Can Crop Productivity Indices Improve Crop Insurance Rates?

Xiaofei Li
Department of Agricultural Economics
Mississippi State University
xiaofei.li@msstate.edu

Jesse B. Tack
Department of Agricultural Economics
Mississippi State University
tack@agecon.msstate.edu

Keith H. Coble
Department of Agricultural Economics
Mississippi State University
coble@agecon.msstate.edu

Barry J. Barnett
Department of Agricultural Economics
Mississippi State University
barnett@agecon.msstate.edu

Selected Paper prepared for presentation at the 2016 Agricultural & Applied Economics Association Annual Meeting, Boston, Massachusetts, July 31-August 2

Can Crop Productivity Indices Improve Crop Insurance Rates?

Abstract

This study explores whether the soil information contributes additional explanatory power beyond the base premium rate to crop insurance loss. By examining a panel of 697 counties from the Corn Belt states in the U.S. over the period of 2005−2015, we find the loss costs are systematically associated with the National Commodity Crop Productivity Index (NCCPI) conditioning on the county base premium rates, weather conditions, and year fixed effects. The loss cost rises first with NCCPI in lower NCCPI quartiles and then decreases in higher NCCPI quartile. In general, counties with medium NCCPI values are expected to have higher relative loss cost comparing with low and high NCCPI counties. The pattern of NCCPI's effects is robust to different model specifications, heteroskedasticity and spatial correlation of data, as well as subsamples by insured acreage thresholds. The finding of this study presents empirical evidence that there is additional information embodied in soil and spatial variables not captured by the current base premium rates that is correlated with loss experiences. It suggests the potential to incorporate soil and spatial information to improve the crop insurance ratemaking.

Keywords: Crop insurance, loss cost, base premium rate, soil productivity index (NCCPI), risk

JEL Codes: Q18, Q14

Introduction

As a Federal agency the USDA/RMA has objectives defined by legislative language found in the Federal Crop Insurance Act which contains the following provisions pertinent to rate making:

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

Thus, RMA is expected to set rates that are actuarially sound. Interestingly, the legislation does not specify at what level the actuarial soundness should be evaluated. Several aspects of Sec 508 merit note. First, this legislation is understood to exclude the cost of sales, loss adjustment, underwriting, and other activities that a private insurance firm would have to cover. Second, the operating cost of the USDA/RMA is not included in premium rates, nor is the administrative and overhead expense to the approved insurance providers (AIP) that deliver the insurance program to producers.

The RMA actuarial process used to generate Approved Production History (APH) rates primarily uses historical loss experience for a crop in a county to derive the rates for an insured unit within that county. The process begins by collecting the observed insurance and loss data for that county/crop combination and using it to derive a base county rate. Extreme loss experience is smoothed by a regional catastrophic pool. The remaining experience is averaged to derive the base county rate using weather-weights derived as described in Rejesus *et al.* (2015). Crop insurance like other lines of property and casualty insurance uses various observable attributes to group insureds within a county into risk pools of similar risk levels. There are several factors used to tailor the rate to an individual producer, depending on utilization of certain farming practices, coverage choices, and the ratio of APH yield history relative to a county reference yield.

However, one fundamental factor related to crop yield risk has not yet been directly utilized in the RMA rating system: the soil. Obviously, soil is a most critical factor determining the crop growth process, as is well documented by crop science and production practices. Soil types and qualities determine both the mean and variation of crop yield. It is widely noted in the literature that the crop risk displays large spatial disparities both across regions (Glauber, 2004; Babcock, 2008; Woodard *et al*., 2012) and within a region (Popp *et al.*, 2005; Lobell *et al.*, 2007; Claassen and Just, 2011). Besides climate conditions, those disparities are to a large extent driven by the underlying soil type and quality variability, especially for the intra-regional disparities where climate conditions are general homogenous.

Some insurance programs, for example, the Saskatchewan crop insurance uses soil classification as a factor in rating. But historically the US Federal Crop Insurance Program has not included soil information in the premium ratemaking. In recent years, a growing interest has emerged about incorporating soil information in rating. Pilot efforts have been undertaken to explore how to incorporate high resolution soil data in rating and the rating error if soil information is not considered (Woodard, 2016; Woodard and Verteramo Chiu, 2016). The efforts of incorporating soil information has been made possible by the increasing availability of high quality insurance and soil data. RMA now requires that common land unit (CLU) shape files be submitted with insurance submissions (Brady, 2013). Prior to the addition of CLU data, RMA captured legal descriptions that could typical identify the 640 acre section in which an insured unit was located. The CLU data then allows more accurate identification of soils when there is heterogeneity within the section. The national wide high resolution gridded soil data (such as the gSSURGO database) are also readily available and being improved with advancing soil survey, remote sensing, and GIS technologies.

This study follows this vein of effort to explore the incorporating of soil data in crop insurance rating. We focus on a basic question first: Is the soil information useful for the rating? If the answer is yes, we then proceed to address how to incorporate soil information. Indeed, the RMA may have a good reason not to incorporate soil in rating, because when the sample history is long enough the APH yield data have already fully captured the soil information. However, in practice the APH yields are usually calculated based on short-period sample. For example, at individual farm level the APH data are typically available for 4 to 10years. These short-period APH yield data are subject to small sample weather fluctuations and noisy. Premium rates derived based on those noisy APH data is likely to deviate substantially from the actual long history risk. In addition, the RMA follows typical actuarial practices of using various observable attributes to adjust APH rating – cropping practice, unit structure, etc. This study then empirical tests whether the current APH premium rates have fully captured the soil information. Following the previous rationales, if APH has truly captured all soil information, adding the soil information *per se* to the risk model would not provide extra explanation power beyond the existing APH rates. But if there is still some remaining information in the soil, then including soil information can add more explanation power for the loss experience. That way we convert the question of whether we should consider soil information in insurance rating into a testable hypothesis.

The hypothesis test is based on a panel sample of 697 corn production counties in the U.S. over the years of 2005−2015. We regress the observed county level losses on APH rates and soil variables, and test for the statistical significance of soil variables conditioning on APH rates. The county level loss is measured by the loss cost ratio aggregated at county level, and APH rate is measured by county base premium rate. Soil properties are represented by the 10-meter National Commodity Crop Productivity Index (NCCPI) developed from the USDA/NRCS, which is averaged to a county level. If the APH base rates have already embodied all risk-related soil information, the effect of soil variable (NCCPI) ought to be insignificant to explain county losses. On the other hand, if NCCPI is found to be systematically related with losses, this provides evidence that there exists additional soil information not currently contained in APH base rates that are related with losses.

Note that the county level data are not ideal for this test as much variability within a county is omitted. The more appropriate data for this test are individual farms or fields. However, due to data confidential policy those individual data are not available to the public. Under this data restriction, we propose to use county units do conduct a pilot test. Fortunately, the county level variabilities in soil and loss are also rich to support a meaningful statistical test. If county level test is significant, then individual level test will be more significant because of noisier APH and heterogeneous soil within county.

The results significantly reject that soil information is irrelevant to crop insurance losses. A nonlinear pattern of effect of the soil productivity index (NCCPI) is found. The county losses increase with NCCPI for the counties with low to medium NCCPI values, and decline for medium to high NCCPI counties. The pattern is statistically significant and robust to various spatially correlated errors, subsamples, and insurance types. This finding provides strong empirical evidence supporting that the current RMA rate does not fully capture soil information.

The finding that soil information explains extra losses beyond APH premium rates has important implications for the crop insurance program. It suggests potential for rating improvement by incorporating soil information as a component of the rating system. With the availability of high resolution soil data, this improved rating can individualize premium rates to farm or even field level which is corresponding to the land's site-specific risk characteristics. It will greatly reduce the adverse selection issue caused by using regional average crop risk rating (Goodwin, 1994; and more), which was previously difficult to address due to the lack of adequate APH data in individual level.

Literature review

There is an extensive discussion assessing the current RMA ratemaking procedures. A mostly documented source of mis-rating of the RMA APH approach is the adverse selection, where higher risk producers are under-charged relative to lower risk producers (Skees and Reed, 1986; Makki and Somwaru, 2001; Glauber, 2004). RMA assumes a simplified constant variability structure of the yield, which determines the premium rate only using average yield and induces adverse selection (Goodwin, 1994), particularly because of the use of aggregate (typically county) measures to estimate individual yields and rates. Another major source of rating error comes from the trending of yield over time (Skees and Reed, 1986; Coble *et al*., 2011; Adhikari *et al.*, 2012), which makes the historical average yield not indicative of future risk. Given the evolution of agricultural technologies and managements, the yield risk is highly likely to be non-stationary over time (Harri *et al*., 2009). Also, in the rating practice the determination of individual farm rates is mostly relied on short period APH yield data that are not long enough to fully represent crop yield risk (Skees and Reed, 1986; Rejesus *et al*., 2015).

Against this background, the improvement of rating is an on-going effort by RMA. Among various improvement strategies, a recommendation is to utilize additional risk-related weather, soil and other locational variables in the rating procedure (Coble et al., 2010). Weather data has already been incorporated into the rating system. For example, Rejesus et al. (2015) develop a methodology to weight historical APH loss data based on long period weather frequencies. As to the soil and spatial information, however, fewer studies have been looking at how to incorporate it into the rating system. To our best knowledge, the most relevant work is done by Woodard (2016) and Woodard and Verteramo Chiu (2016), who incorporate the high resolution gSSURGO soil data into the micro-level (field level) insurance rates for McLean County, Illinois. The general idea of their method is to first use soil information in modeling yield, and then use the estimated yield model and CLU level soil information to predict the CLU level premium rate. Their work has demonstrated a significant improvement of the soil induced field level rates from the APH rates without considering soil. Our study explore the same question towards the feasibility of utilizing soil in insurance rating, but from a different perspective. We look at the losses and rates at the county level, and investigate whether soil data can add extra information to the losses conditioning on the current RMA rates.

Empirical Model

The empirical model for testing soil's effect on crop losses is specified as follows:

$$\begin{aligned} LC_{it} = \alpha + \beta Rate_{it} + f(Soil_i) + \delta_1 Prec_{it} + \delta_2 Prec_{it}^2 + \\ GGDlow_{it} + GGDmed_{it} + GGDhigh_{it} + Year_t + \varepsilon_{it}, \end{aligned} \tag{1}$$

where LC_{it} is the loss cost for county *i* at year *t*, and $Rate_{it}$ is the RMA base premium rate calculated based on APH historical yields. $Soil_i$ represents the quality of soil in county *i*, which is approximated here by a soil productivity index of NCCPI (National Commodity Crop Productivity Index). The soil quality is assumed to be constant over time. To allow for more flexibility of soil's effect, the functional form *f*() takes a splined linear specification. *Prec* represents annual precipitation during the growing seasons, and *GGDlow*, *GGDmed*, and *GGDhigh* are the growing degree days under low, medium, and temperature thresholds respectively. The inclusion of those weather variables is to control for the year-to-year growing conditions that substantially influence the yield risks. $Year_t$ is the year-specific fixed effects which controls for the yearly market crop price fluctuations that influence the losses.

The central idea of this regression model is to empirically test whether the NCCPI index contributes additional explanatory power to the losses beyond the RMA premium rates. Note that the model is not intended to test for the actuarial soundness of the program, i.e., "E(*LC*) = *Rate*". That differs from many actuarial soundness regression models where the dependent variable is the loss ratio (i.e., *LC/Rate*) or the *LC*−*Rate*. Instead, the dependent variable is the loss cost, and RMA rate is treated as a control variable in the regression whose slope β is allowed to deviate from one. We have mainly two reasons for that specification. First, the sample period (11 years) is not long enough to conduct a solid actuarial soundness test. The deviations of the losses from base rates found in the test is very likely due to the short-period unrepresentative weather of the sample years. However, even if the overall losses deviate from base rates, as long as the RMA base rates have really fully captured the soil information, the deviations should be systematically uncorrelated with spatial and soil characteristics, which can be tested by the model specified. Second, as will be discussed in the data section shortly, the loss cost variable in the model is defined based on only the revenue protection, which is determined not only by weather risks but also by price risks. The associated coverage levels are also diversified which differ from the 65% of the RMA base rate. That makes the *LC* and *Rate* not directly comparable for the actuarial soundness test.

It is noteworthy to mention that by this model specification, the effect of soil variable in this model can be only interpreted in a relative sense. It can reveal whether there exists additional information embedded in soil that has power in explaining in losses and is not currently contained in RMA premium rates. But it cannot be interpreted as whether the RMA premium rate is overpriced or underpriced.

The error term (ε_{it}) captures the unobserved risk-influencing county characteristics, such as un-modeled weather variables, pests, diseases, and other shared geographic features. Those characteristics are highly likely to be heteroskedastic and spatially autocorrelated across counties. To account for those types of error terms, we use a non-parametric spatial heteroskedasticity and autocorrelation consistent (spatial HAC) estimator of the variance-covariance matrix. Following Kelejian and Prucha (2007), the OLS spatial HAC variance-covariance matrix is estimated as:

$$\hat{V}ar(\hat{\theta}) = \frac{1}{n}\sum_i \sum_j X_i X_j \hat{\varepsilon}_i \hat{\varepsilon}_j K(\frac{d_{ij}}{d}) , \tag{2}$$

where $\hat{\theta}$ is the vector of coefficient estimates, X_i the i-th row of the matrix of explanatory variables, and $\hat{\varepsilon}_i$ the OLS residuals. The index i and j refer to counties, and n refers to the total number of sample counties. *K()* is a kernel function to assign the inverse-distance weights. The distance d_{ij} denotes the great circle distance between centroids of counties i and j. The cutoff distance d is chosen as the distance between the farthest two counties in the sample, as we believe the unobserved weather as well as other risk-influencing factors interact in very long distance ranges. Also, considering the sample data we use are panel data, the covariance structure of disturbances is actually a stacked block matrix with each year's covariance matrix on the diagonal and zeros on off-diagonals (i.e., assuming spatial correlations only exist in cross section data).

Data

The sample data we use in this analysis are from the counties of 9 Corn Belt states (Illinois, Iowa, Indiana, Michigan, Minnesota, Wisconsin, Nebraska, Missouri, and Ohio), covering an 11-year period from 2005 to 2015. Corn insurance is used as it is the largest insured crop type and the insured counties are relatively continuous over space. To obtain a balanced panel, we exclude the counties with missing data in any of the 11 years, and in total we have 697 counties.

That way we can avoid over-representing the more recent years since there are more missing values in earlier years.

The loss information for corn insurance is obtained from RMA's Summary of Business database (SOB). Given there are different types of insurance plans, we choose the revenue protection (RP) insurance plan to ensure comparability since the RP plan is the dominant crop insurance plan during the sample period. The county level loss cost ratio (*LC*) is defined as the sum of all RP indemnities divided by the sum of all RP liabilities within a county. It should be noted that this loss cost ratio definition is a county level aggregation across various coverage levels (from 50% to 85% by a 5% increment). Ideally, this loss cost ratio should be adjusted according to a common basis of the 65% coverage level which is comparable with the base rate. However, since the SOB loss data do not provide individual insured data, it is not possible to implement this adjustment. We control for the county average coverage level later on in the modeling for robustness check, and find it does not affect the results.

RMA county base premium rates (*Rate*) are obtained from RMA's Actuarial Data Master database (ADM). We set the county base premium rates as the sum of Reference Rate and Fixed Rate Load. The Reference Rate is the premium rate corresponding to the established county average yield (Reference Yield) at the 65% coverage level. Specifically, we choose the rates under Grain purpose (type code 16) and Non-irrigated land (practice code 3). The RMA calculates the base rates on the basis of historical risk by using the Actual Production History (APH) approach and constantly updates them over time, but the amount of change is usually small at most of the time. Note that we use the county base premium rates instead of the county-level aggregated individual premium rate (i.e., sum of individual premiums divided by sum of individual liabilities). The county base rates can be regarded as RMA's predictions for future loss in county level based on past loss information, and the individual rates are calculated based on county base rates and information of individual producers. Because this study is conducted based on sample of county units, using base rate as explanatory variable is advantageous as it allows to directly test RMA's county level prediction of forward-looking risk. The base county rate for each year is used to reflect RMA's expectations of yield losses for a common level of coverage and insurance plans. This data is determined prior to the crop insurance enrollment period and is not affected by producer choices regarding coverage level, insurance plan, or unit structure which are endogenous to the premium or effective premium rate (i.e., premium/liability) determined when insurance is purchased.

The soil information is represented by the National Commodity Crop Productivity Index (NCCPI) that is developed by the USDA/NRCS. The index ranges from 0.01 to 1, which is a rating for the production capacity of dry-land commodity crops based on inherent soil properties, landscape features and climatic characteristics (Dobos, et al., 2012). Higher value of the index represents more desirable traits. Two points merit notice. First, the index is a combined measure of many agro-climatic conditions (soils, landscapes and climates) to foster crop productivity, and for non-irrigated crops only. Second, the NCCPI productivity index is considered invariant over time, while temporary fluctuations in productivity caused by year-to-year weather variations and management practices are not addressed. The NCCPI index data are obtained from the Gridded Soil Survey Geographic (gSSURGO) Database, where the index is arranged in 10-meter resolution. We choose the index for corn and soybean production, and aggregate the index to county level. In particular, we aggregate the NCCPI over only the corn producing lands. The corn land distribution is from the 30-meter resolution Cropland Data Layer (CDL) database released by USDA.

Following Schlenker and Roberts (2009), we select annual precipitation and heat to represent weather information. The yearly precipitation is the total rainfall during the corn growing season. Here we choose the corn growing season as the months of April through September, as northern regions tend to plant later. We realize this choice is a little arbitrary, and the actual planting dates may vary from year to year depending on weather conditions. But the slight difference in the growing season specification usually does not alter the results very much (Schlenker and Roberts, 2009). The annual heat for crop growth is measured by the growing degree days (GGD), which are measures of heat accumulation above certain base temperatures. For example, for a 0 degree Celsius base temperature, a day of 15°C contributes 15 degree days, while a day of 35°C contributes 35 degree days. At the 10°C base temperature, a day of 15°C contributes 5 degree days, while a day of 35°C contributes 25 degree days. Degree days are then summed over the entire growing season. Furthermore, we adopt a more sophisticated modified version definition of growing degree days by Schlenker and Roberts (2009) to get more continuous values. In addition, to allow for the flexibility of the heat effect, we try three different base temperatures: 0°C, 10°C, and 30°C. The three growing degree days then have some overlapping part, and we subtract the overlapping by defining:

$$GGDlow = GGD_{0°C} - GGD_{10°C},$$

$$GGDmed = GGD_{10°C} - GGD_{30°C},$$

$$GGDhigh = GGD_{30°C}.$$

The *GGDlow*, *GGDmed*, and *GGDhigh* represent the accumulated heat under the low, medium, and high temperature respectively. The precipitation and temperature information is obtained from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) database in 4 km grids, which is then aggregated into county level.

The summary statistics of variables are shown in table 1. Figure 1 and 2 show the spatial distributions of the soil, loss, and base rate of the sample counties. Figure 3 shows the statistical variations of the variables both cross sectional and cross temporal.

Results

The effect of NCCPI on county level loss cost follows a nonlinear pattern of first positive and then negative, controlling for the county base premium rates, weather conditions, and yearly dummy variables. As shown by the estimated splines in table 2, the loss cost rises first with NCCPI when NCCPI is low (<0.65), then decreases when NCCPI is high (>0.65). The magnitudes of the NCCPI effect on the extra loss cost are economically significant. Taking the full model (column 4 of table 2) for example, at the first spline (NCCPI<0.38) a county's loss cost is predicted to increase by 0.043 if its NCCPI is 0.1 higher (that is about 10% of the total range of NCCPI values; the number is calculated as 0.043=0.43×0.1). At the second spline (0.38<NCCPI<0.65) the loss cost still increases with NCCPI, though the magnitude reduces to 0.0149 for a 0.1 increase of NCCPI. At the third spline (NCCPI>0.65), the effect turns into negative and an NCCPI value of 0.1 higher will result in decrease of loss cost by 0.0156. Those changes in loss cost are considerable in size given that the sample average loss cost is about 0.09.

The four columns in table 2 correspond to four different model specifications. The column (1) only has base rate as the single regressor. Column (2) adds three linear splines of NCCPI. Column (3) also controls for the year fixed effects. Column (4) further controls for weather variations. The spline knots are selected by trying all possible combinations and pick the one combination with the best fit. The finally chosen knots for the splines are 0.38 and 0.65. But the results are very similar when slightly changing the knots.

OLS is used to estimate the model, while the coefficient standard errors reported in table 2 have been corrected for unknown form of spatial heteroskedasticity and autocorrelation. This robustness correction has resulted in larger standard errors comparing to the OLS standard errors (which are not reported in the paper), and therefore lowered the significance levels of coefficients. Overall, the effect of NCCPI is highly significant in the first spline, while the second and third splines are statistically less insignificant. The significance levels rise as more control variables are added to the models. In the full model (column 4), all the three splines are significant (the second spline is very close to the 10% significance level with a p-value of 11.6%). The joint significance test of the three splines also significantly rejects the hypothesis of zero-relevance of NCCPI for all model specifications.

To better visualize the nonlinear pattern of NCCPI's effects, we plot the predicted loss cost in response to NCCPI values in figure 4 (red solid line). This curve only focuses on the change in loss cost caused by NCCPI, while all other control variables are fixed. The loss cost at NCCPI=0.9 is set as the zero level, and the heights of the curve only have relative meanings.

The graph shows a clear inverse-U shape pattern. In addition to the linear regression spline function whose structure is to some extent preset, we also specify a more flexible step function model that fits a separate effect for each 0.05-interval NCCPI range. The estimation results are also shown in figure 4 with the solid blue line, and the dashed blue lines represents the 95% confidence intervals (after adjusting for spatial heteroskedasticity and autocorrelation). The lowest three steps (0<NCCPI<0.15) and the highest two steps (0.9<NCCPI<1) are combined as one step respectively, due to small number of observations in the two extremes of NCCPI values. As can be noticed, the loss cost increases with NCCPI first in the low and medium NCCPI value range, and then decreases with NCCPI in the higher NCCPI value range. That pattern is consistent with that of the linear spline function result. For comparison purpose the highest NCCPI value group (0.9<NCCPI<1) is set as the base group where the effect is set as zero. But the choice of base group does not affect the shape of the effect curves.

Besides the significance and magnitude of the soil coefficient, it is also of interest to see whether the incorporation of soil variable increases the model predicting power. This can be done by comparing the out-of-sample prediction performance of the models with and without soil variable. Following Schlenker and Roberts (2009), we conduct the cross-validation that leaves out 20% of sample observations. We make a 9999 times simulations, where in each simulation 20% of observations are randomly dropped out of the sample. The model is estimated using the remaining 80% of the sample, and then the estimated model is used to predict the loss costs for the 20% dropped sample. The mean squared error (MSE) statistics between the predicted and actually observed loss costs are calculated as a measure of the closeness of the prediction. The

MSE is 0.0126 for the without-soil model, and 0.0122 for the with-soil model, which is around 3% decrease.

Beyond the main interest of the soil variable, we also look at the other control variables. The county base premium rate is found to be significantly positively related with loss cost, which is consistent with our expectation. Note that in this model the base rate is set as a control variable, which factors out the part of soil information that has already been contained by the APH historical information. Therefore, the estimated effects of NCCPI in the models of this paper capture the remaining soil effects conditional on APH rates. In addition, the coefficient of base rate highly deviates from 1 and an intercept term also exists in the regression, because neither the definition of base rate (for revenue insurance plan only) nor the short period sample size (only 11 years data) can grant the actuarial soundness analysis of the base rate, as mentioned earlier.

Precipitation is found to have a quadratic form impact on losses, that too little or too much rainfall causes more losses. But the effect is not statistically significant. The growing degree days under low and medium temperatures contribute to reduce the losses, though the low temperature GGD's effect is much smaller and insignificant. On the other hand, more growing degree days under the high temperature significantly increase losses, which suggests that extreme heat damages the crop growth.

Various robustness checks of models are conducted, all showing a similar nonlinear relationship between NCCPI and losses, as shown in Appendix tables A1 through A4.

First, table A1 reports the estimation results of models with extra control of insurance coverage level. As discussed earlier, it is concerned that the average coverage levels may differ across counties. Counties with higher coverage level are tended to be paid higher indemnities (hence, loss cost ratios) under the same loss risk, while the base rates are all set at the same 65% coverage level. Therefore, for some counties the high loss cost may simply be due to the higher average coverage levels. To make remedy to that situation, we add the insured area weighted average coverage level to the regression to account for the variation in coverage level. The results find weak significance of the coverage level variable, suggesting that the choice of coverages may be uncorrelated with the key explanatory variables in the model. The coefficients estimations for other variables of interest in the model are also not affected, where the pattern of the NCCPI effects is unchanged.

Second, to account for the possible outliers existing in the sample caused by small acreage of insured land areas, we also estimate the model using subsample of counties with only large insured acres. As shown in table A2, we limit the estimate to counties with larger than 2000 acres, 5000 acres, and 10000 acres of insured land areas under the revenue protection plan. The corresponding number of counties are 506, 363, and 240 respectively. Only the full model (with a full set of controls) is reported. The inverse-U shape nonlinear pattern of NCCPI's effect persists for all subsamples, though the significance levels of the coefficients decline as the size of subsample decreases.

Finally, we also try loss definitions based on the yield protection insurance plan. The yield protection loss is completely determined by physical yield, and prices do not have influence on it. As shown in table A3, the results are very similar to the revenue protection plan results. However, a drawback is that the yield protection plan's share in the entire crop insurance market decreased substantially in recent years, and the insured acreages are much smaller comparing to the more popular revenue protection plan. For more robustness check we also calculate the loss cost by combining the yield and revenue protection plans together. Again, the estimate results are very similar and the inverse-U shape pattern of NCCPI's effects remain unchanged.

In sum, the empirical data reveal that county loss costs are systematically associated with the soil index of NCCPI even after controlling for RMA base rates. The association is significant and robust to various specifications and samples. This finding presents empirical evidence that there is additional information from soil characteristics that is not currently contained in the RMA base premium rate, but is correlated with observed loss cost.

6. Conclusion

This study empirically demonstrates that soil information is systematically associated with loss costs at county level conditional on the current RMA rating. Using a panel of 697 counties of the Corn Belt states in the U.S. over the period of 2005−2015, we are able to significantly reject the hypothesis that the soil productivity index (NCCPI) is uncorrelated with county level loss costs. A nonlinear pattern of the relationship between soil and loss is found for the sample data, where the loss cost rises first with NCCPI in lower NCCPI values and then decreases in higher NCCPI values. The pattern of the effects are robust to various model specifications, variable definitions, and subsamples. This finding provides empirical evidence that the current RMA rating does not fully capture soil information. It suggests there is additional information embodied in the soil which may be useful to improve the base premium rate making, even at the county level.

This study is a first step exploration on incorporating soil and spatial information into the current crop insurance rating system. The model we use is rather descriptive. It finds significant correlation between soil productivity index and losses, which justifies the necessity to add soil information to rating. But in order to further explain the observed pattern of the effect and incorporate soil information in the rating process, more detailed mechanistically explanatory models are needed for the next steps.

We also calls for more spatially disaggregated insurance data and more detailed soil data for the modeling. Due to the current restrictions of access to individual farm data we are only able to use county level data for the analysis. Given the strong spatial heterogeneity of soil conditions within a county, it is more desirable to refine the analysis based on more spatially disaggregated individual insurance data. Certain strategies need to be developed to best handle the confidentiality and privacy issues in using individual farm data. The NCCPI index is also only a highly compacted one-dimension variable, which is easy to use in the modeling but its interpretation is also relatively vague. We suggest collecting more actual soil, landscape and climate characteristics, and directly incorporate those variables in the modeling. Longer time series sample is also needed. The 11 years period is also not long enough to represent historical weather distribution. A potential drawback is that it does not rule out the possibility that the significant effect of NCCPI was due to abnormal short-period weather variability that happened to be spatially correlated with soil distribution. By addressing those issues, the further improvement of this study is expected to lie on a solid methodology foundation for utilize soil and spatial information in premium rating, and finally to allow the individualizing of premium rate to individual farm or even field level.

There are two most relevant policy implications for incorporating soil information in rating. First, it provides a method to downscale the premium rating to the field level (or micro-level), that each individual piece of land can be precisely rated according to its risk characteristics. That can substantially reduce the adverse selection problem caused by the rating based on average risk over large areas. Second, even at the county level, the soil information is useful for the expansion of insurance program into new areas without adequate historical risk data.

References:

Adhikari, S., T.O. Knight, and E.J. Belasco. 2012. Evaluation of Crop Insurance Yield Guarantees and Producer Welfare with Upward Trending Yields. *Agricultural and Resource Economics Review* 41(3): 367−376.

Babcock, B.A. 2008. Corn Belt Contributions to the Crop Insurance Industry. *Iowa Ag. Review* 14 (2) (Spring 2008): 1−3.

Brady, T. 2013. Common Land Unit Acreage Reporting Plan. Crop Insurance Today Magazine. February 46(1), 4–9.

Claassen, R. and R.E. Just. 2011. Heterogeneity and Distributional Form of Farm-level Yields. *American Journal of Agricultural Economics* 93(1): 144−160.

Coble, K.H., T.O. Knight, B.K. Goodwin, M.F. Miller, R.M. Rejesus, and G. Duffield. 2010. A Comprehensive Review of the RMA APH and COMBO Rating Methodology. *Report for the USDA Risk Management Agency by Sumaria Systems, Inc.*

Coble, K.H., M.F. Miller, R.M. Rejesus, B.K. Goodwin, R. Boyles, and T.O. Knight. 2011. Methodology Analysis for Weighting Historical Experience. *Report for the USDA Risk Management Agency by Sumaria Systems, Inc.*

Glauber, J.W. 2004. Crop Insurance Reconsidered. *American Journal of Agricultural Economics* 86(5): 1179−1195.

Goodwin, B.K. 1994. Premium Rate Determination in the Federal Crop Insurance Program: What Do Averages Have to Say about Risk? *Journal of Agricultural and Resource Economics* 19(2): 382−395.

Harri, A., C. Erdem, K.H. Coble, and T.O. Knight. 2009. Crop Yield Distributions: A Reconciliation of Previous Research and Statistical Tests for Normality. *Applied Economic Perspectives and Policy* 31(1): 163−182.

Kelejian, H.H., and I.R. Prucha. 2007. HAC Estimation in a Spatial Framework. *Journal of Econometrics* 140(1): 131−154.

Lobell, D.B., J.I. Ortiz-Monasterio, and W.P. Falcon. 2007. Yield Uncertainty at the Field Scale Evaluated with Multi-year Satellite Data. *Agricultural Systems* 92(1): 76−90.

Makki, S.S., and A. Somwaru. 2001. Evidence of Adverse Selection in Crop Insurance Markets. *Journal of Risk and Insurance* 68(4): 685−708.

Popp, M., M. Rudstrom, and P. Manning. 2005. Spatial Yield Risk across Region, Crop and Aggregation Method. *Canadian Journal of Agricultural Economics* 53(2−3): 103−115.

Rejesus, R.M., K.H. Coble, M.F. Miller, R. Boyles, B.K. Goodwin, and T.O. Knight. 2015. Accounting for Weather Probabilities in Crop Insurance Rating. *Journal of Agricultural and Resource Economics* 40(2): 306−324.

Skees, J.R., and M.R. Reed. 1986. Rate Making for Farm-level Crop Insurance: Implications for Adverse Selection. *American Journal of Agricultural Economics* 68(3): 653−659.

U.S. Department of Agriculture, Risk Management Agency. 2010. Informational Memorandum: IS-10-004 July 9, 2010. Available online at: www.rma.usda.gov/bulletins/info/2010/is-10-004.pdf (accessed May 2016).

——. 2016. *Summary of Business*. Available online at: http://www.rma.usda.gov/data/sob.html (accessed May 2016).

Woodard, J.D., G.D. Schnitkey, B.J. Sherrick, N. Lozano-Gracia, and L. Anselin. 2012. A Spatial Econometric Analysis of Loss Experience in the US Crop Insurance Program. *Journal of Risk and Insurance* 79(1): 261−286.

Woodard, J.D. 2016. Integrating High Resolution Soil Data into Federal Crop Insurance Rates: Actuarial, Policy, and Sustainability Implications. AGreen Working Paper, available online at: http://www.foodandagpolicy.org/sites/default/files/Woodard%20Working%20Paper_SoilCropInsurance_Jan-2016.pdf (accessed May 2016).

Woodard, J.D., and L.J. Verteramo Chiu. 2016. Soil Data Not Considered in Cornerstone U.S. Agricultural Policy. AGreen Working Paper, available online at: http://www.foodandagpolicy.org/sites/default/files/Woodard%20Verteramo_WorkingPaper_SoilData(Feb2016).pdf (accessed May 2016).

Table 1 Descriptive statistics

Variable	Mean	Std. Dev.	Min	Max
Loss cost ratio	0.091	0.141	0	1.208
Base premium rate	0.075	0.052	0.009	0.395
NCCPI	0.617	0.154	0.006	0.915
Precipitation (cm)	608.4	154.7	120.3	1231.8
Growing Degree Days (0°C)	3394.4	326.2	2405.7	4415.5
Growing Degree Days (10°C)	1680.3	275.7	904.1	2592.6
Growing Degree Days (30°C)	19.1	21.3	0	150.8

Note: The sample is a panel of 697 counties over 11 years (2005−2015), with a total of 7,667 observations. The loss cost ratio is calculated based on Revenue Protection (RP) insurance plan.

Table 2 Estimated correlation of NCCPI to county loss cost (OLS model)

	(1)	(2)	(3)	(4)
Base rate	0.423***	0.462***	0.674***	0.410***
	(0.102)	(0.129)	(0.123)	(0.0965)
NCCPI [spline 1]		0.162**	0.221***	0.430***
		(0.0794)	(0.0820)	(0.0834)
NCCPI [spline 2]		0.0523	0.110	0.149
		(0.133)	(0.139)	(0.0951)
NCCPI [spline 3]		-0.135	-0.0973	-0.156**
		(0.121)	(0.115)	(0.0694)
Precipitation				-0.0000819
				(0.00491)
Precipitation squared				0.0000183
				(0.000034)
Low GGD				-0.0000792
				(0.000410)
Medium GGD				-0.000230**
				(0.000117)
High GGD				0.00459***
				(0.00171)
Constant	0.0598**	-0.00818	-0.0979**	0.222
	(0.0238)	(0.0389)	(0.0401)	(0.575)
Year Fixed Effects	No	No	Yes	Yes
Joint significance of 3 splines (p-value)	-	0.011	0.0057	<0.0001
Observations	7,667	7,667	7,667	7,667
R-squared	0.315	0.318	0.479	0.567

Note: Dependent variable is the county-level loss cost ratio (Revenue Protection (RP) insurance plan). The knots for the linear splines are NCCPI=0.38 and NCCPI=0.65. Spatial heteroscedasticity and autocorrelation consistent (HAC) standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The estimation are based on a panel of 697 Corn Belt State counties over 11 years (2005−2015).

Figure 1 Spatial distribution of NCCPI by county

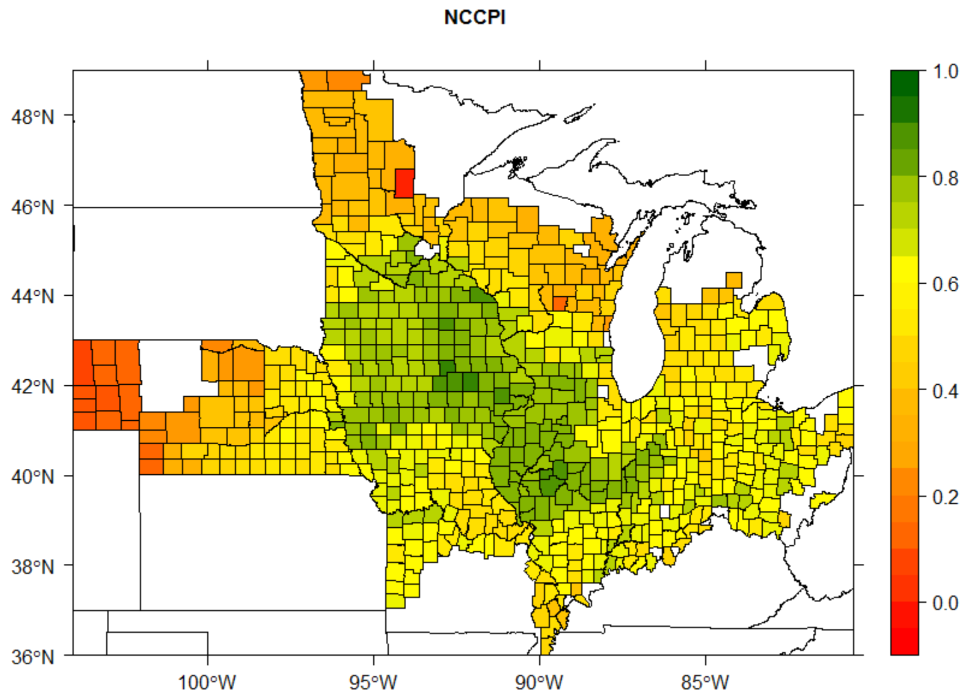


Figure 2 Spatial distribution of county-level base premium rates and loss costs (average over 2005−2015)

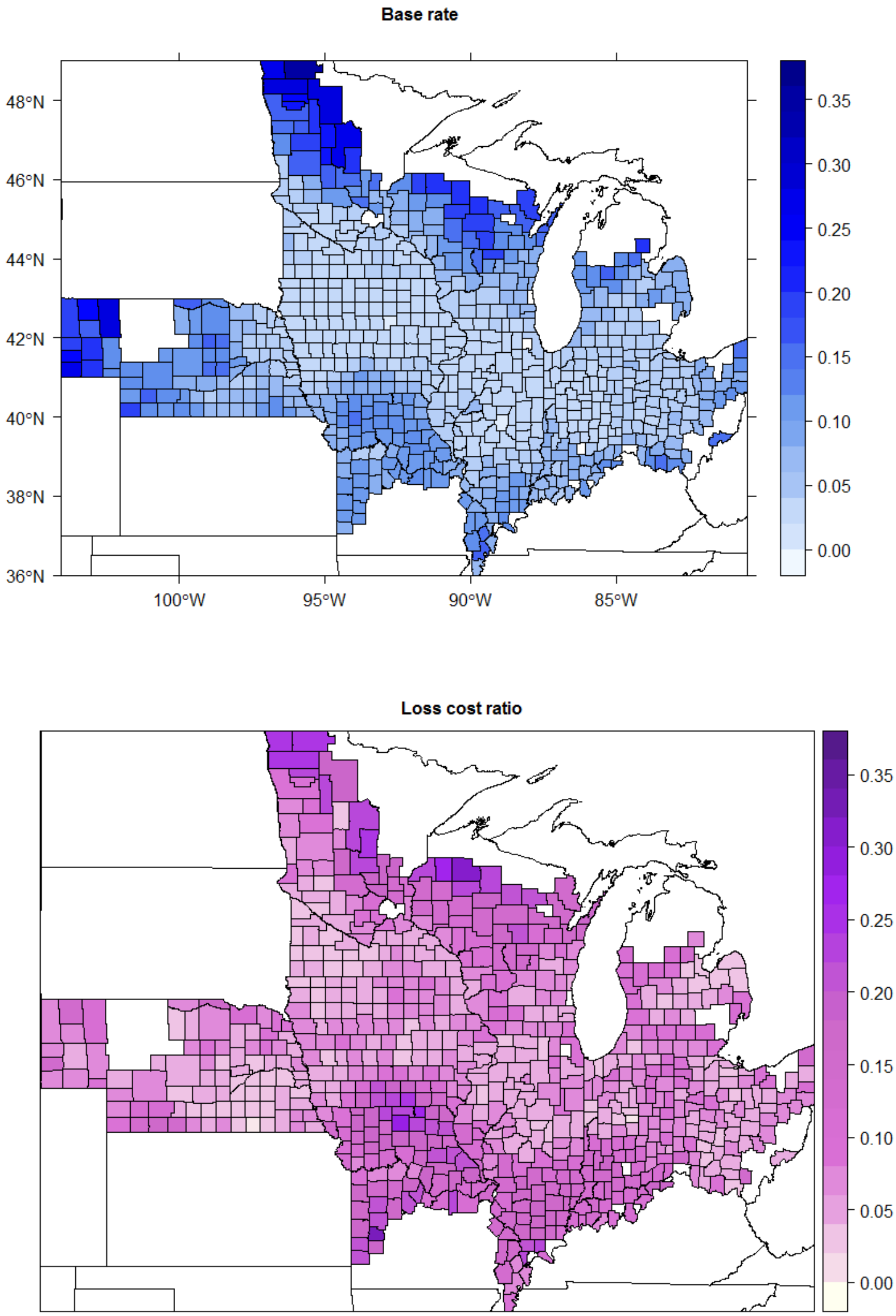


Figure 3 Box plots of variables

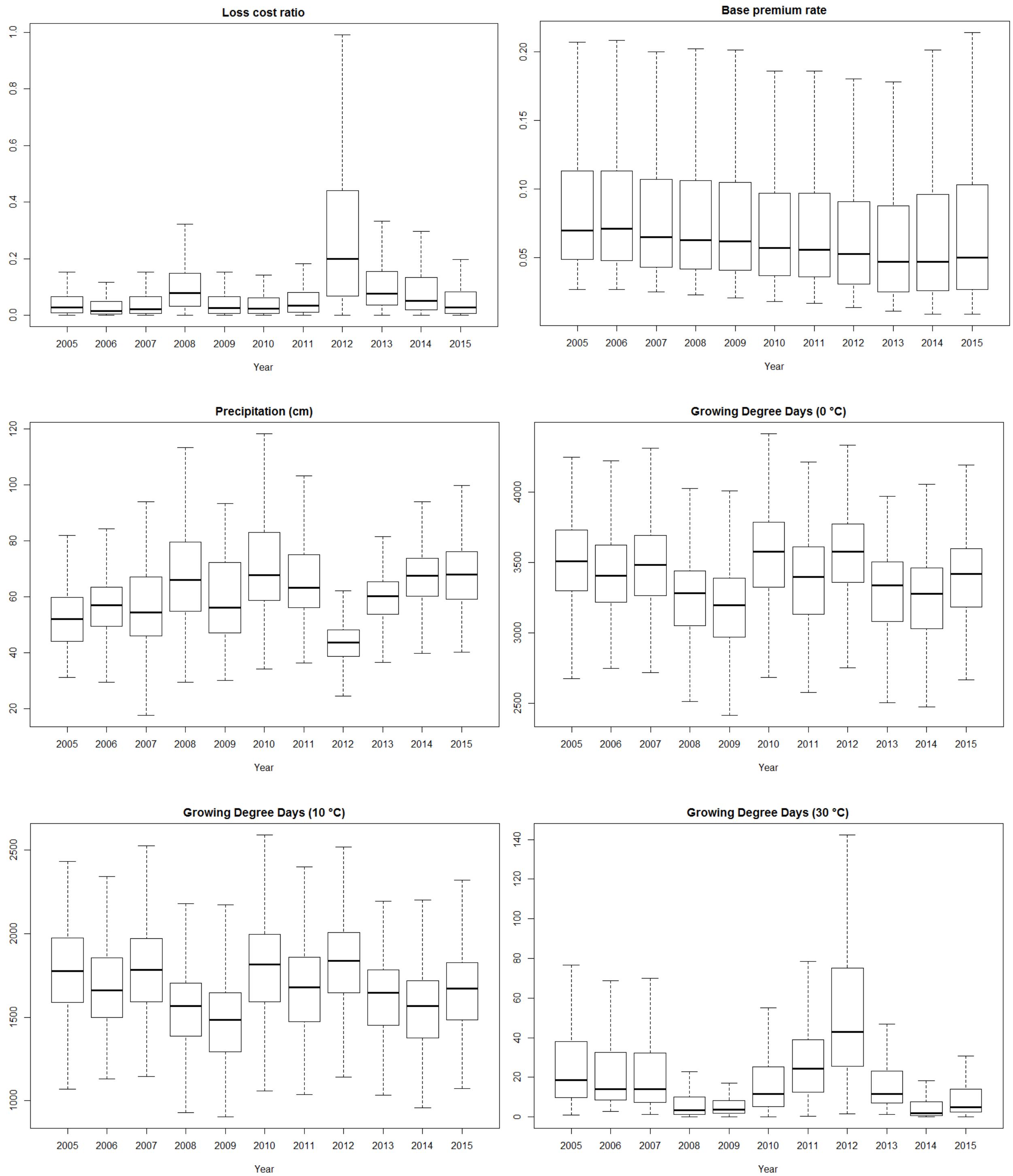
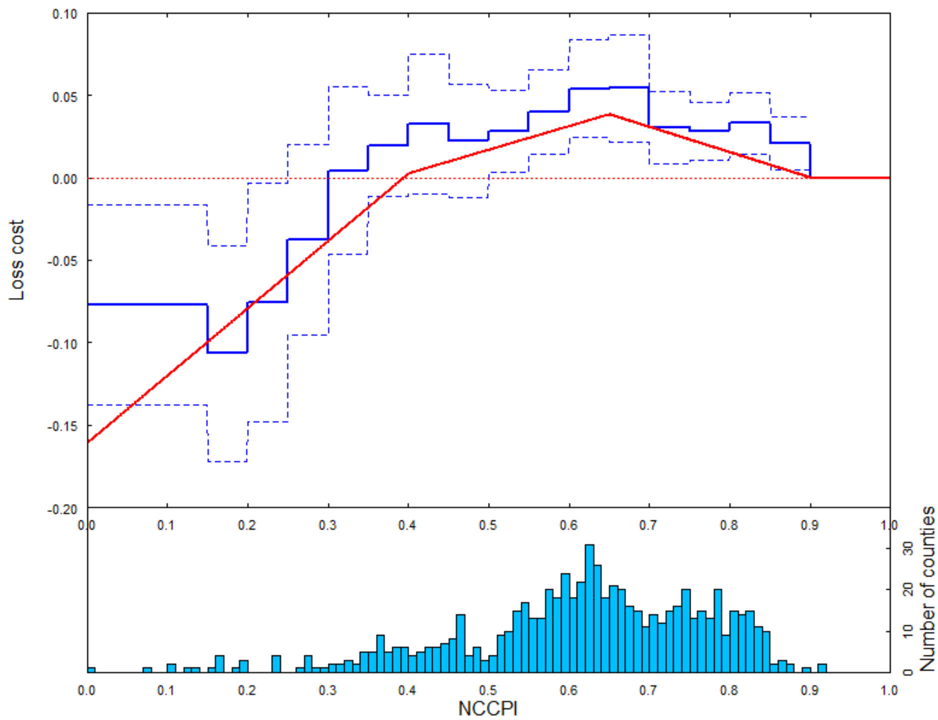


Figure 4 Nonlinear relation between county loss cost and NCCPI. Everything in this graph is in a relative sense. The highest NCCPI group (>0.9) is set as the base group (i.e., zero). The loss cost is *conditional on* base premium rate as well as weather variables and year fixed effects. Histogram on the bottom panel shows the frequency distribution of the NCCPI values.



Appendix: Tables of Robustness Check Results

Table A1 Estimated correlation of NCCPI to county loss cost, controlling for insurance coverage level

	(1)	(2)	(3)	(4)
Base rate	0.486***	0.567***	0.665***	0.566***
	(0.123)	(0.124)	(0.136)	(0.120)
NCCPI [spline 1]		0.180**	0.220***	0.464***
		(0.0798)	(0.0835)	(0.0902)
NCCPI [spline 2]		0.0585	0.110	0.148
		(0.133)	(0.139)	(0.0939)
NCCPI [spline 3]		-0.149	-0.0957	-0.186**
		(0.123)	(0.115)	(0.0806)
Precipitation				-0.000038
				(0.00488)
Precipitation squared				0.0000184
				(0.0000337)
Low GGD				-0.000132
				(0.000401)
Medium GGD				-0.000220**
				(0.000112)
High GGD				0.00470***
				(0.00174)
Coverage	0.0904	0.143	-0.0140	0.271*
	(0.129)	(0.132)	(0.0700)	(0.152)
Constant	-0.0124	-0.130	-0.0867	0.0734
	(0.101)	(0.104)	(0.0729)	(0.574)
Year Fixed Effects	No	No	Yes	Yes
Joint significance of 3 splines (p-value)	-	0.0081	0.0084	<0.0001
Observations	7,667	7,667	7,667	7,667
R-squared	0.315	0.319	0.479	0.570

Note: Dependent variable is the county-level loss cost ratio (Revenue Protection (RP) insurance plan). The knots for the linear splines are NCCPI=0.38 and NCCPI=0.65. OLS model estimates with Spatial heteroscedasticity and autocorrelation consistent (HAC) standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The estimation are based on a panel of 697 Corn Belt State counties over 11 years (2005−2015).

Table A2 Estimated correlation of NCCPI to county loss cost, for subsamples of various insured acreage thresholds

	Full sample	**>2,000 acres**	**>5,000 acres**	**>10,000 acres**
Base rate	0.410***	0.312***	0.243**	0.127
	(0.0965)	(0.111)	(0.115)	(0.137)
NCCPI [spline 1]	0.430***	0.532***	0.509***	0.313***
	(0.0834)	(0.123)	(0.118)	(0.113)
NCCPI [spline 2]	0.149	0.0702	0.0105	0.0410
	(0.0951)	(0.0803)	(0.0728)	(0.0924)
NCCPI [spline 3]	-0.156**	-0.0933	-0.0540	-0.0439
	(0.0694)	(0.0637)	(0.0644)	(0.0658)
Precipitation	-0.0000819	0.0000753	0.000837	0.00104
	(0.00491)	(0.00491)	(0.00525)	(0.00530)
Precipitation squared	0.0000183	0.0000133	0.00000567	0.0000015
	-0.0000339	-0.0000331	-0.000035	-0.0000362
Low GGD	-0.0000792	-0.000205	-0.000289	-0.000215
	(0.000410)	(0.000411)	(0.000419)	(0.000405)
Medium GGD	-0.000230**	-0.000160	-0.000123	-0.000145
	(0.000117)	(0.000107)	(0.000108)	(0.000107)
High GGD	0.00459***	0.00391**	0.00343**	0.00317**
	(0.00171)	(0.00165)	(0.00172)	(0.00154)
Constant	0.222	0.320	0.422	0.415
	(0.575)	(0.575)	(0.588)	(0.556)
Year Fixed Effects	Yes	Yes	Yes	Yes
Joint significance of 3 splines (p-value)	<0.0001	<0.0001	<0.0001	0.0007
Observations	7,667	5,566	3,993	2,640
R-squared	0.567	0.574	0.564	0.574

Note: Dependent variable is the county-level loss cost ratio (Revenue Protection (RP) insurance plan). The knots for the linear splines are NCCPI=0.38 and NCCPI=0.65. OLS model estimates with Spatial heteroscedasticity and autocorrelation consistent (HAC) standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The full sample is a balanced panel of 697 Corn Belt State counties over 11 years (2005−2015). All subsamples are 11-year balanced panels, too, where counties of any year's RP insured area less than the thresholds are excluded from the sample. The 2000 acres, 5000 acres, and 10000 acres thresholds are corresponding to 506 counties, 363 counties, and 240 counties respectively.

Table A3 Estimated correlation of NCCPI to county loss cost, by insurance type

	RP	**YP**	**RP+YP**
Base rate	0.410***	0.393***	0.340***
	(0.0965)	(0.0628)	(0.0902)
NCCPI [spline 1]	0.430***	0.208***	0.376***
	(0.0834)	(0.0584)	(0.0754)
NCCPI [spline 2]	0.149	0.0904*	0.156*
	(0.0951)	(0.0502)	(0.0920)
NCCPI [spline 3]	-0.156**	-0.0742*	-0.142**
	(0.0694)	(0.0397)	(0.0669)
Precipitation	-0.0000819	-0.000979	0.000148
	(0.00491)	(0.00264)	(0.00472)
Precipitation squared	0.0000183	0.0000176	0.0000161
	(0.0000339)	(0.0000182)	(0.0000326)
Low GGD	-0.0000792	0.0000142	-0.000130
	(0.000410)	(0.000189)	(0.000379)
Medium GGD	-0.000230**	-0.000118*	-0.000213*
	(0.000117)	(0.0000685)	(0.000114)
High GGD	0.00459***	0.00241**	0.00442***
	(0.00171)	(0.000973)	(0.00169)
Constant	0.222	0.0271	0.289
	(0.575)	(0.253)	(0.530)
Year Fixed Effects	Yes	Yes	Yes
Joint significance of 3 splines (p-value)	<0.0001	0.0003	<0.0001
Observations	7,667	7,655	7,667
R-squared	0.567	0.434	0.576

Note: Dependent variable is the county-level loss cost ratio for different insurance plans. "RP" is for the Revenue Protection insurance plan (or the Crop Revenue Coverage (CRC) plan). "YP" is for the Yield Protection insurance plan (or the Actual Production History (APH) plan). "RP+YP" is the two plans combined. The knots for the linear splines are NCCPI=0.38 and NCCPI=0.65. OLS model estimates with Spatial heteroscedasticity and autocorrelation consistent (HAC) standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample is a balanced panel of 697 Corn Belt State counties over 11 years (2005−2015).