



# The Implication of Copula-Based Models for Crop Insurance and Reinsurance Under Systemic Risk

Yan Sun<sup>1\*</sup> and Ke Wang<sup>2</sup>

<sup>1</sup>School of Economics and Management, Institute of Disaster Prevention, Langfang, China, <sup>2</sup>China Agriculture Reinsurance Corporation Beijing, Beijing, China

The purpose of this paper is to estimate county-level aggregate crop insurance and reinsurance losses under systematic risk. The effect of dependence risk on losses assessment and insurance pricing is quantified by establishing joint distribution functions between county-level yields using different forms of multivariate copulas. The research also stresses the importance of selecting the appropriate copula form for estimating losses. This article highlights several significant findings. The estimated aggregate losses for related counties are not significantly different between the model assuming dependence (copula-based) and the model assuming independence (individual) that adheres to the equivalence principle. On the other hand, the copula-based model has a discernible effect on the estimated Value-at-Risk and Expected Shortfall for related counties. Additionally, for the different layers of the Standard Reinsurance Agreement policy, the copula-based model can measure the aggregate losses more accurately than the individual models. Furthermore, when there is obvious tail dependence in the related counties' yields, the vine copula function form, which provides a more flexible description of the dependence, is more suitable for quantifying tail risk. As a result, insurers and governments should conduct a more comprehensive risk assessment of yield dependence when rate-making and allocating subsidies.

**Keywords:** copula function, crop insurance, reinsurance, actuarial fair pricing, effective analysis

## OPEN ACCESS

### Edited by:

Benjamin Lamptey,  
University of Leeds, United Kingdom

### Reviewed by:

Yahua Xu,  
Central University of Finance and  
Economics, China  
Cocou Amegnaglo,  
National University of Agriculture,  
Benin

### \*Correspondence:

Yan Sun  
kkfll@sina.com

### Specialty section:

This article was submitted to  
Environmental Economics and  
Management,  
a section of the journal  
Frontiers in Environmental Science

**Received:** 09 April 2022

**Accepted:** 22 June 2022

**Published:** 18 July 2022

### Citation:

Sun Y and Wang K (2022) The  
Implication of Copula-Based Models  
for Crop Insurance and Reinsurance  
Under Systemic Risk.  
Front. Environ. Sci. 10:916494.  
doi: 10.3389/fenvs.2022.916494

## 1 INTRODUCTION

The federal crop insurance program in the United States, which currently provides yield-based coverage and revenue insurance to help agricultural producers avoid losses due to low crop yields or lower-than-expected crop prices, is one of the most extensive support programs for agricultural producers.

Under the federal crop insurance program, producers can purchase insurance policies at a subsidized rate. These insurance policies compensate producers for current losses caused by either below-average yields (yield-based coverage) or below-average revenue (revenue-based coverage) (revenue insurance). Under a Standard Reinsurance Agreement (SRA), policies are sold through private insurance companies. USDA's Risk Management Agency (RMA) subsidized the policy premium and a portion of the companies' administrative and operating expenses and shares in the companies' underwriting gains and losses. Premium subsidy rates have increased in recent decades for a variety of policy types so that producers typically pay only about 40% of their premiums. It cost

the federal government \$5 billion in 2016 and nearly \$9 billion on an annual average basis over the last 5 years.

Many studies cast doubt on the government's support for crop insurance schemes, claiming that the substantial subsidies offered to farmers and insurance corporations would result in a variety of inefficiencies in the aggregate economy (Smith and Goodwin, 1996; Goodwin and Vado, 2007; Goodwin and Smith, 2013; Smith, 2011). Otherwise, proponents of such schemes argue that agricultural damage from possible systemic risks such as droughts and floods is too great for commercial insurers and reinsurers to cover adequately (Miranda and Glauber, 1997; Coble et al., 2003).

Crop insurance policies are designed to protect against income losses caused by fluctuations in crop prices and yields. Premium pricing for such contracts is an actuarial issue that involves describing the joint distribution of prices and yields (Coble et al., 2010). A density function of crop yields under proper information circumstances must be determined as part of the actuarial fair pricing procedure. Unlike most other types of insurance, crop insurance lacks the necessary information to accurately estimate the likelihood and magnitude of losses. One of the most significant issues is a lack of relevant data. Today's standard method regresses historical yield time series data and then uses the predicted and adjusted residuals to estimate a crop yield conditional density function. Numerous studies have been conducted to determine the marginal distribution of yields. For example, Atwood, Shaik, and Watts (2003) argue that calculating trends with short-term panel data biases the analysis in a Type II direction, failing to reject normality when the underlying distribution is nonnormal. This research reveals that yield parameterization can influence crop insurance payouts. Unchecked yield distribution specifications may result in economic inaccuracies in crop insurance rating and expected payouts (Sherrick et al., 2004). Goodwin et al. (2000) compared price risk estimation and insurance rates and used specification tests to analyze distributional assumptions. The reality that the best-fitting distribution in-sample is not necessarily the best choice out-of-sample is critical for determining crop yield distributions. Woodard and Sherrick (2011) present a method for estimating flexible and efficient mixed models using cross-validation that alleviates many of these model selection difficulties. Nelson and Preckel (1989) proposed the conditional beta distribution as a parameterized model for the probability distribution of agricultural output by a two-stage maximum likelihood estimation method.

In addition, inaccurate predictions and inferences of a parametric may result from incorrect distributional choices. Due to these restrictions, nonparametric approaches for estimating yield distributions have been developed (Goodwin and Ker, 1998). Tolhurst and Ker (2015) propose estimating crop yields using mixtures with embedded trend functions to account for potentially different technical development rates. When distinct trend functions are incorporated into mixture components, projected conditional yield densities differ from those derived from detrended data. The focus of these researches has been on estimating the marginal distribution of

crop yields and determining whether an optimal marginal distribution can be established. Even though years of research have failed to determine which distribution provides the most accurate yield estimate, the current trend is toward employing more flexible parameter distributions that accommodate for skewness and excess kurtosis (Ramsey and Goodwin, 2019).

For the discussion of utilizing the copula function to measure the dependence of variables in contract design and pricing for government revenue insurance, Ramsey et al. (2019) conclude that the economic impact of copula choice on pricing for individual policies was minor, and changes in marginal distributions significantly influence rates. At the moment, the revenue insurance pricing model makes two assumptions regarding the structure of the dependence. The first assumption is that the Gaussian copula accurately captures the relationship between price and yield, and the second is that the Gaussian copula function is deterministic for each state (Ramsey et al., 2019). The second implication is that crop insurance rate determination ignores the county-level yield dependence, which may result in some mistakes or omissions of systemic risk.

Because yield risk is spatially correlated, a measure of the spatial dependence of yield is required. The empirical investigation has indicated that the systemic risk tends to be much greater when extreme weather circumstances such as drought occur (e.g., Goodwin, 2001). Quantifying the degree of systemic risk is critical for addressing public policy issues regarding the need for large crop insurance subsidies, especially relevant given the reliance on county-level yield hazards. For example, Wang and Zhang (2003) use spatial statistics to investigate the efficiency of risk pooling for crop insurance in the presence of correlation. When modeling the insured losses dependence structure, the size of the buffer load is determined by the value-at-risk (VaR) of aggregate losses between states (Xu et al., 2010; Okhrin et al., 2012). Goodwin and Hungerford (2015) use a variety of copula models to assess the degree to which weather and natural disaster risks in agriculture are systemic and state-dependent. Their findings indicate that the approach taken to quantify multiple, correlated sources of risk may have significant implications for the accurate measurement of portfolio risks. Studies have also argued for improved pricing accuracy based on yield data from surrounding counties and that the dependent yield information can make insurance pricing more accurate when the dependency structure is known (Racine and Ker, 2006; Zhu, Goodwin, and Ghosh, 2014).

The critical challenge of estimating and pricing the loss risk of a portfolio of crop insurance policies has attracted significant attention. The same component that creates dependencies between yields in different locations in crop yield modeling can also lead to systematic risk in a crop insurance portfolio. To model a portfolio of county-level contracts, we must consider the inter-dependencies between the portfolio's county-level yields. A copula can be used in the yield model to estimate the dependence's losses, and then the contract can be priced actuarially based on the mean of the loss distribution. The copula function that represents the dependence of random variables and generates the distribution of random variables must be estimated in

modeling this yield structure. Goodwin and Hungerford (2015) propose that RMA may use the deterministic Gaussian copula assumption to fit the dependent structure of price and yield when the data sample is small and the effects of different copula function forms on rate-making are negligible.

So far as we know, no objective criterion exists to compare the accuracy of copula-based approaches to those that do not account for yield dependence. Moreover, no research has been undertaken to ascertain the efficacy of the copula-based approach for crop insurance indemnity in the presence of systemic risk. This research aims to determine the extent to which systematic risk directly or indirectly affects indemnification by developing aggregate losses models for county-level yields using various copula forms. We are particularly interested in the tail risk of the losses distribution, as the possibility of experiencing large losses is critical for getting reinsurance funds and subsidies. We also measure the tail risk here by simulating the aggregate losses' VaR and Expected Shortfall (ES). Stochastic simulation approaches can calculate the aggregate losses, ES, and VaR using the estimated copula functions and marginal distributions of county-level yields. By comparing the simulation results, we could determine whether the pricing model with a dependence assumption is more effective than the individual model with no dependence assumption.

This research conducts an empirical study of the efficacy of risk measurement in two particular areas. We test the effectiveness of the estimation approach for crop insurance indemnity of risk portfolio based on copula-based and individual models. Then, we examine the accuracy of those two models for the aggregate losses assessment at various layers of SRA policy. We conclude that, compared to the individual models, the copula-based models have a negligible effect on estimating aggregate losses under the equivalence principle while beating the individual models in estimating VaR and ES values. Otherwise, the copula-based models have significant consequences for appropriately estimating portfolio risk at various layers of SRA policy at various layers of SRA policy. The Vine copula function approach, which allows for a more flexible definition of dependence risk, is more suitable for loss measurement in crop insurance when the tail dependence can be captured.

The rest of the paper is structured in the following. **Section 2** provides an overview of the data sources. **Section 3** discusses using copula functions as the Vine copula in modeling county-level dependent risk, the model-building procedure, and the effectiveness assessment. In **Section 4**, this approach is used to county-level data sets of the United States to conduct empirical analysis, and the findings are reported. The paper is ended by exploring the effectiveness of copula-based models for estimating dependent risk in crop insurance.

## 2 DATA SOURCES

We obtain the average county-level corn yield per acre from the US Department of Agriculture's National Agricultural Statistics

Service (NASS) (USDA). SRA established three state-based reinsurance pools<sup>1</sup>. Because our objective is to quantify yield dependence under systemic risk for each group, we restrict our analysis to the rainfed region of the United States (east of the 100th meridian) from 1977 to 2007. Our model consists of three states (Illinois (IL), Indiana (IN), and Iowa (IA) in Group 1, Alabama (AL), Arkansas (AR), and Florida (FL) in Group 2, and Pennsylvania (PA), New Jersey (NJ), and Maryland (MD) in Group 3) each with three related counties. As a result, the data sets include the aggregated yields of 27 counties (**Figure 1**). Notably, we randomly select the counties in each state, which either contain adjoining or nonadjacent counties, to test the effectiveness of the copula-based model for losses estimation in different situations.

Each county's raw yields were detrended using robust regression procedures similar to those used in RMA. The regression method we use, M-estimation, is a more generalization of MM-estimation (Huber, 1973). Finger (2010) notes that the errors associated with this detrending method are likely to be small or moderate in magnitude.

## 3 METHODOLOGY

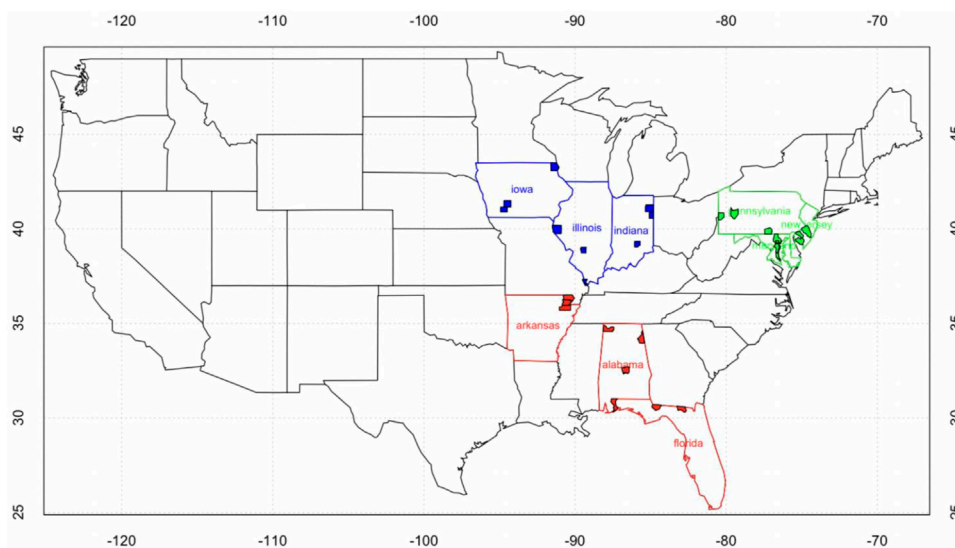
### 3.1 The Methods for Measuring Aggregate Losses

The SRA decides the government's losses/gain split with Approved Insurance Providers (AIPs, the RMA's designation for insurance companies), which means that the government will bear all unaffordable underwriting losses. The FCIC provides both proportional and non-proportional reinsurance under the present SRA. Insurance companies are permitted to commercially reinsure any portion of their liability that has not been ceded to the FCIC, as long as they disclose all facts in their operations plan. The SRA's non-proportional reinsurance provision reduces insurers' liability exposure on their retained book of business. The FCIC's share of losses on an insurer's retained book of business varies with each fund and is contingent on the insurer's losses ratio. The state losses ratio of an individual AIP determines the AIP's gain/losses share. Each state fund's losses ratio is determined separately<sup>2</sup>. The FCIC employs a graded structure in which insurers are held accountable for diminishing percentages of eventual net losses as their losses ratio improve.

Under the SRA, a company's retention levels and potential gain/losses are greatest when policies are placed in the Commercial Fund and lowest in the Assigned Risk Fund.

<sup>1</sup>"State Group 1" means Illinois, Indiana, Iowa, Minnesota, and Nebraska. "State Group 2" means Alabama, Arizona, Arkansas, California, Colorado, Florida, Georgia, Idaho, Kansas, Kentucky, Louisiana, Michigan, Missouri, Mississippi, Montana, North Carolina, North Dakota, New Mexico, Ohio, Oklahoma, Oregon, South Carolina, South Dakota, Tennessee, Texas, Virginia, Washington, and Wisconsin. "State Group 3" means Alaska, Connecticut, Delaware, Hawaii, Maine, Massachusetts, Maryland, Nevada, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Utah, Vermont, West Virginia, and Wyoming.

<sup>2</sup>The losses ratio that determines the AIP percent share of the underwriting results is based on their entire state book of insurance, i.e. it is the combined losses ratio for all crops and all contracts for the state.



**FIGURE 1 |** The selected states and counties for Group 1, Group 2 and Group 3.

**TABLE 1 |** Shares of underwriting gains and losses to insurance companies under the 2021 Standard Reinsurance Agreement.

| Losses ratio | Reinsurance fund          |                 |         |         |
|--------------|---------------------------|-----------------|---------|---------|
|              | Assigned risk             | Commercial fund |         |         |
|              | Percentage of losses/Gain |                 |         |         |
|              |                           | Group 1         | Group 2 | Group 3 |
| Losses       |                           |                 |         |         |
| 1.0–1.6      | 7.5                       | 65              | 42.5    | 42.5    |
| 1.6–2.2      | 6.0                       | 45.0            | 40.0    | 40.0    |
| 2.2–5.0      | 3.0                       | 10.0            | 5.0     | 5.0     |
| >5.0         | 0.0                       | 0.0             | 0.0     | 0.0     |
| Gains        |                           |                 |         |         |
| 0.65–1.0     | 22.5                      | 75.0            | 97.5    | 97.5    |
| 0.5–0.65     | 13.5                      | 40.0            | 40.0    | 40.0    |
| <0.5         | 3.0                       | 5.0             | 5.0     | 5.0     |

FCIC receives 80% of the premium and associated liability in the Assigned Risk Fund, while the company retains 20%. Companies must keep a minimum of 30% (and may retain up to 100%) of the premium and associated liabilities in the commercial fund. FCIC pays increasing shares of indemnities from the liability retained by the company, based on the company’s state-level losses ratio (indemnities divided by total premium) in the fund, with FCIC paying the total losses once the losses ratio surpasses 5.0. The proportions of income and losses that fall to private insurance companies vary significantly by the fund (as shown in **Table 1**). For Assigned Risk policies, for example, the highest potential underwriting losses is 16.5 percent ( $0.6 \times 7.5 + 0.6 \times 6.0 + 2.8 \times 3$ ) of company-retained premium. However, the company’s potential for underwriting gains on policies put in the Assigned Risk Fund is likewise relatively small, as 1.4 percent of retained premium at most. In comparison, on policies put in the Commercial Fund, Group 1 might earn up to 34.75 percent of the retained premium on Commercial Fund policies. However, the downside risks are

also greater, and the maximum permissible underwriting losses is 94 percent of the retained premium.

The main problem addressed here is evaluating the effectiveness of measuring insurance payment methods based on the dependence risk model. We must determine whether there is any potential for dependence between the portfolio’s county-level yields, and this dependence could be captured in the insurance losses model to reflect the more accurate fair rates. Then, from a reinsurance perspective, losses evaluation exposed to systematic risk is crucial for determining underwriting terms. Risk assessment at various layers enables us to determine the effect of yield dependence on rate-making.

Crop insurance rate-making involves the simultaneous distribution of prices and yields. To keep things simple, we will focus exclusively on the indemnity under systemic risk and ignore the volatility of the corn price. The indemnity for county  $i$ ,  $indem_i$ , for the area yield insurance contract that guarantees the coverage level, say  $\lambda \hat{y}_i$ , (Ker and Coble 2003) is:

$$indem_i = Pr(y_i < \lambda \hat{y}_i) (\lambda \hat{y}_i - E[(y_i | y_i < \lambda \hat{y}_i)]) \quad (1)$$

Where  $y_i$  is the realized yield,  $\hat{y}_i$  is the predicted yield in county  $i$  expected yield, and  $\lambda$  is a coverage level between 0 and 1. In this study, we focus on  $\lambda = 0.9$ , which is the most commonly selected coverage level, and it accounts for 95% of the policies (Ker, Tolhurst, and Liu, 2015). Using **Eq. 1**, we can calculate the aggregate losses based on the copula-based and individual models for the related counties portfolio.

The objective of reinsurance is to provide protection for aggregate losses  $L$  that exceed some trigger level  $T$  based on the level of losses that the insurer can absorb. By contract design, once the aggregate losses exceed this trigger, the reinsurance pays some fixed proportion of aggregate losses  $L$  usually up to some predetermined cap  $C$  on the reinsurer’s exposure. Therefore, if losses are less than the trigger, the contract pays nothing. The



contract pays out  $L - T$  if losses fall in the range between the trigger and the cap. If losses exceed the cap, the contract pays out the difference between the cap and the trigger,  $C - T$ . Using this basic structure, one can specify the payout of a layer (a, b] of payout, denoted by  $\text{Pay}_{[a,b]}$  of the reinsurance as follows:

$$\text{Pay}_{[a,b]} = \text{Max} [0, \text{Min}(L - T, C - T)] \quad (2)$$

Therefore, a layer contract covers part of the losses between trigger level  $T$  and predetermined cap  $C$ , with the maximum payout limited to the difference  $b - a$ . It is also referred to as an excess of losses (XOL) insurance policy. The reinsurance program of SRA is based on a modified version of the XOL insurance policy, which covers a certain percentage of the standard layer policy. However, based on minimizing the moral hazard and adverse-selection problems associated with writing company reinsurance. The payout function on these XOL contracts would also be a function of aggregate losses for the different groups. Thus, the expected payout of the contract has the form  $\delta \text{Pay}_{[a,b]}$ ,  $0 < \delta < 1$  representing the payout proportion, which depends on the contract. Under the SRA program, there are three different coverage layers: losses ratios between 1.00–1.60, 1.60–2.20, and 2.20–5.00, separately.

Other than estimating the aggregate losses for the regular contract and SRA layers contract, we also estimate the VaR of the aggregate losses, which can be applied to measure the systemic risk inherent to an insurance portfolio. The VaR of aggregate losses for the risk portfolio in each State is:

$$\text{VaR}_{AL}(\alpha) = \inf \left\{ al: P \left( \sum_{i=1}^3 L(y_i) > al \right) \leq \alpha \right\} \quad (3)$$

Herein  $L(y_i)$  denotes the losses of county  $i$  in the state, which depends on the yield  $y_i$  in each county. And  $1 - \alpha$  is the confidence level, the  $\alpha = 0.1$  at here, it means that 90% of losses are less than the VaR value.

We also consider Expected Shortfall, which measures the average losses over the VaR value. ES takes the average of all the losses greater than the VaR value. Since it is an expectation calculated by integrating over an entire region, it is a much more robust statistic compared to VaR, which is just a single value.

### 3.2 Simulation of Aggregate Losses Using Copulas

We use the copula-based models to measure the dependence between county-level yields for aggregate losses quantification. Due to the “curse of dimension” associated with an excessive number of variables, a yield portfolio including three counties from each of the nine states is chosen for testing for simplicity and practical reasons. The data sets, which cover the yields of three counties in nine states, are summarized in **Table 2**. We can find that county Allamakee in Group 1 and county Clay in Group 2 has obvious leptokurtic distribution, and most of the counties have negative skewness. **Figure 1** shows the geographical location of the selected counties in each state. The counties in AR, NJ, and

MD are adjacent to each other, and the state AL, FL, PA has the most scattered counties portfolio.

Copulas are very flexible with regard to representing the dependent structure of variables. The copula function’s fundamental concept is to link the marginal distribution to the joint distribution (Sklar, 1959), avoiding direct estimation of the multivariate distribution function and fitting the multivariate distribution function using parametric or non-parametric methods to estimate the marginal distributions of the variables independently. Copulas have grown in popularity in recent years and have been used to a variety of financial difficulties (Embrechts et al., 1999; Cherubini et al., 2004). They have also been used extensively in agricultural economics, mostly to model spatial dependence in yields or the dependence of random variables in crop revenue insurance, margin insurance, and whole-farm insurance (Vedenov and Power, 2008; Zhu et al., 2008; Bozic et al., 2014; Bulut and J. Collins, 2014; Ahmed and Serra, 2015; Goodwin and Hungerford, 2015; Feng and Hayes, 2016).

Let  $F$  be a joint distribution function with univariate marginal distribution functions  $F_1, \dots, F_d$ . Then there exists a copula function  $C: [0, 1]^d \rightarrow [0, 1]$  such that

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)), \quad (4)$$

where  $x_1, \dots, x_d \in \mathbb{R}$  are random variables. By inversion of the joint distribution in **Eq. 4**, the copula function can be written as

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)), \quad (5)$$

where  $F_1^{-1}, \dots, F_d^{-1}$  are 1-dimensional quantile functions and  $u_1, \dots, u_d \in [0, 1]$ . The copula function is parameterized by a vector  $\rho$  consisting of dependence parameters. Given that the copula is itself a joint distribution function, it satisfies all of the criteria for a joint distribution function.

The copula function’s fundamental properties are symmetry and tail dependency (Li and Genton, 2013). Insurance company and financial experts are often interested in the tail characteristic, since they represent the degree to which a portfolio’s losses occur in the worst-case scenario (Nelsen, 1999). Tail dependency is an asymptotic concept, and an asymmetric copula function distribution might have either lower or upper dependence, depending on the empirical phenomenon being described. The coefficient of upper-tail dependence is defined as follows:

$$\lambda_U = \lim_{u \rightarrow 1} \Pr(U_1 \geq u | U_2 \geq u) = \lim_{u \rightarrow 1} \frac{1 - 2u + C(u, u)}{1 - u}, \quad (6)$$

While the lower-tail dependence coefficient is defined as:

$$\lambda_L = \lim_{u \rightarrow 0} \Pr(U_1 \leq u | U_2 \leq u) = \lim_{u \rightarrow 0} \frac{C(u, u)}{u}, \quad (7)$$

Copulas exhibit tail dependence if the respective coefficients in **Eqs 6, 7** are nonzero, and when the tail dependence coefficient is close to 1, large values of the variables occur together.

The copula family chosen is inherently a decision of theoretical tail dependence. The parametric estimating procedure in an empirical investigation involves an *a priori* assumption about the data sets’ relationship. The five most often used parametric

**TABLE 2** | Summary statistics of average county-level yield (bushel per acre) of corn, by counties.

| Group   | State | Observations |       | Summary statistics of crop yield |        |        |         |                 |          |          |
|---------|-------|--------------|-------|----------------------------------|--------|--------|---------|-----------------|----------|----------|
|         |       | County       | Years | Min.                             | Median | Max.   | Std.Dev | CV <sup>3</sup> | Skewness | Kurtosis |
| Group 1 | IL    | Adams        | 30    | 85.72                            | 151.91 | 199.96 | 24.3110 | 0.1620          | -0.5094  | 3.1834   |
|         |       | Alexander    | 30    | 93.68                            | 140.60 | 178.00 | 22.4498 | 0.1631          | -0.3542  | 2.4093   |
|         |       | Bond         | 30    | 69.72                            | 133.41 | 173.96 | 23.3853 | 0.1780          | -0.5013  | 2.9523   |
|         | IN    | Adams        | 30    | 102.0                            | 151.9  | 171.1  | 17.1990 | 0.1164          | -0.8245  | 2.9605   |
|         |       | Allen        | 30    | 119.3                            | 156.8  | 170.3  | 13.0258 | 0.0859          | -0.8696  | 2.7811   |
|         |       | Bartholomew  | 30    | 96.42                            | 149.68 | 178.92 | 20.5502 | 0.1387          | -0.6109  | 2.5246   |
|         | IA    | Adair        | 30    | 99.97                            | 152.71 | 187.06 | 20.0485 | 0.1350          | -0.6756  | 2.9488   |
|         |       | Adams        | 30    | 99.82                            | 145.33 | 174.26 | 20.9762 | 0.1476          | -0.6395  | 2.2976   |
|         |       | Allamakee    | 30    | 97.54                            | 159.71 | 180.71 | 17.7399 | 0.1132          | -1.8518  | 6.5536   |
| Group 2 | AL    | Autauga      | 30    | 40.00                            | 80.12  | 126.62 | 21.5820 | 0.2740          | 0.0680   | 2.3087   |
|         |       | Cherokee     | 30    | 35.00                            | 92.09  | 147.96 | 28.0576 | 0.2896          | 0.0100   | 2.1373   |
|         |       | Colbert      | 30    | 68.38                            | 104.17 | 167.62 | 26.6857 | 0.2444          | 0.4541   | 2.0615   |
|         | AR    | Clay         | 30    | 73.68                            | 154.01 | 179.02 | 21.5321 | 0.1439          | -1.4126  | 5.8428   |
|         |       | Craighead    | 30    | 68.68                            | 140.55 | 184.00 | 26.0007 | 0.1897          | -0.4766  | 2.8372   |
|         |       | Greene       | 30    | 68.68                            | 138.50 | 172.00 | 20.3717 | 0.1522          | -1.0445  | 4.5724   |
|         | FL    | Escambia     | 30    | 64.66                            | 117.38 | 149.70 | 22.1617 | 0.1928          | -0.4251  | 2.3057   |
|         |       | Gadsden      | 30    | 64.48                            | 96.83  | 133.06 | 18.1695 | 0.1830          | 0.2534   | 2.2667   |
|         |       | Hamilton     | 30    | 73.03                            | 99.24  | 135.00 | 14.2299 | 0.1394          | 0.4311   | 3.1189   |
| Group 3 | PA    | Adams        | 30    | 63.48                            | 120.13 | 188.96 | 27.5589 | 0.2415          | 0.0687   | 3.1202   |
|         |       | Armstrong    | 30    | 76.44                            | 122.91 | 155.06 | 22.9851 | 0.1890          | -0.3821  | 2.1000   |
|         |       | Beaver       | 30    | 73.43                            | 124.48 | 150.75 | 21.3783 | 0.1761          | -0.6073  | 2.3726   |
|         | NJ    | Burlington   | 30    | 61.54                            | 134.41 | 149.82 | 21.5385 | 0.1705          | -1.1812  | 3.9779   |
|         |       | Cumberland   | 30    | 57.27                            | 127.79 | 160.06 | 25.4670 | 0.2037          | -0.7255  | 3.0111   |
|         |       | Gloucester   | 30    | 50.27                            | 126.22 | 153.06 | 26.7836 | 0.2221          | -0.8964  | 3.1647   |
|         | MD    | Anne arundel | 30    | 86.77                            | 129.01 | 169.28 | 22.9579 | 0.1837          | -0.0216  | 1.7190   |
|         |       | Baltimore    | 30    | 86.15                            | 137.34 | 171.18 | 20.5206 | 0.1543          | -0.5514  | 2.5109   |
|         |       | Calvert      | 30    | 52.07                            | 116.12 | 167.98 | 26.4296 | 0.2282          | -0.3819  | 2.7586   |

<sup>3</sup>CV is the Coefficient of Variance.

copulas (Gaussian, t, Gumbel, Clayton, and Frank) and their various rotations may capture a wide range of bivariate dependent behavior. However, for aggregate losses of coverage yields greater than two counties, the archimedean copula (Gumbel, Clayton, and Frank) cannot accurately simulate multivariate dependence due to its single parameter. In addition to elliptical copulas, the vine and hierarchical copulas can also capture multivariate dependence.

By combining pair-wise copulas, the vine copula enables much more flexibility in multivariate modeling (Bedford and Cooke, 2002; Kurowicka and Cooke, 2006). According to Aas et al. (2009), the factored form of a joint multivariate density function for a set of  $k$  random variables is:

$$f(x_1, x_2, \dots, x_k) = f_k(x_k) \cdot f(x_{k-1}|x_k) \cdot f(x_{k-2}|x_{k-1}, x_k) \dots \cdot f(x_1|x_2, \dots, x_k). \quad (8)$$

This density is unique for a given ordering of variables. The joint density can also be expressed in terms of a copula function, as:

$$f(x_1, x_2, \dots, x_k) = c_{1..k}(F_1(x_1), \dots, F_k(x_k)) \cdot \prod_{i=1}^k f_i(x_i). \quad (9)$$

Joe (1996) shown that each term in Eq. 9 may be decomposed into the product of a pair-wise copula and a conditional marginal density:

$$f(x|v) = c_{x,v_k|v_{-k}}(F(x|v_{-k}), F(v_k|v_{-k})) \cdot f(x|v_{-k}). \quad (10)$$

Aas et al. (2009) described how to represent a multivariate density as a product of pair-wise copulas. The vine copula functions are best described as a set of “trees,” with each variable’s distribution represented by conditional distributions at a higher tree level. Bedford and Cooke (2002) presented a “regular vine” representation that enables considerable flexibility in representing multivariate densities in terms of pair-wise copula combinations. Kurowicka and Cooke (2006) identified two distinct types of vine copulas, dubbed the “C-vine” and the “D-vine.” When variables have a particular ordering, a D-vine is appropriate, whereas a C-vine is good when variables may be arranged according to their effect on other variables (Aas et al., 2009).

Following Goodwin and Hungerford (2015), we establish county-level yield dependence models using elliptical copulas (Gaussian, t) and vine copulas. To begin, we estimated the corresponding Gaussian and t-copula from detrended county-level crop yield data using the empirical marginal distribution approach. Goodwin and Hungerford (2015) point out that this approach results in less variability than estimation methods based on parametric marginal distributions. We need to estimate three separate coefficients for the Gaussian and t-copula dependence models since each state has three random variables representing county-level yield. Additionally, the t-copula must estimate the

**TABLE 3** | Elliptical (Normal and t) copula estimation results for the nine datasets.

| Parameter   | IL       |          | IN       |          | IA       |          |
|-------------|----------|----------|----------|----------|----------|----------|
|             | Normal   | t        | Normal   | t        | Normal   | t        |
| $\rho_{12}$ | 0.6862   | 0.6162   | 0.8228   | 0.7254   | 0.8937   | 0.8949   |
| $\rho_{13}$ | 0.7757   | 0.8209   | 0.8706   | 0.8223   | 0.7301   | 0.7263   |
| $\rho_{23}$ | 0.5843   | 0.5818   | 0.7710   | 0.6494   | 0.7637   | 0.7596   |
| df          | NA       | 2.7526   | NA       | 1.6125   | NA       | 33.8831  |
| AIC         | -33.0568 | -36.4327 | -62.0138 | -61.3223 | -59.8349 | -57.8301 |
|             | AL       |          | AR       |          | FL       |          |
| Parameter   | Normal   | t        | Normal   | t        | Normal   | t        |
| $\rho_{12}$ | 0.7816   | 0.7869   | 0.9011   | 0.9003   | 0.7020   | 0.6996   |
| $\rho_{13}$ | 0.5671   | 0.6060   | 0.8705   | 0.8692   | -0.4843  | -0.4819  |
| $\rho_{23}$ | 0.7963   | 0.7914   | 0.9042   | 0.9035   | -0.4731  | -0.4708  |
| df          | NA       | 5.2821   | NA       | 37.0054  | NA       | 138.7882 |
| AIC         | -44.8526 | -44.0452 | -87.7323 | -85.5402 | -17.3834 | -15.2785 |
|             | PA       |          | NJ       |          | MD       |          |
| Parameter   | Normal   | t        | Normal   | t        | Normal   | t        |
| $\rho_{12}$ | 0.5176   | 0.5187   | 0.8104   | 0.8014   | 0.7328   | 0.7751   |
| $\rho_{13}$ | 0.4244   | 0.4265   | 0.8312   | 0.8211   | 0.8926   | 0.9069   |
| $\rho_{23}$ | 0.8463   | 0.8455   | 0.9211   | 0.9280   | 0.7510   | 0.7582   |
| df          | NA       | 94.9907  | NA       | 4.5455   | NA       | 3.9828   |
| AIC         | -34.2009 | -32.1208 | -77.2285 | -77.7075 | -58.7623 | -62.0529 |

Note: where 1, 2 and 3 indicate the three selected county, respectively. The estimation method to be used to estimate the dependence parameter is the inversion of Kendall's tau estimator, and df is the degree of freedom.

degree of freedom parameter, which converges gradually to the Gaussian copula as the degree of freedom grows (the typical cut-off degree of freedom is 30).

The Gaussian and t-copula maximum likelihood estimation results for the yield data sets are reported in **Table 3**. The first conclusion is that in nine states, the dependencies between county-level yields are all-powerful, and the data sets from IL, IN in Group 1, AL in Group 2, NJ, MD in Group 3 exhibit rather substantial tail dependence. However, the fitted t-copula has a degree of freedom of more than 30 for several states, such as IA in Group 1, AR and FL in Group 2, PA in Group 3, indicating that the tail dependence of these data sets is minor, with essentially no difference from Gaussian copula. Moreover, the county-level yields negatively depend on each other for FL in Group 3. As a result, we cannot determine which copula between Gaussian and t is superior based on the AIC values of Gaussian and t-copulas since the better-fitting copula varies between data sets.

We then estimate the R-vine copulas using a sequential maximum likelihood procedure. Vine copulas are formed of Gaussian, t, Clayton, Gumbel, and Frank copula and their rotating type. Maximum spanning trees are used to pick R-vine trees with respect to certain edge weights. The appropriate pair-wise copula families are selected according to the Akaike Information Criteria (using maximum likelihood estimation) and estimated sequentially (forward selection of trees). The estimation results and AIC value statistics are presented in **Table 4**, and when compared to other vine copula forms (R-vine and D-vine), this approach selects the C-vine copula as the dependence structure for all data sets.

Additionally, we perform a Goodness-of-fit (GOF) test on the fitted results of each portfolio's various copula functions, and the results are provided in **Table 5**. The  $p$ -values of the GOF test for the fitted copula function are more than 0.05 for all analyzed risk portfolios, indicating that the Gaussian and t-copula assumptions are not rejected. **Table 4** also compares the AIC values of several copula forms, and we see that the C-vine copula's AIC values are much fewer than those of the Gaussian and t-copula for each portfolio, indicating that the C-vine copula is favored over the Gaussian and t-copula.

Then, we need to fit the yield marginal distribution to simulate crop yield losses based on the determined copula-based model. Compared to using a non-parametric marginal distribution to fit the yield distribution, we choose a parametric marginal distribution since it can give an explicit explanation. We estimate the parametric distribution individually for each county yield and then combine it with the copula-based copula model to generate random variables. Numerous parametric distributions have been used to model yield distributions in crop insurance, and we fit the yield using a two-component mixture normal and a Weibull. The distribution hypothesis test and AIC values for the county yield data sets are provided in **Table 6**. As predicted, the Weibull and mixture normal distribution hypotheses are both accepted.

After all, the aggregate losses are fitted based on the estimated copula function and marginal distribution using the approaches outlined above. We emphasize the likelihood of significant indemnity resulting from systemic risk in county-level risk portfolios. Notably, we do not adjust the portfolio for

**TABLE 4 |** Vine copula estimation results of data sets in each state.

|    | Factorization                                   | Copula family        | Parameter 1 | Parameter 2 |
|----|-------------------------------------------------|----------------------|-------------|-------------|
| IL | C <sub>1</sub> , C <sub>2</sub>                 | Survival Gumbel      | 1.83        | NA          |
|    | C <sub>3</sub> , C <sub>1</sub>                 | t                    | 0.81        | 2.00        |
|    | C <sub>3</sub> , C <sub>2</sub>  C <sub>1</sub> | Clayton              | 0.40        | NA          |
|    | C-Vine AIC                                      | -39.62               |             |             |
| IN | C <sub>1</sub> , C <sub>2</sub>                 | Survival Gumbel      | 2.61        | NA          |
|    | C <sub>3</sub> , C <sub>1</sub>                 | Gumbel               | 3.15        | NA          |
|    | C <sub>3</sub> , C <sub>2</sub>  C <sub>1</sub> | Normal               | 0.22        | NA          |
|    | C-Vine AIC                                      | -67.91               |             |             |
| IA | C <sub>2</sub> , C <sub>1</sub>                 | Survival Gumbel      | 3.33        | NA          |
|    | C <sub>3</sub> , C <sub>2</sub>                 | Survival Gumbel      | 2.18        | NA          |
|    | C <sub>3</sub> , C <sub>1</sub>  C <sub>2</sub> | Survival Gumbel      | 0.25        | NA          |
|    | C-Vine AIC                                      | -62.34               |             |             |
| AL | C <sub>2</sub> , C <sub>1</sub>                 | Survival Gumbel      | 2.36        | NA          |
|    | C <sub>3</sub> , C <sub>2</sub>                 | Gumbel               | 2.43        | NA          |
|    | C <sub>3</sub> , C <sub>1</sub>  C <sub>2</sub> | Normal               | -0.06       | NA          |
|    | C-Vine AIC                                      | -48.71               |             |             |
| AR | C <sub>2</sub> , C <sub>1</sub>                 | Normal               | 0.90        | NA          |
|    | C <sub>3</sub> , C <sub>2</sub>                 | Normal               | 0.90        | NA          |
|    | C <sub>3</sub> , C <sub>1</sub>  C <sub>2</sub> | Clayton              | 0.53        | NA          |
|    | C-Vine AIC                                      | -88.4839             |             |             |
| FL | C <sub>1</sub> , C <sub>2</sub>                 | Normal               | 0.70        | NA          |
|    | C <sub>3</sub> , C <sub>1</sub>                 | Normal               | -0.48       | NA          |
|    | C <sub>3</sub> , C <sub>2</sub>  C <sub>1</sub> | Rotated Gumbel (90°) | -1.14       | NA          |
|    | C-Vine AIC                                      | -17.4168             |             |             |
| PA | C <sub>2</sub> , C <sub>1</sub>                 | Frank                | 3.69        | NA          |
|    | C <sub>3</sub> , C <sub>2</sub>                 | Frank                | 9.37        | NA          |
|    | C <sub>3</sub> , C <sub>1</sub>  C <sub>2</sub> | Normal               | -0.07       | NA          |
|    | C-Vine AIC                                      | -36.6100             |             |             |
| NJ | C <sub>3</sub> , C <sub>1</sub>                 | Survival Gumbel      | 2.63        | NA          |
|    | C <sub>3</sub> , C <sub>2</sub>                 | Survival Gumbel      | 4.11        | NA          |
|    | C <sub>2</sub> , C <sub>1</sub>  C <sub>3</sub> | Normal               | 0.25        | NA          |
|    | C-Vine AIC                                      | -82.7749             |             |             |
| MD | C <sub>3</sub> , C <sub>1</sub>                 | Survival Clayton     | 5.12        | NA          |
|    | C <sub>3</sub> , C <sub>2</sub>                 | Survival Clayton     | 2.25        | NA          |
|    | C <sub>2</sub> , C <sub>1</sub>  C <sub>3</sub> | Gumbel               | 1.33        | NA          |
|    | C-Vine AIC                                      | -78.7577             |             |             |

Note: where 1, 2 and 3 indicate whether the yields of three counties in each state, respectively. Using the inversion of Kendall's tau estimator to estimate the dependence parameter.

variations in exposure (i.e., various amounts of acreage) between counties and hence assume that all counties have the same level of corn acreage, then we can adjust the total losses by aggregating the average county-level yield losses per acre for each portfolio. Additional loading factors that account for administrative costs are disregarded.

### 3.3 The Efficacy Tests of Copula-Based Models

The AIC criteria and *p*-value can only be used to determine the fitted copula's goodness-of-fit (Goodwin and Hungerford, 2015). To assess the efficacy of the copula-based model on the accuracy of losses

**TABLE 5 |** GOF test results of the data sets in different states.

| State | Gaussian        |          | t               |           | C-vine   |
|-------|-----------------|----------|-----------------|-----------|----------|
|       | <i>p</i> -value | AIC      | <i>p</i> -value | AIC       | AIC      |
| IL    | 0.6628          | -33.0568 | 0.8257          | -36.4327  | -39.6190 |
| IN    | 0.3641          | -62.0138 | 0.1444          | -61.3223  | -67.9088 |
| IA    | 0.3032          | -59.8349 | 0.2692          | -57.8301  | -62.3418 |
| AL    | 0.3601          | -44.8526 | 0.6089          | -44.0452  | -48.7106 |
| AR    | 0.3521          | -87.7323 | 0.3212          | -85.5402  | -88.4839 |
| FL    | 0.3561          | -17.3834 | 0.3611          | -15.2785  | -17.4168 |
| PA    | 0.1164          | -34.2009 | 0.0894          | -32.1208  | -36.6100 |
| NJ    | 0.9286          | -77.2285 | 0.9356          | -77.7075  | -82.7749 |
| MD    | 0.1703          | -58.7623 | 0.5420          | -62.05290 | -78.7577 |

assessment and insurance pricing, we design a simulation experiment using the Mean Squared Error (MSE), which quantifies the difference between the copula-based model's estimated losses and the "true" losses as defined by Harri et al. (2011). Ker et al. (2016), Yvette Zhang (2017), and Yi et al. (2020) have all employed the MSE measure in the context of agricultural commodity hedging.

We use nonparametric kernel techniques to estimate county-level yield density functions for each portfolio in the 9 states, highlighting that these densities account for county-level yield dependence. Then, assuming that these densities accurately represent the "true" density function, we can generate dependent data samples from these densities. To create data sets based on these "true" densities, an aggregate of 500 samples of sizes *n* = 30 and *n* = 50 is drawn from the "true" distributions for each county-level data set. The mean squared error (MSE) in calculating aggregate losses compared to "true" aggregate losses is used to evaluate the effectiveness of pricing. Premiums are derived for each sample using a Monte Carlo approach based on the individual model with the non-dependence assumption and an estimated copula-based model (Gaussian, t, and C-vine copula). The Weibull and mixture normal distributions obtained in Section 3 are used to fit the yield distributions of the individual and copula-based models. The following equation is used to obtain their appropriate MSE:

$$MSE(L) = \frac{1}{500} \sum_{i=1}^{500} [\hat{\pi}_i - \pi(L)]^2 \quad (11)$$

Where  $\pi$  is the "true" losses, and  $\hat{\pi}_i$  is the corresponding estimate in the *i*th experiment. The MSE is calculated by comparing the estimated aggregate losses from the various models to the "true" aggregate losses in each state. The primary role of this simulation experiment is to determine if the copula-based model has an effect on losses measurement by comparing the estimated losses for these two types of models (individual and copula-based) to the "true" losses. We ensure consistency with this large sample size. It makes our estimate consistent and produces the correct result on average.

## 4 EMPIRICAL RESULTS AND DISCUSSION

Using the procedure outlined in Section 3, we collect yield data sets from three counties in each selected states for three groups. We



**TABLE 6 |** Results of Kolmogorov-Smirnov test (at 5% significance level) and the AIC value of the goodness-of-fit test.

| Group 1 | IL        |          |            | IN          |          |          | IA           |          |          |
|---------|-----------|----------|------------|-------------|----------|----------|--------------|----------|----------|
|         | County    | Weibull  | Mixture    | County      | Weibull  | Mixture  | County       | Weibull  | Mixture  |
| p-value | Adams     | 0.9921   | 0.9706     | Adams       | 0.9433   | 0.9674   | Adair        | 0.4624   | 0.3968   |
|         | Alexander | 0.9238   | 0.9647     | Allen       | 0.4586   | 0.9940   | Adams        | 0.7824   | 0.8879   |
|         | Bond      | 0.9990   | 0.9995     | Bartholomew | 0.8015   | 0.6246   | Allamakee    | 0.7296   | 0.2401   |
| AIC     | Adams     | 278.2256 | 275.1835   | Adams       | 253.9654 | 250.2059 | Adair        | 265.0451 | 260.4891 |
|         | Alexander | 273.6248 | 271.2883   | Allen       | 236.0602 | 229.0293 | Adams        | 267.0493 | 253.3133 |
|         | Bond      | 275.6055 | 271.5169   | Bartholomew | 266.2210 | 257.9925 | Allamakee    | 249.0293 | 244.8312 |
| Group 2 | AL        |          |            | AR          |          |          | FL           |          |          |
| p-value | County    | Weibull  | Mixture    | County      | Weibull  | Mixture  | County       | Weibull  | Mixture  |
|         | Autauga   | 0.9966   | 0.9955     | Clay        | 0.8303   | 0.3127   | Escambia     | 0.9027   | 0.9524   |
|         | Cherokee  | 0.8343   | 0.9308     | Craighead   | 0.9750   | 0.7012   | Gadsden      | 0.8805   | 0.9964   |
| AIC     | Colbert   | 0.7330   | 0.9995     | Greene      | 0.8286   | 0.9469   | Hamilton     | 0.4201   | 0.4335   |
|         | Autauga   | 272.2961 | 268.1752   | Clay        | 265.9092 | 262.5262 | Escambia     | 272.2098 | 269.4284 |
|         | Cherokee  | 287.8535 | 285.2474   | Craighead   | 282.0591 | 272.9506 | Gadsden      | 263.8013 | 258.8797 |
| Colbert | 286.2712  | 277.1186 | Greene     | 264.9158    | 257.7720 | Hamilton | 251.5291     | 237.1732 |          |
| Group 3 | PA        |          |            | NJ          |          |          | MD           |          |          |
| p-value | County    | Weibull  | Mixture    | County      | Weibull  | Mixture  | County       | Weibull  | Mixture  |
|         | Adams     | 0.4246   | 0.4854     | Burlington  | 0.4640   | 0.8507   | Anne arundel | 0.6400   | 0.9909   |
|         | Armstrong | 0.9818   | 0.9195     | Cumberland  | 0.5786   | 0.8057   | Baltimore    | 0.7759   | 0.9861   |
| AIC     | Beaver    | 0.8211   | 0.9900     | Gloucester  | 0.7634   | 0.4254   | Calvert      | 0.9977   | 0.9616   |
|         | Adams     | 287.7737 | 281.2121   | Burlington  | 266.1494 | 255.0774 | Anne arundel | 276.0907 | 264.7787 |
|         | Armstrong | 274.4310 | 269.1059   | Cumberland  | 279.5334 | 276.3238 | Baltimore    | 266.9404 | 260.6883 |
| Beaver  | 268.7325  | 262.4391 | Gloucester | 281.9373    | 282.8471 | Calvert  | 283.5852     | 280.0709 |          |

estimate SRA policy losses using the random numbers generated by each fitted model. Then we examine if the tail dependency of yield risks affects the aggregate indemnity of crop insurance, such as joint losses due to severe weather occurrences. The asymmetric lower-tail dependency is described by the Rotated Gumbel (180°) and Clayton copulas, which are used to define pair copulas in vine copula. Additionally, t-copula exhibits symmetric tail dependency, while the Gaussian copula does not<sup>4</sup>.

The MSE of the aggregate losses between the “true” and the fitted losses (based on the copula and the individual model) is calculated, emphasizing comparing the differences between models with and without dependence. The losses distributions are calculated using 10,000 random samples generated from the estimated losses model. Notably, we use the Weibull and mixture normal as the marginal distributions, and the copula model’s dependence structure varies for each state’s yield data sets. The following are the findings of this experiment:

The mean value of aggregate yield losses is not sensitive to the individual or copula-based model in various risk regions and sample sizes, as shown in **Tables 7–9**. We can determine that the copula-based approach has a negligible effect on the unbiased

**TABLE 7 |** The average Mean Squared Error of aggregate losses for states in Group 1.

|            | n = 30  |         | n = 50  |         |
|------------|---------|---------|---------|---------|
|            | Weibull | Mixture | Weibull | Mixture |
| Individual | 19.2354 | 20.7218 | 13.8076 | 13.2086 |
| Gaussian   | 19.2320 | 20.6700 | 13.7932 | 13.3622 |
| t          | 19.3448 | 20.8181 | 13.8149 | 13.2851 |
| Vine       | 19.2873 | 20.7303 | 13.7933 | 13.2876 |

**TABLE 8 |** The average Mean Squared Error of aggregate losses for states in Group 2.

|            | n = 30  |         | n = 50  |         |
|------------|---------|---------|---------|---------|
|            | Weibull | Mixture | Weibull | Mixture |
| Individual | 9.1846  | 8.5403  | 6.9524  | 4.9648  |
| Gaussian   | 9.2053  | 8.5509  | 7.0002  | 4.9989  |
| t          | 9.2302  | 8.5746  | 6.9867  | 5.0089  |
| Vine       | 9.2216  | 8.5578  | 6.9315  | 4.9918  |

estimator of aggregate losses accuracy through stochastic simulations. The aggregate losses density distribution plots for the states in **Figures 2–4** help to explain this conclusion. As seen in **Figures 2–4**, the copula-based or individual model may be the most well-fitting model at different losses layers. Due to the fact that the fair premium for crop insurance is determined on the expectation of losses, there is no statistically significant difference between the copula-based and individual models.

<sup>4</sup>In the later application we take the following copula families into consideration (some properties are given in brackets): Gaussian (tail-symmetric, no tail dependence). Student-t (tail-symmetric, tail dependence). Gumbel (tail-asymmetric, upper tail dependence) and survival Gumbel (tail-asymmetric, lower tail dependence). Rotated Gumbel by 90° and 270° (tail-asymmetric, no tail dependence). Frank (tail-symmetric, no tail dependence). Clayton (tail-asymmetric, lower tail dependence).

**TABLE 9 |** The average Mean Squared Error of aggregate losses for states in Group 3.

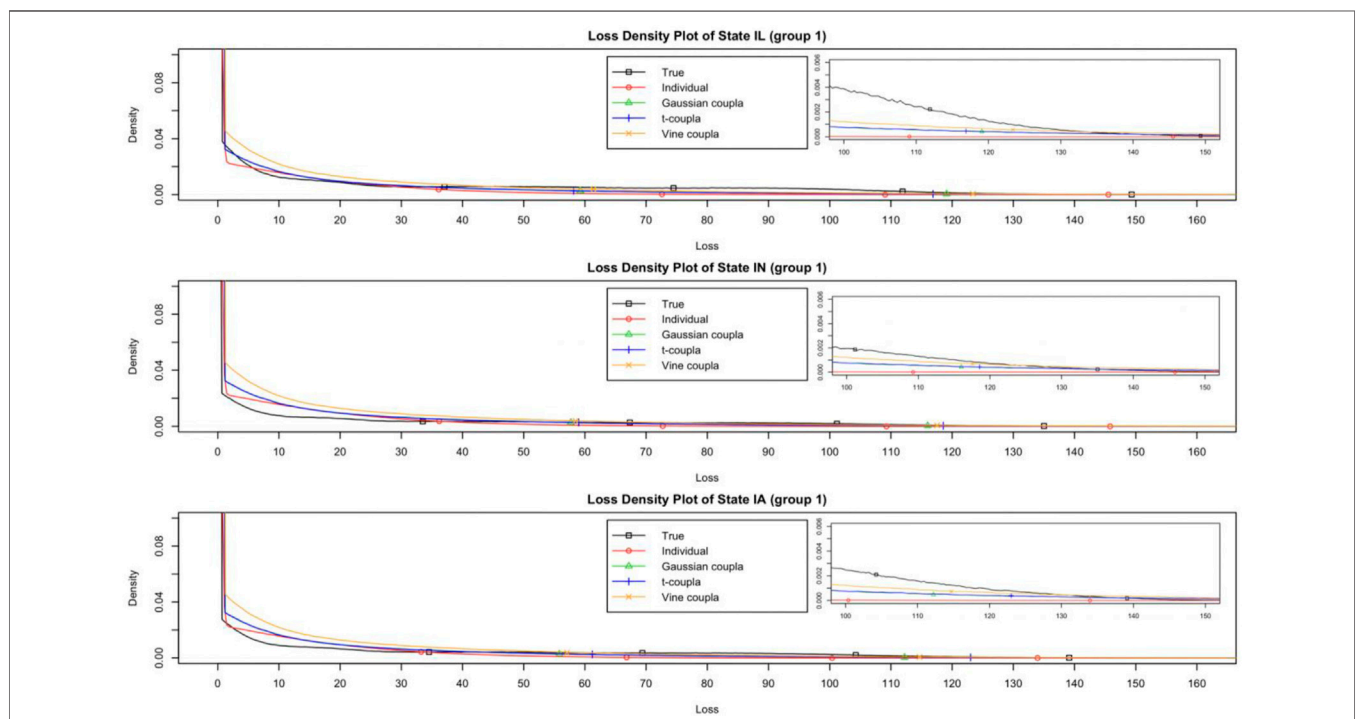
|            | n = 30  |         | n = 50  |         |
|------------|---------|---------|---------|---------|
|            | Weibull | Mixture | Weibull | Mixture |
| Individual | 26.0669 | 22.1998 | 15.7114 | 14.5431 |
| Gaussian   | 26.1958 | 22.2409 | 15.7976 | 14.6718 |
| t          | 26.2791 | 22.2665 | 15.8409 | 14.7108 |
| Vine       | 26.1050 | 22.3048 | 15.8187 | 14.6373 |

Nonetheless, as shown in **Tables 10–12**, all copula-based models estimated higher VaR and ES values than the individual model. In Groups 1 and 3, the copula-based model's VaR and ES values are much greater than those of the individual model. VaR is used to measure tail risk by determining the maximum possible losses value for a given probability, and ES takes the average of all the possible losses greater than the VaR value. By comparing the VaR and ES values of “true” distribution, we may infer that the copula-based model is the better-fitting model for VaR and ES values than the individual model in Groups 1 and 3. On the other hand, there is no discernible difference in VaR and ES values between the individual model and the copula-based model in Group 2, which may be due to no obvious tail dependence between the county-level yields in Group 2. It also can be explained in **Figure 3** that the individual model is the most well-fitting model around the tail of the losses for Group 2.

Second, consider the payment situations for various layers. The findings in **Tables 13–15** demonstrate that the individual model's losses MSE is greater than the copula-based model at various levels for all three groups, except for the MSE estimated

by using Weibull distribution for the layer of the losses ratio 2.20–5.00 in Group 2. The MSE results reveal that in Groups 1 and 3, the copula-based model may significantly reduce the MSE of losses compared to the individual model, indicating that dependence strongly influences this layer's payment. In Group 2, however, there is no substantial reduction in the MSE of losses between the copula-based model and the individual model since there is no evident tail dependence between the county-level yields (as shown in **Table 4**). Thus, when considerable yield reliance exists in the risk portfolio, the copula-based model beats the individual model for calculating SRA payments.

Additionally, the losses estimations vary across by using the various copula models. For example, in **Table 13**, the measured losses MSE of data sets with a duration of 30 years and Weibull distribution based on the Gaussian, t, and C-vine copula models for the layer of the losses ratio 2.20–5.00 are 7.7005, 7.6656, and 7.2505, respectively. The above findings demonstrate that the accuracy of the losses estimation for the layer of the losses ratio 2.20–5.00 changes based on the using of various alternative copula structures, and the C-vine copula surpasses the Gaussian and t-copula in Group 1, and the reason may be that the IL, IN and IA states' dependence structure has the lower-tail characteristic (survival Gumbel has the upper-tail properties). In comparison, in **Table 14**, the measured losses MSE with a 30-year duration and Weibull distribution based on Gaussian, t, and C-vine copula models for the layer of 2.20–5.00 is 3.4782, 3.4103, and 3.9801, respectively. While the losses MSE for the layer of 1.00–1.60 is 4.8743, 4.9417, and 4.3894, respectively. Thus, adopting the C-vine copula model would result in a more accurate losses measurement at the layer of the losses ratio



**FIGURE 2 |** The losses density plot for states in Group 1 (Weibull distribution with a duration of 30 years).

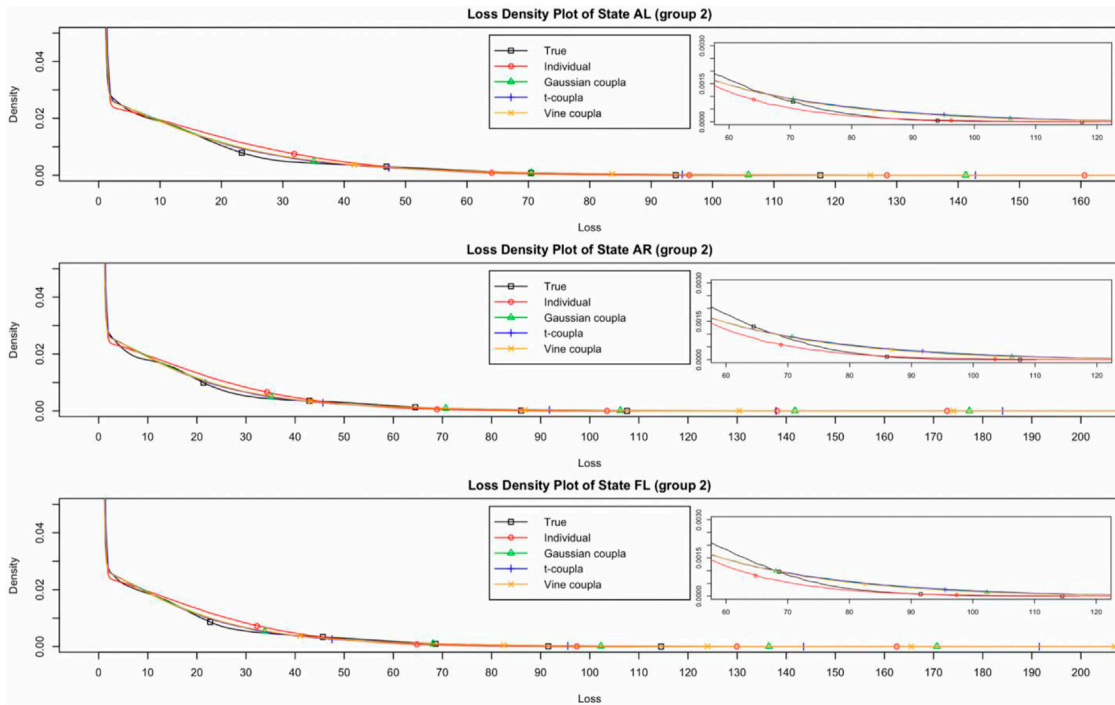


FIGURE 3 | The losses density plot for states in Group 2 (Weibull distribution with a duration of 30 years).

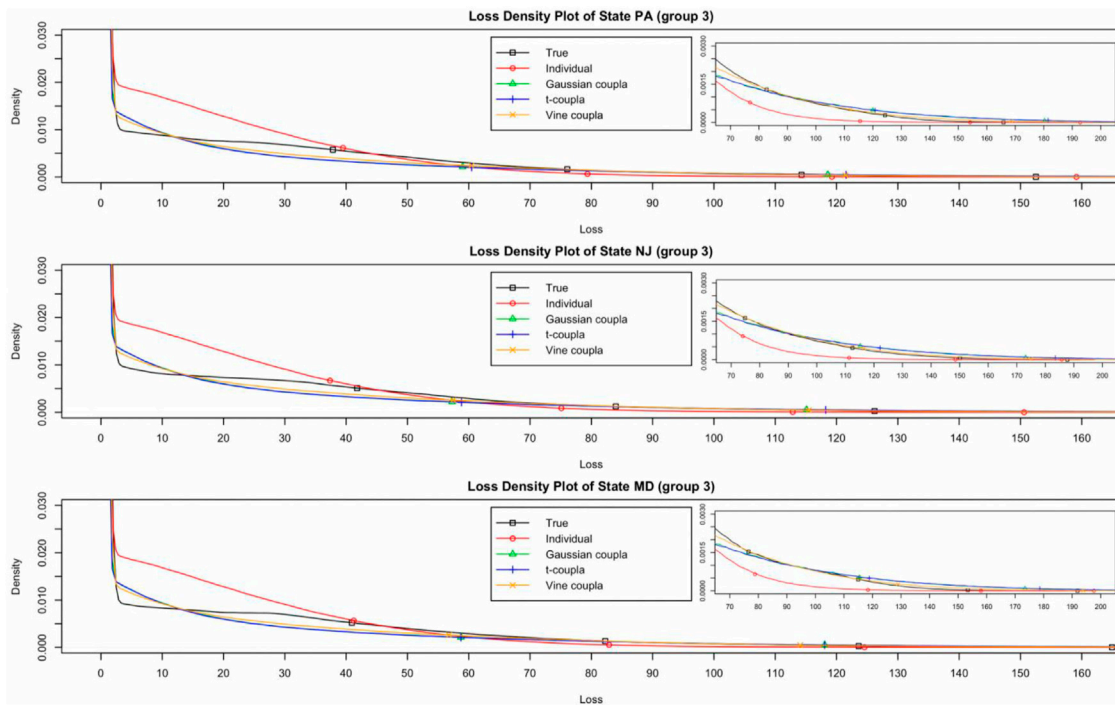


FIGURE 4 | The losses density plot for states in Group 3 (Weibull distribution with a duration of 30 years).

**TABLE 10 |** The average VaR and ES of aggregate losses for states in Group 1.

|     |                     | n = 30  |         | n = 50  |         |
|-----|---------------------|---------|---------|---------|---------|
|     |                     | Weibull | Mixture | Weibull | Mixture |
| VaR | "True" distribution | 46.9562 |         | 48.2575 |         |
|     | Individual          | 26.5994 | 31.6233 | 26.9795 | 32.9377 |
|     | Gaussian            | 28.9666 | 39.1346 | 29.4282 | 39.3472 |
|     | t                   | 28.7839 | 38.9980 | 29.2539 | 39.1394 |
|     | C-vine              | 28.9486 | 39.3872 | 29.2249 | 38.6265 |
| ES  | "True" distribution | 79.1993 |         | 79.7089 |         |
|     | Individual          | 40.0049 | 43.5855 | 40.4974 | 45.6066 |
|     | Gaussian            | 59.0655 | 66.3426 | 59.7508 | 69.1085 |
|     | t                   | 59.2127 | 66.5895 | 59.8756 | 69.3212 |
|     | C-vine              | 60.4425 | 68.6596 | 61.3516 | 71.7302 |

**TABLE 11 |** The average VaR and ES of aggregate losses for states in Group 2.

|     |                     | n = 30  |         | n = 50  |         |
|-----|---------------------|---------|---------|---------|---------|
|     |                     | Weibull | Mixture | Weibull | Mixture |
| VaR | "True" distribution | 34.3501 |         | 33.7988 |         |
|     | Individual          | 33.9879 | 28.4978 | 34.4625 | 28.9639 |
|     | Gaussian            | 34.9967 | 30.7408 | 35.3239 | 30.9064 |
|     | t                   | 35.128  | 30.7782 | 35.3823 | 30.958  |
|     | C-vine              | 35.1499 | 30.9079 | 35.5104 | 30.7988 |
| ES  | "True" distribution | 49.3004 |         | 49.1049 |         |
|     | Individual          | 47.1956 | 39.0047 | 47.767  | 39.759  |
|     | Gaussian            | 53.8759 | 45.4534 | 54.1538 | 46.2313 |
|     | t                   | 54.7911 | 45.9685 | 54.7368 | 46.5877 |
|     | C-vine              | 53.717  | 45.3896 | 54.1409 | 46.1026 |

**TABLE 12 |** The average VaR and ES of aggregate losses for states in Group 3.

|     |                     | n = 30  |         | n = 50  |         |
|-----|---------------------|---------|---------|---------|---------|
|     |                     | Weibull | Mixture | Weibull | Mixture |
| VaR | "True" distribution | 52.4282 |         | 52.6927 |         |
|     | Individual          | 41.0873 | 40.4925 | 41.8217 | 40.9976 |
|     | Gaussian            | 51.5542 | 54.1355 | 52.4565 | 54.5344 |
|     | t                   | 51.3275 | 54.0712 | 52.3422 | 54.5641 |
|     | C-vine              | 51.4176 | 53.2053 | 52.0784 | 53.4731 |
| ES  | "True" distribution | 76.3729 |         | 76.694  |         |
|     | Individual          | 56.6024 | 54.5455 | 57.5033 | 55.4088 |
|     | Gaussian            | 85.9036 | 81.1081 | 86.9555 | 82.5066 |
|     | t                   | 86.5183 | 81.6033 | 87.7802 | 83.1246 |
|     | C-vine              | 79.8021 | 76.1397 | 80.1034 | 76.956  |

between 1.00 and 1.60, and the reason may be that the MD state's dependence structure has the upper-tail characteristic (survival Clayton has the upper-tail properties). From the empirical results of those two groups, we may conclude that if the portfolio contains a substantial tail risk, the C-vine copula model, which provides a more flexible description of yield dependence, is better suited to measure this layer of losses.

Finally, since systemic risks associated with weather will result in concurrent agricultural production losses throughout a large geographic area, for example, the 1988 and 2012 droughts and the

**TABLE 13 |** The average Mean Squared Error of aggregate losses at different layers for states in Group 1.

|                                   | n = 30  |         | n = 50  |         |
|-----------------------------------|---------|---------|---------|---------|
|                                   | Weibull | Mixture | Weibull | Mixture |
| Losses Ratio Between 100 and 160% |         |         |         |         |
| Individual                        | 1.6129  | 4.0302  | 1.0581  | 3.0915  |
| Gaussian                          | 0.9985  | 1.2076  | 0.6347  | 0.7103  |
| t                                 | 1.0066  | 1.2113  | 0.6465  | 0.7146  |
| C-vine                            | 1.0370  | 1.1618  | 0.6922  | 0.6980  |
| Losses Ratio Between 160 and 220% |         |         |         |         |
| Individual                        | 3.2074  | 4.1928  | 2.3939  | 2.5521  |
| Gaussian                          | 2.0057  | 1.9788  | 1.4758  | 1.1166  |
| t                                 | 2.0144  | 1.9765  | 1.4949  | 1.1218  |
| C-vine                            | 1.9763  | 1.9117  | 1.4836  | 1.1275  |
| Losses Ratio Between 220 and 500% |         |         |         |         |
| Individual                        | 14.3698 | 11.9700 | 13.4862 | 9.4323  |
| Gaussian                          | 7.7005  | 6.8211  | 6.5387  | 4.3115  |
| t                                 | 7.6656  | 6.7546  | 6.5470  | 4.2838  |
| C-vine                            | 7.2505  | 6.4902  | 6.1143  | 4.0955  |

**TABLE 14 |** The average Mean Squared Error of aggregate losses at different layers for states in Group 2.

|                                   | n = 30  |         | n = 50  |         |
|-----------------------------------|---------|---------|---------|---------|
|                                   | Weibull | Mixture | Weibull | Mixture |
| Losses Ratio Between 100 and 160% |         |         |         |         |
| Individual                        | 1.4352  | 1.5407  | 1.1887  | 0.9001  |
| Gaussian                          | 0.9680  | 1.1981  | 0.6955  | 0.7269  |
| t                                 | 0.9183  | 1.1930  | 0.6697  | 0.7305  |
| C-vine                            | 1.0249  | 1.2179  | 0.7280  | 0.7339  |
| Losses Ratio Between 160 and 220% |         |         |         |         |
| Individual                        | 1.6392  | 1.9401  | 1.1451  | 1.2569  |
| Gaussian                          | 1.1760  | 1.3799  | 0.7963  | 0.8822  |
| t                                 | 1.1210  | 1.3609  | 0.7766  | 0.8783  |
| C-vine                            | 1.2244  | 1.4196  | 0.8261  | 0.8903  |
| Losses Ratio Between 220 and 500% |         |         |         |         |
| Individual                        | 1.8717  | 2.3430  | 1.1887  | 1.7412  |
| Gaussian                          | 1.9153  | 1.7647  | 1.3531  | 1.1366  |
| t                                 | 1.9080  | 1.7289  | 1.3648  | 1.1130  |
| C-vine                            | 1.9296  | 1.8503  | 1.3642  | 1.1674  |

1993 floods in the central and western areas all resulted in widespread crop yield losses. During times of significant yield losses caused by natural catastrophes, there may be greater interdependence between yields in different locations. A generally concerning scenario is that Gaussian copula has a tail reliance of zero, which cannot reflect the tail dependence of risk portfolios as t and vine copula. Nonetheless, the findings of this paper may lead to an intriguing conclusion. According to the above empirical results for Groups 1, 2, and 3, the C-vine copula model can provide a more accurate measurement of losses for layers with tail risk. In addition, the simulation results based on the Gaussian copula model may not differ significantly from those based on the C-vine copula and t-copula model for the layers that lack strong tail dependence between counties.

**TABLE 15** | The average Mean Squared Error of aggregate losses at different layers for states in Group 3.

|                                   | n = 30  |         | n = 50  |         |
|-----------------------------------|---------|---------|---------|---------|
|                                   | Weibull | Mixture | Weibull | Mixture |
| Losses Ratio Between 100 and 160% |         |         |         |         |
| Individual                        | 6.1879  | 6.1112  | 3.5517  | 3.8775  |
| Gaussian                          | 4.8743  | 3.6330  | 3.0748  | 2.4397  |
| t                                 | 4.9417  | 3.6584  | 3.1490  | 2.4429  |
| C-vine                            | 4.3894  | 3.5051  | 2.6419  | 2.2253  |
| Losses Ratio Between 160 and 220% |         |         |         |         |
| Individual                        | 7.7127  | 8.0662  | 5.1362  | 5.6581  |
| Gaussian                          | 3.1998  | 2.9090  | 1.8722  | 1.8003  |
| t                                 | 3.2092  | 2.8849  | 1.8952  | 1.7698  |
| C-vine                            | 3.4685  | 3.3838  | 2.0679  | 2.1381  |
| Losses Ratio Between 220 and 500% |         |         |         |         |
| Individual                        | 8.4005  | 9.0640  | 6.6696  | 7.4172  |
| Gaussian                          | 3.4782  | 4.5983  | 2.5739  | 3.3544  |
| t                                 | 3.4103  | 4.5170  | 2.5572  | 3.3033  |
| C-vine                            | 3.9801  | 5.0259  | 2.8126  | 3.4737  |

## 5 CONCLUSION

The data sets of crop yield in the neighborhood county are affected by systemic risks. Accurate modeling of these systemic risks is essential for determining reliable premium rates of the publicly-subsidized federal crop insurance program. This article detects whether the copula-based models significantly affect losses assessment and insurance pricing. Specifically, we apply the different copula-based models to the losses evaluation of crop insurance. The

## REFERENCES

- Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula Constructions of Multiple Dependence. *Insur. Math. Econ.* 44 (2), 182–198. doi:10.1016/j.insmatheco.2007.02.001
- Ahmed, O., and Serra, T. (2015). Economic Analysis of the Introduction of Agricultural Revenue Insurance Contracts in Spain Using Statistical Copulas. *Agric. Econ.* 46 (1), 69–79. doi:10.1111/agec.12141
- Atwood, J., Shaik, S., and Watts, M. (2003). Are Crop Yields Normally Distributed? A Reexamination. *Am. J. Agric. Econ.* 85 (4), 888–901. doi:10.1111/1467-8276.00495
- Bedford, T. J., and Cooke, R. M. (2002). Vines—A New Graphical Model for Dependent Random Variables. *Ann. Statistics* 30 (4), 1031–1068. doi:10.1214/aos/1031689016
- Bulut, H., and J. Collins, K. (2014). Designing Farm Supplemental Revenue Coverage Options on Top of Crop Insurance Coverage. *Agric. Finance Rev.* 74 (3), 397–426. doi:10.1108/afir-08-2013-0032
- Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula Methods in Finance*. England: John Wiley & Sons.
- Coble, K. H., Hanson, T., Miller, J. C., and Shaik, S. (2003). Agricultural Insurance as an Environmental Policy Tool. *J. Agric. Appl. Econ.* 35 (02), 391–405. doi:10.1017/s1074070800021350
- Coble, K. H., Knight, T. O., Goodwin, B. K., Miller, M. F., and Rejesus, R. M. (2010). *A Comprehensive Review of the RMA APH and COMBO Rating Methodology: Final Report*. Washington, DC: Report prepared for the USDA Risk Management Agency by Sumaria Systems, Inc., U.S. Department of Agriculture, Risk Management Agency.

copulas consider the dependence of yields for adjacent counties—especially the upper and lower tails of the distribution.

To put this into perspective, the unbiased estimation of claims under systemic risk measured by individual and copula-based methods does not significantly differ. Nevertheless, for the different layers of SRA policy, the measuring of dependence significantly impacts aggregate claims. Given the massive scale of the federal crop insurance program, the copula-based model will significantly impact crop reinsurance contracts' pricing, feasibility, and profitability for the insurers, the government, and taxpayers. The vine copula provides a more flexible property in the aggregate losses estimation process in the regions with obviously estimated tail dependence.

## DATA AVAILABILITY STATEMENT

All data used in this study are publicly available. The processed data needed to reproduce this study are available at USDA's datasets (<https://www.usda.gov/>).

## AUTHOR CONTRIBUTIONS

YS conceived and designed the study, provided the idea, performed the data analysis, analyzed the results, and wrote the manuscript. KW discussed the idea with YS, and helped design the study and check the results.

- Embrechts, P., McNeil, A. J., and Straumann, D. (1999). Correlation: Pitfalls and Alternatives. *Risk* 12, 69–71. doi:10.1017/CBO9780511615337.008
- Feng, X., and Hayes, D. (2016). Diversifying Systemic Risk in Agriculture. *Afr* 76 (4), 512–531. doi:10.1108/afir-06-2016-0061
- Finger, R. (2010). Revisiting the Evaluation of Robust Regression Techniques for Crop Yield Data Detrending. *Am. J. Agric. Econ.* 92, 205–211. doi:10.1093/ajae/aap021
- Goodwin, B. K., and Hungerford, A. (2015). Copula-Based Models of Systemic Risk in U.S. Agriculture: Implications for Crop Insurance and Reinsurance Contracts. *Am. J. Agric. Econ.* 97 (3), 879–896. doi:10.1093/ajae/aa086
- Goodwin, B. K., and Ker, A. P. (1998). Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance Contracts. *Am. J. Agric. Econ.* 80 (1), 139–153. doi:10.2307/3180276
- Goodwin, B. K. (2001). Problems with Market Insurance in Agriculture. *Am. J. Agric. Econ.* 83 (3), 643–649. doi:10.1111/0002-9092.00184
- Goodwin, B. K., Roberts, M. C., and Coble, K. H. (2000). Measurement of Price Risk in Revenue Insurance: Implications of Distributional Assumptions. *J. Agric. Resour. Econ.* 25 (1), 195–214. doi:10.22004/ag.econ.30830
- Goodwin, B. K., and Smith, V. H. (2013). What Harm Is Done by Subsidizing Crop Insurance? *Am. J. Agric. Econ.* 95 (2), 489–497. doi:10.1093/ajae/aas092
- Goodwin, B. K., and Vado, L. (2007). Public Responses to Agricultural Disasters: Rethinking the Role of Government. *Can. J. Agric. Econ.* 55 (4), 399–417. doi:10.1111/j.1744-7976.2007.00099.x
- Harri, A., Coble, K. H., Ker, A. P., and Goodwin, B. J. (2011). Relaxing Heteroscedasticity Assumptions in Area-Yield Crop Insurance Rating. *Am. J. Agric. Econ.* 93 (3), 707–717. doi:10.1093/ajae/aar009



- Huber, P. J. (1973). Robust Regression: Asymptotics, Conjectures and Monte Carlo. *Ann. Statistics* 1, 799–821. doi:10.1214/aos/1176342503
- Ker, A. P., Tolhurst, T. N., and Liu, Y. (2016). Bayesian Estimation of Possibly Similar Yield Densities: Implications for Rating Crop Insurance Contracts. *Am. J. Agric. Econ.* 98, 360–382. doi:10.1093/ajae/aav065
- Kurowiczka, D., and Cooke, R. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. Chichester: Wiley.
- Li, B., and Genton, M. G. (2013). Nonparametric Identification of Copula Structures. *J. Am. Stat. Assoc.* 108 (502), 666–675. doi:10.1080/01621459.2013.787083
- Miranda, M. J., and Glauber, J. W. (1997). Systemic Risk, Reinsurance, and the Failure of Crop Insurance Markets. *Am. J. Agric. Econ.* 79, 206–215. doi:10.2307/1243954
- Nelsen, R. B. (1999). *An Introduction to Copulas*. New York: Springer.
- Nelson, C. H., and Preckel, P. V. (1989). The Conditional Beta Distribution as a Stochastic Production Function. *Am. J. Agric. Econ.* 71 (2), 370–378. doi:10.2307/1241595
- Okhrin, O., Odening, M., and Xu, W. (2012). Systemic Weather Risk and Crop Insurance: The Case of China. *J. Risk Insur.* 80 (2), 351–372. doi:10.1111/j.1539-6975.2012.01476.x
- Ramsey, A. F., and Goodwin, B. K. (2019). Value-at-Risk and Models of Dependence in the U.S. Federal Crop Insurance Program. *J. Risk Financial Manag.* 12 (2), 1–21. doi:10.3390/jrfm12020065
- Ramsey, A. F., Goodwin, B. K., and Ghosh, S. (2019). How High the Hedge: Relationships between Prices and Yields in the Federal Crop Insurance Program. *J. Agric. Resour. Econ.* 44 (2), 227–245. doi:10.22004/ag.econ.287967
- Sherrick, B. J., Zanini, F. C., Schnitkey, G. D., and Irwin, S. H. (2004). Crop Insurance Valuation under Alternative Yield Distributions. *Am. J. Agric. Econ.* 86 (2), 406–419. doi:10.1111/j.0092-5853.2004.00587.x
- Sklar, A. (1959). *Fonctions de répartition à n dimensions et leurs marges*, 8. Paris: Publications de l'Institut de Statistique de l'Université de Paris, 229–231.
- Smith, V. H., and Goodwin, B. K. (1996). Crop Insurance, Moral Hazard, and Agricultural Chemical Use. *Am. J. Agric. Econ.* 78 (2), 428–438. doi:10.2307/1243714
- Smith, V. (2011). “Premium Payments: Why Crop Insurance Costs Too Much,” in *American Boondoggle—Fixing the 2012 Farm Bill* (Washington, DC: American Enterprise Institute).
- Tolhurst, T. N., and Ker, A. P. (2015). On Technological Change in Crop Yields. *Am. J. Agric. Econ.* 97 (1), 137–158. doi:10.1093/ajae/aau082
- Vedenov, D. V., and Power, G. J. (2008). Risk-Reducing Effectiveness of Revenue versus Yield Insurance in the Presence of Government Payments. *J. Agric. Appl. Econ.* 40 (2), 443–459. doi:10.1017/s1074070800023737
- Wang, H., and Zhang, H. (2003). On the Possibility of a Private Crop Insurance Market: A Spatial Statistics Approach. *J. Risk Insur.* 70, 111–124. doi:10.1111/1539-6975.00051
- Woodard, J. D., and Sherrick, B. J. (2011). Estimation of Mixture Models Using Cross-Validation Optimization: Implications for Crop Yield Distribution Modeling. *Am. J. Agric. Econ.* 93 (4), 968–982. doi:10.1093/ajae/aar034
- Xu, W., Filler, G., Odening, M., and Okhrin, O. (2010). On the Systemic Nature of Weather Risk. *Agric. Finance Rev.* 70 (2), 267–284. doi:10.1108/00021461011065283
- Yi, F., Zhou, M., and Zhang, Y. Y. (2020). Value of Incorporating ENSO Forecast in Crop Insurance Programs. *Am. J. Agric. Econ.* 102, 439–457. doi:10.1002/ajae.12034
- Yvette Zhang, Y. (2017). A Density-Ratio Model of Crop Yield Distributions. *Am. J. Agric. Econ.* 99, 1327–1343. doi:10.1093/ajae/aax021
- Zhu, Y., Ghosh, S. K., and Goodwin, B. K. (2008). Modeling Dependence in the Design of Whole Farm Insurance Contract, a Copula-Based Model Approach. Orlando, FL, USA: Proceedings of the Annual Meetings of the American Agricultural Economics Association.

**Conflict of Interest:** KW was employed by the company China Agriculture Reinsurance Corporation.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Sun and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.