Assessing the Effectiveness of the Actuaries Climate Index for Estimating the Impact of Extreme Weather for Crop Yield and Insurance Applications

by

Qimeng Pan

A thesis submitted to
The Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements
of the degree of

MASTER OF SCIENCE

I.H. Asper School of Business
The University of Manitoba
Winnipeg, Manitoba, Canada
January 2021

Abstract

For crop production, extreme weather risk is a concern for farmers, governments, insurers, reinsurers, and other stakeholders. Some research estimates that as much as 90% of crop loss may be due to adverse weather for some crops in some locations. From an insurance/reinsurance perspective, which is the focus of this study, the ability to accurately estimate the frequency and severity of crop loss is critical for producing actuarial models. The Actuaries Climate Index (ACI) was launched recently by organizations representing the actuarial profession and is designed to help better understand the potential effects of climate trends and extreme weather events for actuarial applications. The ACI measures weather variables and sea levels within the United States and Canada at a scale of 2.5 degrees longitude (about 278.3km) by 2.5 degrees latitude (about 277.5km).

The objective of this research is to examine the effectiveness and feasibility of utilizing the ACI to improve estimates for crop yield and insurance/reinsurance applications through considering the impact of extreme weather. Three crop yield models are designed based on both linear and probit regression models, using corn yields from 8 midwestern states in the United States from 1961 to 2018 as the dependent variable and weather variables derived from the ACI as the independent variables. Model 1 considers the monthly ACI as the independent variable. Model 2 derives several individual components from the monthly ACI as independent variables, including temperature above the 90th percentile and below the 10th percentile, maximum monthly 5-day rainfall period, annual consecutive dry days, and wind speed above the 90th percentile. Model 3 considers the similar individual ACI components (excluding wind speed), but, based on an annual, rather than monthly, temporal scale. As a comparison, a high-level spatial resolution climate dataset including global monthly county-level climate information is employed to develop a high-level resolution ACI. Models 4 and 5 derive similar weather variables from high-level resolution data, including minimum and maximum temperature, precipitation, PDSI, and soil moisture, and also contains linear and probit regression models for Iowa and Midwest, respectively.

The results suggest that the ACI can help to explain corn yield. The linear regression Model 1 and probit Model 1 both include July and September ACI as two significant variables and produce an R^2 of 0.8917 and a pseudo R^2 of 0.1617, respectively. The simplest individual components of ACI linear regression, Model 2, presents an R^2 of 0.9535. Comparatively, the Probit Model 2 produces a pseudo R^2 of 0.3451, with only three significant variables, including May and September minimum and July maximum temperature. The R^2 and pseudo R^2 of the linear regression Model 3 and probit Model 3 is reduced to 0.7504 and 0.0739, respectively, both with minimum and maximum temperature as the only two significant variables. The simplest linear regression Model 4 with an R^2 of

0.7982 excludes July PDSI and June precipitation, two insignificant variables, while the simplest probit Model 4 presents a pseudo R^2 of 0.2864 and does not contain three more other insignificant variables, April minimum and June maximum temperature and June soil moisture. The R^2 of linear regression Model 5 is 0.7126 with all variables significant, while the pseudo R^2 of probit Model 5 is 0.2341 excluding two insignificant variables, July PDSI and August precipitation.

The results suggest that the ACI can provide some important insights when modelling crop yield, however, more research is needed as the model fit, and significant variables change depending on the temporal and spatial resolution of the data. In addition, the high-level resolution data might imply that crop yield and insurance are related to extreme weather events, even though R^2 decreases moderately when the research area extended, and high-level resolution ACI could be a reliable weather factor in pricing modelling.

Acknowledgement

Firstly, I wish to express my deepest gratitude to my thesis advisor, Dr. Lysa Porth, for her insightful advice and continuous support in completing this thesis. She consistently encouraged and allowed this paper to be my own work but steered me in the right direction whenever she thought I needed help. Without her persistent guidance and patience, the goal of my thesis would not have to be realized.

Besides my advisor, I would like to sincerely thank the rest members of my thesis committee, Dr. Hong Li, Dr. Milton Boyd, and Dr. Xikui Wang, for their valuable assistance and constructive suggestions throughout the completion of this thesis.

I also want to thank Dr. Lysa Porth and the I.H. Asper School of Business at the University of Manitoba for having provided me with the financial support for my M.Sc. studies.

Additionally, I appreciate the support and advice I have received from my classmates and friends to complete my thesis, particularly the insights from Mitchell Roznik. I also wish to show my gratitude to all professors who taught me invaluable theoretical knowledge and practical research tools during my M.Sc. studies.

Last but not least, I would like to thank my beloved parents for their unconditional mountains of love, support, and encouragement during the process of pursuing my M.Sc. studies and my life in general.

Contents

List of tables

List of figures

1. Introduction

1.1 Agriculture and extreme weather risks

In recent decades, global climate change has become one of the most complicated and serious challenges facing humans. Extreme weather risk has the potential to impact many people and affect resource shortage, food consumption, and human health (Lee & Nadolnyak, 2012). The Intergovernmental Panel on Climate Changes (IPCC) reports that the observed global mean surface temperature for the decade 2006 - 2015 was 0.87℃ higher than the average temperature over the last half of the 19th century. The National Oceanic and Atmospheric Administration (NOAA) reports that since 2001, seven of the eight warmest years on record have happened, and all of the ten warmest years have taken place since 1995. This means that even though the temperature may not have increased significantly, extreme climate events appear to be frequent.

Agriculture is the foundation of human development and social stability. The yields of many crops, and in many locations, are strongly related to weather variability, and the impact of severe weather conditions on crops can threaten agricultural production (Wheeler & Von Braun, 2013). The increase of extreme climate events exacerbates the volatility of agricultural production and severe agricultural loss. Lobell et al. (2011) posit that rising temperatures tend to lower maize and wheat yields in some countries. The trend is enough to offset the increased portion of crop yields contributed by technology, such as biotechnology, farming equipment, and practices, etc. Schlenker and Roberts (2009) find that extreme high temperature can result in predicted corn yields declining by approximately 7% in the U.S.

Moreover, Wheeler and Von Braun (2013) indicate that extreme weather events may have more negative impacts on crops planted in tropical areas and higher latitude regions, and influences become more severe with an increasing degree of extreme climate. There is significant research that focuses on the impact of extreme weather risks on crop yield using specific climate variables or weather index variables, such as the Palmer Drought Severity Index (PDSI), Cooling Degree Day (CDD), temperature, precipitation, etc. Although other research tends to focus on changes in the average over time, it is the frequency of severe weather that matters to insurers in agriculture (Lesk et al., 2016). In particular, for insurance purposes, it is not just frequency, but also severity that impacts agricultural loss.

1.2 Introduction to the Actuaries Climate Index

In 2016, the American Academy of Actuaries (AAA), the Casualty Actuarial Society (CAS), Canadian Institute of Actuaries (CIA), and the Society of Actuaries (SOA) launched the Actuaries Climate Index (ACI) that measures the observations of extreme weather and sea levels within the continent of the United States and Canada. The ACI is intended to help actuaries, insurance companies, governments, and the general public understand the potential effects of climate trends and extreme weather events better. Actuaries not only focus on how to measure the risks facing people in their everyday lives, but they also assist in mitigating, managing, and predicting the future risks people will face. As the climatologists, environmentalists, and agricultural scientists build models to assess the potential climate changes and the impacts of climate changes on their fields, actuaries establish models to analyze the effects of uncertain climate events on financial issues. Therefore, the ACI may provide actuaries with reliable information regarding the frequency and severity of extreme weather events, which is important for estimating and modelling risk probability.

1.3 Objective

To examine the effectiveness and feasibility of the ACI, this paper adopts the components of the ACI and the combined ACI variables as weather variables to build a statistical model to estimate the impacts of extreme weather events on crop yields in the Midwest region in the U.S. In addition, this research extends the ACI and develops a methodology for assessing and predicting the agricultural loss for crop insurance and reinsurance applications. The result is a better understanding of the key components of the ACI that are most important for estimating crop yields for insurance or reinsurance applications. Since the ACI was recently launched on November 30, 2016, there is very little research examining the importance and feasibility of the index for forecasting and managing risks. Therefore, to the best of my knowledge, this paper provides the first discussion on the strengths and weaknesses of the ACI for agriculture applications.

1.4 Actuaries Climate Index Background

The ACI has six components, including warm temperatures (T90: based on 90th percentile of both daily maximum and minimum temperature), cool temperatures (T10: based on 10th percentile of both daily maximum and minimum temperature), precipitation (P), drought (D), wind power (WP), and sea-level changes (S). In practice, ACI employs the maximum number of consecutive dry days in a year (CDD) and maximum consecutive 5-day precipitation amount in a month (Rx5day) to represent drought and precipitation, respectively. All components are presented as standardized anomalies and summarized on a monthly and seasonal basis. The standardized anomalies are determined by taking the difference between the component value for the current period and its mean value for the 1961-1990 reference period and divided by the standard deviation of the reference period. Then ACI is the average of the standardized anomalies of six components (T10 is subtracted in the index). Agriculture is highly sensitive to climate variability and weather extremes, such as heavy drought, floods, severe storms, heatwaves, and untimely freezes.

Therefore, ACI focuses on the effects of the extreme value of climate data. Although other published data tend to focus on changes in the average over time, it is the frequency of severe weather that matters to insurers.

1.5 The Actuaries Climate Index and Risk Classification

Insurance companies develop the premium for a given individual based on the information of the groups or regions expected to have the same costs as the individual. Actuaries generally group individuals with the same anticipated costs together as one block and then derive a product premium for them. Many challenges exist in estimating risk probability and calculating prices. For example, premiums may vary if the underlying costs vary, and risks may vary among the different insured individuals, and there is not a single method to solve all of them. Statistical models are typically used to predict the potential costs of providing coverage and other necessary additional expenses, and statistical approaches may produce better results when measuring large scales. Risk grouping is an important step to utilize statistical methods more effectively and efficiently (On Risk Classification, 2011). Risk classification is a kind of risk grouping method that can observe the risks or groups with substantially similar probabilities and characteristics over time. The ACI data may provide reliable information on the frequency and severity of extreme weather events, which could be the basis for estimating risk probability. Thus, actuaries could use historical climate data from the ACI dataset to group insureds and classify risk categories in pricing.

1.6 The Actuaries Climate Index and Reinsurance

Reinsurance is a type of insurance for insurance companies. Extreme climate events represent a possible failure to catastrophe reinsurance companies. By far, North America has the largest proportion of the reinsurance premiums capital (Murnane, 2004). In the past 30 to 40 years, the number of high-cost weather events in the U.S. has increased four times. In fact, in 2011, the U.S. experienced fourteen extreme climate-related events, each resulting in more than a $1 billion loss (Climate Science Watch). The occurrence of unusual extreme climate events and their potentially large losses are of concern to reinsurance companies (Della-Marta et al. 2010). Based on measurements from meteorological and coastal tide stations in the United States and Canada, the ACI is generated from observations of climate changes in temperature, sea level, precipitation, and wind speed. It is hypothesized that the ACI may improve the accuracy of estimating the frequency of rare events and contribute to building risk models that could be used by reinsurance companies in future business decisions.

1.7 The Actuaries Climate Risk Index

The Actuaries Climate Risk Index (ACRI) is under development to represent the relationships between economic losses and climate variables using regression analysis. The intent is to develop several ACRI's that can be used for a wide range of fields related to climate risk. Therefore, by employing the ACRI, actuaries only need to consider the risk measurements when measuring the risks of climate changes rather than using climate trending historical data that may not be sufficient to estimate the risks (Collins et al., 2020). Currently, ACRI has not been developed for agriculture applications. As a result, one of the outcomes of this research is an insight into a potential preliminary ACRI that may apply to the Midwest in the US for assessing climate risks in agriculture.

2. Literature

2.1 Statistical models for estimating impacts of weather for crop yields

Regression models are commonly adopted to assess the extreme climate risk probability distribution and impacts of extreme weather events on agricultural production based on historical climate data (Song, 2016). Production function models usually use cross-section or time-series data to evaluate the impacts of temperature, precipitation, labour, and fertilization for crop yields. In addition, regression analysis is employed to quantify the marginal effects of these factors and predict the crop yields in different climates. Compared to agronomic process-based models and economic models, statistical models are easy to implement and have fewer data requirements (Ward et al., 2014). For example, the most basic statistical model only needs historical yields and weather data. Furthermore, statistical models are more transparent in dealing with uncertainties compared to many other models. If a statistical model is unable to reflect the way that weather affects crop yields, it will have relatively low goodness of fit and insignificant estimation performance (Lobell & Burke, 2010).

Coble et al. (2010) recommend the use of a reweighting method applied to weather data over a long period, such as over 100 years, by employing a Generalized Linear Model (GLM) linked with a fractional logit regression model to build a weather index to estimate crop yield loss more reliably. Similarly, Woodard (2014) shows that time horizon is an essential factor in crop insurance ratemaking and develops a conditional Weibull model to determine the effects of sample period span and weather variation on yield risk and pricing estimation.

On the other hand, many researchers adopt three main types of statistical approaches, time series, cross-section, and panel data methods, to predict the crop yields and related insurance losses (Lobell & Burke, 2010; Verón et al., 2015; Ward et al., 2014). Lobell and Burke (2010) discuss the advantages and disadvantages of the most common statistical models, including time series, cross-section, and panel data methods and compare the process-based simulation methods to statistical models. When using the time series model, the most important is the choice of the spatial and temporal range. The cross-section model is capable of capturing how farmers adopt different climate conditions, and the panel data model can capture omitted factors and control unobserved heterogeneity by adding fixed effects variables. Verón et al. (2015) use panel regression to assess the impacts of temperature and precipitation on crop yields in the Pampas since 1971, and find that climate has relatively small but significant negative effects. Lesk et al. (2016) employ Superposed Epoch Analysis (SEA) in a time-series regression to assess the impact of extreme weather events on crop production, and observe that both drought and extreme heat lead to national production reduction globally.

Crop yields are seasonal and are affected by a variety of weather variables. Variable selection is necessary to reduce the complexity of regression models and decide which variables contribute most to explaining crop yields. Three regularization techniques based on the Ordinary Least Squares (OLS) regression, including ridge regression, lasso regression, and elastic net, can be applied to get the optimum regression models. All of the approaches shrink the regression coefficients, ridge regression shrinks the parameters close to zero, while lasso regression shrinks them exactly to zero. Therefore, ridge regression may prevent multicollinearity and reduce the model complexity, while lasso regression can select the most significant variables and reduce the model complexity. In other words, ridge regression works well when most parameters affect the result, while lasso works well when only a few parameters are related to the response. Elastic net combines the penalty approaches of lasso and ridge regression.

2.2 Machine learning-based methods and simulation approaches for predicting crop yields

The process of variable selection is also a machine learning (ML)-based method. ML methods are widely employed to assess and forecast the impacts of climate signals on crop yields. Hoffman et al. (2018) use the random forest approach to explore how climate and other variables affect crop yields and the relationship between crop yields and scale-compatible climate data from 1962 to 2014 in sub-Saharan Africa. This model captures the negative response of crop yields to increasing maximum temperature and low precipitation. Cai et al. (2019) adopt the lasso regression model with three ML approaches that include support vector machine, random forest, and neural network to build various empirical models for forecasting wheat yields from 2000 to 2014 in Australia. In general, they find precipitation and wet day frequency have a positive correlation with crop yields, while maximum temperature and potential evapotranspiration have a negative correlation with crop yields.

The empirical approach usually assumes that climate does not have a trend, so the estimation and forecasting process may disregard the sensitivity of crop yields to technology and climate trends. Burke and Lobell (2010) demonstrate that statistical approaches tend to overestimate yield losses from warming, especially in dry years, while the Crop Environment Resource Synthesis Maize (CERES-Maize) model (a simulation method) captures the interaction effects of warming and moisture. The simulation model simulates dynamic weather factors to establish a mathematical model to predict the crop yields under specific conditions. At present, representative crop simulation models include CERES used in the U.S., Agricultural Production Systems Simulator (APSIM) employed in Australia, and World Food Studies (WOFOST) developed in Netherland (Guo, 2015).

2.3 Weather related variables used to explain the impacts of extreme weather events for estimating crop yields

Except for typical temperature and precipitation variables (Blanc, 2012; Hoffman et al., 2018; Lobell et al., 2011; Sheehy et al., 2006; Verón et al., 2015), the PDSI is another weather variable commonly considered in many pieces of research (Rejesus et al., 2015; Woodard, 2014; Lee & Nadolnyak, 2012). PDSI subsumes the effects of both precipitation and temperature and provides a locally relative scale ranging from very wet to very dry conditions. Rejesus et al. (2015) use positive PDSI values to represent wet spells and negative PDSI values to represent drought conditions to assess whether longer time series weather data can provide a more accurate weather index to capture the impacts of weather events on crop losses. They separate the growing season (May to August) into two parts, including May to June and July to August, and then combine the period with positive and negative PDSI to get four parameters: May-June PDSI for positive values, May-June PDSI for negative values, July-August PDSI for positive values, and July-August PDSI for negative values. Woodard (2014) develops a summer average PDSI index (June, July, and August) and employs it to monitor significant weather events affecting crop yield growth during the critical growing season. While other weather variables (such as temperature, precipitation, and various degrees of integration or disaggregation levels, etc.) may have been used, the analysis does not suggest that using other PDSI index types, months or month combinations, or direct temperature and precipitation measures have a qualitative impact on the outcomes.

The normalized difference vegetation index (NDVI) and crop moisture index (CMI) are also used to represent drought conditions. NDVI is a satellite-derived remote sensing measure that can estimate the health of the vegetation during any given period by measuring greenness, which is thought to be closely correlated to crop yields. NDVI is easy to measure on a regular basis and is not subject to manipulation by agricultural producers or insurers. Most research focuses on the applicability of NDVI in index-based crop insurance, not indemnity-based crop insurance. Turvey and McLaurin (2012) investigate whether NDVI can be a reliable indicator in developing index-based agricultural insurance products and conclude there is a wide variation in relationships between crop yields and NDVI depending on locations and crop types. Makaudze and Miranda (2010) show that while NDVI, as a factor of catastrophic drought index insurance in Zimbabwe, has a relatively high correlation with corn and cotton yield loss, the relationship between NDVI and crop yield vary across crops and districts. The CMI indicates general conditions and not local variations caused by isolated rain. It is the sum of the evapotranspiration anomaly (which is generally negative or slightly positive) and excess moisture (either zero or positive), and is developed from some of the moisture accounting procedures used in computing the PDSI. The CMI gives the short-term or current status of purely agriculture drought or moisture surplus and can change rapidly from week to week.

Cooling degree day (CDD), growing degree days (GDD), and diurnal temperature range (DTR) are also widely considered to assess the impacts of temperature or extreme heat (Rejesus et al., 2015; Deschênes & Greenstone, 2007; Huang & Khanna, 2012; Turvey & McLaurin, 2012; Verón et al., 2015; Vogel et al., 2019; Ward et al., 2014). CDD refers to the degrees that a day's average temperature above 65 Fahrenheit degrees, and GDD represents the mean daily temperature above a certain temperature during the growing season. The DTR means the range of temperature between the high of the day and the cool of the night. Rejesus et al. (2015) develop the weather index using whole season CDD (from May to September) and June-July total CDD. The reason for using the June-July periods is that crop growth is frequently adversely affected by heat units. Turvey and McLaurin (2012) use the GDD standard of 80 Fahrenheit degrees to represent the instances of extreme heat. Deschênes and Greenstone (2007) calculate the GDD regarding a base of 46.4 Fahrenheit degrees and a ceiling of 89.6 Fahrenheit degrees to measure the impacts of temperature on agricultural productivity. Verón et al. (2015) calculate monthly means of DTR and then average it across the growing season. They then estimate the impacts of precipitation, temperature, and DTR on crop yields on a county basis through time series regression models and conclude that compared to temperature, DTR has relatively small negative impacts on corn, wheat, and soy in the Pampas since 1971. Vogel et al. (2019) demonstrate that a greater DTR indicates a higher warm temperature and a lower cool temperature in a day, both of which have negative effects on crop yields.

2.4 Other variables affect crop yields

The impact of technology is one of the main reasons for growing yields for many crop types in many countries. Technology contributes a positive effect on the long-term crop yields by making the crops more resistant to extreme weather events, such as drought and strong wind (Porth et al., 2019). Hoffman et al. (2018) use time and country as fixed effect variables to represent technology changes and conclude that technological advances especially explained the increasing yields in sub-Saharan Africa from 1962 to 2014. Goodwin and Piggott (2020) demonstrate that under similar drought conditions in 1988 and 2012 in 12 states in Corn Belt, the yield losses, as well as loss-cost ratios and loss ratios that reflect crop insurance risk, are significantly lower in the second period. In 2012, 91% of corn planted in Iowa was a genetically engineered variety that is more resistant to heat and moisture stresses, but no genetically engineered variety was available in 1988.

In addition, other factors, for instance, soil quality, management skill of farmers, government policy, and solar radiation are considerable consequences of crop insurance (Priest, 1996; Rejesus et al., 2015; Tollenaar et al., 2017; Ward et al., 2014; Woodard, 2014). Woodard (2014) uses soil productivity as an independent variable to control the farm heterogeneity through time and show that better soil results in higher yields under the same weather and technology conditions. Ward et al. (2014) document that failing to locate the account for the spatial dependence of each dataset would mislead the estimation of some coefficients. The government also plays a vital role whose responses to unusual and unexpected natural disasters will decline catastrophe losses and relieve the farmer's financial stress (Priest, 1996). If the government encourages farmers to plant a specific crop with agricultural subsidies, the yields of this crop in that period will be higher than normal. Tollenaar et al. (2017) find that solar radiation is considered accounting for the increase of corn yields in the U.S. Corn Belt from 1984 to 2013.

2.5 Extreme weather events and insurance applications

Extreme weather incidents result in huge losses for insurance and reinsurance firms, and crop insurance is bound to suffer as a result. Many large payments caused by extreme events depend on the risk loading included in the premium since the risk loading takes into account the likelihood and severity of extreme events (Roebber et al., 2017). Measuring the risk of extreme events provides more information to actuaries to climate insurance modelling. Jin and Erhardt (2018) use simulated Cooling Degree Day to build up a weather call option and measure tails of the payments of call option using conditional tail expectation (CTE) and Value-at-Risk (VaR). Then they utilize simulation-based methods to forecast the distributions of future daily temperature and insurance payments in California. The statistical ensemble models are essential to simulate and predict future indemnities or costs for insurance products so that actuaries could develop a better and more reasonable premium rate.

The following sections 3 and 4 will describe datasets and methodologies used in this research paper. The results will be exhibited and interpreted in section 5. Section 6 concludes this paper and includes discussion and implication. Section 7 indicates the next steps for further research.

3. Data

3.1 Crop yields data

One of the 12 subregions identified by the Actuaries Climate Index is the Midwest (MID) U.S., which contains eight states. It includes Illinois, Indiana, Iowa, Michigan, Minnesota, Missouri, Ohio, and Wisconsin, which is the sub-region of focus for the empirical analysis in this research. A set of state-level data for yields (bushel/acre) and areas planted (acres) for corn covering a 58 year (from 1961 to 2018) historical period in the eight states corresponding to the Midwest, U.S. is gathered from the United States Department of Agriculture (USDA)'s National Agricultural Statistical Service (NASS). The weighted average corn yields are aggregated from state to calculate the regional yields (bushel/acre) for all of the Midwest. Next, the corn yields are matched to the Midwest monthly ACI dataset for each year from 1961 to 2018.

3.2 The Actuaries Climate Index data

The corresponding standardized ACI data, including the monthly and yearly data for five components of the ACI (no sea-level changes records in the Midwest region) and the individual combined ACI in the Midwest, is obtained from the Actuaries Climate Index official website (https://actuariesclimateindex.org/data/). Variable selection is based on the growing season of corn (from April to September). The combined ACI variables selected in this paper are ACI_4, ACI_5, ACI_6, ACI_7, ACI_8, and ACI_9. In addition, individual components of ACI are June to July CDD (CDD_6 and CDD_7), May to August Rx5Days (Rx5Day_5, Rx5Day_6, Rx5Day_7, and Rx5Day_8), April, May and September T10 (T10_4, T10_5, and T10_9), June to August T90 (T90_6, T90_7, and T90_8), and June WP90 (WP90_6). Summer is the growing season of corn, so most of the variables utilized in this research are summer variables. The reason for adopting the June wind speed is that tornados usually happen in June in the Midwest. Frozen nights often take place in the spring and autumn, so low-temperature variables in April, May, and September are selected.

3.3 The high-level resolution climate data

The ACI data is composed of grid-level data, where each grid is 2.5 degrees longitude (about 278.3km) by 2.5 degrees latitude (about 277.5km). Based on this, each grid is developed to state and region level data. The grid level scale is relatively large since station level data is commonly used in climate-related research, and the concern is that large scale data might eliminate extreme observations and lead to underestimating or overestimating results. Therefore, another climate dataset employed in this research is a high-level spatial resolution global climate dataset called TerraClimate (more details in doi: 10.1038/sdata.2017.191). This dataset is based on an approximate 4km by 4km scale and contains monthly climate data, including historical weather data for global land surfaces, such as precipitation, PDSI, temperature, vapour pressure, etc. According to the ACI dataset, similar variables are selected, including PDSI, precipitation, minimum and maximum temperature, and soil moisture. Section 5.4 outlines each of the variables used in the estimation. This research contributes to developing a better understanding of the impact of extreme weather for estimating crop yield and insurance applications based on ACI data at first, then comparing the results from ACI data with the outcomes obtained from high-level spatial resolution data to discover the potential improvement of ACI.

4. Methodology

This paper applies the Linear regression and Probit regression models as two main statistical methods to estimate corn yields in the Midwest region in the U.S. Linear regression is widely used in estimation and prediction, measuring correlation and model fit. The probit regression model uses a binary response variable and belongs to the Generalized Linear Model (GLM), which makes it an appropriate model to estimate the occurrence of crop yield losses and is widely applied in insurance applications. Five crop yield models are designed based on both linear and probit regression models, using different scale-level corn yields as the dependent variable, and weather variables derived from the ACI and TerraClimate data as the independent variables. Details and results for each model will be explained in section 5.

4.1 Linear regression model

The linear regression is the first model applied, and is expressed as follows:

$$\boldsymbol{Y}_{ik} = \beta_{0k} + \boldsymbol{\beta}_j \boldsymbol{X}_{jik} + \gamma t + \boldsymbol{\epsilon}_{ik},$$

where $\boldsymbol{Y}_{ik}$ represents the weighted average corn yields of the whole Midwest for year i (i = 1961, …, 2018) or corn yields of state k in the Midwest for year i (i = 1961, …, 2016), $\boldsymbol{X}_{jik}$ is the group of j explanatory variables described in section 3 for year i and state k, and $\boldsymbol{\beta}_j$ is the estimated coefficient of each explanatory variable. In addition, β_{0k} is the intercept for state k, t is time effect, and $\boldsymbol{\epsilon}_{ik}$ represents the error terms following a normal distribution with (0, σ^2).

The distributions of the different aggregation level corn yields are shown in Figures 1-5. In addition, use the Jarque-Bera test and the Durbin-Watson test to check the normality and independence of residual items in five linear regression models, respectively. Results are in Table 16 and Table 17.

4.2 Generalized Linear Model – Probit regression model

The second model is probit regression, which estimates corn yield losses with the selected variables. Since the corn yields have an increasing trend overall, detrending is necessary to prepare the data and ensure a more accurate and representative result. After detrending, set the whole corn yield dataset to follow a binomial distribution in the following two steps:

Step 1: Compute the 25th percentile of detrended corn yields;

Step 2: Set the data that is less than the 25th percentile as “1”, which means there is a loss, and the part above the 25th percentile data is “0”, which means there is no loss.

Increased temperatures, drought conditions, and extreme weather disasters (EWDs) could reduce crop yields. For instance, corn yields decrease about 10% with each degree Celsius increase in the U.S. (Zhao et al., 2017), and EWDs would produce an average of 19.9% reduction in national cereal production in North America (Lesk et al., 2016). However, choosing the 25th percentile of detrended corn yields as the threshold of losses is a consideration for sample effectiveness and efficiency. A 10th percentile threshold would make the number of losses too small to generate a convincible result, particularly for Model 1, which includes only 58 observations. In addition, probit regression focuses on demonstrating that different resolution databases might produce inconsistent results for estimating crop yield losses.

The linear regression model is written as follows:

$$\boldsymbol{Y}_{ik} = \beta_{0k} + \boldsymbol{\beta}_j \boldsymbol{X}_{jik} + u.$$

Then probit regression model is:

$$P(\boldsymbol{Y}_{ik} = 1 | \boldsymbol{X}_{jik}) = \Phi(\cdot),$$

where $\boldsymbol{Y}_{ik}$ represents the weighted average corn yields of the whole Midwest for year i (i = 1961, …, 2018) or corn yields of state k in the Midwest for the year i (i = 1961, …, 2016), $\boldsymbol{X}_{jik}$ is the group of j explanatory variables described in section 3 for year i and state k, and $\boldsymbol{\beta}_j$ is the estimated coefficient of each explanatory variable. Also, β_{0k} is intercept for state k, and Φ (·) is the cumulative standard normal distribution function.

5. Results

Model 1 and Model 2 use the region-level annual weighted average corn yields as dependent variables, and monthly combined ACI and individual components of ACI as independent variables, respectively. Model 3 employs state-level annual corn yields and annual individual components of ACI as dependent and independent variables. Model 4 utilizes the county-level annual corn yields in Iowa as dependent variables and monthly standardized high-level spatial resolution weather variables as independent variables. Model 5 is the same as Model 4, but the research area expands to the whole Midwest region.

5.1 Midwest analysis with the combined Actuaries Climate Index

In this sub-section, the effectiveness of the individual combined ACI variables is assessed for estimating the corn yields in the whole Midwest region. The linear regression results are shown in Table 1 column 2, which has an R^2 value of 0.8917 and two significant variables, July and September ACI. The coefficient of July ACI illustrates that the extreme weather conditions may have negative effects on corn yields. However, the coefficient of the September ACI is positive, and it might be reasonable as well. The high daytime warm temperatures might offset the adverse effects of cold nights in September. The results suggest that the coefficients of ACI_4, ACI_5, ACI_6, and ACI_8 are not significant may due to multicollinearity. Column 3 is the linear regression results without the insignificant variables, dropping variables starting from the one with the highest p-value step by step. All remaining variables are significant, and the R^2 value is as high as 0.8901. The fit of the linear regression model implies that changes in ACI variables explain about 89% of the changes in corn yields across the Midwest. Overall, the model is significant, as shown by the significance of the F-test statistic at a 0.1% significant level.

Column 4 is the results of probit regression and contains two significant variables as well. The pseudo R^2 value of 0.1617 is calculated by the formula of

$1-\frac{loglikelihood\ (full\ model)}{loglikelihood\ (null\ model)}$, which is also known as McFadden R^2. The low pseudo R^2 might imply that the combined ACI might have fewer insights when modelling corn yield losses since a pseudo R^2 value between 0.2 and 0.4 could represent a very good fit.

Table 1 Midwest analysis with combined ACI variables

Variables	Linear regression	Linear regression	Probit regression
Intercept	66.8829***	67.15695***	-0.76085
ACI_4	3.5330		-0.44542
ACI_5	-1.5629		0.09833
ACI_6	-3.2987		0.80559
ACI_7	-6.9927*	-6.24302.	0.76382.
ACI_8	-5.6327	-6.16164.	0.55087
ACI_9	8.2069*	7.80479*	-1.11165*
time	1.8737***	1.86356***	
N	58	58	58
R^2 (Pseudo R^2)	0.8917	0.8901	0.1617
AIC			69.582
F-test statistic	58.82***	84.22***	

Note: The outcome variable is annual weighted average corn yields of the farms in the Midwest reported to USDA. Descriptions of other variables are in Table 10. The second column is the result of the linear regression model. Column 3 is the result of the linear regression with fewer variables. Column 4 is the result of the probit regression model. '.', '*', '**', and '***' indicate significance at 10%, 5%, 1%, and 0.1% levels, respectively.

5.2.1 Midwest analysis with individual component variables of Actuaries Climate Index using Linear regression model

The results of the linear regression model with individual components of the ACI are shown in Table 2. Although the R^2 value is as high as 0.9548 and the F-test statistic is high enough to imply that all explanatory variables together significantly explain corn yields, only four individual variables are statistically significant, and one of them (T10_9) is significant at a 10% level. Using the Farrar-Glauder test to diagnose multicollinearity among variables and the results show CDD_6 and CDD_7 are highly collinear variables. Therefore, the next step is to remove CDD_6 and CDD_7 from linear regression. The results shown in Table 2 column 3 are similar to the first model. Drop variables with the highest p-value step by step to get the simplest model with variables that are all significant. The final simplest model results in an R^2 of 0.9535 and five significant variables, shown in Table 2, column 4.

The backward stepwise selection method is employed to be a comparison to check the accuracy of the manual stepwise selection results. The results shown in Table 13 column 2 have the same estimated coefficients and significance levels as the manual selection results. Variable changes during each dropping step of the manual selection method could be observed, such as the estimated coefficients, level of significance, and the value of R^2 and pseudo R^2. This paper keeps using the manual selection method in the following analysis but provides the results from backward stepwise selection for comparison and reference in the appendix.

Table 2 Midwest analysis with individual component variables of ACI using Linear regression model

Variables	Linear regression	Linear regression without CDD	Simplest Linear regression
Intercept	67.11794***	67.17533***	67.32980***
CDD_6	2.14337		
CDD_7	-1.86621		
Rx5Day_5	0.41072	0.40784	
Rx5Day_6	-0.46413	-0.47983	
Rx5Day_7	1.98230	1.98341	1.98864.
Rx5Day_8	2.86024*	2.83325*	2.79095*
T10_4	-0.45676	-0.44421	
T10_5	0.11886	0.01343	
T10_9	-2.54756.	-2.61742.	-2.85146*
T90_6	0.72151	0.66719	
T90_7	-5.77264***	-5.78398***	-5.45714***
T90_8	-6.13024***	-6.09450***	-5.76785***
WP90_6	-0.62446	-0.57842	
Time	1.84387***	1.83969***	1.83177***
N	58	58	58
R^2	0.9548	0.9547	0.9535
AIC	425.5799	421.6728	411.2117
F-test statistic	64.82***	79.01***	174.2***

Note: The outcome variable is annual weighted average corn yields of the farms in the Midwest reported to USDA. Descriptions of other variables are in Table 10. The second column is the result of the original linear regression model. Column 3 is the result of the linear regression model without CDD variables. Column 4 is the result of the linear regression model that removed all insignificant variables. '.', '*', '**', and '***' indicate significance at 10%, 5%, 1%, and 0.1% levels, respectively.

5.2.2 Midwest analysis with individual component variables of Actuaries Climate Index using Probit regression model

The results of the probit regression model with individual component ACI variables are shown in Table 3. Use the same dropping variables approaches in section 5.2.1 to get the simplest model, and the simplest model is shown in column 4. Only three significant variables are in the simplest probit model, and two of them, T10_5 and T10_9 (May and September minimum temperature), are significant at a 10% level of significance. This result is not as expected as the results from the linear regression model in section 5.2.1. However, the pseudo R^2 value of the probit model could represent it might be a good fitness since it is between 0.2 and 0.4 and even close to 0.4. Furthermore, even though pseudo R^2 is gradually decreasing, the AIC value and p-value of the Chi-square test are both declining as well. It might demonstrate that the simplest model, as a whole, has a better fit compared to the original probit model (ANOVA chi-square test is based on the current model and the null model). The backward stepwise selection results in the same as the simplest model and is shown in Table 13, column 3.

Table 3 Midwest analysis with individual component variables of ACI using Probit regression model

Variables	Probit regression	Probit regression without CDD	Simplest Probit regression
Intercept	-0.83975**	-0.84186**	-0.7595***
CDD_6	-0.52693		
CDD_7	0.66512		
Rx5Day_5	0.12407	0.10113	
Rx5Day_6	0.09689	0.08025	
Rx5Day_7	-0.22456	-0.21207	
Rx5Day_8	-0.08674	-0.11425	
T10_4	0.21453	0.18391	
T10_5	0.35710	0.35282	0.4222.
T10_9	0.42315	0.39679	0.4200.
T90_6	0.15687	0.18821	
T90_7	0.75301*	0.76816*	0.7353**
T90_8	0.48152	0.48877	0.4278
WP90_6	0.24637	0.22502	
N	58	58	58
Pseudo R^2	0.4015	0.3973	0.3451
AIC	67.686	63.963	53.426
ANOVA Chi-square test (P-value)	0.01402*	0.005767**	0.0001338***

Note: The outcome variable is a binary variable transformed from the annual weighted average corn yields of the farms in the Midwest reported to USDA. Descriptions of other variables are in Table 10. Column 2 is the result of the original probit regression model. Column 3 is the result of the probit regression model without CDD variables. Column 4 is the result of the probit regression model that removed almost all insignificant variables. '.', '*', '**', and '***' indicate significance at 10%, 5%, 1%, and 0.1% levels, respectively.

5.3 Midwest analysis with annual state-level individual component variables of Actuaries Climate Index

The results in section 5.2 might imply that a relatively large spatial scale dataset might cause unexpected results, the absence of significant variables, since some extreme records may be averaged in some places. The following analysis employs the Midwest state-level raw data of the ACI.

ACI's state-level data is a yearly dataset rather than monthly data and only contains four independent variables, Rx5days, Tn10, Tx90, and CDD, from 1961 to 2016. Similarly, use linear regression and probit regression model to estimate the impacts of extreme weather on the Midwest state-level corn yields. State-level corn yields are also detrended and transformed to follow a binomial distribution to prepare for the probit regression. Among the four variables, Tn10 and Tx90 represent the percentage of days when the daily minimum temperature lower than the 10th percentile of the base period (1961-1990) and maximum temperature higher than the 90th percentile of the base period (1961-1990), respectively. These definitions are slightly different from the terms of T10 and T90.

Before running the regression, standardize the state-level raw data is necessary. 30 years (from 1961 to 1990) are selected as the reference period, and then the two steps below are followed (use CDD as an example):

Step 1: Calculate mean and standard deviation of CDD in the reference period 1961 to 1990, written as μ_f and σ_f;

Step 2: Calculate the standardized CDD for year i in each state k within Midwest from 1961 to 2016, using the formula as $\frac{CDD_{(i,\ k)}-\mu_f}{\sigma_f}$.

The results in Table 4, columns 2 and 3, illuminate that all of the explanatory variables together might be meaningful evidence since the F-test statistics are significant. Two variables out of four variables are significant at least a 1% level of significance. The R^2 value of the linear regression model is 0.7504, and the pseudo R^2 value of the probit regression model is 0.0739. After adding a state effect variable to the linear regression model, the R^2 value grows to 0.7607, and the significance level of the state variable might imply that locations are highly correlated to crop yields. However, all of these R^2's are still smaller than the results in section 5.2. This result might be due to that yearly data is an averaged evidence of the entire year and ignores the monthly or daily extrema. In addition, state-level might still be a large spatial scale and may eliminate local extrema.

Table 4 Midwest analysis with annual state-level individual component variables of ACI

Variables	Linear regression	Linear regression	Probit regression
Intercept	67.87581***	76.29208***	-0.62538***
Rx5Day	0.10005	-0.63556	0.06065
Tn10	-2.32393**	-2.51860**	0.09919.
Tx90	-5.73952***	-5.80187***	0.29302***
CDD	1.47515	0.77219	-0.06731
Time	1.67055***	1.66590***	
State		-0.29164***	
N	448	448	448
R^2 (Pseudo R^2)	0.7504	0.7607	0.0739
AIC	3809.78	3792.93	492.89
F-test statistic	265.8***	233.7***	

Note: The outcome variable in columns 2 and 3 is state-level corn yields and in column 4 is a binary variable transformed from the state-level corn yields of the farms in the Midwest reported to USDA. Descriptions of other variables are in Table 10. Column 2 is the result of the linear regression model. Column 3 is the result of the linear regression model added with a state effect variable. And column 4 is the result of the probit regression model. '.', '*', '**', and '***' indicate significance at 10%, 5%, 1%, and 0.1% levels, respectively.

5.4.1 Iowa analysis with standardized high-level resolution climate data using Linear regression model

Since the results using the ACI data are not as expected, this step uses the TerraClimate dataset, which is generated based on approximate 4km by 4km grid-level data and contains monthly county-level historical weather data. June to July PDSI (pdsi_6 and pdsi_7), May to August precipitation (pr_5, pr_6, pr_7, and pr_8), June to July soil moisture (soil_6 and soil_7), June to August maximum temperature (tmmx_6, tmmx_7, and tmmx_8), and April, May and September minimum temperature (tmmn_4, tmmn_5, and tmmn_9) data from 1961 to 2018 in Iowa are selected based on the corn growing season to estimate the corresponding county-level yearly corn yields.

Replicate the ACI development approach by selecting a 30-year period from 1961 to 1990 as the reference period and standardizing Iowa high-level resolution climate data in the following steps (using PDSI as an example):

Step 1: Calculate mean and standard deviation of PDSI for each month m in the reference period 1961 to 1990, written as $\mu_f(m)$ and $\sigma_f(m)$;

Step 2: Calculate the standardized PDSI for month m of year i in each county k within Iowa from 1961 to 2018, using the formula as $\frac{pdsi_{(m,\ i,\ k)}-\mu_f(m)}{\sigma_f(m)}$.

Same as the estimation using ACI data, linear regression is applied:

$$\boldsymbol{Y}_{ik} = \beta_{0k} + \boldsymbol{\beta}_j \boldsymbol{X}_{jik} + \gamma t + \boldsymbol{\epsilon}_{ik}$$

where $\boldsymbol{Y}_{ik}$ represents the corn yields of county k in Iowa for year i (i = 1961, …, 2018), $\boldsymbol{X}_{jik}$ is the group of j explanatory variables described in the first paragraph for year i and county k, and $\boldsymbol{\beta}_j$ is the estimated coefficient of each explanatory variable. β_{0k} is the intercept of county k, t is time effect, and $\boldsymbol{\epsilon}_{ik}$ represents the error terms following a normal distribution with (0, σ^2).

The results are shown in Table 5, column 2. The value of R^2 is 0.7982. It could demonstrate that this model employed with standardized high-level resolution climate data is appropriate, even though it is not as high as the R^2 in section 5.1 and 5.2.1. Except for two insignificant variables, July PDSI and June precipitation, all other variables are statistically significant. In addition, the F-test statistic is high enough to imply that all explanatory variables together significantly explain corn yields. Moreover, replicating the approach in section 5.2.1, remove pr_6 and pdsi_7 in the next two steps to get better fit models as a result (shown in Table 5 columns 3 and 4), and the significance level of the estimated coefficients is the same as the second column, and also the R^2 value of 0.7982 does not change. Comparing to the results in section 5.2.1, this simplest linear regression model contains more effective coefficients. The backward stepwise selection results in the same as the simplest model and is shown in Table 14, column 2.

Table 5 Iowa analysis with standardized high-level resolution climate data using Linear regression model

Variables	Linear regression	Linear regression without pr_6	Simplest Linear regression
Intercept	65.01353***	65.01403***	65.01863***
pdsi_6	3.07422*	3.06977*	2.68789***
pdsi_7	-0.46299	-0.45688	
pr_5	-2.78124***	-2.77376***	-2.77632***
pr_6	-0.02197		
pr_7	4.14047***	4.13523***	4.03047***
pr_8	-2.83129***	-2.83299***	-2.83344***
tmmn_4	-0.61862*	-0.62150*	-0.61710*
tmmn_5	-0.54592.	-0.56722.	-0.56257.
tmmn_9	5.08470***	5.08345***	5.08391***
tmmx_6	0.60512.	0.60769.	0.60951.
tmmx_7	-5.01798***	-5.01431***	-5.00016***
tmmx_8	-9.61604***	-9.61887***	-9.62043***
soil_6	3.12003***	3.09629***	3.04412***
soil_7	-8.29489***	-8.28874***	-8.29728***
Time	1.89845***	1.89844***	1.89820***
N	5742	5742	5742
R^2	0.7982	0.7982	0.7982
AIC	49197.69	49195.69	49193.75
F-test statistic	1506***	1614***	1738***

Note: The outcome variable is the annual county-level corn yields of the farms in Iowa reported to USDA. Descriptions of other variables are in Table 10. Column 2 is the result of the linear regression model. Column 3 is the result of the linear regression model without June precipitation. Column 4 is the result of the same linear regression model without another insignificant variable, July PDSI. ‘.’, ‘*’, ‘**’, and ‘***’ indicate significance at 10%, 5%, 1%, and 0.1% levels, respectively.

5.4.2 Iowa analysis with standardized high-level resolution climate data using Probit regression model

The results of the probit regression model using standardized high-level resolution climate data in Iowa are shown in Table 6. Column 3 is the probit regression model removing the most insignificant variable, June soil moisture. Then apply the same approach in section 5.2.1 to drop other insignificant variables, June precipitation, June maximum temperature, April minimum temperature, and July PDSI, step by step, to get the simplest probit model and display in column 4. All remaining variables are statistically significant at a 0.1% level of significance except for the May minimum temperature. The pseudo R^2 values of the original probit model and simplest probit model, 0.2867 and 0.2864, are very close and could represent both models are a good fit.

Furthermore, from column 2 to column 4, the AIC value has a tiny decrease, and the p-value of ANOVA Chi-square test is decreased but still insignificant. This demonstrates that the first model might represent good fitness, and the final simplest probit model could slightly improve the goodness of fit (ANOVA chi-square test is based on the current model and the original probit regression model). The backward stepwise selection results in the same as the simplest model and is shown in Table 14, column 3.

Table 6 Iowa analysis with standardized high-level resolution climate data using Probit regression model

Variables	Probit regression	Probit regression without soil_6	Simplest Probit regression
Intercept	-0.837520***	-0.83781***	-0.83610***
pdsi_6	-0.192399	-0.18535	-0.25164***
pdsi_7	-0.060811	-0.07146	
pr_5	0.265546***	0.26434***	0.26529***
pr_6	-0.009730	-0.01235	
pr_7	-0.214757***	-0.20945***	-0.21372***
pr_8	0.213755***	0.21389***	0.20966***
tmmn_4	0.031181	0.03109	
tmmn_5	0.043589.	0.04375.	0.05174*
tmmn_9	-0.239507***	-0.23952***	-0.23903***
tmmx_6	-0.020039	-0.01969	
tmmx_7	0.362900***	0.36229***	0.37875***
tmmx_8	0.602808***	0.60295***	0.59262***
soil_6	-0.009438		
soil_7	0.395052***	0.39028***	0.38068***
N	5742	5742	5742
Pseudo R^2	0.2867	0.2867	0.2864
AIC	4785.19	4783.22	4777.72
ANOVA Chi-square test (P-value)		0.8769	0.7723

Note: The outcome variable is a binary variable transformed from the annual county-level corn yields of the farms in Iowa reported to USDA. Descriptions of other variables are in Table 10. Column 2 is the result of the original probit regression model. Column 3 is the result of the probit regression model without June soil moisture. Column 4 results from the probit regression model without all other insignificant variables, July PDSI, June precipitation, and April minimum and June maximum temperature. '.', '*', '**', and '***' indicate significance at 10%, 5%, 1%, and 0.1% levels, respectively.

5.5.1 Midwest analysis with standardized high-level resolution climate data using Linear regression model

Expand the research area to the whole Midwest region and apply the ACI development approach to the Midwest high-resolution climate data. In the same way, use the same self-selected variables with a linear regression model. Table 7 column 2 shows that all variables are statistically significant at a 0.1% level except for June soil moisture. The R^2 value of 0.7126 is acceptable, although it is smaller than the R^2 values of the models using ACI data. It might imply that changes in high-level resolution weather variables could explain about 71.26% of the corn yields changes across the Midwest. F-test statistic is significant and demonstrates that all these variables together fit and explain corn yields well.

Least Absolute Shrinkage and Selection Operator (Lasso) regression approach is introduced to select variables and explore whether other else variables could be collected and increase the effectiveness of the high-level resolution climate data for estimating corn yields. Excluding the variables that are not in the growing season (broadly April to September), variables selected to be supplementary variables in the linear regression model are May PDSI (pdsi_5), September precipitation (pr_9), June to August minimum temperature (tmmn_6, tmmn_7, and tmmn_8), and May and September maximum temperature (tmmx_5 and tmmx_9). The results are in Table 7, columns 3 and 4. All variables are statistically significant, while the May minimum temperature is not. After dropping it from regression, no other insignificant variables show up again. The R^2 value remains as 0.7297 and is slightly larger than 0.7126. It might reveal that all these variables together are fitted better than the regression with only self-selected variables based on experience. The F-test statistics and AIC values display the same conclusion in like manner. The backward stepwise selection method is applied to the linear regression model with the variables selected based on both the Lasso approach and experience. The results are the same as the simplest linear model and are shown in Table 15, column 2.

Table 7 Midwest analysis with standardized high-level resolution climate data using Linear regression model

Variables	Linear regression	Linear regression (Lasso selection)	Simplest Linear regression (Lasso selection)
Intercept	62.41806 ***	63.696265***	63.593053***
pdsi_5		-7.358564 ***	-7.356485***
pdsi_6	4.87800***	14.477384***	14.471175***
pdsi_7	-3.88711***	-6.178563***	-6.178364***
pr_5	1.26769***	0.676166***	0.663937***
pr_6	5.45665***	2.114477***	2.120822***
pr_7	9.70496***	8.912569***	8.913359***
pr_8	0.49006***	-0.733426***	-0.730487***
pr_9		1.170014***	1.169836***
tmmn_4	1.80134***	-0.418122*	-0.425323*
tmmn_5	4.53892***	-0.097777	
tmmn_6		4.952973***	4.916094***
tmmn_7		6.389840***	6.375701***
tmmn_8		7.222588***	7.210296***
tmmn_9	2.99060***	-4.724990***	-4.729296***
tmmx_5		4.479994***	4.394181***
tmmx_6	7.97478***	4.156479***	4.189060***
tmmx_7	-7.27719***	-11.281069***	-11.271298***
tmmx_8	-9.84868***	-17.532366***	-17.524793***
tmmx_9		7.555767***	7.564631***
soil_6	0.52865.	0.909000**	0.910250**
soil_7	-8.86023***	-8.340331***	-8.339541***
Time	1.67237***	1.566900***	1.566997***
N	39567	39567	39567
R^2	0.7126	0.7297	0.7297
AIC	352752.4	350338.1	350336.1
F-test statistic	6539***	4854***	5085***

Note: The outcome variable is annual county-level corn yields of the farms in the Midwest reported to USDA. Descriptions of other variables are in Table 10. Columns 2 is the result of the linear regression models with variables selected based on experience. Columns 3 and 4 result from the linear regression models with variables determined by both the Lasso approach and experience, and column 4 removes one insignificant variable, May minimum temperature. ‘.’, ‘*’, ‘**’, and ‘***’ indicate significance at 10%, 5%, 1%, and 0.1% levels, respectively.

5.5.2 Midwest analysis with standardized high-level resolution climate data using Probit regression model

The results of the probit regression model with standardized high-level resolution climate data are shown in Table 8. Column 3 is the result of the simplest probit model without the insignificant variables, July PDSI and August precipitation. All remaining variables are statistically significant at a 0.1% level of significance, except for the June PDSI. The pseudo R^2 of the simplest model, 0.2341, is the same as the pseudo R^2 of the original probit regression and could represent a good fitness.

Columns 4 and 5 are the results of probit regressions using the variables selected through the Lasso approach. September precipitation and June minimum temperature are removed due to insignificance, and the simplest model is shown in column 5. With more variables in the probit regression, the pseudo R^2 of these two models are moderately improved to 0.2536. In addition, the AIC value and p-value of the ANOVA Chi-square test of the probit regression model with more variables (Lasso selection) are smaller than the values of the probit model only with self-selected variables, and demonstrate that the probit model with more variables could better explain the effectiveness of the high-level resolution ACI for estimating corn yield losses (ANOVA Chi-square test is based on the current model and the original probit regression). The backward stepwise selection method is applied to the original probit model and the probit model with the variables selected based on both the Lasso approach and experience. The results are the same as the simplest models and are shown in Table 15, columns 3 and 4.

Table 8 Midwest analysis with standardized high-level resolution climate data using Probit regression model

Variables	Probit regression	Simplest Probit regression	Probit regression (Lasso selection)	Simplest Probit regression (Lasso selection)
Intercept	-0.757158***	-0.756608***	-0.724359***	-0.726714***
pdsi_5			0.625632***	0.626712***
pdsi_6	-0.074283*	-0.019064.	-0.900663***	-0.906566***
pdsi_7	0.064830		0.267352***	0.269595***
pr_5	-0.107747***	-0.105575***	-0.064896***	-0.065205***
pr_6	-0.283505***	-0.280150***	-0.104547***	-0.111829***
pr_7	-0.455160***	-0.439158***	-0.445186***	-0.443344***
pr_8	0.001954		0.040559***	0.041241***
pr_9			-0.014300	
tmmn_4	-0.136050***	-0.137153***	-0.060383***	-0.063977***
tmmn_5	-0.152961***	-0.153493***	-0.051977.	-0.066655*
tmmn_6			-0.043355	
tmmn_7			-0.319458***	-0.328385***
tmmn_8			-0.226305***	-0.226689***
tmmn_9	-0.156244***	-0.155197***	0.051380*	0.034804.
tmmx_5			-0.086972**	-0.075420**
tmmx_6	-0.349582***	-0.352383***	-0.372841***	-0.411686***
tmmx_7	0.324255***	0.321780***	0.544415***	0.555092***
tmmx_8	0.480141***	0.479800***	0.711735***	0.708800***
tmmx_9			-0.151694***	-0.134287***
soil_6	0.067226***	0.073061***	0.064598**	0.066103**
soil_7	0.432060***	0.430947***	0.410868***	0.410685***
N	39567	39567	39567	39567
Pseudo R^2	0.2341	0.2341	0.2537	0.2536
AIC	35294	35293	34407	34406
ANOVA Chi-square test (P-value)		0.2529	<2e-16***	< 2.2e-16***

Note: The outcome variable is a binary variable transformed from the annual county-level corn yields of the farms in the Midwest reported to USDA. Descriptions of other variables are in Table 10. Columns 2 and 3 are the results of the probit regression models with the variables selected based on experience, and column 3 removes two insignificant variables, July PDSI and August precipitation. Columns 4 and 5 result from the probit regression models with variables selected based on both the Lasso approach and experience, and column 5 removes two insignificant variables, September precipitation and June minimum temperature. ‘.’, ‘*’, ‘**’, and ‘***’ indicate significance at 10%, 5%, 1%, and 0.1% levels, respectively.

6. Conclusion and Implication

This paper shows that the Actuaries Climate Index might be used to estimate crop yields and yield losses, even though the results are not as significant as the results obtained using historical data. The relatively low spatial resolution of ACI may be the reason, and it might also result in underestimating the occurrence and impacts of extreme weather events. The results become to be significant after the original ACI data is substituted by a high-level resolution dataset, the TerraClimate dataset. Therefore, if the ACI could be developed to be a higher-resolution index with a shorter time scale and narrower spatial scale, it might be more effective and applicable. However, the TerraClimate dataset is not detrended and not recorded as extreme weather data, which may cause overestimation. Also, county-level aggregation might still weaken the relationships between extreme weather events and crop yields.

The analysis displayed in this paper is about crop yield loss estimation, which could be a reference in insurance ratemaking. In insurance pricing, the net premium (total expected loss of claims) and a risk loading that is proportional to the net premium are usually the product price compositions. Generally, the total expected claim amount is composed of the severity and the frequency of the claims. A simple decomposition could be displayed by a continuous variable and a binary variable (such as the response variables in section 4.2). The binary variable could illustrate whether a claim happened, and the continuous variable could indicate the amount of a claim. GLM is a widely used method to estimate the expectation of these variables (Heras et al., 2018; Goldburd et al., 2019). The advantages of GLM compared to OLS have been addressed in section 4, which could verify that GLM is a more practical model. Poisson regression is widely applied in modelling the claim frequency, and Gamma distribution is employed for estimating claim severity (Ohlsson & Johansson, 2010). In addition, Logit and Probit models are typical models utilized in insurance pricing modelling and are commonly used to estimate and predict the possibility of claims (Frees, 2010). In general, the logit model is preferred in researches. However, when the number of failures (frequency of claims; also, number of ones) is far less than the occurrence of success (number of zeros in the response variable), and the independent variables are not only binary variables, the Probit model is preferred rather than the Logit model under a non-weighted version test (Jin et al., 2005).

High-level resolution climate data and Actuaries Climate Index both are historical data that could be utilized as a pricing factor in pricing modelling. Leeuw et al. (2014) indicate that risk is not uniformly distributed among insureds (or exposures), and insurance companies could accordingly adjust the premiums for different territories. To achieve this objective, insurers could use the historical claim and payment records of each insured. They also could use spatial climate data to ascertain the regions where encountered extreme weather events and would have increased risk to occur claims to address higher premiums. Within a smaller division, pricing for insurance products would be more targeted and accurate.

7. Further Research

Prediction is critical in insurance pricing, as mentioned earlier, so actuaries could develop and renew the premium of insurance products rely on previous claims or losses information. To evaluate the effectiveness of ACI and high-level resolution ACI for predicting crop yields, this paper employs rolling window regression and the Diebold & Mariano test to compare the predictive accuracy and forecast ability of four linear regression models described in section 5. Select a 30-year window width and run rolling window regression. Then compare each regression with a corresponding simple random walk model. The preliminary results are shown in Table 9, and p-values could reveal the predictive accuracy of the two methods in each model. Based on $\alpha = 0.05$, the p-values suggest that two methods in Model 1 and Model 2 have the same forecast accuracy, while the linear regressions of Model 3 and Model 5 are more accurate than the simple random walk model, which might illustrate that the higher-level resolution ACI is better at predicting corn yields than ACI.

Table 9 Predictive ability of different spatial scale ACI

	Model 1	Model 2	Model 3	Model 5
P-value	0.2616	0.0748	1.332e-08	<2.2e-16

Model 1: Linear regression with monthly combined ACI in the Midwest.

Model 2: Linear regression with monthly individual component variables of ACI in the Midwest.

Model 3: Linear regression with annual state-level individual component variables of ACI in the Midwest.

Model 5: Linear regression with monthly high-level (county-level) resolution ACI in the Midwest.

Moreover, exploring the relationship between insurance/reinsurance and ACI and high-level resolution ACI using loss-cost ratios and loss ratios information will be an essential next research step. In the future, insurance companies and the agriculture department could develop individually targeted Actuaries Climate Index in detail and apply it in insurance pricing, valuation, prediction, and agroeconomics analysis.

Appendix

Table 10 Abbreviations and descriptions of variables included in the dataset

Variable abbreviation	Variable description
CDD	Maximum number of consecutive dry days in a year with precipitation less than one millimeter
Rx5Day	Maximum rainfall per month in five consecutive days
T10	Change in frequency of cooler temperatures below the 10th percentile
Tn10	Percentage of days when the daily minimum temperature less than 10th percentile of reference period
T90	Change in frequency of warmer temperatures above the 90th percentile
Tx90	Percentage of days when the daily maximum temperature greater than 90th percentile of reference period
WP90	Frequency of wind speed above the 90th percentile
psdi	Palmer Drought Severity Index
pr	Precipitation
tmmn	Average monthly minimum temperature
tmmx	Average monthly maximum temperature
soil	Soil moisture
time	Estimation year

Table 11 Descriptive summary of Actuaries Climate Index

Midwest ACI data N = 58				
Variables	Min.	Max.	Mean	Std. dev.
CDD_6	-1.5200	1.6200	-0.2228	0.826686
CDD_7	-1.5100	1.5800	-0.2228	0.837129
Rx5Day_5	-2.3600	3.1700	0.2216	1.075954
Rx5Day_6	-2.8600	2.7900	0.3293	1.086263
Rx5Day_7	-1.9500	2.4500	0.1081	1.013655
Rx5Day_8	-1.8600	2.6100	0.0603	1.026407
T10_4	-1.5800	5.3400	0.0966	1.257771
T10_5	-1.5000	2.0600	-0.0107	0.931239
T10_9	-1.7300	2.9100	-0.1960	0.947877
T90_6	-1.4900	3.4200	0.0419	0.869365
T90_7	-1.4200	3.3100	-0.0452	1.072289
T90_8	-1.1400	3.2300	-0.0090	0.916306
WP90_6	-2.0000	2.4700	-0.0519	0.959081

Midwest state-level ACI data N = 448				
Variables	Min.	Max.	Mean	Std. dev.
Rx5Days	-3.0069	3.3768	0.1416	7.055175
Tn10	-3.4417	2.6049	-0.6697	2.834646
Tx90	-2.6841	3.5244	-0.0742	3.126361
CDD	-1.7955	3.2369	-0.0929	5.162458

Table 12 Descriptive summary of standardized TerraClimate high-resolution data

	Iowa N = 5742				Midwest N = 39567			
Variables	Min.	Max.	Mean	Std. dev.	Min.	Max.	Mean	Std. dev.
pdsi_5	-2.2089	2.4257	0.2369	0.9235	-2.8968	2.7909	0.2197	0.9634
pdsi_6	-2.1255	2.8809	0.2810	0.9640	-3.0956	3.2119	0.2450	0.9805
pdsi_7	-2.0579	3.1031	0.2855	1.0049	-3.3342	3.4489	0.2567	1.0043
pr_5	-2.1374	4.1245	0.2233	1.0853	-2.0385	4.9146	0.1300	1.0406
pr_6	-1.9785	4.4164	0.2075	1.1000	-2.0814	5.6135	0.1683	1.0733
pr_7	-2.1185	4.7036	-0.0063	1.0682	-2.1026	5.3021	0.0378	1.0226
pr_8	-1.6730	3.9577	0.0393	1.0099	-2.0023	4.9549	-0.0004	0.9564
pr_9	-1.6081	3.4128	-0.1191	0.8677	-1.7247	5.2737	-0.0442	0.9430
tmmn_4	-4.2684	3.0962	0.0633	1.0847	-3.7637	3.3897	0.0744	1.0185
tmmn_5	-2.3074	3.3706	0.1385	0.9498	-3.5180	3.3233	0.1413	0.9930
tmmn_6	-3.2213	3.9836	0.3211	1.0449	-5.1971	3.4176	0.2040	0.9881
tmmn_7	-3.2254	3.5331	0.0854	1.0876	-4.7108	3.1205	0.1064	1.0149
tmmn_8	-2.9774	3.8303	0.1778	1.0520	-3.6873	3.3509	0.1649	1.0146
tmmn_9	-3.0840	3.0610	0.1807	1.0536	-3.3779	3.0657	0.0988	0.9565
tmmx_5	-2.3430	2.9852	-0.0271	0.9485	-3.9462	2.9478	0.0162	0.9473
tmmx_6	-3.6915	3.0492	-0.0573	0.9724	-5.0084	2.8003	0.0058	0.9499
tmmx_7	-3.8262	3.5587	-0.1971	1.1176	-5.0404	4.2726	-0.0735	1.0334
tmmx_8	-2.6017	3.7864	-0.0768	0.9583	-3.6573	3.6903	0.0115	0.9895
tmmx_9	-3.9046	2.7694	0.1893	1.0727	-4.0176	2.6574	0.0828	0.9721
soil_6	-1.4541	3.6462	0.3125	1.0964	-1.5698	4.4239	0.1609	1.0362
soil_7	-1.3344	4.8123	0.2354	1.1375	-1.3969	6.0008	0.1532	1.0864

Table 13 Backward stepwise selection for Model 2

Variables	Simplest Linear regression	Simplest Probit regression
Intercept	67.32980***	-0.7595***
CDD_6		
CDD_7		
Rx5Day_5		
Rx5Day_6		
Rx5Day_7	1.98864.	
Rx5Day_8	2.79095*	
T10_4		
T10_5		0.4222.
T10_9	-2.85146*	0.4200.
T90_6		
T90_7	-5.45714***	0.7353**
T90_8	-5.76785***	0.4278
WP90_6		
Time	1.83177***	
N	58	58
R^2 (Pseudo R^2)	0.9535	0.3451
AIC	411.2117	53.426
F-test statistic /ANOVA Chi-square test (P-value)	174.2***	0.0001338***

Note: Results from backward stepwise selection show the same significant variables as the results from the manual selection approach. While the results are the same, the process of the manual selection method would provide the information about the changes in estimated coefficients, significance levels, R^2 (Pseudo R^2) values, and F-test statistics during each step.

Table 14 Backward stepwise selection for Model 4

Variables	Simplest Linear regression	Simplest Probit regression
Intercept	65.01863***	-0.83610***
pdsi_6	2.68789***	-0.25164***
pdsi_7		
pr_5	-2.77632***	0.26529***
pr_6		
pr_7	4.03047***	-0.21372***
pr_8	-2.83344***	0.20966***
tmmn_4	-0.61710*	
tmmn_5	-0.56257.	0.05174*
tmmn_9	5.08391***	-0.23903***
tmmx_6	0.60951.	
tmmx_7	-5.00016***	0.37875***
tmmx_8	-9.62043***	0.59262***
soil_6	3.04412***	
soil_7	-8.29728***	0.38068***
Time	1.89820***	
N	5742	5742
R^2 (Pseudo R^2)	0.7982	0.2864
AIC	49193.75	4777.72
F-test statistic /ANOVA Chi-square test (P-value)	1738***	0.7723

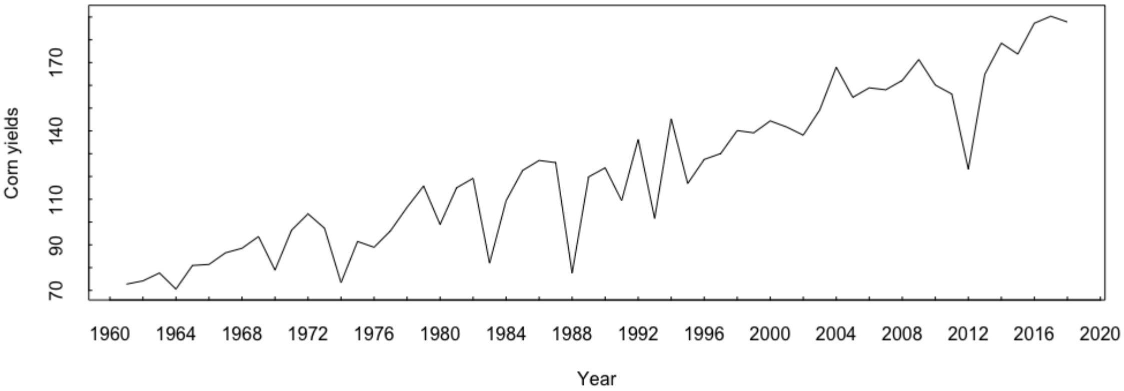
Note: Results from backward stepwise selection show the same significant variables as the results from the manual selection approach. While the results are the same, the process of the manual selection method would provide the information about the changes in estimated coefficients, significance levels, R^2 (Pseudo R^2) values, and F-test statistics during each step.

Table 15 Backward stepwise selection for Model 5

Variables	Simplest Linear regression (Lasso selection)	Simplest Probit regression	Simplest Probit regression (Lasso selection)
Intercept	63.593053***	-0.756608***	-0.726714***
pdsi_5	-7.356485***		0.626712***
pdsi_6	14.471175***	-0.019064.	-0.906566***
pdsi_7	-6.178364***		0.269595***
pr_5	0.663937***	-0.105575***	-0.065205***
pr_6	2.120822***	-0.280150***	-0.111829***
pr_7	8.913359***	-0.439158***	-0.443344***
pr_8	-0.730487***		0.041241***
pr_9	1.169836***		
tmmn_4	-0.425323*	-0.137153***	-0.063977***
tmmn_5		-0.153493***	-0.066655*
tmmn_6	4.916094***		
tmmn_7	6.375701***		-0.328385***
tmmn_8	7.210296***		-0.226689***
tmmn_9	-4.729296***	-0.155197***	0.034804.
tmmx_5	4.394181***		-0.075420**
tmmx_6	4.189060***	-0.352383***	-0.411686***
tmmx_7	-11.271298***	0.321780***	0.555092***
tmmx_8	-17.524793***	0.479800***	0.708800***
tmmx_9	7.564631***		-0.134287***
soil_6	0.910250**	0.073061***	0.066103**
soil_7	-8.339541***	0.430947***	0.410685***
Time	1.566997***		
N	39567	39567	39567
R^2 (Pseudo R^2)	0.7297	0.2341	0.2536
AIC	350336.1	35293	34406
F-test statistic /ANOVA Chi-square test (P-value)	5085***	0.2529	< 2.2e-16***

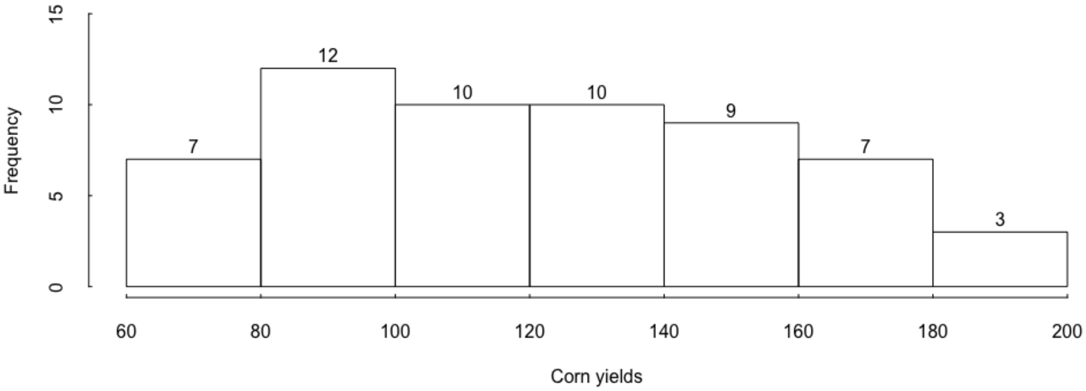
Note: Results from backward stepwise selection show the same significant variables as the results from the manual selection approach. While the results are the same, the process of the manual selection method would provide the information about the changes in estimated coefficients, significance levels, R^2 (Pseudo R^2) values, and F-test statistics during each step.

Figure 1: Annual weighted average corn yields in the Midwest (N=58)



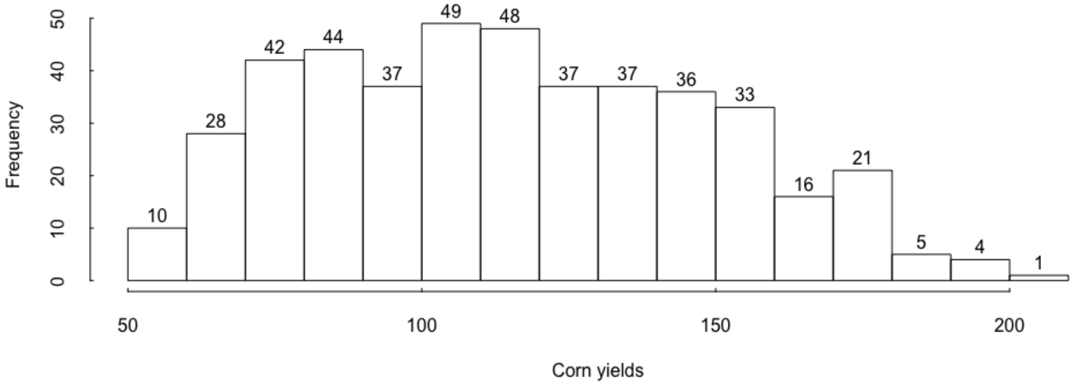
Note: The 1958-2018 annual weighted average corn yields in the Midwest show an upward trend.

Figure 2: Distribution of annual weighted average corn yields in the Midwest (N=58)



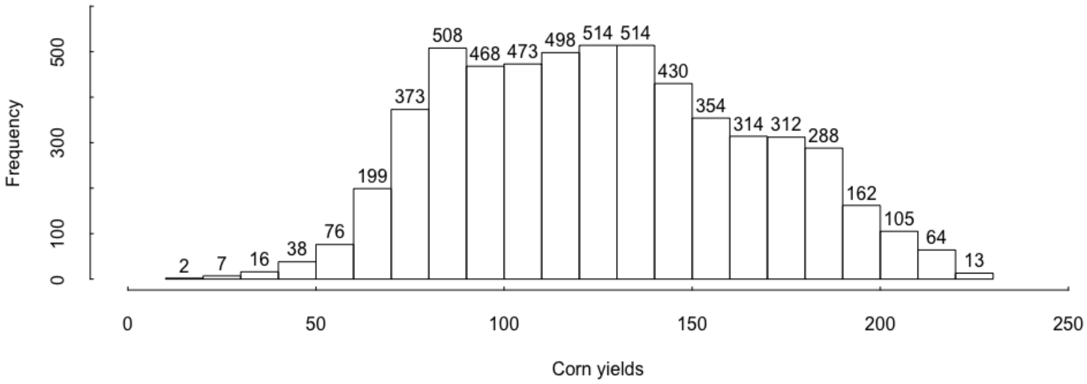
Note: The 1958-2018 annual weighted average corn yields in the Midwest show a roughly right-skewed distribution.

Figure 3: Distribution of annual state-level corn yields in the Midwest (N=448)



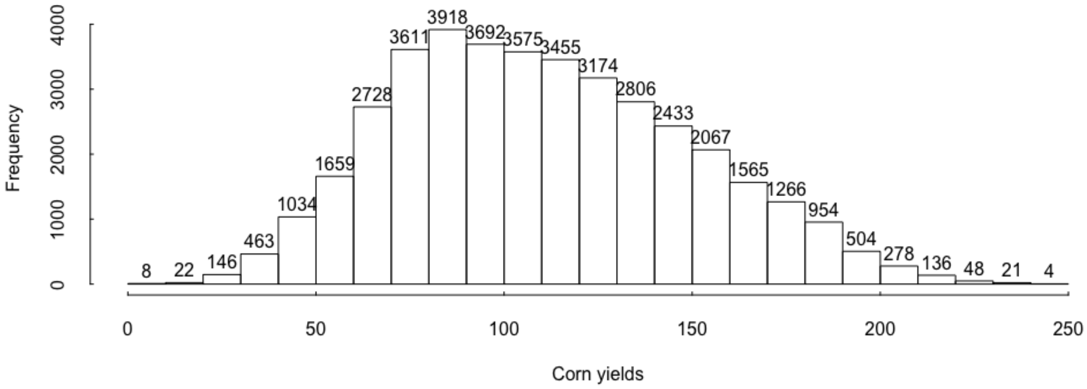
Note: The annual corn yields from 1958 to 2018 of all states in the Midwest show an approximately right-skewed distribution.

Figure 4: Distribution of annual county-level corn yields in Iowa (N=5742)



Note: The annual corn yields from 1958 to 2018 of all counties in Iowa show a roughly normal distribution.

Figure 5: Distribution of annual county-level corn yields in the Midwest (N=39567)



Note: The annual corn yields from 1958 to 2018 of all counties in the Midwest show a right-skewed distribution.

Table 16 Jarque-Bera test for normality assumption of the Linear regression models

	Model 1	Model 2	Model 3	Model 4	Model 5
Original Linear	4.655e-07	0.0002398	4.971e-09	< 2.2e-16	< 2.2e-16
Simplest Linear	4.497e-09	7.909e-06		< 2.2e-16	< 2.2e-16

Note: The Jarque-Bera test is applied to the linear regressions of the five models considered. Based on $\alpha = 0.05$, the p-value is less than 0.05 illustrating the residuals in the linear regression are not normally distributed. All five models do not meet the normality assumption.

Table 17 Durbin-Watson test for independence assumption of the Linear regression models

	Original Linear regression		Simplest Linear regression	
	D-W Statistic	P-value	D-W Statistic	P-value
Model 1	1.976757	0.806	1.98921	0.774
Model 2	1.74731	0.234	1.757969	0.312
Model 3	1.458714	0		
Model 4	1.156747	0	1.156669	0
Model 5	1.170486	0	1.20001	0

Note: The Durbin-Watson (D-W) test is applied to the linear regressions of the five models. The Original model refers to the linear regression based on all selected variables, while the Simplest model refers to the linear regression based only on the significant variables. The simplest linear regression, Model 5, selects variables based on both the Lasso approach and experience. The D-W test statistic has results between 0 and 4, where a value equal to 2 indicates no autocorrelation. Based on $\alpha = 0.05$, the p-value is less than 0.05 illustrating the residuals in the regression model are autocorrelated. Model 1 and Model 2 meet the independence assumption while the other three models are not.

Bibliography

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, *5*(1). https://doi.org/10.1038/sdata.2017.191

ACI (2016). *Actuaries Climate Index: Development and Design.* American Academy of Actuaries, Canadian Institute of Actuaries, Casualty Actuarial Society, Society of Actuaries. Available at actuariesclimateindex.org (accessed June 23, 2019).

Blanc, E. (2012). The Impact of Climate Change on Crop Yields in Sub-Saharan Africa. *American Journal of Climate Change*, *01*(01), 1–13. https://doi.org/10.4236/ajcc.2012.11001

Burke, M., & Lobell, D. (2010). Climate change and food security: adapting agriculture to a warmer world. *Climate Change and Food Security: Adapting Agriculture to a Warmer World*, 201. https://doi.org/10.1007/978-90-481-2953-9

Cai, Y., Guan, K., Lobell, D., Potgieter, A., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., You, L., & Peng, B. (2019). Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agricultural and Forest Meteorology*, *274*, 144–159. https://doi.org/10.1016/j.agrformet.2019.03.010

Coble, K.H., Miller, M.F., Rejesus, R.M., Goodwin, B.K., Knight, T.O. (2010). A Comprehensive Review of the RMA APH and COMBO Rating Methodology Final Report.

Collins, D., Gibson, R., Kolk, S., Lindman, C., Mathewson, S., Hall, D., & Guérard, Y. (2020, January). Actuaries Climate Risk Index Preliminary Findings. *American Academy of Actuaries*. Retrieved from https://www.actuary.org/sites/default/files/2020-01/ACRI.pdf.

Della-Marta, P. M., Liniger, M. A., Appenzeller, C., Bresch, D. N., KöLlner-Heck, P., & Muccione, V. (2010). Improved estimates of the European winter windstorm climate and the risk of reinsurance loss using climate model data. *Journal of Applied Meteorology and Climatology*, *49*(10), 2092–2120. https://doi.org/10.1175/2010JAMC2133.1

Deschênes, O., & Greenstone, M. (2007). The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather. *American Economic Review*, *97*(1), 354–385. https://doi.org/10.1257/aer.97.1.354

Frees, E. W. (2010). *Regression Modeling with Actuarial and Financial Applications* (pp. 1–565). Cambridge University Press. https://doi.org/10.1017/CBO9780511814372

Goldburd, M., Khare, A., Tevet, D., & Guller, D. (2019). *Generalized Linear Models for Insurance Rating*. 2nd edition. *Casualty Actuarial Society*.

Goodwin, B. K., & Piggott, N. E. (2020). Has Technology Increased Agricultural Yield Risk? Evidence from the Crop Insurance Biotech Endorsement. *American Journal of Agricultural Economics*, *102*(5), 1578–1597. https://doi.org/10.1002/ajae.12087

Guo, J. (2015). Advances in impacts of climate change on agricultural production in China. *Journal of Applied Meteorological Science*, *26*(1): 1-11

Heras, A., Moreno, I., & Vilar-Zanón, J. (2018). An application of two-stage quantile regression to insurance ratemaking. *Scandinavian Actuarial Journal*, *2018*(9), 753–769. https://doi.org/10.1080/03461238.2018.1452786

Hoffman, A. L., Kemanian, A. R., & Forest, C. E. (2018). Analysis of climate signals in the crop yield record of sub-Saharan Africa. *Global Change Biology*, *24*(1), 143–157. https://doi.org/10.1111/gcb.13901

Huang, H., & Khanna, M. (2012). An Econometric Analysis of U.S. Crop Yield and Cropland Acreage: Implications for the Impact of Climate Change. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1700707

IPCC, 2018: Summary for Policymakers. In: Global Warming of 1.5°C. *An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty* [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. *World Meteorological Organization, Geneva, Switzerland, 32 pp.*

Jin, Y., Rejesus, R. M., & Little, B. B. (2005). Binary choice models for rare events data: A crop insurance fraud application. *Applied Economics*, *37*(7), 841–848. https://doi.org/10.1080/0003684042000337433

Jin, Z., & Erhardt, R. J. (2020). Incorporating Climate Change Projections into Risk Measures of Index-Based Insurance. *North American Actuarial Journal*. https://doi.org/10.1080/10920277.2019.1690525

de Jong, P., & Heller, G. Z. (2008). *Generalized linear models for insurance data*. *Generalized Linear Models for Insurance Data* (pp. 1–196). Cambridge University Press. https://doi.org/10.1017/CBO9780511755408

Lee, J., & Nadolnyak, D. (2012). The Impacts of Climate Change on Agricultural Farm Profits in the U.S. In *Agricultural & Applied Economics Association's 2012 AAEA Annual Meeting* (Vol. 42, pp. 405–416).

de Leeuw, J., Vrieling, A., Shee, A., Atzberger, C., Hadgu, K. M., Biradar, C. M., … Turvey, C. (2014). The potential and uptake of remote sensing in insurance: A review. *Remote Sensing*. MDPI AG. https://doi.org/10.3390/rs61110888

Lesk, C., Rowhani, P., & Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. *Nature*, *529*(7584), 84–87. https://doi.org/10.1038/nature16467

Lobell, D. B., & Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*, *150*(11), 1443–1452. https://doi.org/10.1016/j.agrformet.2010.07.008

Lobell, D. B., Schlenker, W., & Costa-Roberts, J. (2011). Climate trends and global crop production since 1980. *Science*, *333*(6042), 616–620. https://doi.org/10.1126/science.1204531

Makaudze, E. M., & Miranda, M. J. (2010). Catastrophic drought insurance based on the remotely sensed normalised difference vegetation index for smallholder farmers in Zimbabwe. *Agrekon*, *49*(4), 418–432. https://doi.org/10.1080/03031853.2010.526690

Murnane, R. J. (2004). Climate research and reinsurance. *Bulletin of the American Meteorological Society*, *85*(5), 697–707. https://doi.org/10.1175/BAMS-85-5-697

Ohlsson, E., & Johansson, B. (2010). Non-Life Insurance Pricing with Generalized Linear Models. In *EAA Lecture Notes* (pp. 71–99). Berlin Heidelberg: Springer-Verlag. Retrieved from http://www.springerlink.com/index/10.1007/978-3-642-10791-7

On Risk Classification. (2011). American Academy of Actuaries. Retrieved from https://www.actuary.org/sites/default/files/files/publications/RCWG_RiskMonograph_Nov2011.pdf

Porth, L., Roznik, M., Tan, K., Zhu, W., & Porth, C. B. (2019) Predictive Analysis: The Effects of Technology and Weather on Crop Yield. *Society of Actuaries*. Retrieved from https://www.soa.org/globalassets/assets/files/resources/research-report/2019/predictive-analysis-effects.pdf

Priest, G. L. (1996). The government, the market, and the problem of catastrophic loss. *Journal of Risk and Uncertainty*, *12*(2–3), 219–237. https://doi.org/10.1007/BF00055795

Rejesus, R. M., Coble, K. H., Miller, M. F., Boyles, R., Goodwin, B. K., & Knight, T. O. (2015). Accounting for weather probabilities in crop insurance rating. *Journal of Agricultural and Resource Economics*, *40*(2), 306–324. https://doi.org/10.22004/ag.econ.206598

Roebber, P., Brazauskas, V., & Kravtsov, S. (2017). The Actuarial Utility of Weather and Climate Predictions. *Submitted for publication*. Retrieved fromhttps://cpb-us-w2.wpmucdn.com/sites.uwm.edu/dist/a/122/files/2016/05/2017rbk-1yl41el.pdf

Schlenker, W., & Roberts, M.J. (2009). Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields under Climate Change. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(37), 15594–15598. https://doi.org/10.1073/pnas.0906865106

Sheehy, J. E., Mitchell, P. L., & Ferrer, A. B. (2006). Decline in rice grain yields with temperature: Models and correlations can give different estimates. *Field Crops Research*, *98*(2–3), 151–156. https://doi.org/10.1016/j.fcr.2006.01.001

Song, J. (2016). Advances in research methods on the impacts of climate change on agricultural production. *Science & Technology for Development, 12(6)*, 765-776. https://doi.org/10.11842/2016.06.014

Tollenaar, M., Fridgen, J., Tyagi, P., Stackhouse, P. W., & Kumudini, S. (2017). The contribution of solar brightening to the US maize yield trend. *Nature Climate Change*, *7*(4), 275–278. https://doi.org/10.1038/nclimate3234

Turvey, C. G., & McLaurin, M. K. (2012). Applicability of the normalized difference vegetation index (NDVI) In index-based crop insurance design. *Weather, Climate, and Society*, *4*(4), 271–284. https://doi.org/10.1175/WCAS-D-11-00059.1

Verón, S. R., de Abelleyra, D., & Lobell, D. B. (2015). Impacts of precipitation and temperature on crop yields in the Pampas. *Climatic Change*, *130*(2), 235–245. https://doi.org/10.1007/s10584-015-1350-1

Vogel, E., Donat, M. G., Alexander, L. V., Meinshausen, M., Ray, D. K., Karoly, D., … Frieler, K. (2019). The effects of climate extremes on global agricultural yields. *Environmental Research Letters*, *14*(5). https://doi.org/10.1088/1748-9326/ab154b

Ward, P. S., Florax, R. J. G. M., & Flores-Lagunes, A. (2014). Climate change and agricultural productivity in Sub-Saharan Africa: a spatial sample selection model. *European Review of Agricultural Economics*, *41*(2), 199–226. https://doi.org/10.1093/erae/jbt025

Wheeler, T., & Von Braun, J. (2013). Climate change impacts on global food security. *Science*. American Association for the Advancement of Science. *341*(6145), 508–513. https://doi.org/10.1126/science.1239402

Woodard, J. D. (2014). Impacts of Weather and Time Horizon Selection on Crop Insurance Ratemaking: A Conditional Distribution Approach. *North American Actuarial Journal*, *18*(2), 279–293. https://doi.org/10.1080/10920277.2014.887476

Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., … Asseng, S. (2017). Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(35), 9326–9331. https://doi.org/10.1073/pnas.1701762114

Climate Science Watch, www.climatesciencewatch.org.

NASA- Goddard Institute for Space Studies and the National Climatic Data Center, www.nasa.gov

National Oceanic and Atmospheric Administration, www.noaa.gov

United State Department of Agriculture (USDA), www.usda.gov