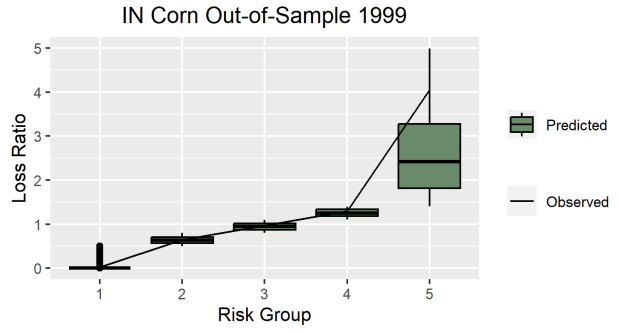

IL Corn In-sample 1999
Loss Ratio
Risk Group
Predicted
Observed

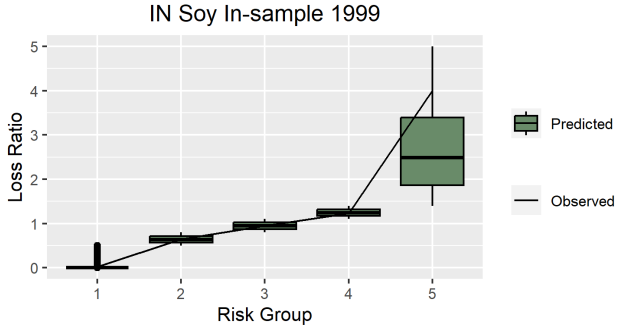

IL Corn Out-of-Sample 1999
Loss Ratio
Risk Group
Predicted
Observed

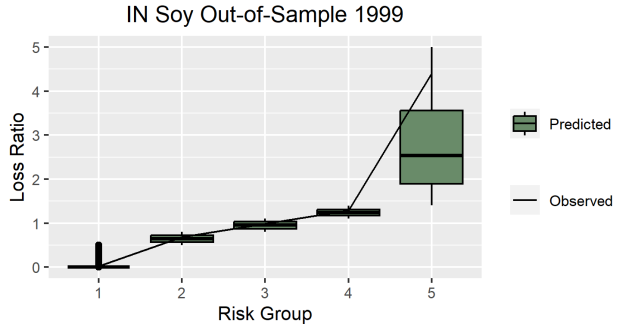

IL Soy In-sample 1999
Loss Ratio
Risk Group
Predicted
Observed

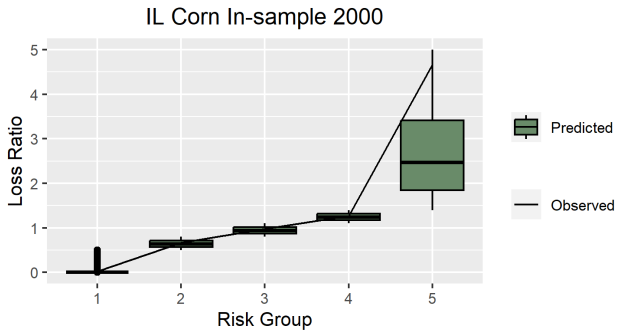

IL Soy Out-of-Sample 1999
Loss Ratio
Risk Group
Predicted
Observed

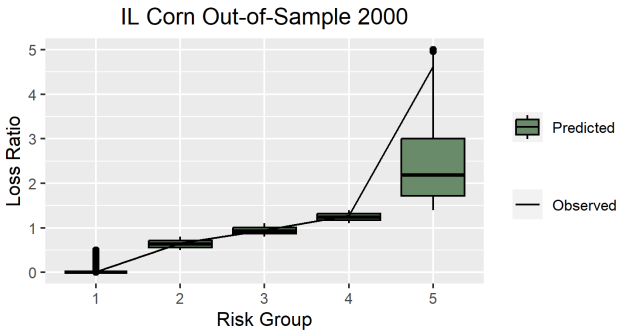

IN Corn In-sample 1999
Loss Ratio
Risk Group
Predicted
Observed

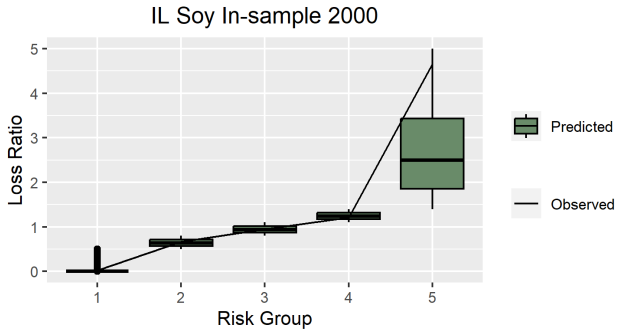

IN Corn Out-of-Sample 1999
Loss Ratio
Risk Group
Predicted
Observed

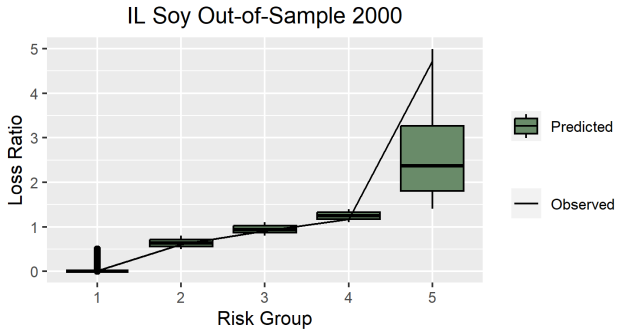

IN Soy In-sample 1999
Loss Ratio
Risk Group
Predicted
Observed

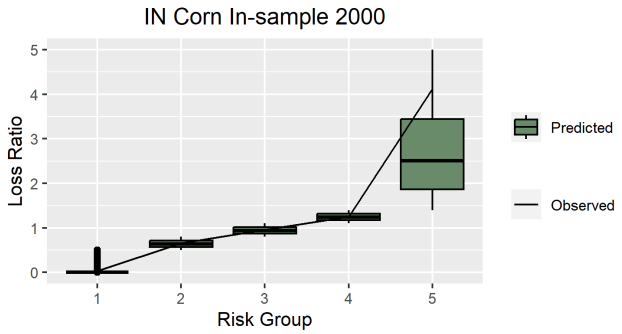

IN Soy Out-of-Sample 1999
Loss Ratio
Risk Group
Predicted
Observed

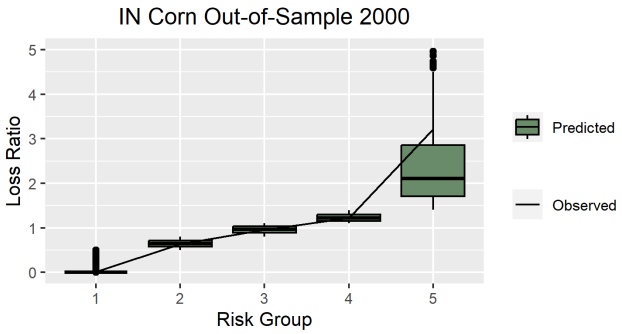

IL Corn In-sample 2000
Loss Ratio
Risk Group
Predicted
Observed

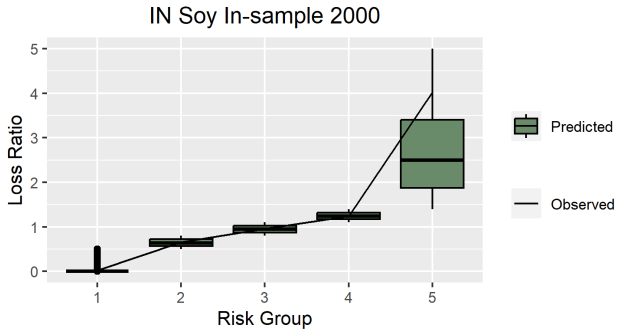

IL Corn Out-of-Sample 2000
Loss Ratio
Risk Group
Predicted
Observed

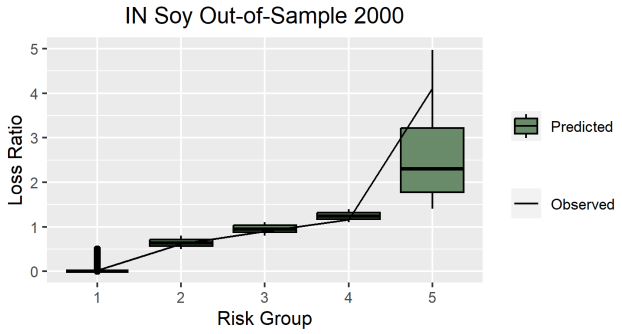

IL Soy In-sample 2000
Loss Ratio
Risk Group
Predicted
Observed

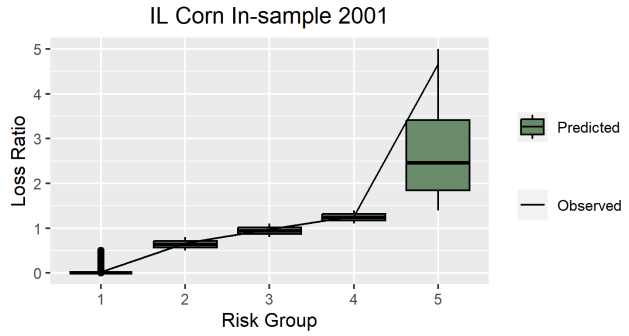

IL Soy Out-of-Sample 2000
Loss Ratio
Risk Group
Predicted
Observed

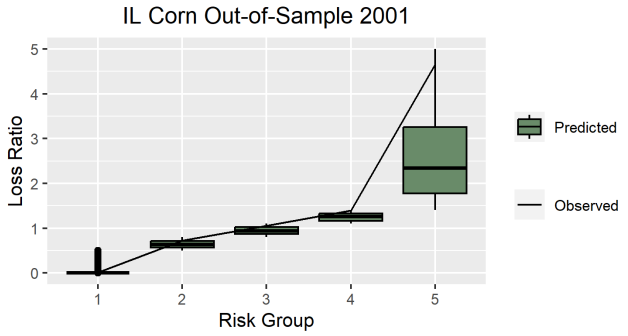

IN Corn In-sample 2000
Loss Ratio
Risk Group
Predicted
Observed

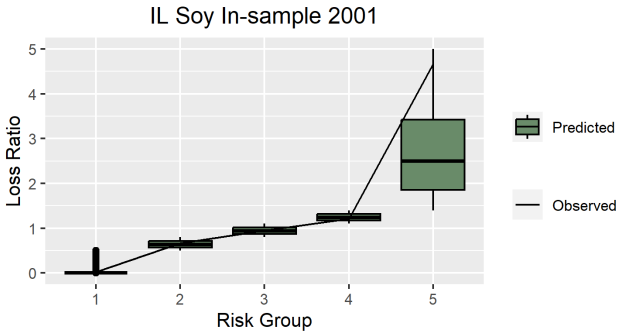

IN Corn Out-of-Sample 2000
Loss Ratio
Risk Group
Predicted
Observed

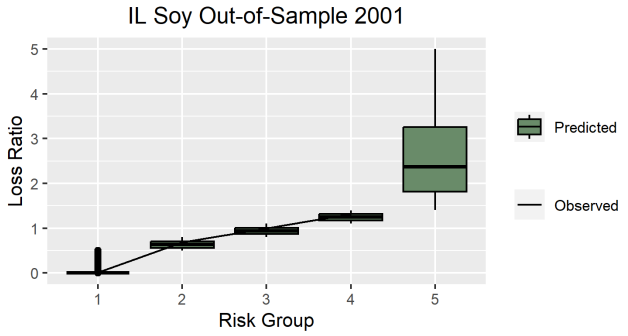

IN Soy In-sample 2000
Loss Ratio
Risk Group
Predicted
Observed

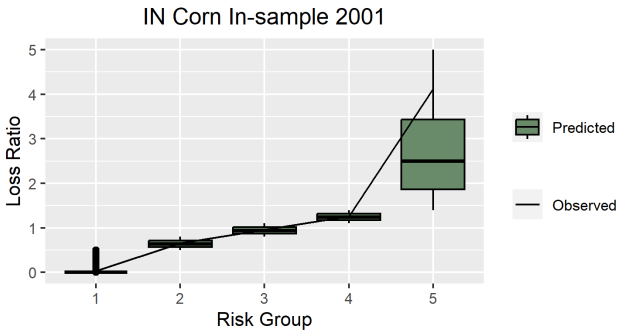

IN Soy Out-of-Sample 2000
Loss Ratio
Risk Group
Predicted
Observed

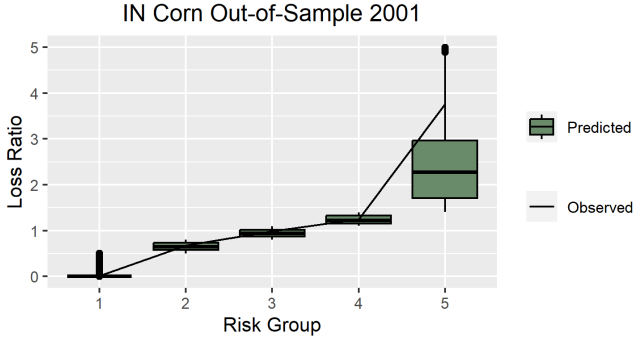

IL Corn In-sample 2001
Loss Ratio
Risk Group
Predicted
Observed

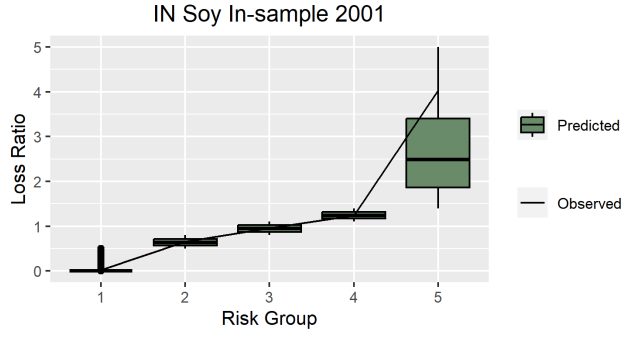

IL Corn Out-of-Sample 2001
Loss Ratio
Risk Group
Predicted
Observed

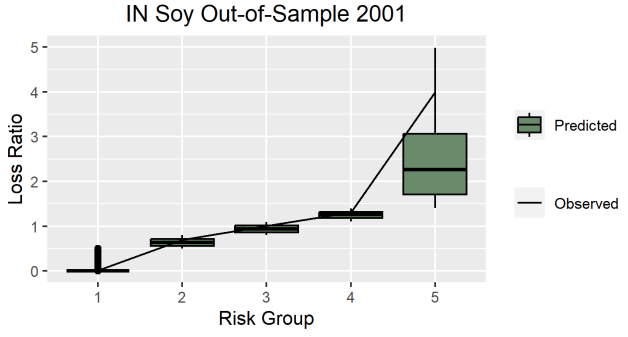

IL Soy In-sample 2001
Loss Ratio
Risk Group
Predicted
Observed

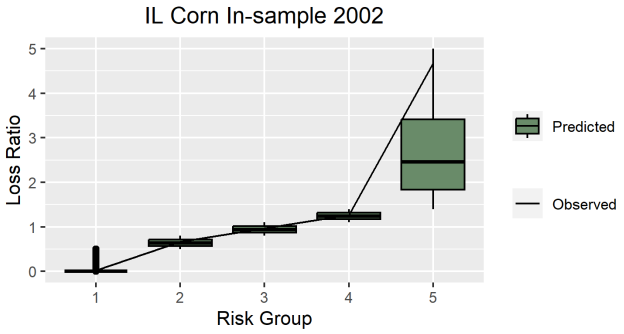

IL Soy Out-of-Sample 2001
Loss Ratio
Risk Group
Predicted
Observed

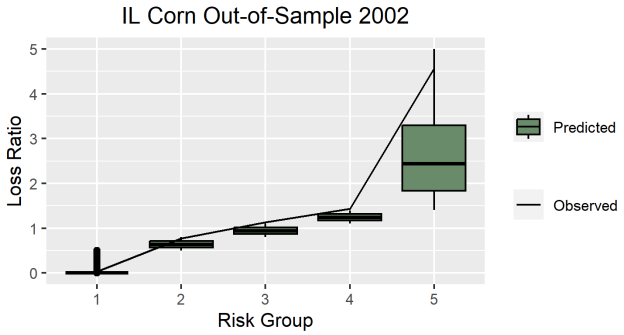

IN Corn In-sample 2001
Loss Ratio
Risk Group
Predicted
Observed

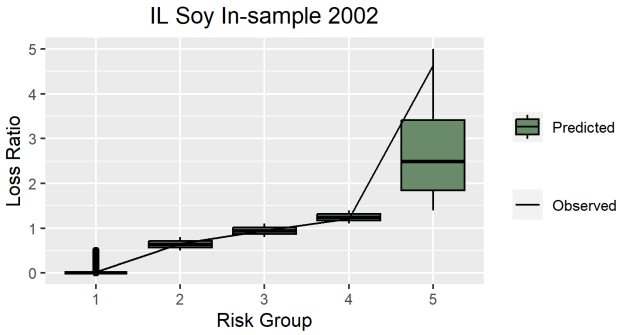

IN Corn Out-of-Sample 2001
Loss Ratio
Risk Group
Predicted
Observed

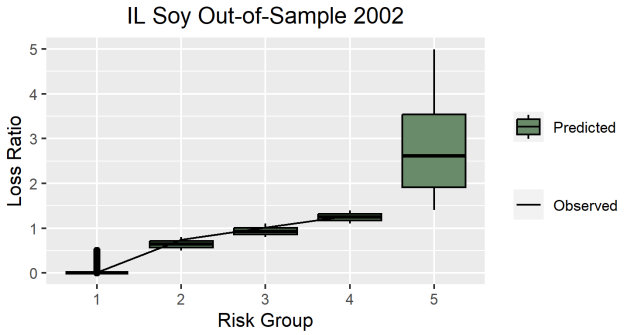

IN Soy In-sample 2001
Loss Ratio
Risk Group
Predicted
Observed

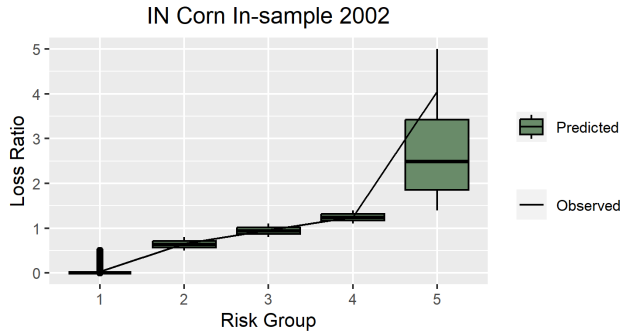

IN Soy Out-of-Sample 2001
Loss Ratio
Risk Group
Predicted
Observed

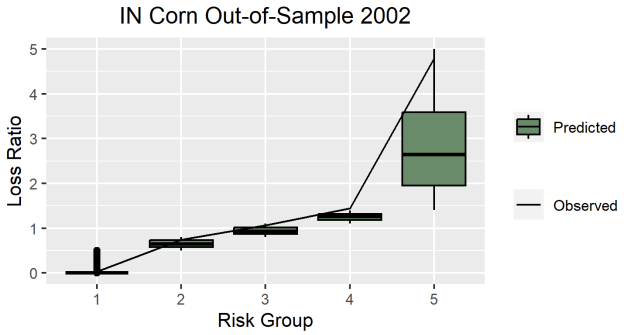

IL Corn In-sample 2002
Loss Ratio
Risk Group
Predicted
Observed

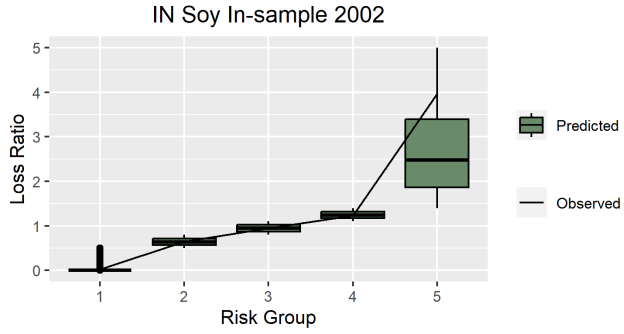

IL Corn Out-of-Sample 2002
Loss Ratio
Risk Group
Predicted
Observed

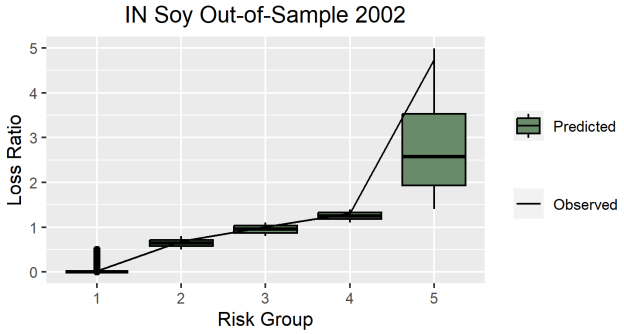

IL Soy In-sample 2002
Loss Ratio
Risk Group
Predicted
Observed


IL Soy Out-of-Sample 2002
Loss Ratio
Risk Group
Predicted
Observed

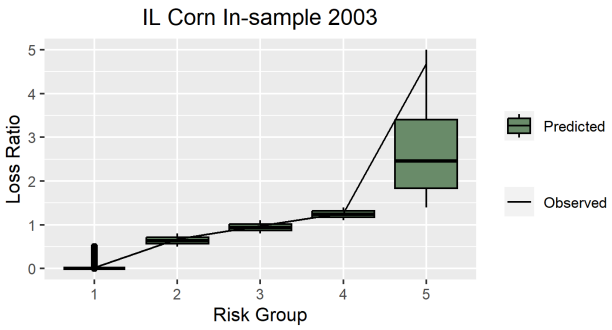

IN Corn In-sample 2002
Loss Ratio
Risk Group
Predicted
Observed

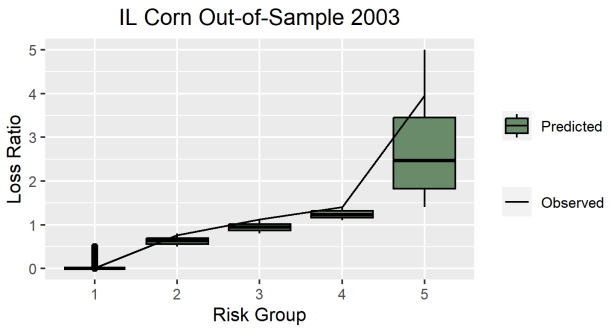

IN Corn Out-of-Sample 2002
Loss Ratio
Risk Group
Predicted
Observed

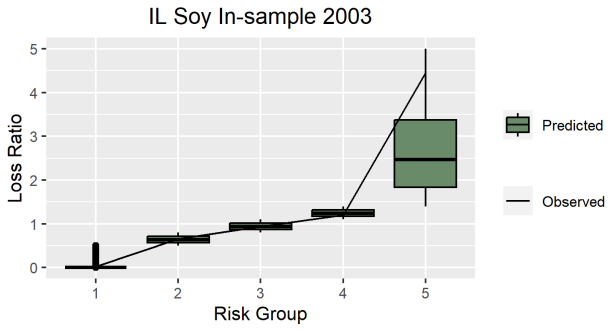

IN Soy In-sample 2002
Loss Ratio
Risk Group
Predicted
Observed

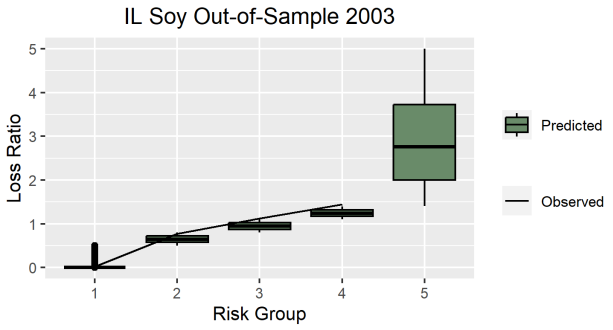

IN Soy Out-of-Sample 2002
Loss Ratio
Risk Group
Predicted
Observed

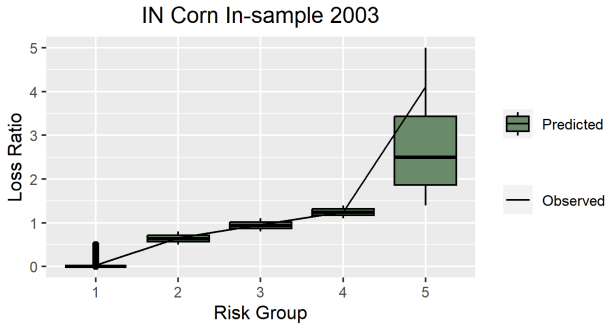
2003

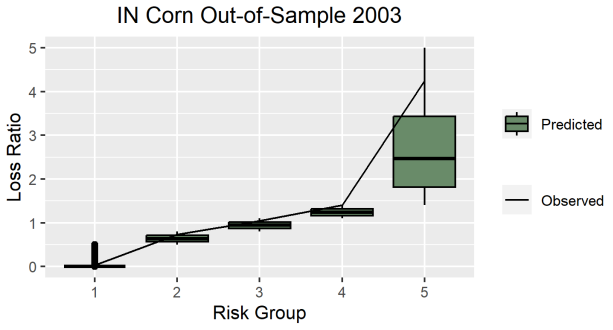

IL Corn In-sample 2003
Loss Ratio
Risk Group
Predicted
Observed

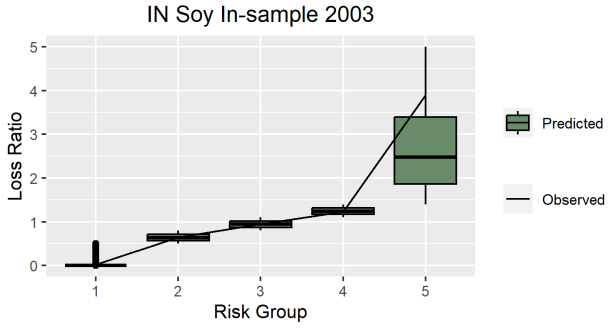

IL Corn Out-of-Sample 2003
Loss Ratio
Risk Group
Predicted
Observed

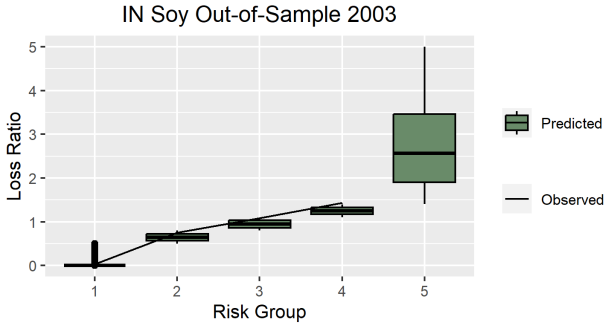

IL Soy In-sample 2003
Loss Ratio
Risk Group
Predicted
Observed

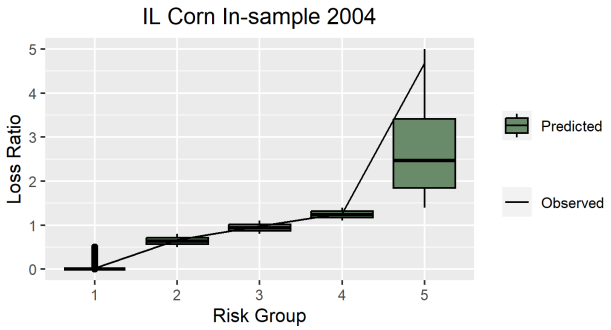

IL Soy Out-of-Sample 2003
Loss Ratio
Risk Group
Predicted
Observed

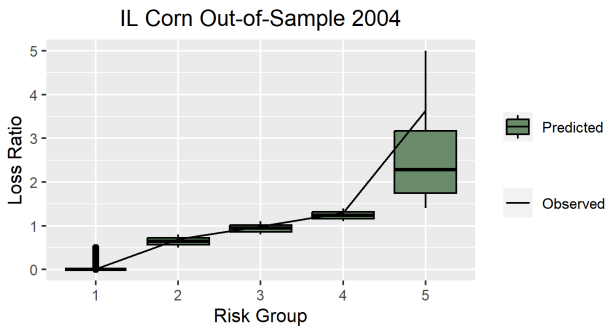

IN Corn In-sample 2003
Loss Ratio
Risk Group
Predicted
Observed

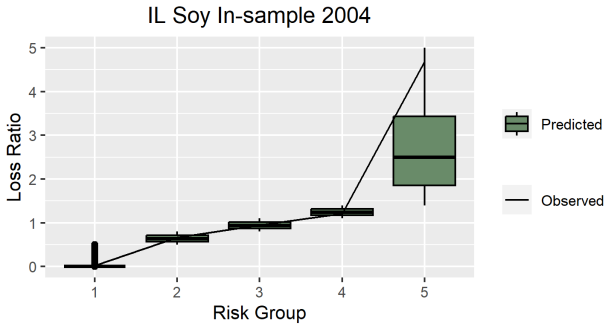

IN Corn Out-of-Sample 2003
Loss Ratio
Risk Group
Predicted
Observed

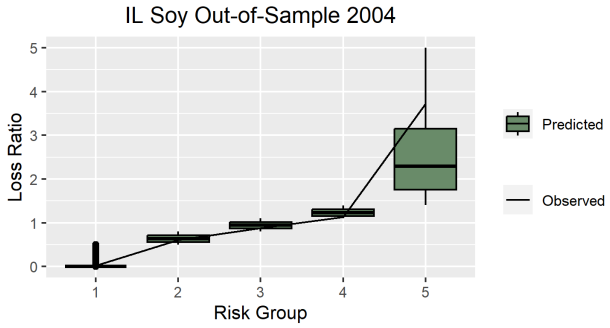

IN Soy In-sample 2003
Loss Ratio
Risk Group
Predicted
Observed

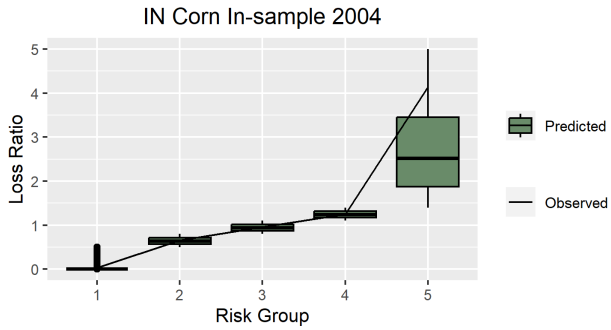

IN Soy Out-of-Sample 2003
Loss Ratio
Risk Group
Predicted
Observed

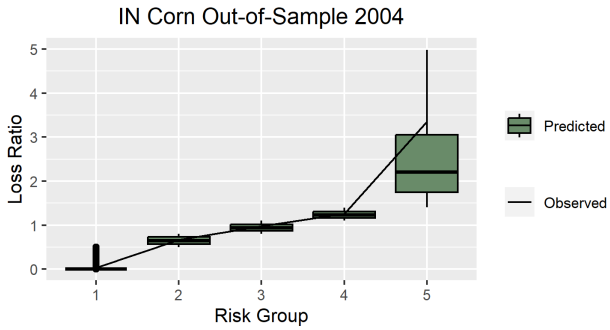

IL Corn In-sample 2004
Loss Ratio
Risk Group
Predicted
Observed

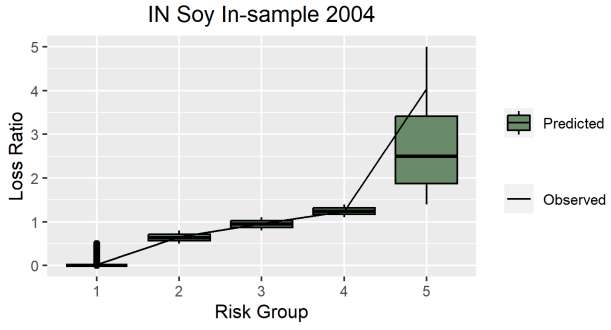

IL Corn Out-of-Sample 2004
Loss Ratio
Risk Group
Predicted
Observed

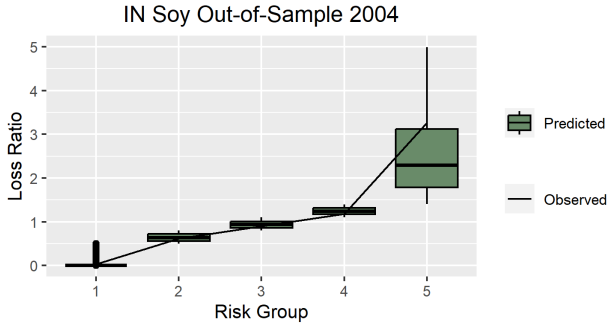

IL Soy In-sample 2004
Loss Ratio
Risk Group
Predicted
Observed

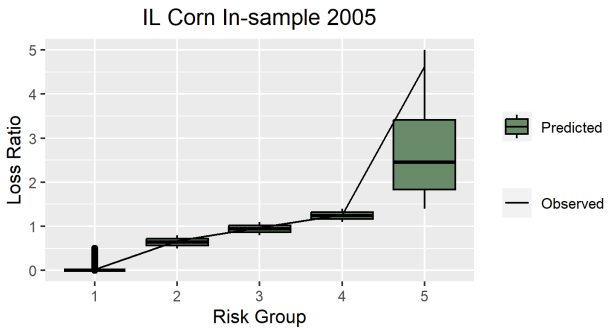

IL Soy Out-of-Sample 2004
Loss Ratio
Risk Group
Predicted
Observed

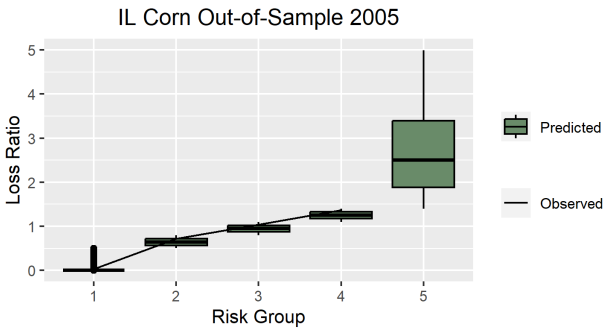

IN Corn In-sample 2004
Loss Ratio
Risk Group
Predicted
Observed

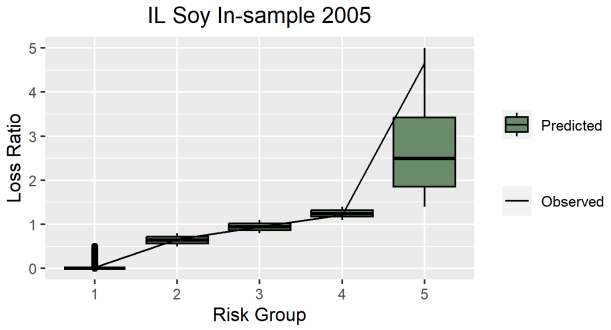

IN Corn Out-of-Sample 2004
Loss Ratio
Risk Group
Predicted
Observed

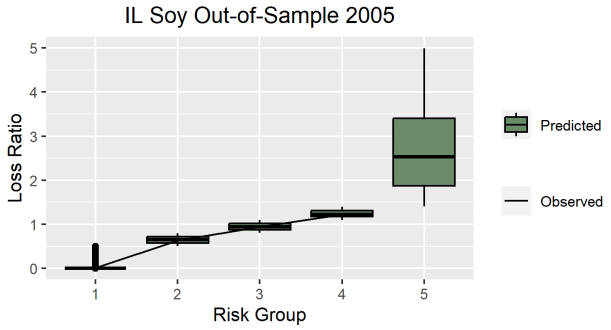

IN Soy In-sample 2004
Loss Ratio
Risk Group
Predicted
Observed

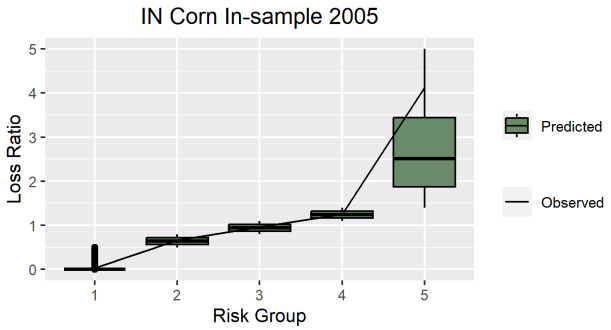

IN Soy Out-of-Sample 2004
Loss Ratio
Risk Group
Predicted
Observed

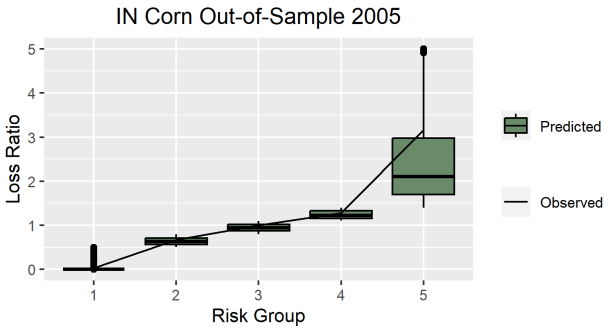

IL Corn In-sample 2005
Loss Ratio
Risk Group
Predicted
Observed

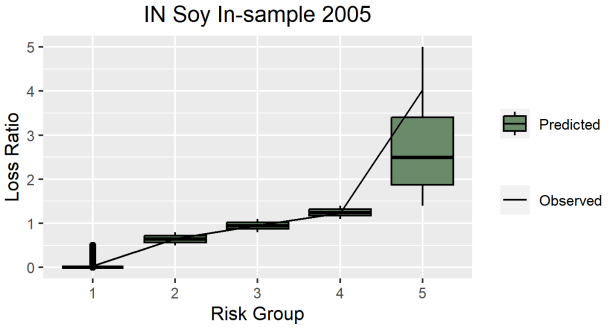

IL Corn Out-of-Sample 2005
Loss Ratio
Risk Group
Predicted
Observed

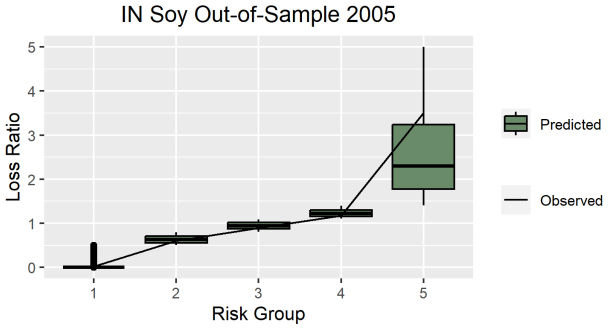

IL Soy In-sample 2005
Loss Ratio
Risk Group
Predicted
Observed

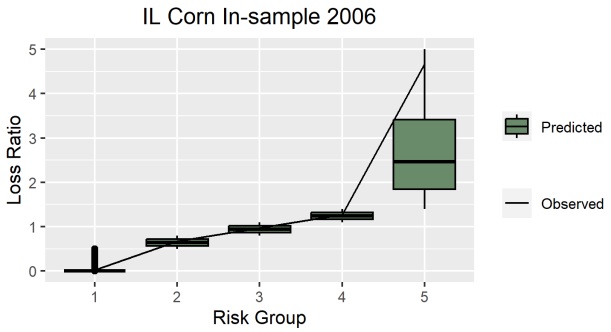

IL Soy Out-of-Sample 2005
Loss Ratio
Risk Group
Predicted
Observed

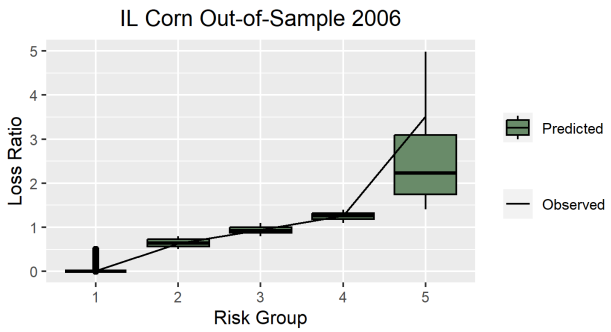

IN Corn In-sample 2005
Loss Ratio
Risk Group
Predicted
Observed

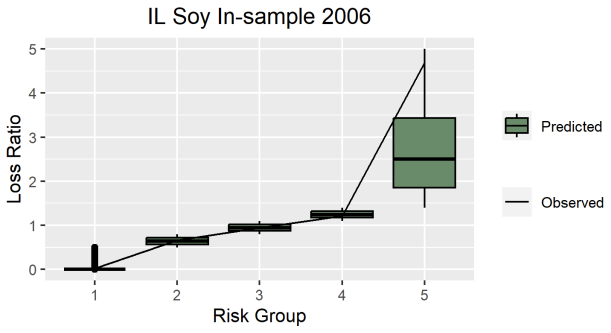

IN Corn Out-of-Sample 2005
Loss Ratio
Risk Group
Predicted
Observed

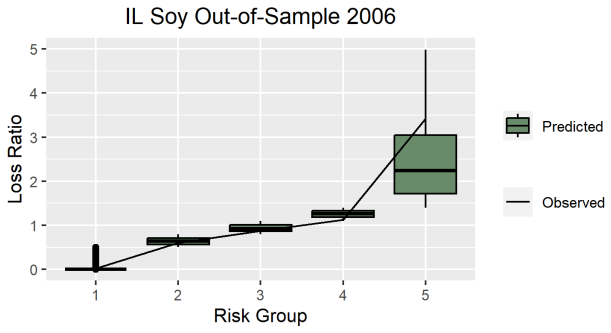

IN Soy In-sample 2005
Loss Ratio
Risk Group
Predicted
Observed

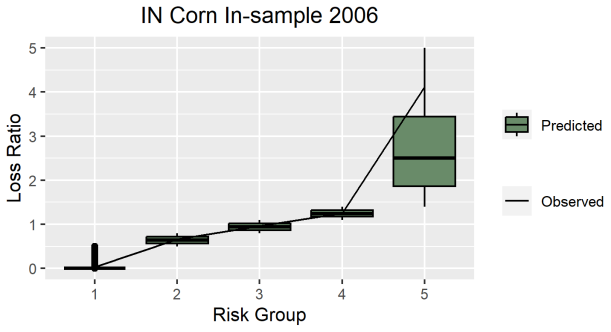

IN Soy Out-of-Sample 2005
Loss Ratio
Risk Group
Predicted
Observed

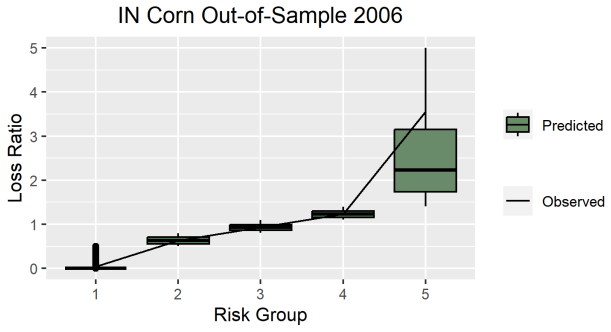

IL Corn In-sample 2006
Loss Ratio
Risk Group
Predicted
Observed

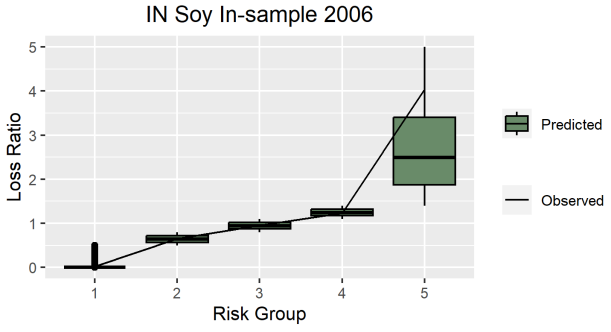

IL Corn Out-of-Sample 2006
Loss Ratio
Risk Group
Predicted
Observed

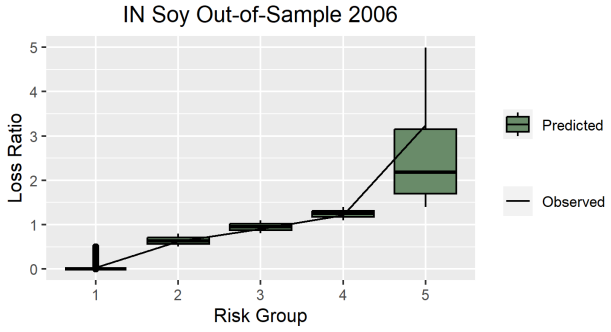

IL Soy In-sample 2006
Loss Ratio
Risk Group
Predicted
Observed

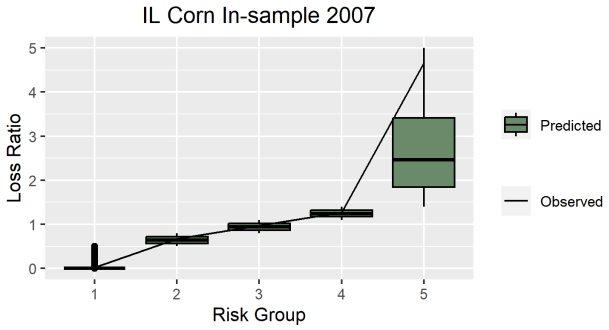

IL Soy Out-of-Sample 2006
Loss Ratio
Risk Group
Predicted
Observed

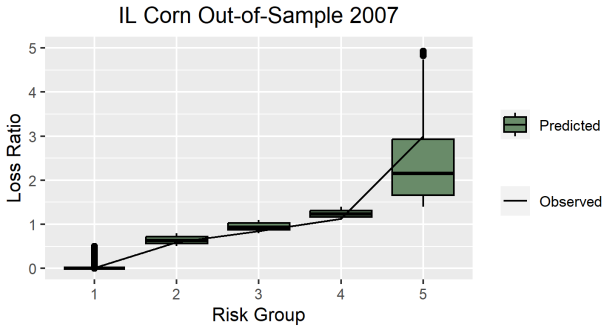

IN Corn In-sample 2006
Loss Ratio
Risk Group
Predicted
Observed

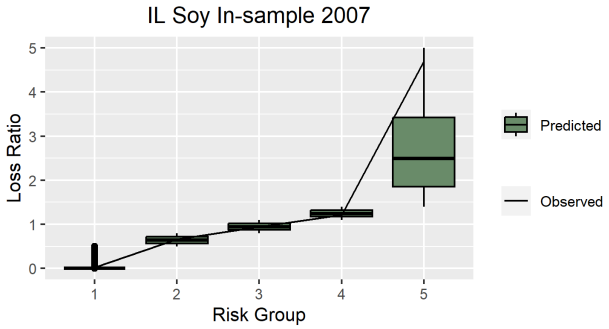

IN Corn Out-of-Sample 2006
Loss Ratio
Risk Group
Predicted
Observed

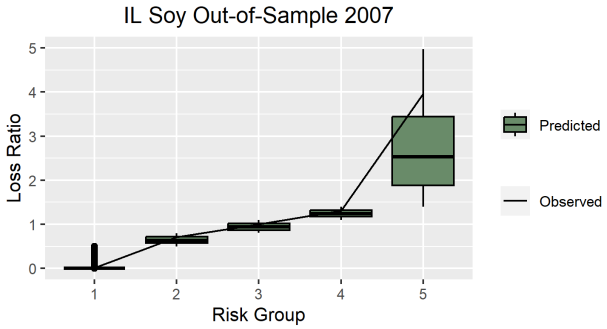

IN Soy In-sample 2006
Loss Ratio
Risk Group
Predicted
Observed

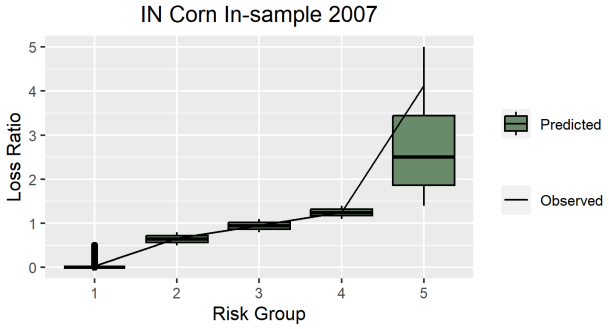

IN Soy Out-of-Sample 2006
Loss Ratio
Risk Group
Predicted
Observed

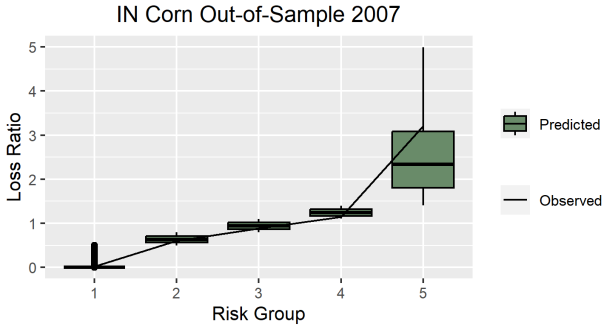

IL Corn In-sample 2007
Loss Ratio
Risk Group
Predicted
Observed

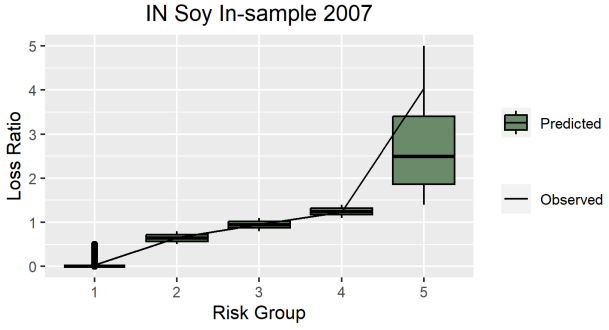

IL Corn Out-of-Sample 2007
Loss Ratio
Risk Group
Predicted
Observed

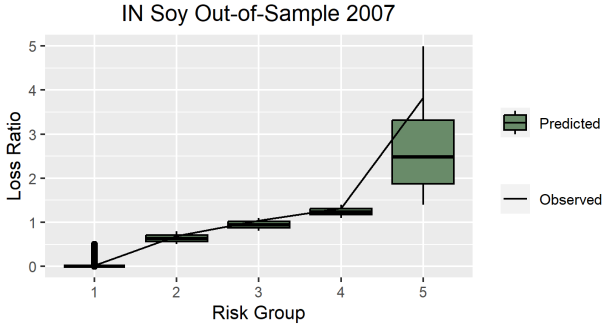

IL Soy In-sample 2007
Loss Ratio
Risk Group
Predicted
Observed

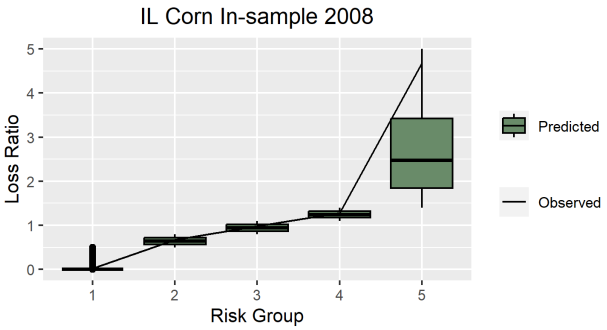

IL Soy Out-of-Sample 2007
Loss Ratio
Risk Group
Predicted
Observed

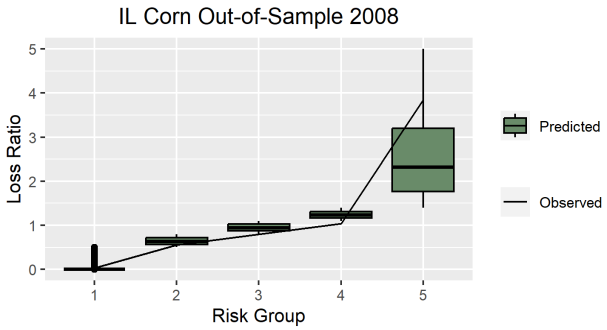

IN Corn In-sample 2007
Loss Ratio
Risk Group
Predicted
Observed

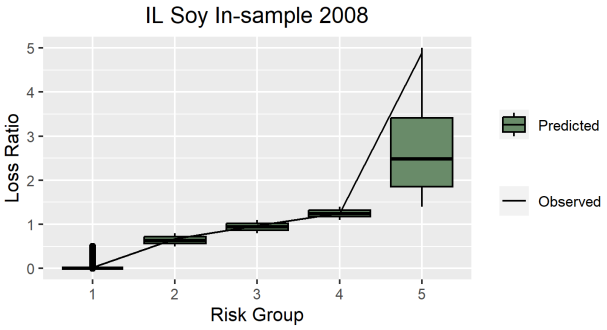

IN Corn Out-of-Sample 2007
Loss Ratio
Risk Group
Predicted
Observed

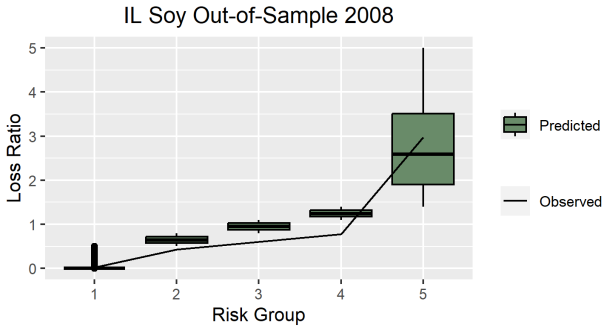

IN Soy In-sample 2007
Loss Ratio
Risk Group
Predicted
Observed

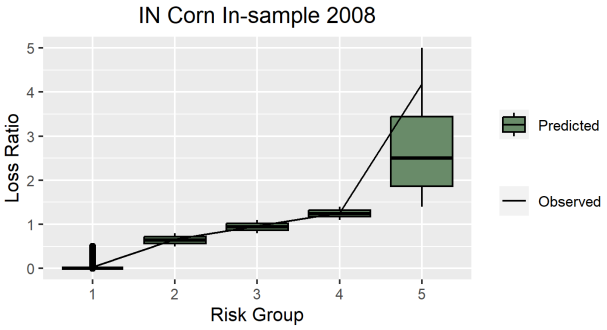

IN Soy Out-of-Sample 2007
Loss Ratio
Risk Group
Predicted
Observed

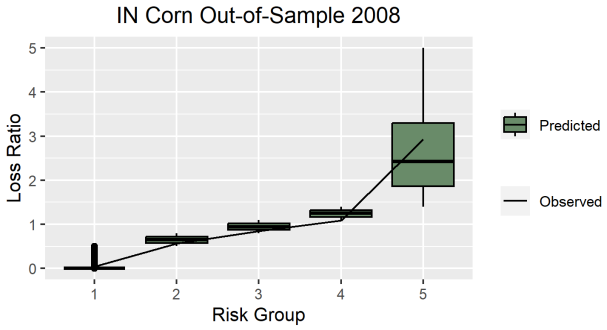

IL Corn In-sample 2008
Loss Ratio
Risk Group
Predicted
Observed

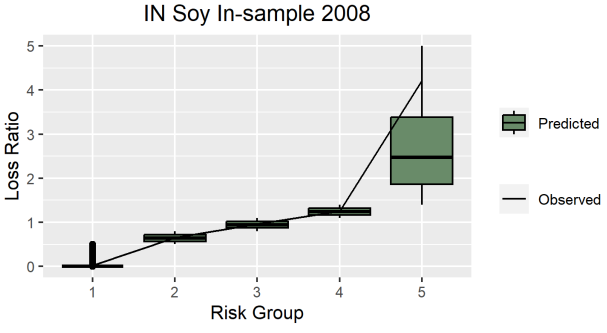

IL Corn Out-of-Sample 2008
Loss Ratio
Risk Group
Predicted
Observed

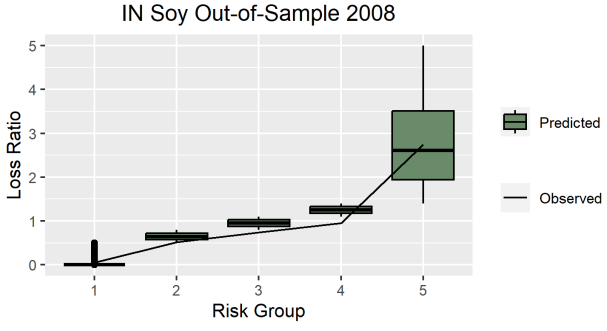

IL Soy In-sample 2008
Loss Ratio
Risk Group
Predicted
Observed

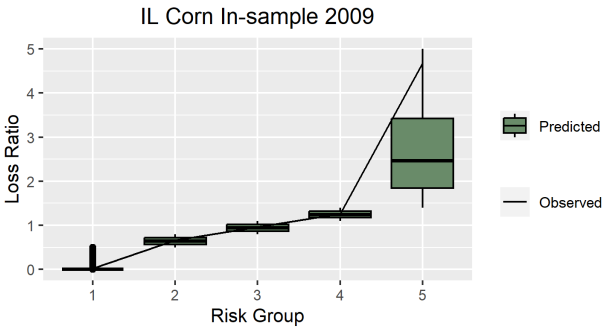

IL Soy Out-of-Sample 2008
Loss Ratio
Risk Group
Predicted
Observed

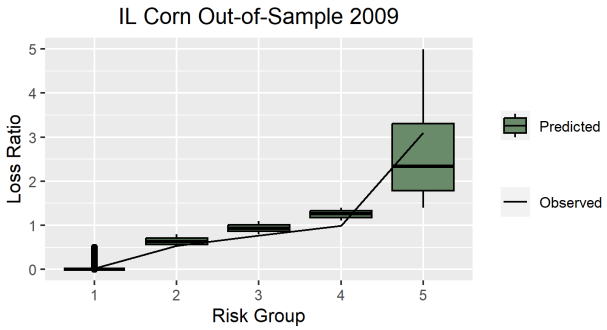

IN Corn In-sample 2008
Loss Ratio
Risk Group
Predicted
Observed

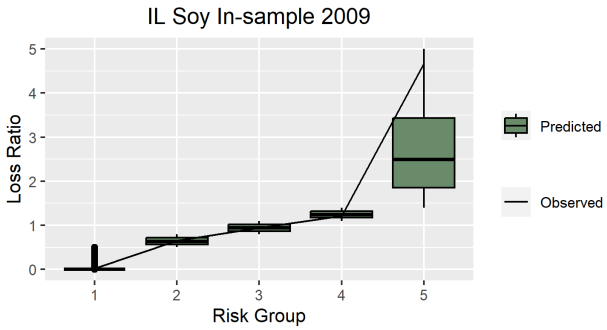

IN Corn Out-of-Sample 2008
Loss Ratio
Risk Group
Predicted
Observed

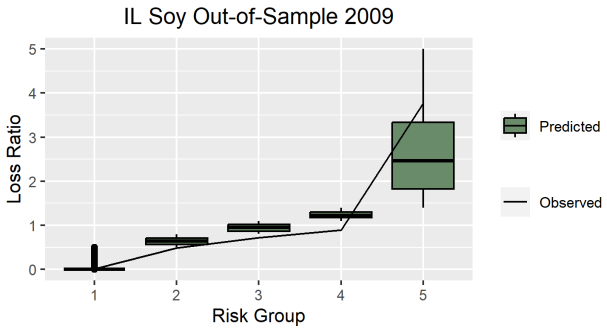

IN Soy In-sample 2008
Loss Ratio
Risk Group
Predicted
Observed

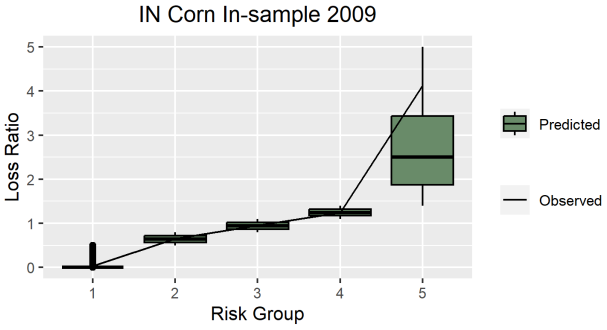

IN Soy Out-of-Sample 2008
Loss Ratio
Risk Group
Predicted
Observed

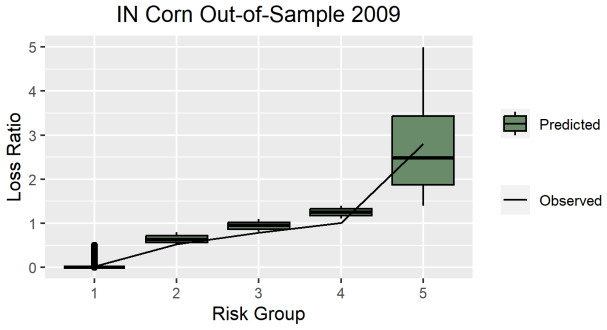

IL Corn In-sample 2009
Loss Ratio
Risk Group
Predicted
Observed

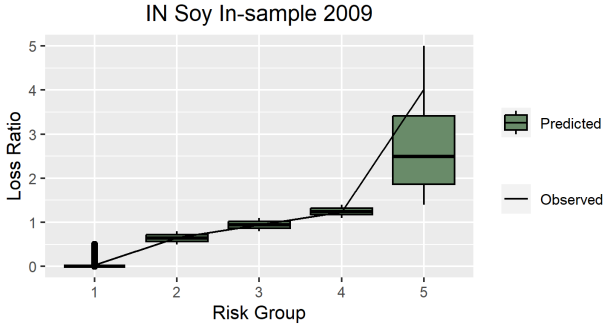

IL Corn Out-of-Sample 2009
Loss Ratio
Risk Group
Predicted
Observed

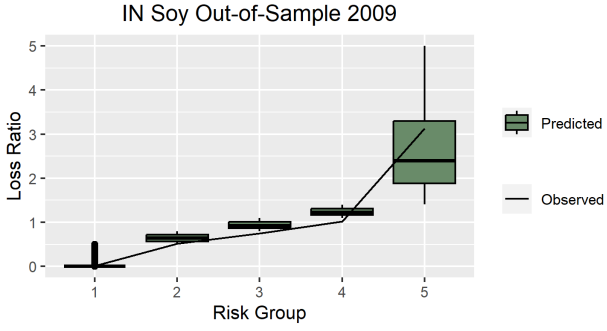

IL Soy In-sample 2009
Loss Ratio
Risk Group
Predicted
Observed

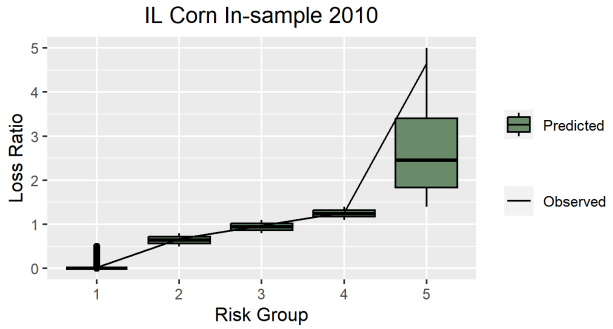

IL Soy Out-of-Sample 2009
Loss Ratio
Risk Group
Predicted
Observed

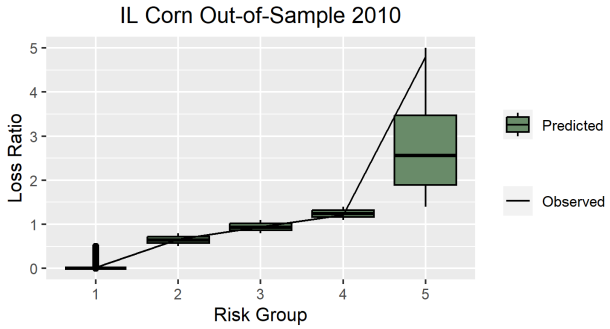

IN Corn In-sample 2009
Loss Ratio
Risk Group
Predicted
Observed

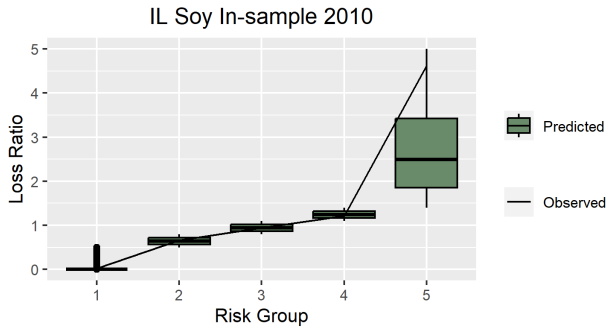

IN Corn Out-of-Sample 2009
Loss Ratio
Risk Group
Predicted
Observed

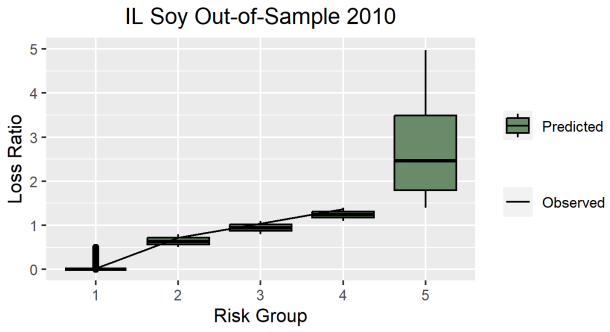

IN Soy In-sample 2009
Loss Ratio
Risk Group
Predicted
Observed

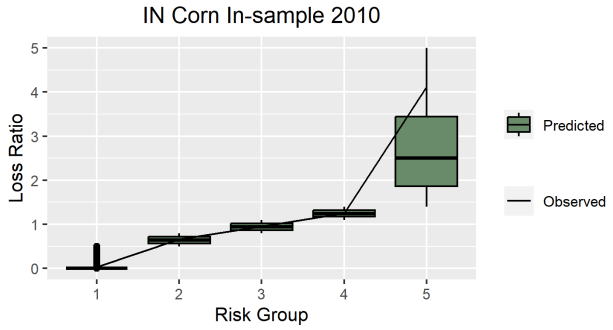

IN Soy Out-of-Sample 2009
Loss Ratio
Risk Group
Predicted
Observed

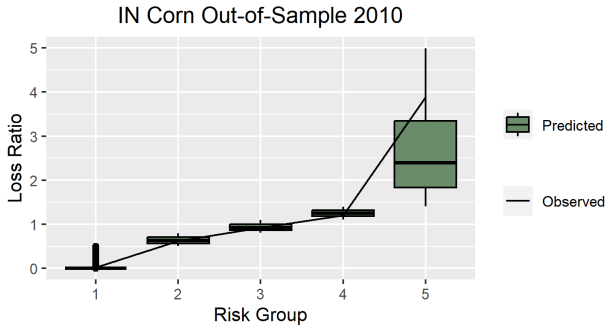

IL Corn In-sample 2010
Loss Ratio
Risk Group
Predicted
Observed

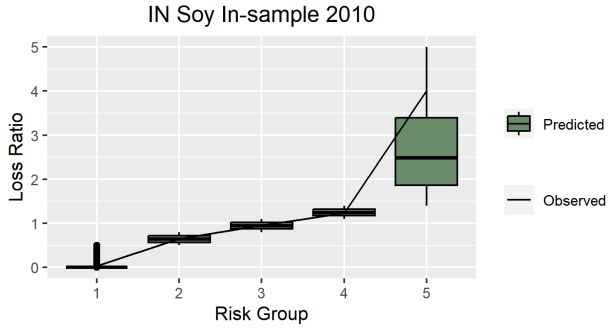

IL Corn Out-of-Sample 2010
Loss Ratio
Risk Group
Predicted
Observed

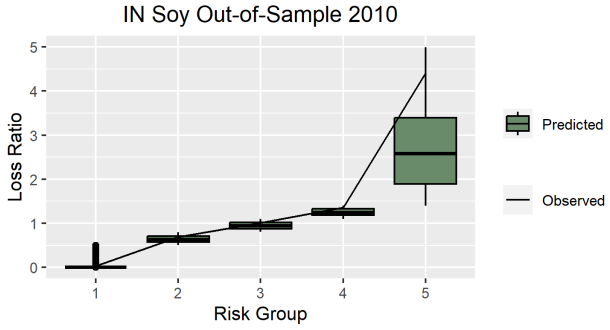

IL Soy In-sample 2010
Loss Ratio
Risk Group
Predicted
Observed

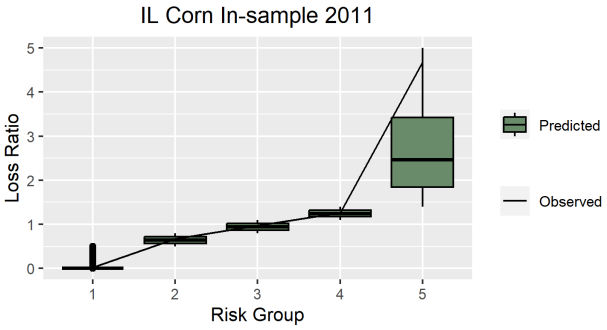

IL Soy Out-of-Sample 2010
Loss Ratio
Risk Group
Predicted
Observed

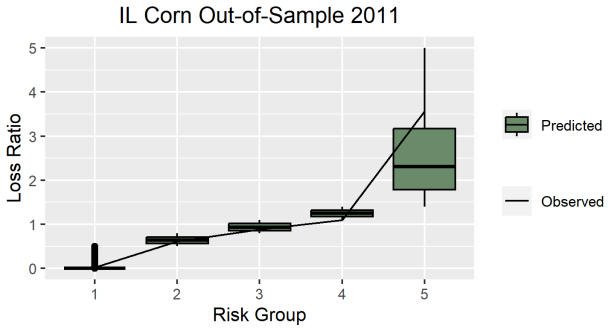

IN Corn In-sample 2010
Loss Ratio
Risk Group
Predicted
Observed

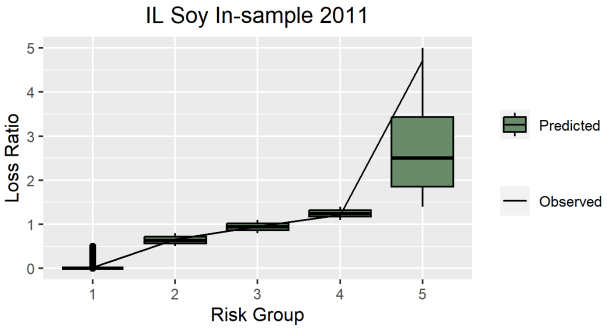

IN Corn Out-of-Sample 2010
Loss Ratio
Risk Group
Predicted
Observed

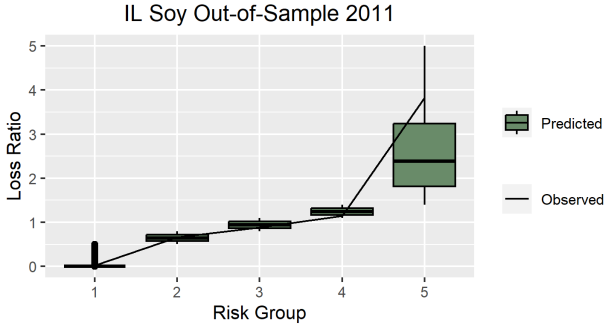

IN Soy In-sample 2010
Loss Ratio
Risk Group
Predicted
Observed

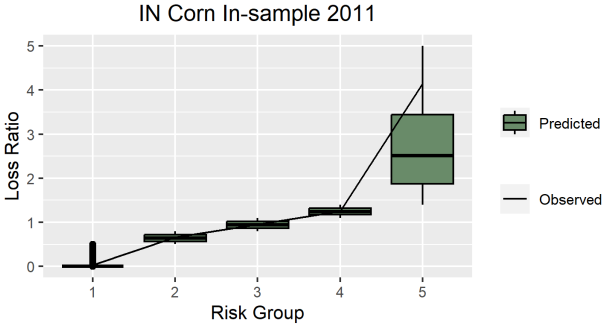

IN Soy Out-of-Sample 2010
Loss Ratio
Risk Group
Predicted
Observed

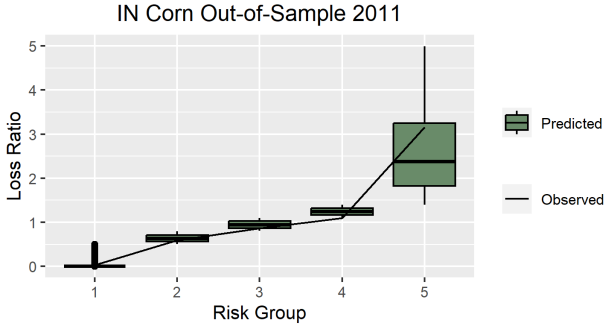

IL Corn In-sample 2011
Loss Ratio
Risk Group
Predicted
Observed

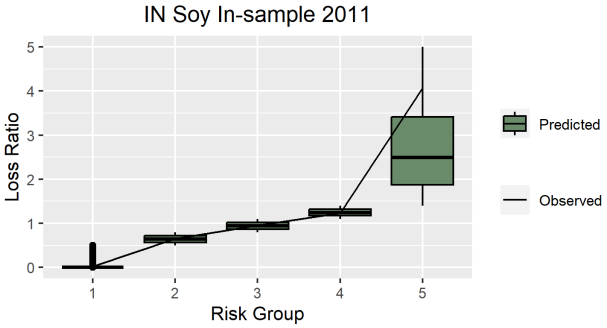

IL Corn Out-of-Sample 2011
Loss Ratio
Risk Group
Predicted
Observed

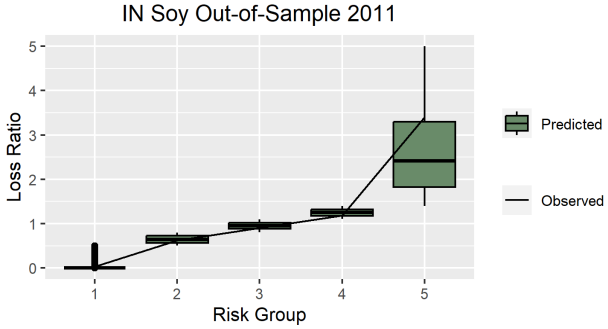

IL Soy In-sample 2011
Loss Ratio
Risk Group
Predicted
Observed


IL Soy Out-of-Sample 2011
Loss Ratio
Risk Group
Predicted
Observed

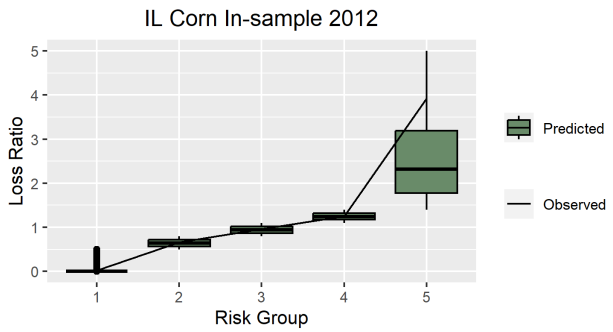

IN Corn In-sample 2011
Loss Ratio
Risk Group
Predicted
Observed

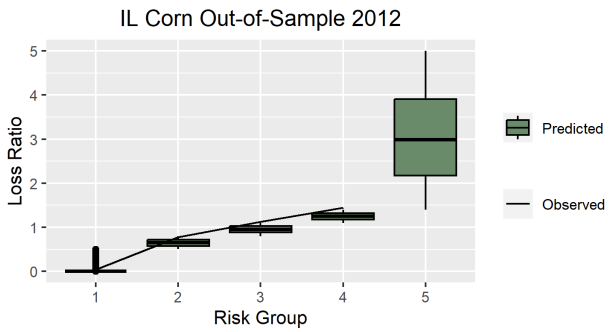

IN Corn Out-of-Sample 2011
Loss Ratio
Risk Group
Predicted
Observed

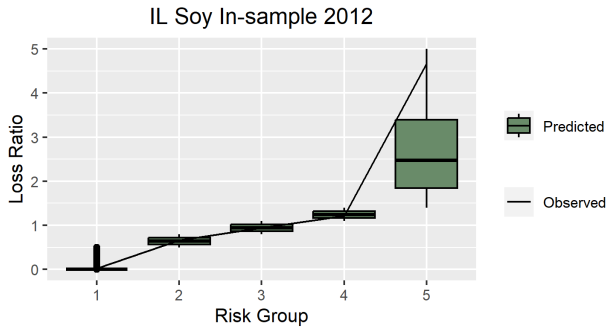

IN Soy In-sample 2011
Loss Ratio
Risk Group
Predicted
Observed

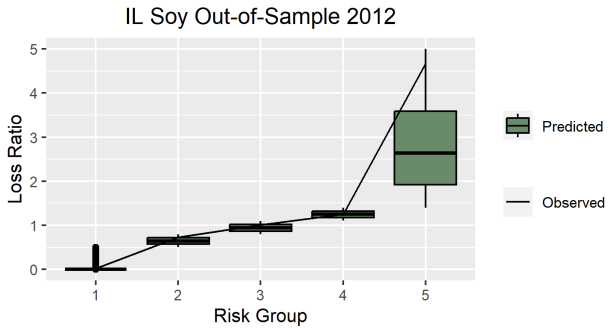

IN Soy Out-of-Sample 2011
Loss Ratio
Risk Group
Predicted
Observed

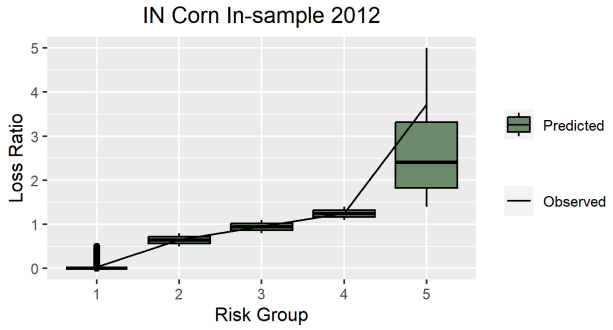
2012

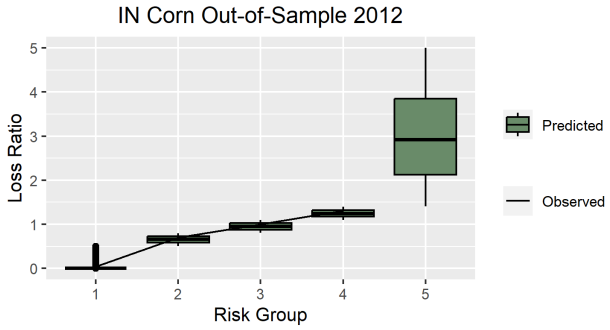

IL Corn In-sample 2012
Loss Ratio
Risk Group
Predicted
Observed

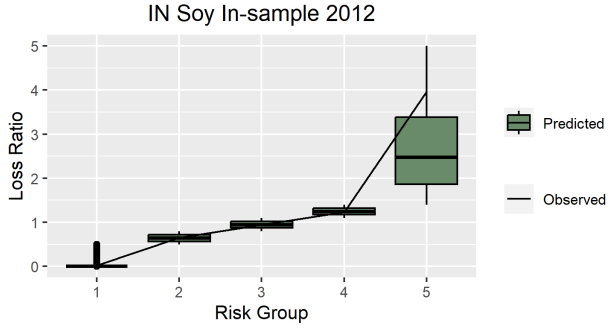

IL Corn Out-of-Sample 2012
Loss Ratio
Risk Group
Predicted
Observed

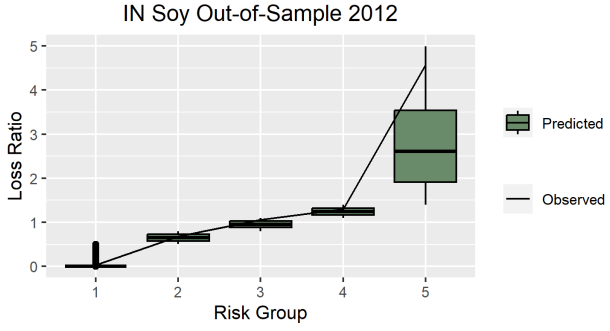

IL Soy In-sample 2012
Loss Ratio
Risk Group
Predicted
Observed

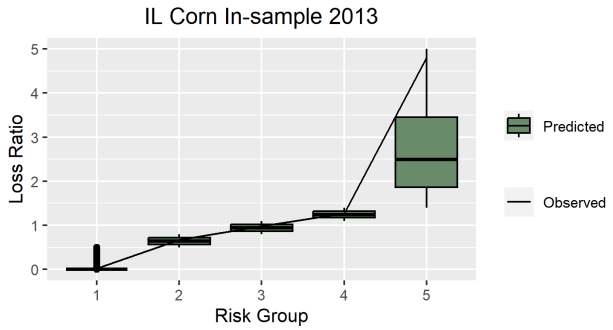

IL Soy Out-of-Sample 2012
Loss Ratio
Risk Group
Predicted
Observed

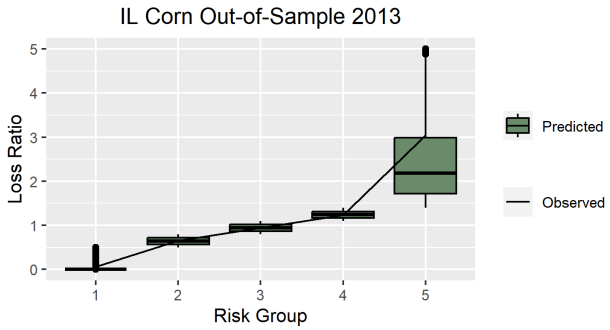

IN Corn In-sample 2012
Loss Ratio
Risk Group
Predicted
Observed

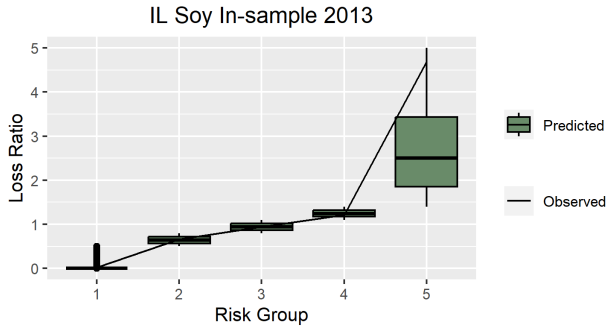

IN Corn Out-of-Sample 2012
Loss Ratio
Risk Group
Predicted
Observed

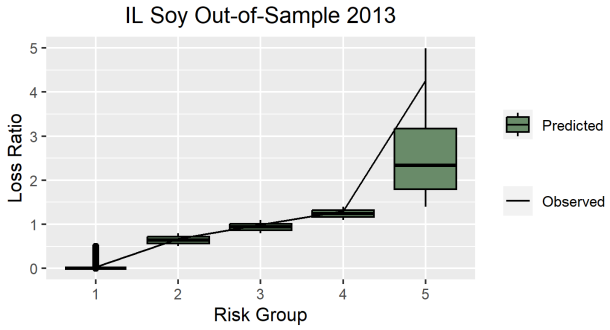

IN Soy In-sample 2012
Loss Ratio
Risk Group
Predicted
Observed

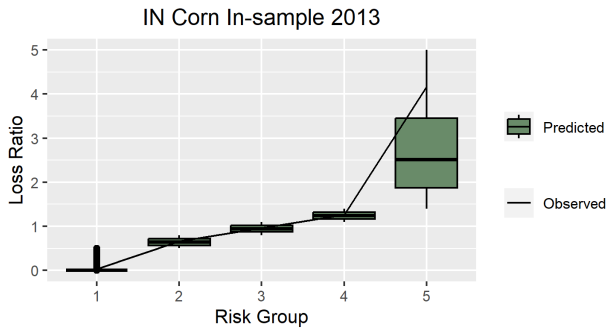

IN Soy Out-of-Sample 2012
Loss Ratio
Risk Group
Predicted
Observed

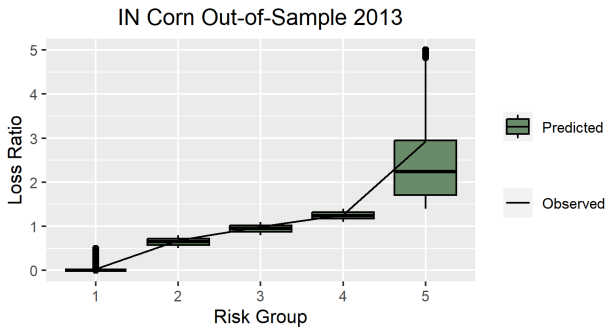

IL Corn In-sample 2013
Loss Ratio
Risk Group
Predicted
Observed

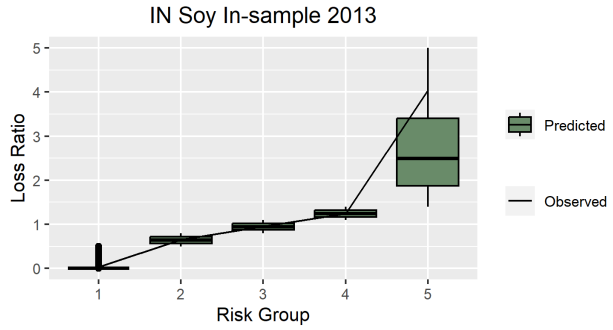

IL Corn Out-of-Sample 2013
Loss Ratio
Risk Group
Predicted
Observed

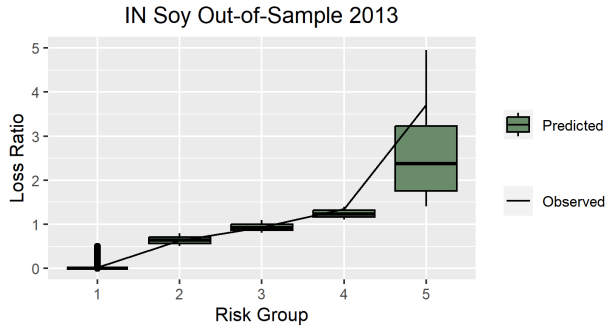

IL Soy In-sample 2013
Loss Ratio
Risk Group
Predicted
Observed

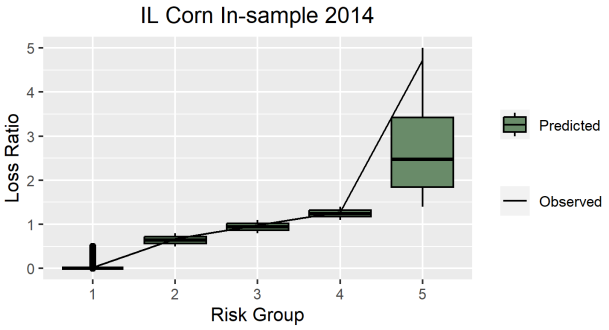

IL Soy Out-of-Sample 2013
Loss Ratio
Risk Group
Predicted
Observed

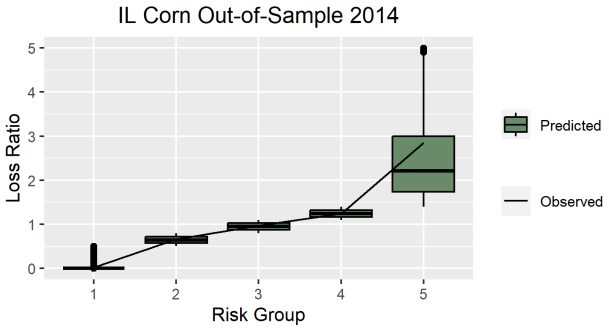

IN Corn In-sample 2013
Loss Ratio
Risk Group
Predicted
Observed

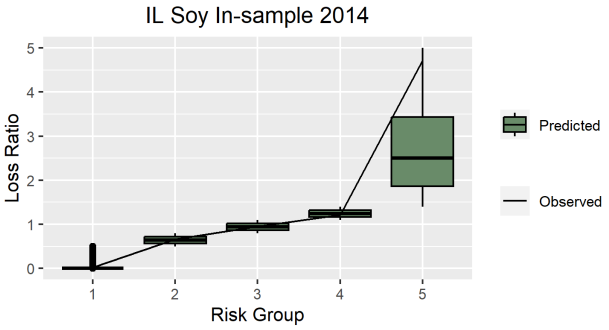

IN Corn Out-of-Sample 2013
Loss Ratio
Risk Group
Predicted
Observed

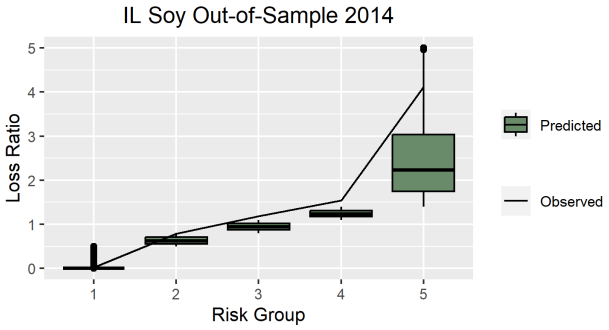

IN Soy In-sample 2013
Loss Ratio
Risk Group
Predicted
Observed

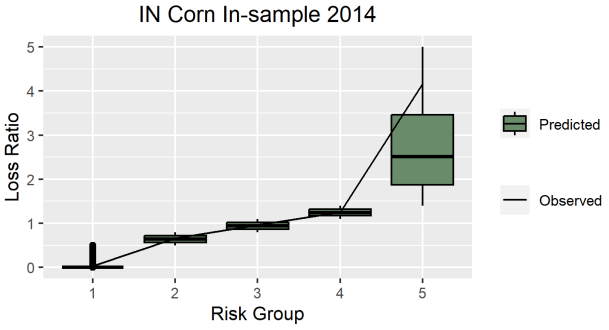

IN Soy Out-of-Sample 2013
Loss Ratio
Risk Group
Predicted
Observed

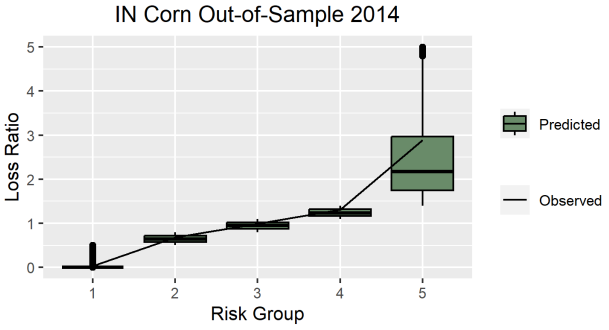

IL Corn In-sample 2014
Loss Ratio
Risk Group
Predicted
Observed

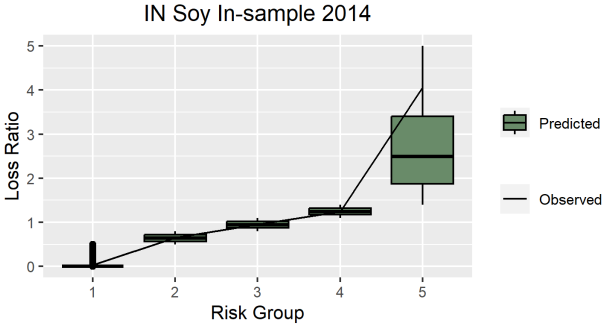

IL Corn Out-of-Sample 2014
Loss Ratio
Risk Group
Predicted
Observed

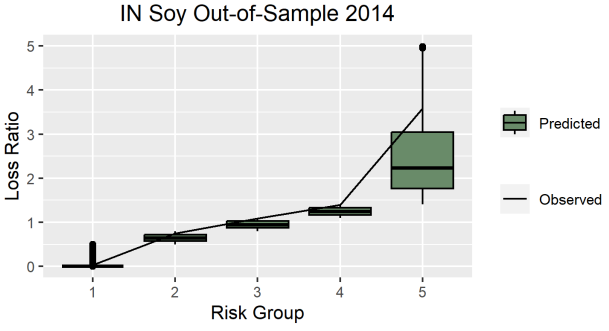

IL Soy In-sample 2014
Loss Ratio
Risk Group
Predicted
Observed

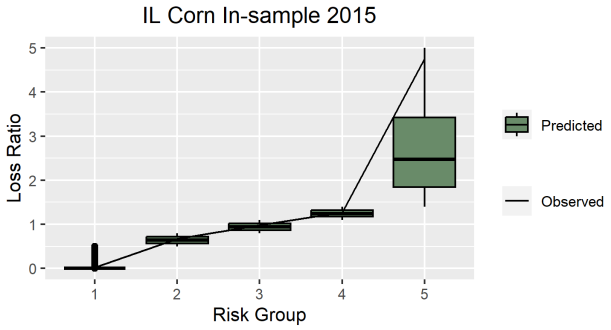

IL Soy Out-of-Sample 2014
Loss Ratio
Risk Group
Predicted
Observed

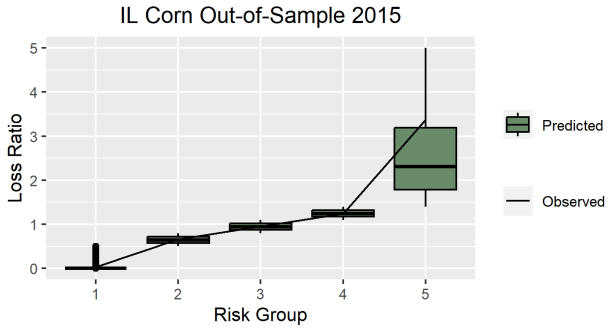

IN Corn In-sample 2014
Loss Ratio
Risk Group
Predicted
Observed

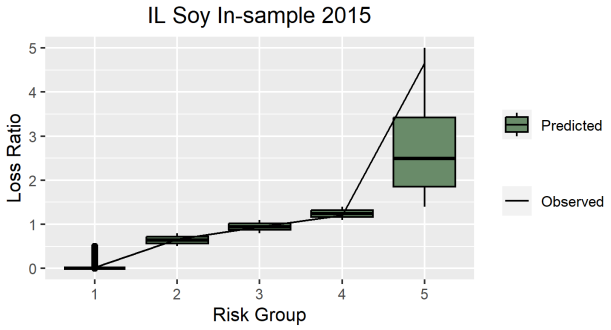

IN Corn Out-of-Sample 2014
Loss Ratio
Risk Group
Predicted
Observed

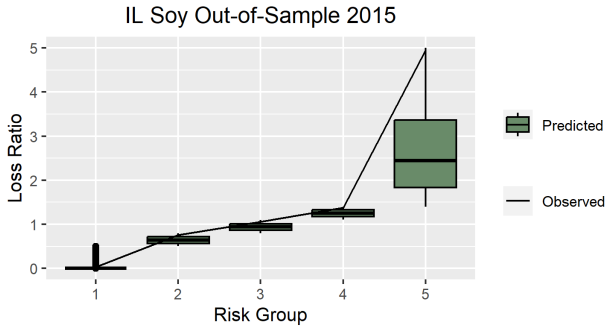

IN Soy In-sample 2014
Loss Ratio
Risk Group
Predicted
Observed

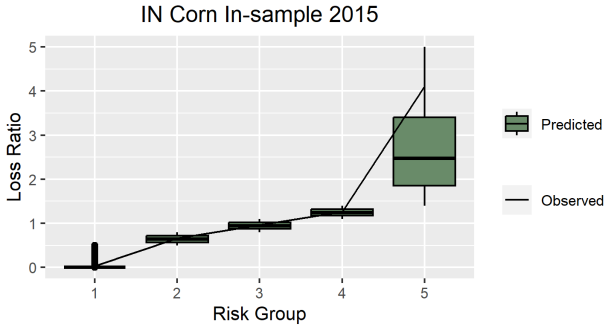

IN Soy Out-of-Sample 2014
Loss Ratio
Risk Group
Predicted
Observed

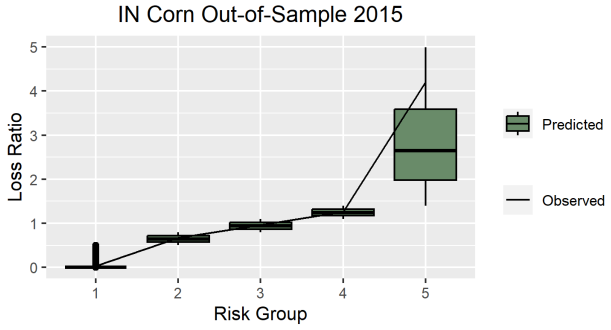

IL Corn In-sample 2015
Loss Ratio
Risk Group
Predicted
Observed

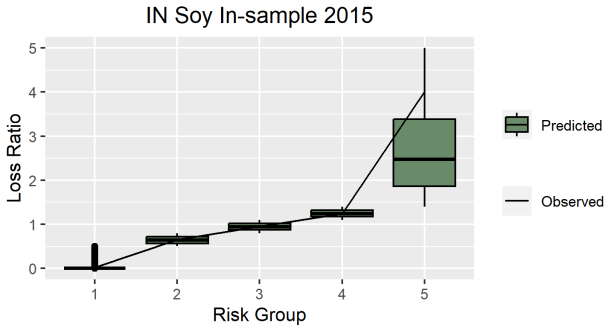

IL Corn Out-of-Sample 2015
Loss Ratio
Risk Group
Predicted
Observed

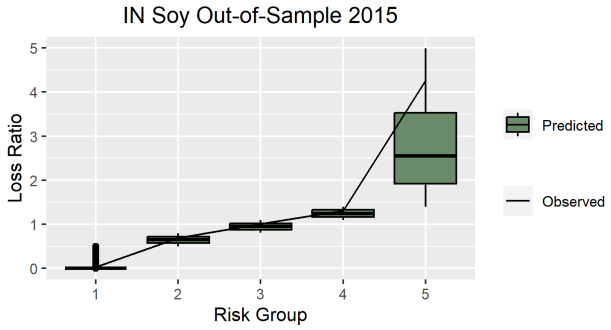

IL Soy In-sample 2015
Loss Ratio
Risk Group
Predicted
Observed

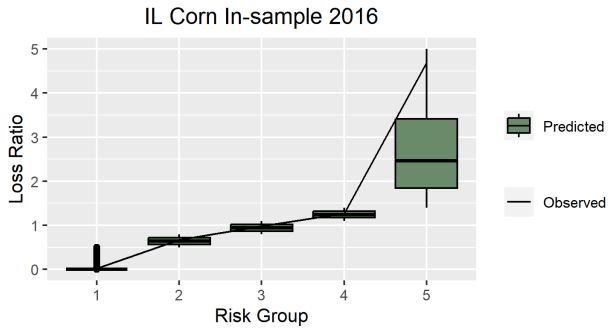

IL Soy Out-of-Sample 2015
Loss Ratio
Risk Group
Predicted
Observed

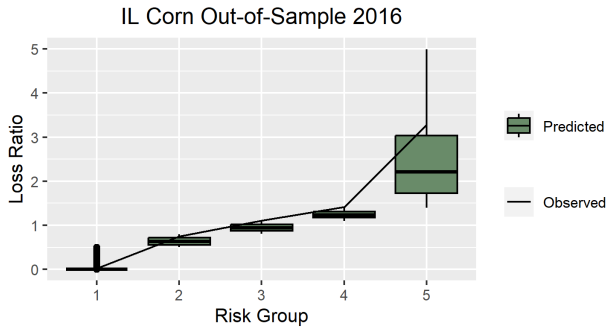

IN Corn In-sample 2015
Loss Ratio
Risk Group
Predicted
Observed

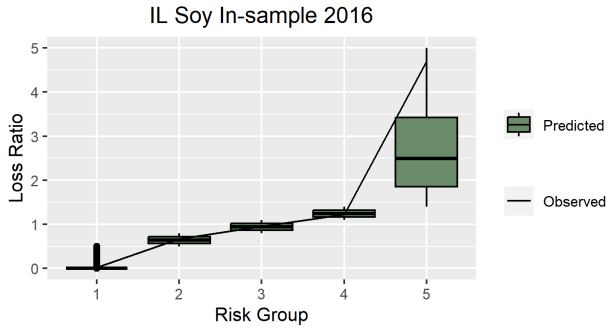

IN Corn Out-of-Sample 2015
Loss Ratio
Risk Group
Predicted
Observed

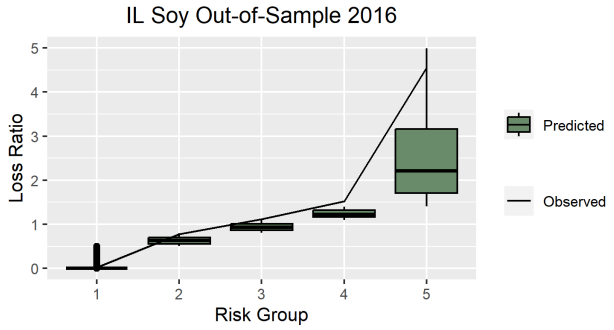

IN Soy In-sample 2015
Loss Ratio
Risk Group
Predicted
Observed

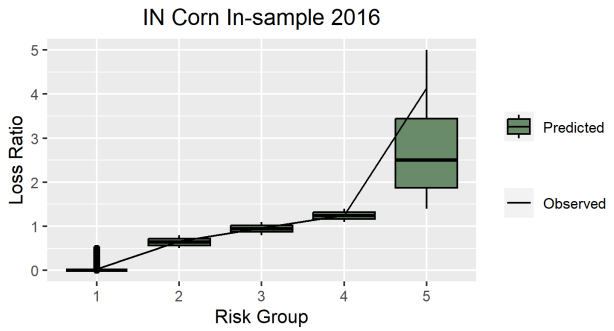

IN Soy Out-of-Sample 2015
Loss Ratio
Risk Group
Predicted
Observed

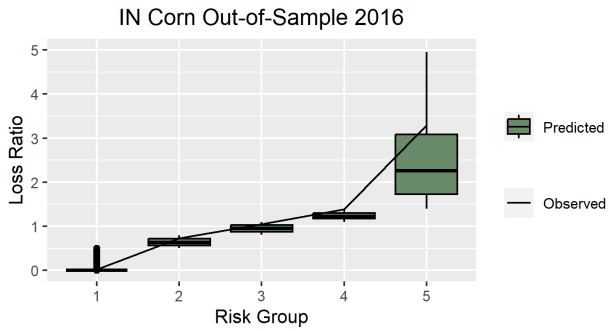

IL Corn In-sample 2016
Loss Ratio
Risk Group
Predicted
Observed

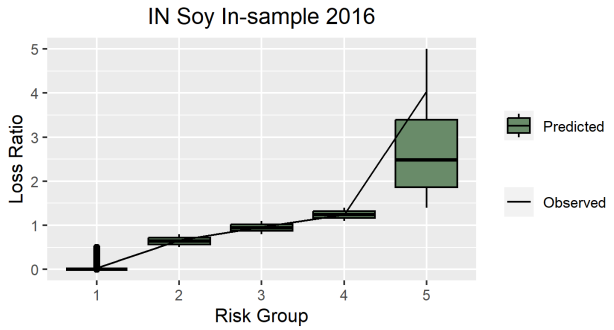

IL Corn Out-of-Sample 2016
Loss Ratio
Risk Group
Predicted
Observed

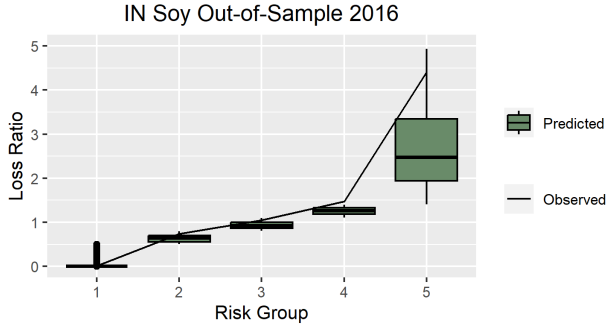

IL Soy In-sample 2016
Loss Ratio
Risk Group
Predicted
Observed

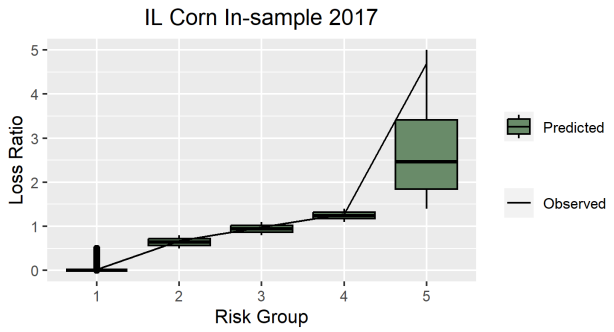

IL Soy Out-of-Sample 2016
Loss Ratio
Risk Group
Predicted
Observed

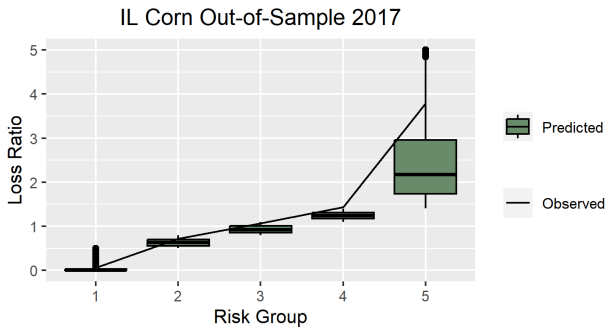

IN Corn In-sample 2016
Loss Ratio
Risk Group
Predicted
Observed

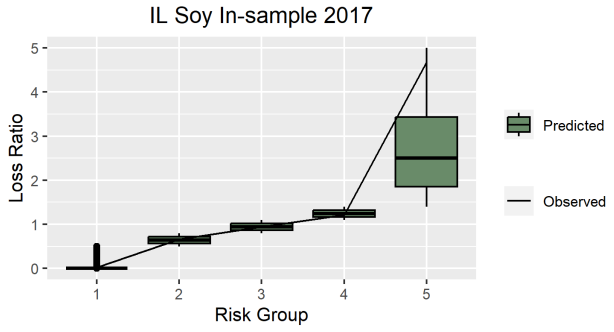

IN Corn Out-of-Sample 2016
Loss Ratio
Risk Group
Predicted
Observed

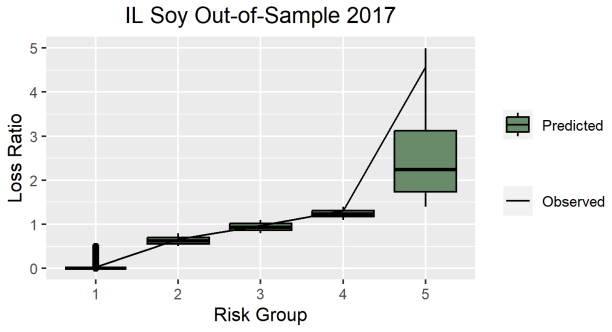

IN Soy In-sample 2016
Loss Ratio
Risk Group
Predicted
Observed

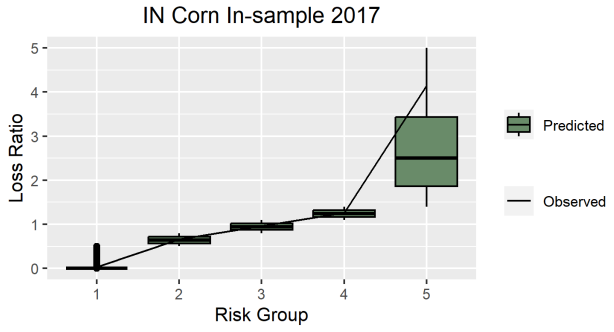

IN Soy Out-of-Sample 2016
Loss Ratio
Risk Group
Predicted
Observed

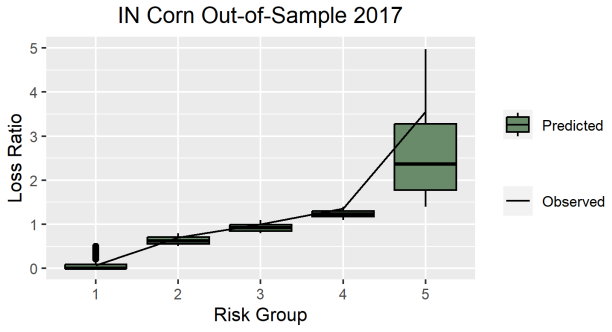

IL Corn In-sample 2017
Loss Ratio
Risk Group
Predicted
Observed

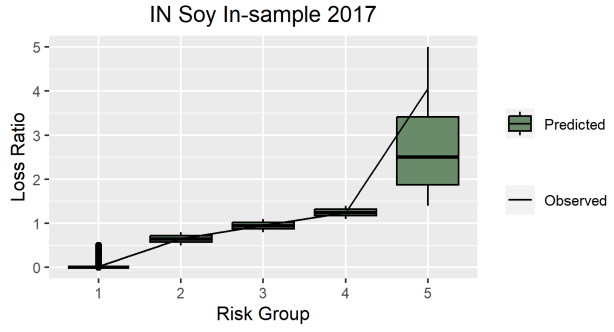

IL Corn Out-of-Sample 2017
Loss Ratio
Risk Group
Predicted
Observed

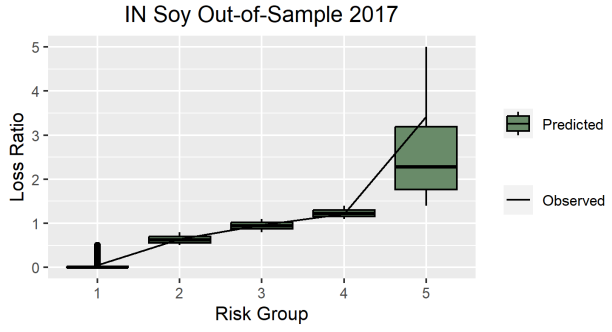

IL Soy In-sample 2017
Loss Ratio
Risk Group
Predicted
Observed


IL Soy Out-of-Sample 2017
Loss Ratio
Risk Group
Predicted
Observed


IN Corn In-sample 2017
Loss Ratio
Risk Group
Predicted
Observed


IN Corn Out-of-Sample 2017
Loss Ratio
Risk Group
Predicted
Observed


IN Soy In-sample 2017
Loss Ratio
Risk Group
Predicted
Observed


IN Soy Out-of-Sample 2017
Loss Ratio
Risk Group
Predicted
Observed