Systemic Risk, Geography and Area Insurance

Xuche Gong
Department of Agricultural, Food, and Resource Economics
Michigan State University
Email: gongxuch@anr.msu.edu

David A. Hennessy
Department of Agricultural, Food, and Resource Economics
Michigan State University
Email: hennes64@anr.msu.edu

Hongli Feng
Department of Agricultural, Food, and Resource Economics
Michigan State University
Email: hennes65@msu.edu

Selected paper prepared for presentation at the 2018 Agricultural & Applied Economics Association Annual meeting, Washington, D.C., August 5-August 7

Copyright 2018 by Xuche Gong, David A. Hennessy, and Hongli Feng. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies
Abstract: The federal Crop Insurance Program has long been the most expensive agricultural commodity program in the U.S.. Recent studies argue that it is systemic risk, not asymmetric information that causes the failure of private crop insurance markets, but rare empirical studies have been completed to qualify the systemic risk in crop insurance market. Utilizing unit-level RMA corn yield historical data, our study measures the systemic risk in unit-level corn yield by county Fixed Effects models. We further model the systemic risk in terms of growing conditions, such as precipitation, temperature and land quality. Results show that systemic risk in corn yield is common and relatively large in major corn production counties, and is higher in counties that has better growing conditions.

Keywords: Crop insurance, corn yield, systemic risk, weather conditions, land quality

JEL Classification: G22, Q14, Q18
Systemic Risk, Geography and Area Insurance

1. Introduction

The Federal Crop Insurance Program (FCIP) has long been the most expensive agricultural commodity program in the U.S., with around $10 billion in expected taxpayer costs annually on $100 billion dollars of coverage (Woodard 2016), and thus have long been criticized for its poor actuarial fairness performance. However, historical experience strongly suggests that government subsidies are necessary for the existence of crop insurance markets (Kramer 1983), and private crop insurance market without government subsidies are doomed to fail (Miranda and Glauber 1997).

Numerous studies have examined causes of the failure of private crop insurance market. Most of them conclude that the failure mainly stems from asymmetric information problems, particularly the adverse selection and moral hazard problems (Chamber 1989; Skees and Reed 1986; Nelson and Loehman 1987; Goodwin and Smith 1995; Just et al. 1999). Asymmetric information problems increase the expected indemnity payment relative to premiums and result in actuarial losses of the insurer, forcing private insurance provider to quit the market. Thus the government is called to take over the high actuarial losses of private insurers if it wants to induce high participation rate in crop insurance market among crop growers and huge amount of subsidies are required. However, as pointed out by Miranda and Glauber (1997), the observed adverse selection and moral hazard may not be intrinsic to crop insurance, but are consequences of government program design, and the impact of asymmetric information problems may not be significant because the bulk of crop losses are caused by natural disasters, which are beyond the control of the insured and are observable by the insurer.

In recent years, a growing number of studies start to argue that it is systemic risk, not asymmetric information, that is the source of the failure of private crop insurance market. Systemic risk of crop insurance stems from the fact that, unlike other types of
property insurance, crop insurance losses are often driven by natural disasters which systemically affect a large amount of farms in a given region (Miranda and Glauber 1997). Thus there is a substantial correlation of yield losses across individual farms. The existence of systemic risk undermines insurer’s ability to diversify risk across individuals, forces the insurer to increase insurance premiums considerably for a targeted ruin probability and eventually results in a breakdown of the insurance market (Cummins and Trainar 2009). But unlike the intractability of asymmetric problems, costs induced by systemic risk in crop insurance market can be reduced by the design of insurance products based on some off-farm information, such as area yield and weather index (Halcrow 1949; Miranda 199; Miranda and Glauber 1997; Miranda and Farrin 2012).

Though the impact of systemic risk on insurance market has been theoretically well established, there are few empirical studies qualifying the systemic risk in crop insurance market. As the first attempt, Miranda and Glauber (1997) measures the systemic risk in terms of correlations across realized individual indemnity payments. Using a detrended farm-level data, they find that U.S. crop insurer portfolio are twenty to fifty times riskier than they would be otherwise if yield were stochastically independent across farms. Mason et al. (2003) also find a substantial level of systemic risk after further considering the systemic risk passed on to the federal government by means of reinsurance agreements. Other studies mainly study the systemic risk in terms of weather dependence across regions (Woodard and Garcia 2008; Okhrin et al. 2013; Goodwin and Hungerford 2014). However, possibly due to the lack of farm-level yield data, none of these studies characterizes the systemic risk in terms of the correlation of crop yield across individual farms, which is the most informative measurement of systemic risk in crop yield for insurers. The only exception is Claassen and Just (2010). They isolate systematic variation from random variation of yields at the farm or subfarm level and find that for Corn Belt Corn, 66% of total standardized variation is from the random component, and 64% of total systematic
variation is within-county variation. But their data are only from 69 Illinois counties, which limits the spatial extension of their conclusion.

In this paper, we directly measure the systemic risk of crop yield within a given county using the R-squared obtained by county Fixed Effects model. R squared, by its construction, measures the average fraction of total variation of individual farm yield that can be explained by the variation of the county average yield for all farms within the given county. The advantages of using R-squared of historical yield data to measure the systemic risk over using the correlation across individual indemnity payments and weather conditions are that 1) county average yield data combines all information that systemically affects individual yield; 2) it offers insurers a better reference to price on than what any other measurement could offer since most crop insurance products are priced based on yield information. Using the Risk Management Agency (RMA) insurance unit-level corn yield data and the National Agricultural Statistical Service (NASS) county average corn yield data, we measure the systemic risk of corn yield for 742 counties across the Midwest and Northern Great Plains, thus our measurement results are presentative for the major corn production areas. Our results suggest that systemic risk for corn yield is common and relatively large among those counties.

After obtaining the R-squared for each county, we further model it in terms of county weather variables to check the correlation between systemic risk and weather conditions. Following previous studies (Xu et al. 2013; Woodard and Garcia 2008; Okhrin et al. 2013), our weather condition variables include sum of precipitation over growing season, sum of suitable degrees over growing season and sum of stressful degrees over growing season. Since Woodard and Verteramo-Chiu (2017) find the importance of including land information for government to price crop insurance products, we also model R-squared in terms of county land quality. We find that systemic risk of corn yield significantly increases as growing conditions improve.
This paper begins with a brief literature review of systemic risk in crop insurance market in section 2. In section 3 we present our methodology and data information. We show our empirical results in section 4 and present our conclusion in section 5.

2. Literature review of measurement of systemic risk in crop insurance market

Starting with Miranda and Glauber (1997), there are several studies measuring the systemic risk in crop insurance market. As the first attempt, Miranda and Glauber (1997) measures the systemic risk by the ratio of the insurer’s portfolio risk, which is then measured by the coefficient of variation of total indemnities paid, when individual indemnity payments are correlated over the portfolio risk when individual indemnity payments are independent. A ratio equals to one suggests no systemic risk is present, and a ratio larger than one suggests the indemnities are systemically correlated. Thus Miranda and Glauber (1997) define the systemic risk in terms of correlations across realized individual indemnity payments, not in terms of correlations across individual crop yields. They use a detrended farm-level data to specify a joint distribution of yields and indemnities to simulate portfolio risks if individual indemnity payments are independent. Their simulation results found that U.S. crop insurers face portfolio risks 22 to 49 times larger than if indemnities were independent.

Following Miranda and Glauber (1997), Mason et al. (2003) further calculate the systemic risk passed on to the federal government by means of current Standard Reinsurance Agreement (SRA) and the total value of the SRA to the insurance industry. They find that in 1997 there was a 5% probability that RMA would have had to reimburse at least $1 billion to insurance companies and the fair value of RMA’s insurance service to insurance firms in 1997 was $78.7 million.

Some other studies measure systemic risk in crop insurance market by the correlation of weather conditions across spaces based on the argument that the stochasticity of crop yields is mainly driven by weather events. Employing copula
techniques and utilizing data from seven provinces, Okhrin et al. (2013) find that in China there is a significant spatial diversification effect of weather conditions measured by temperature indexes. Woodard and Garcia (2008) find that systemic weather risk increases as the spatial aggregate level increases, which justifies for the potential for weather derivative in agriculture for aggregators of risk.

The study closest to ours is by Claassen and Just (2010). Using corn data from 69 Illinois counties and wheat data from 43 counties in North and South Dakota, they decompose the total variation into systemic variation and random variation, where the systemic variation is defined as the expected value of unit-level level variance, farm-level yield variance or county-level variance. They find that 66% of total standardized variation of corn yield and 70% of total standardized variation of wheat yield are from the random component, and 64% of total systemic variation for corn yield and 66% of total systemic variation for wheat yield are from within-county variation. Their results are consistent with our findings that county average yield variation can explain about 35% to 39% of unit-level corn yield variation.

3. Methodology and data

We employ county-level Fixed Effects (FE) model to characterize unit-level systemic risk among corn yield within a given county. The model is given by:

\[ y_{i,c,t} - \bar{y}_{i,c} = \beta_0 + \beta_c \times (x_{c,t} - \bar{x}_c) + \epsilon_{i,c,t}, \quad \epsilon_{i,c,t} = u_{i,c,t} - \bar{u}_{i,c}; \]

where subscript \( i \) denotes insurance unit, \( c \) denotes county, \( t \) denotes yield year. \( y_{i,c,t} \) is insurance unit \( i \)'s yield in county \( c \) in year \( t \), \( \bar{y}_{i,c} \) is the insurance unit’s ten year average yield, \( x_{c,t} \) is the average yield for county \( c \) in year \( t \), \( \bar{x}_c \) is the ten-year average yield for the county, and \( \epsilon_{i,c,t} \) is unit \( i \)'s residual in year \( t \) net of the unit’s ten-year average residual. By running the Fixed Effects model for each county, we get the R-squared for each county, which measures the fraction of the ten-year unit corn
yield variation that explained by the ten-year county yield variation. We denote it as $R_c^2$. Thus counties with larger $R_c^2$ have larger systemic risk than counties with smaller $R_c^2$.

Since the extent of residual risk is likely to vary with overall growing conditions, we allow for heteroscedasticity as follows:

$$\text{var}(\varepsilon_{c,t} | x_{c,t} - \bar{x}_c) = h(x_{c,t} - \bar{x}_c); \quad h(x_{c,t} - \bar{x}_c) = e^{\delta_0 + \delta_1(x_{c,t} - \bar{x}_c)}.$$  \hspace{1cm} (2)

which claims that the variance of residuals vary with county attributes in an exponential form.

Feasible Generalized Least Squares (FGLS) is then employed to estimate the parameter set $\{\beta_c, \delta_0, \delta_1\}$. The weighting that FGLS places on each residual may be viewed as a measure of the unit’s importance as a source of variation.

After obtaining the $R_c^2$ for each county, we then turn to model systemic risk in terms of geographic attributes, including some weather and land quality variables by an Ordinary Least Square (OLS) model, which is given by:

$$R_c^2 = \alpha_0 + \alpha_P P_c + \alpha_s G_c + \alpha_s S_c + \alpha_Q Q_c + \eta_c;$$  \hspace{1cm} (3)

where $P_c$ is the thirty-year (1978-2007) average of the sum of precipitation over growing season (1st May to 31th August) for county $c$, $G_c$ is the thirty-year average of the sum of suitable degrees in growing season for county $c$, $S_c$ is the thirty-year average of sum of stressful degrees in growing season for county $c$, $Q_c$ is the fraction of land that is in land capability categories I or II for county $c$.

Unit corn yield data are obtained from the 2008 unit-level Risk Management Agency (RMA) records, which contains 1504706 insurance units from 1138 counties.

---

1 For a more detailed discussion of how the sum of suitable degrees and the sum of stressful degrees during the growing season are constructed, see Xu et al. (2013)
We only keep the traditional major corn production states in the Midwest Great Plains, including Illinois, Indiana, Iowa, Kansas, Kentucky, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin. Each insurance unit has a record of yield history for ten years. To avoid any distortion of our results caused by incorrect data induction made by data collector when actual yield history is not available, we only keep those units that has ten actual yield records. We further drop units who have history record year that is dated prior to 1990 to obtain a serial of yield history as continuous as possible. Thus only insurance units that have their ten history yield records all located between year 1990 and 2007 are adopted in our analysis. To make our conclusion more representative, we further drop counties that have less than 30 insurance units. After matching with NASS county average yield data and weather and land data, we finally get 4950820 observations of 495082 insurance units from 742 counties. Figure 1 maps the number of insurance units in each county. The minimum unit number is 31 and the maximum unit number is 3658, so our data is not only representative for the overall corn production region, but is also representative for each selected county.

County average corn yield data are obtained from National Agricultural Statistical Service (NASS), which are widely used in other crop insurance studies. Since the 2018 American Farm Bureau Federation platform includes seeking to use RMA county average yield data as the measure of county yields when computing area yield indemnities, we also calculate the county average corn yield using the RMA unit-level yield data as a robustness check. The RMA county average corn yield is calculated by first summing up all unit’s yield in that county and then, dividing it by the sum of planted acres in that county. As we will show below, analysis results using RMA county average data are very close to results using NASS county average data.

Weather data are from National Oceanic and Atmospheric Administration files, while land quality data are from National Resource Inventory files. We use thirty-year (1978 to 2007) average of weather variables to reflect the long-run weather condition
in a given county. Since there is little variation of land capability across years, we use the 2010 land capability information as the fixed measure of land quality for all sample years.

4. Empirical results

Panel A of Table 1 presents the descriptive statistic of unit yield and county average yield variables. The mean of unit yield is 141.10 bushels per acre and the standard deviation is 44.28. The mean of NASS county average yield is 126.80 bushels per acre, smaller than but close to the mean of RMA county average yield, which is 131.11 bushels per acre. The standard deviation and range of NASS county average yield are also smaller than than RMA county average yield, but there is no significant difference.

Panel B of Table 1 presents the descriptive statistics of thirty-year county average weather condition variables and county land quality variable. For weather data, the mean of county average precipitation over growing season is 386.6 millimeters, the mean of sum of suitable degrees over growing season and the mean of sum of stressful degrees over growing season are 1272 degrees Celsius and 20.51 degrees Celsius, respectively. The mean of the fraction of good land for counties, measured by the fraction of land in a county that is in land capability categories I or II is 44.7%. All weather condition and land quality variable have sufficient variation for us to identify the correlation between systemic risk and growing conditions.

Table 2 lists the descriptive statistics of $Rc^2$. The first two rows present $Rc^2$ obtained by county Fixed Effects models. The mean of $Rc^2$ obtained by NASS county Fixed Effects model is 0.349, suggesting that about 35% of individual unit yield variation can be explained by the county yield variation. The mean of $Rc^2$ obtained by RMA county Fixed Effects model is 0.390, close to former one, suggesting that our measure of systemic risk is robust to the choice of county average yield data. The median of $Rc^2$ obtained by NASS county Fixed Effects model and RMA county Fixed
Effects model are 0.364 and 0.397, respectively, and the maximum of $R_c^2$ using the two county average yield data are 0.756 and 0.768, respectively. The last two rows present $R_c^2$ obtained by FGLS models, and the results are very similar to those obtained by Fixed Effects model. To conclude, regardless of which model and which county average yield data are used, we find a common and relatively large systemic risk in unit-level corn yield, and the magnitude is similar to the magnitude found by Claassen and Just (2010).

Besides the descriptive statistics, we also map the geographical distribution of $R_c^2$ in Figure 2 to Figure 5. Four figures are very similar in patterns and all suggest that systemic risk generally increases as one moves south and west. This trend lines up with the trend of improving growing conditions, such as more precipitation, larger fraction of good land, larger sum of suitable degrees and smaller sum of stressful degrees over growing season, as shown by Figure 6 to Figure 9. Our results also show that the systemic risk in corn yield is low in Nebraska and Kansas where irrigated corn is prominent and is high in poorer land in the Eastern Corn Belt.

Results of OLS model listed in Table 3 confirm findings in Figure 2 to Figure 5. Column (1) of Table 3 shows that when precipitation over growing season increases by 100 mm, $R_c^2$ obtained by NASS county Fixed Effects model will increase by 0.063, which is about an 0.417 standard deviation increment in $R_c^2$, and is significant at 1% level. When sum of suitable degrees over growing season increases by 100 degrees Celsius, $R_c^2$ obtained by NASS county Fixed Effects model will significantly increase by 0.017. If sum of stressful degrees over the growing season increase by 10 degrees Celsius, $R_c^2$ obtained by NASS county Fixed Effects model will significantly decrease by 0.014. When the fraction of land in land capability category I or II increases by 10 percentage points, $R_c^2$ obtained by NASS county Fixed Effects model will increase by 0.020. Results for $R_c^2$ obtained by RMA county Fixed Effects model in column (2) of Table 3 are similar to those obtained by NASS county Fixed Effects in both coefficient signs and statistical significance, except for that the sum of
stressful degrees over growing season does not have significant impact on $R_c^2$ when we adopt RMA county average yield data. Results for $R_c^2$ obtained by FGLS models in column (3) and column (4), again, are consistent with results obtained by county Fixed Effects models. All these results suggest that systemic risk is higher in geographic area that has better weather and land conditions.

5. Conclusion

This paper contributes to the growing body of literature studying the systemic risk in crop insurance market in two aspects: we directly measure the systemic risk in terms of yield data, and we model the measured systemic risk in terms of weather condition and land quality variables to examine the correlation between the systemic risk and growing conditions.

Using a representative set of RMA unit-level corn yield data and NASS county average corn yield data, our study find that county-level systemic risk is common and large in extent for corn yield. This finding provides direct empirical evidence for the existence of systemic risk in crop insurance market and provide reference for the design of FCIP products such as area yield insurance. Our study further finds that better growing conditions lead to higher systemic risk. An important implication of this finding may be that providing area insurance products may be helpful to reduce subsides in area with good growing conditions.

References


Figure 1. Number of units in each county
Figure 2. Geographical distribution of $R_c^2$ using NASS county average yield data of FE model.
Figure 3. Geographical distribution of $R^2_c$ using RMA county average yield data of FE model.
Figure 4. Geographical distribution of $R_c^2$ using NASS county average yield data of FGLS model.
Figure 5. Geographical distribution of $R_c^2$ using RMA county average yield data of FGLS model.
Figure 6. Geographical distribution of sum of precipitation in growing season
Figure 7. Geographical distribution of sum of suitable degrees in growing season
Figure 8. Geographical distribution of sum of stressful degrees in growing season
Figure 9. Geographical distribution of fraction of land in land capability category I and II
Table 1. Descriptive statistics of unit yield, county average yield and weather and land condition variables.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>Std. Dev</td>
<td>Min</td>
<td>Max</td>
</tr>
<tr>
<td><strong>Panel A: yield variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMA unit yield</td>
<td>4950820</td>
<td>141.10</td>
<td>44.28</td>
<td>0</td>
<td>396</td>
</tr>
<tr>
<td>NASS county average yield</td>
<td>13241</td>
<td>126.80</td>
<td>30.07</td>
<td>18.40</td>
<td>204</td>
</tr>
<tr>
<td>RMA county average yield</td>
<td>13241</td>
<td>131.11</td>
<td>34.48</td>
<td>0</td>
<td>227</td>
</tr>
<tr>
<td><strong>Panel B: weather and land condition variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precipitation (mm)</td>
<td>742</td>
<td>386.6</td>
<td>55.90</td>
<td>168.6</td>
<td>516.3</td>
</tr>
<tr>
<td>Sum of suitable degrees (Celsius)</td>
<td>742</td>
<td>1,272</td>
<td>164.6</td>
<td>10.60</td>
<td>1,675</td>
</tr>
<tr>
<td>Sum of stressful degrees (Celsius)</td>
<td>742</td>
<td>20.51</td>
<td>19.21</td>
<td>0.000</td>
<td>111.6</td>
</tr>
<tr>
<td>Fraction of good land</td>
<td>742</td>
<td>0.447</td>
<td>0.227</td>
<td>0.000</td>
<td>0.935</td>
</tr>
</tbody>
</table>
Table 2. Descriptive statistics of $R_c^2$

<table>
<thead>
<tr>
<th></th>
<th>$R_c^2$ _NASS</th>
<th>$R_c^2$ _RMA</th>
<th>$R_c^2$ _NASS</th>
<th>$R_c^2$ _RMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>FE</td>
<td>N</td>
<td>742</td>
<td>0.349</td>
<td>0.151</td>
</tr>
<tr>
<td>FGLS</td>
<td>N</td>
<td>742</td>
<td>0.340</td>
<td>0.146</td>
</tr>
</tbody>
</table>
Table 3. Regression result $R_c^2$ on weather and land variables

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>From FE</td>
<td>From FGLS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_c^2$_NASS</td>
<td>$R_c^2$_RMA</td>
<td>$R_c^2$_NASS</td>
<td>$R_c^2$_RMA</td>
<td></td>
</tr>
<tr>
<td>Precipitation (100mm)</td>
<td>0.063***</td>
<td>0.053***</td>
<td>0.061***</td>
<td>0.061***</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Sum of suitable degrees/100</td>
<td>0.017***</td>
<td>0.011***</td>
<td>0.019***</td>
<td>0.019***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Sum of stressful degrees/10</td>
<td>-0.014***</td>
<td>-0.005</td>
<td>-0.012***</td>
<td>-0.012***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Fraction of good land (percentage)</td>
<td>0.002***</td>
<td>0.001***</td>
<td>0.002***</td>
<td>0.002***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.169***</td>
<td>-0.008</td>
<td>-0.184***</td>
<td>-0.181***</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.044)</td>
<td>(0.045)</td>
<td>(0.045)</td>
</tr>
<tr>
<td>Observations</td>
<td>742</td>
<td>742</td>
<td>742</td>
<td>742</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.220</td>
<td>0.149</td>
<td>0.213</td>
<td>0.211</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1