Review of Crop Revenue Insurance in China and A New Estimation of Premium

Abstract

The main goal of our research is to put forward an innovative thought to study the development of "Insurance + Futures" mode in China from a perspective of small area estimation of crop yield distributions. We focus on crop revenue insurance based on rainfall index where the agricultural futures prices are incorporated and realize the combination of insurance and futures. In this paper, a density ratio model is adopted to estimate the crop yield distributions of 16 cities in Shandong Province, which has been proved to be a more reliable method in the case of limited sample size. Then we take a city as an example and construct the joint distribution of yield, price and rainfall with the help of Vine Copula to calculate the premium of crop revenue insurance based on rainfall index. Through these steps, we not only make a more reasonable estimation of crop yield distribution using pooled information from adjacent areas, but also achieve the target of using futures to diversify price risk. Our results show that the yield distributions obtained by a small area estimation are more reasonable than those by a direct estimation. Then we verify the practicability of our work. The method we proposed to calculate premium is more effective than the other ones. Meanwhile, compared with traditional revenue insurance, the new insurance is more affordable. Based on theoretical analyses and reality, we put forward some relevant suggestions which can contribute to the further research and the sustainable development of crop insurance under the "Insurance + Futures" background.

Key Words: Crop yield distribution, Premium, Rainfall Index

1. INTRODUCTION

For a long time, agricultural insurance has provided effective safeguard for agricultural production to deal with different kinds of risks, which is also an important means for governments to support agricultural development. With the constant changes of agricultural production pattern, traditional agricultural insurance cannot satisfy farmers’ demands, whose production features, risk attitudes and insurance demands are quite different. Insurance for those who have a relatively small farming scale is dispensable. They are very sensitive to the premium due to their indifference towards losses. In contrast, farmers with large production scale hate the heavy losses. There are urgent needs for agricultural insurance for them and they can accept a slightly higher premium. However, most of the agricultural insurance in China is cost insurance at present. The low protection level of insurance fails to meet the needs of large-scale farmers. The unattractive insurance premium and inadequate coverage level reduce the enthusiasm of farmers to participate in insurance.

The financial crisis, global climate changes and the COVID-19 outbreak increase the uncertainty of farmers' income, which makes it more difficult for farmers to make decisions on agricultural production. It is necessary to explore innovative insurance patterns. Compared with traditional agricultural insurance, crop revenue insurance can provide comprehensive protection for both agricultural yield risk and price risk. From the micro point of view, launching agricultural revenue insurance provides farmers a new opportunity to avoid various risks.

In addition, under the mode of “Insurance + Futures”, the pricing of crop revenue insurance is based on the agricultural futures price. This helps insurance companies to avoid collecting data lacking of timeliness, which may be also discontinuous and subjective. And by connecting the insurance market with the futures market, agricultural production risks can be dispersed in a larger market scope.

Considering the advantages of revenue insurance and risk hedging function of futures market, in this paper, we focus on the crop revenue insurance based on rainfall index, which is also a kind of insurance designed according to the agricultural futures price.

In agricultural insurance, yield risk is related to the probability that the actual yield is lower than a certain threshold (target yield). Therefore, a reasonable estimation of crop yield distributions is of fundamental importance in the process of agricultural policy making and insurance design. Most of the existing studies estimate the regional crop yield distribution respectively on the base of a short panel data with a few observations of each region, which only use information from the small-area samples. The Individual yield density estimated neither through a parametric nor non-parametric method tend to suffer from large sampling variations due to limited observations. So, it is urgent to find a feasible way to achieve an accurate estimation of crop yield distributions.

The development of small area estimation is inspirational for our research. In recent years, the demand for small area estimators which are widely used in various fields has increased rapidly. The idea of small area estimation is “borrowing strength” —there is little direct information from a small area, we can effectively use the samples or data in similar areas as auxiliary information to achieve a reliable estimation. Since the farming methods and climatic conditions in geographically adjacent areas are similar, the crop yield distributions in these areas tend to resemble each other. Encouraged by such an interesting fact, a crop yield distribution of can be gained through small area estimation using pooled samples from adjacent regions.

On these foundations, we can calculate the premiums of crop revenue insurance based on rainfall index. With the tremendous development of "insurance+ futures", we refer to the method of using futures market to disperse agricultural risks earlier, and take the agricultural futures price as the agricultural products price in the process of insurance design. Understanding the links between three variables: price, yield, and rainfall index in the designed insurance is a crucial to premium calculation, a key element of an insurance policy. Rainfall, for example, affects crop yield, which in turn affects crop price and ultimate payments. Traditional linear correlation has some limitations, so we draw support from Vine Copula to measure the dependency, which can ensure a more precise calculation of insurance premium.

In this paper, we adopt a flexible semi-parametric density ratio method developed by Zhang (2017) and Vine Copula to study the development of combining insurance and futures. We first introduce the basic situation of "Insurance + Futures" mode and the demand for various agricultural insurance in China. Next, taking soybean, wheat and corn as examples, we estimate the yield distributions of seventeen cities in Shandong Province using a small area estimation model. Then we choose a city to calculate the premiums of the designed insurance based on related structure of price, crop yield, and standard rainfall constructed through Vine copula. Finally, we evaluate the effectiveness of our work from the accuracy of premium calculation and premium level. Based on the above analysis, we also put forward some policy recommendations and possible future expansion of this study.

There are some innovations in our research. Our method can solve two problems at the same time: the inaccurate estimation of yield distribution caused by too small sample size and the unrealistic estimation caused by individual estimation of the distribution of each variable. These are the two aspects that need to be improved in the process of premium calculation. On this basis, we design a new type of insurance specifically for drought risk, which is more attractive in terms of price than traditional revenue insurance.

We also have made several major contributions. Firstly, our study is an early application of small area statistical method to the estimation of crop yield distribution in China. Secondly, the combination of a density ratio model with Vine Copula ensures a more realistic premium calculation as a result of the reliable estimation of a yield distribution and the utilization of the relationships between factors, which is proved to be a more efficient way. Various tests can ensure the reliability of the results in this paper.

2. LITERATURE

The existing studies on the estimation of crop yield distributions can be divided into two parts according to their method: a parametric or non-parametric estimation.

In a parametric method, it is necessary to assume a distribution that crop yield follows and determine the specific function form by estimating the parameters. Common parametric distributions used in most studies include Normal distribution, Log-Normal distribution, Beta distribution, Gamma distribution and their extension. For instance, we assume that crop yield follows the log-normal distribution, in which case we only need to get the mean and variance to determine the specific function of yield. The parametric method is simple and asymptotically stable under the correct distribution assumption. However, it requires that the assumed yield distribution is accurate, which is difficult to realize. A series of studies (Harri et al. 2009; Hennessy, 2009; Koundouri and Kourogenis 2011; Richard and Claassen 2011) have offered the important insight into the defect of a parametric method in estimating yield distributions that there are skewness and heterogeneity in a distribution of crop yield. It has been proved that the large variation in the expected payout of crop insurance products was largely due to the parameter yield distribution chosen (Sherrick et al. 2004). The misspecification of yield distribution may lead to the miscalculation of the expected payout in an insurance.

Unlike a parametric method which assumes the distribution form in advance, the non-parametric method chooses a yield distribution according to data. Instead of seeking the true model, a non-parametric estimation is to find an approximate unknown curve which can better describe the distribution of yield, while it is usually biased. In addition, since a non-parametric estimation is usually less efficient than a parametric estimation, a larger sample size is required. The most common non-parametric methods include kernel density estimation, log-spline density estimation, series density estimation and local maximum likelihood estimation. Goodwin and Ker (1998) adopted the non-parametric kernel density estimation to estimate yield distributions of different crops and calculate the insurance premium rates of wheat and barley based on the yield distributions. Ker and Goodwin (2000) shed new light on the yield distribution estimation by using the Bayesian kernel density estimation method to estimate the crop yield distribution and proved that this method is efficient in premium calculation. Norwood et al. (2004) compared the efficiency of the six methods used to estimate the yield distribution and finally found that the best model to describe the yield distribution out-of-sample is a model with a kernel smoother.

In recent years, the more detailed small area estimators have found broad application over the world. For governments, small area statistical estimation is of great significance in formulating policies, regional planning and fund allocation (Fay and Herriot 1979; Schaible 1993). Meanwhile, the demand from the private sector has the same trend due to the heavy dependence on the regional economy and environment in the process of making decisions (Kriegler and Berk 2010). More specifically, some studies have used small area statistical methods to estimate crop yields and crop acreage (Flores et al. 2000; Singh et al. 2000). All these have shown the feasibility of small area estimators in various fields.

Copula has been attracting a lot of interest to in constructing related structures of different factors. Although linear correlation is popular in applied economics, this assumption may not be satisfied. More and more scholars consider using copula to describe the relationship between variables. Some studies (Bokusheva 2011; Goodwin and Hungerford 2015) have confirmed that the use of Copula in insurance to measure the relationship between different variables, such as price and yield, has a significant impact on efficiency of crop insurance premium calculation. Compared with a linear correlation, it is shown that Copula can capture the relationship between random variables more reliably. Using copula to take the relationship between weather and yield into account, Bokusheva (2018) designed a weather index insurance designed to provide protection against crop losses from agricultural droughts. Copula has gradually received considerable critical attention in the fields of economics.

3. METHOD

3.1 SMALL AREA ESTIMATION: A DENSITY RATIO MODEL

Usually, if the sample size of a specific domain is not sufficient to support an accurate direct estimation, it will be considered as a small area. Under this circumstance, it is necessary to use values of the variable of interest from related areas or time periods to increase the "effective" sample size. In this paper, we assume that the error term of the nested error regression model follows a density ratio model. We can estimate the error distribution of each small region, which can be used to deduce the distribution of the target variable.

According to a density ratio model, an individual density $f_i(x)$ consists of two parts: the baseline density f_0 and a deviation function $e^{\alpha_i+\beta_i{}'h(x)}$. In this way, an individual density can be seen as a deviation from the baseline density.

$$f_i(x) = f_0(x) \times e^{\alpha_i+\beta_i'\varphi(x)} \tag{1}$$

For *(i, j)*, we have:

$$\frac{f_i(x)}{f_j(x)} = \frac{e^{\alpha_i+\beta_i'\varphi(x)}}{e^{\alpha_i+\beta_j'\varphi(x)}} \tag{2}$$

Since there is a common baseline density, the ratio of the densities of two individuals equals the ratio of their deviation functions. Hence, it is called a "density ratio" model.

Using a density ratio model to estimate individual density is usually a two-stage process. In the first stage, we need to estimate the baseline density of the pooled samples by a parametric or non-parametric means. In the second stage, we adopted the adjusted method developed by Zhang (2016) which proposes to transform the data via the Probability Integral Transformation

(PIT) instead of directly modeling individual distortion function. As for the distortion parts, we use a smooth deviation model from the uniform distribution proposed by Neyman (1937).

$$e^{\alpha+\beta_1\varphi_1(x)+\beta_2\varphi_2(x)+\cdots\beta_k\varphi_k(x)} \quad x \in [0,1] \tag{3}$$

Where β_0 is a normalization constant and $(\varphi_1,..,\varphi_k)$ is a K-th order orthonormal polynomials on [0, 1].

Next, we can estimate individual densities as deviation from the baseline with the help of Poisson regression approach which can be easily implemented with statistical computer programs. Dividing unit interval [0, 1] into *J* equal length sub-intervals, we treat the frequency of transformed data u_{it} falling into these subintervals Y_{ij} $j = 1,\ldots,J$ as the dependent variable in a Poisson regression, and use the mid-point of these subintervals $j^* = (j-1/2)/J$ as the independent variable. And the Poisson regression model is given by:

$$lnE(Y_{ij}) = ln\mu_{ij} = lnT + lnf_0(j^*) + \alpha + \beta_i'\varphi(j^*) \tag{4}$$

According to the model, the coefficients β_i of deviation function can be estimated by Poisson regression. Since the evaluation of the normalizing constant is not required in the Poisson regression, we need to estimate α by the maximum likelihood estimation.

$$\alpha = ln\left\{\int \widehat{f_0}(x) \times e^{\widehat{\beta}_i'\varphi(\widehat{F}_0(x))}\right\}^{-1} \tag{5}$$

Thus, the estimated *i-th* individual density is then given by：

$$\hat{f}_i(x) = \widehat{f_0}(x) \times e^{\widehat{\alpha}+\widehat{\beta}_i'\varphi(\widehat{F}_0(x))} \tag{6}$$

As a typical approach of small area estimation, the benefits of a density ratio model are twofold. First, it allows us access to the utilization of pooled information, by which a common baseline density can be estimated from the pooled sample. Second, flexibility in an individual density can be accommodated via the configuration of an individual distortion function. When the dimension of φ in the distortion function is small, the individual density deviates from the baseline only in a few directions, which effectively shrinks the density to the common baseline. That is to say, the non-parametric estimation of the common baseline makes use of the large capacity of the set sample, and the relatively simple distortion function is more suitable for the small sample size of a single cell.

3.2 VINE COPULA

Copula can describe the non-linear correlation between different elements which is a better measure of dependency between variables than other methods. But it is not appropriate to build the linked structure for higher dimensions. The existence of Vine Copula, a structure that illustrates the pair copula construction (PCC) provides a new way for researchers to consider the relationships between variables. As shown in Figure 1, there are $n-1$ copula trees in n-dimensional Vine Copula. In the first tree, there are n nodes connected by $n-1$ edges. One edge represents a pair of copula density between two nodes adjacent to each other. The nodes in the next tree come from the edges of the previous tree. And so on, in the *j-th* copula tree, there are $n-j$ edges which represent the conditional copula pair density between $i+1-j$ nodes. For example, the edge $C_{1,2}$ represents the copula density between node 1 and node 2 in Tree 1. The edge $C_{1,3|2}$ is the copula density between the first two nodes in Tree 2.

Our goal is to obtain the joint density of the nodes in the first tree. With this formula, we can fit different parametric distributions, such as Lognormal distribution, Weibull distribution, Gamma distribution, Beta distribution to these nodes respectively to find their own proper distribution $F(x_i)$ according to Akaike Information Criterion (AIC) or Bayesian Information

Criterion (BIC). To fit the elements of copula∈[0, 1], we carry on the probability integral transformation to the data of each node. And we can easily find the best choice of copula for each pair of nodes from Gaussian copula, t-copula, Clayton copula, Gumbel copula, Frank copula and Joe copula with the help of Vine Copula R package. Meanwhile, we can get conditional distribution functions $F_{i|i-1,\ldots\ldots,1}(x_i|x_{i-1},\ldots\ldots,x_1)$. Similarly, the optimal copula will be selected for the conditional pairs. After choosing proper marginal functions and copula between the nodes in each tree, we can get joint density according to the vine structure we have constructed.

For example, in the Vine Copula with three variables in the basal tree, we can represent a joint density as follows:

$$
\begin{aligned}
f(x_1,x_2,x_3) &= f_3(x_3)\times f_2(x_2)\times f_1(x_1)\ (marginals) \qquad (7)\\
&\times C_{1,2}\big(F_1(x_1),F_2(x_2)\big)\times C_{2,3}\big(F_2(x_2),F_3(x_3)\big)(unconditional\ pairs)\\
&\times C_{1,3|2}(F_{2|1}(x_1|x_2),F_{3|2}(x_3|x_2))(conditional\ pair)
\end{aligned}
$$

3.3 REVENUE INSURANCE BASED ON RAINFALL INDEX

Agricultural insurance premium is closely related to the value of payments. In order to calculate premium, we now pay attention to the insurance expected loss payment. Different from traditional insurance, in revenue insurance with rainfall index, indemnity depends on rainfall. The less the actual rainfall in the insurance period is, the greater defense against loss can insurance provide. When the agreed revenue per unit area is *T*, the indemnity (I_P) under different actual rainfall can be calculated as below:

$$
I_P = T * \begin{cases} \dfrac{IND_\alpha - IND_t}{IND_\alpha} & if\ IND_t \le IND_\alpha \\ 0 & if\ IND_t > IND_\alpha \end{cases} \qquad (8)
$$

If the actual rainfall (IND_t) during the crop growing period meets the rainfall index standard (IND_α) we set in advance, then the insurance will not play a protective role. If the actual rainfall is not up to our standard, we will determine indemnity based on the ratio of the difference between actual rainfall and standard relative to the numerical value of standard. It means that whether a farmer can receive a loss payment from the insurance company depends on more than one condition. No matter how much the actual income is, he will not receive a payment from the insurance when the accumulated rainfall reaches the standard. Even when the actual rainfall reaches the standard, payment still depends on whether the farmer suffers a loss (L), In other words, whether the actual income of farmers (I_t) is lower than the agreed one (T).

$$L = \begin{cases} T - I_t & if\ T \geq I_t \\ 0 & if\ T < I_t \end{cases} \tag{9}$$

$$P^L = \begin{cases} I_P & if\ L > I_P \\ T - I_t & if\ 0 < L \leq I_P \\ 0 & if\ L = 0 \end{cases} \tag{10}$$

When the agreed revenue is less than actual income which equals the product of actual yield and actual price, the farmer will not receive a payment (P^L) because the loss is equal to 0. If T is more than I_t, in other words, if the loss occurs, the farmer will receive a payment which equals to the difference between them but not more than I_P. Because the expected loss payment reflects the expected value that an insurance company should offer to a farmer, pure premium R of crop insurance can be obtained according to its expectation:

$$R = E(P^L) \tag{11}$$

4. DATA AND RESULT

4.1 DATA

Soybean, wheat and corn are important grain crops in China, a country has five thousand years of history of cultivation. As we all know, rainfall has a decisive influence on crop yield. It is of great practical significance to design an insurance against drought risk. Shandong Province is one of the main crop planting regions in China. All the cities in Shandong are in the monsoon climate zone where rainfall changes greatly in different years. Therefore, the annual crop yield is highly uncertain. Finding a way to accurately estimate crop yield distributions and designing a kind of insurance against drought risk can both improve the efficiency of the insurance market and help farmers cope with uncertainty. In this paper, we use the annual average yield of three crops from 2006 to 2016 of 16 cities in Shandong Province to estimate their yield distributions. Because these agricultural futures contracts have been listed and kinds of possibilities that cause the price fluctuation of agricultural products are taken into consideration in the futures price, it will be closer to the real price when the crop is harvested, which can avoid the adverse selection of farmers for insurance. Hence, we use the settlement price in October of the year when the insurance contract was signed of the soybean futures contract due in January of next year as the price data to calculate revenue. Similarly, the settlement price of the wheat futures in June and the one of corn in September are used. Crop yield and price data are from Wind Economic Database. We choose the value of accumulated precipitation during the period of soybean growth (from June to September), wheat growth (from October to May) and corn growth (from April to August) as the rainfall index separately. Precipitation data come from Shandong Statistic Year Book.

4.2 CROP YIELD DISTRIBUTIONS

In this section, we apply the density ratio model to estimate the crop yield distributions of 16 cities in Shandong province.

A common approach to crop yield distribution estimation is to use historical yields for the area of interest and estimate the conditional yield density based on the trend-adjusted residuals. Because crop yields have been gradually trending up during the past few decades, we apply a flexible model to we need to apply a flexible model to take account of this trend before estimating an individual yield distribution.

$$W_{it} = m(t) + C_i + e_{it} \tag{12}$$

W_{it} is the average yield of city *i* in year *t*. $m(t)$ is the time trend. Following a common practice in the literature of crop yield modeling, we use a two-knot linear spline function to divide the sample period into three equal sub-periods. C_i represents the individual effect at city level. e_{it} is an error term with mean zero and finite variance. Taking the potential heteroscedasticity of the error term into account, we further normalize the estimated error terms.

$$\hat{X}_{it} = \frac{\hat{e}_{it}}{\hat{W}_{it}} \tag{13}$$

In the two-stage process, a common baseline density is estimated using data of the pooled cities through a non-parametric log-spline density estimation. At the same time, we can get its corresponding distribution. Then we can set out to estimate the individual yield density of each city. We first use PIT to transform $\hat{X}_{it}$ into $\hat{u}_{it}$ with a range of [0, 1]. The standard orthogonal polynomials commonly used are shifted Legendre polynomials which are orthonormal with respect to the uniform distribution. By comparing AIC value from $K = 1$ to 6, we find that is the lowest one when the most appropriate number of Legendre polynomials in the distortion function $K= 2$. Next, we divide unit interval into six equal length sub-intervals and use Poisson regression to estimate the deviation function and get the individual density for each city.

In Figure 2, we report the estimated results of soybean. The black solid line is the common baseline density of all samples, and the dotted line is the individual density of each city. Visually, the densities of different cities are similar because they come from a common baseline density. However, due to the existence of the deviation function, a reasonable deviation is exhibited between individual density of each city and the baseline density.

For comparison, we estimate individual soybean yield density of each city separately. Specifically, a log-spline estimation is directly used to their own soybean yield data. As a reference, the same baseline density of the pooled samples is also plotted in Figure 4. The results of this method are quite different from the previous ones. First, the density of each city is irregular which reflects the problem of over-fitting. In addition, there is a considerably large difference between the individual density of each city and the baseline density. These illustrate the individual density of each city estimated respectively suffers from the small sample size.

We also report the yield distributions of wheat and corn of small area estimation and direct estimation in supplementary figures. Predictably, we get the same conclusion as before that the severe shortage of direct information from individual areas may bring about poor performance of the direct estimation while the density ratio model can conquer the barrier. Therefore, the yield distribution estimated based on small area estimation is more reliable than a traditional method. On this basis, agricultural planning and insurance design will be more efficient.

4.3 PURE PREMIUM

In the process of premium calculation, there are three pivotal variables in revenue insurance based on rainfall index we need to pay attention to: price (P_t), yield (Y_t) and rainfall index (IND_t). We now introduce Vine Copula to study the dependency links between variables.

We take Jinan City as an example and use empirical data to test Kendall's correlation between price, yield and rainfall index of each crop. According to the relationship between Kendall correlation coefficients, we can choose two pairs of variables: Y_t and P_t as well as Y_t and IND_t who are more relevant as the fundamental information and construct the copula trees represented in Figure 4.

Next, we need to use Bayesian Information Criterion and Kolmogorov-Smirnov test to select the optimal distribution of price and rainfall index. After fitting Weibull distribution, lognormal distribution and Gamma distribution to P_t and IND_t respectively, we find the lowest BIC value when fitting the Weibull distribution to both the two variables of soybean and the price of wheat. However, the rainfall index during wheat growth period and the price of corn are more in line with lognormal distribution. In addition, the Lognormal distribution is more suitable for corn's rainfall index. The results of K-S test also illustrate that all the variables satisfy the corresponding distribution with p-value significantly larger than 0.05, showing goodness of fit. We then carry on the probability integral transformation to the empirical data of three crops to get $u_1 = F_1(P)$, $u_2 = F_2(Y)$ and $u_3 = F_3(IND)$. Based on the copula trees built in Figure 6, we need to find dependency relationships between u_1 and u_2, u_2 and u_3 as well as $F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)$ which are the nodes in the second tree. Referring to the existing research, we take advantage of BIC to find the optimal copula for each pair. All the results above are reported in the supplementary tables.

In the light of Vine Copula, we form 10,000 sets of simulation data u_1, u_2, u_3 and transform them back to their own distribution to obtain interrelated P_t , Y_t , IND_t . We performed this simulation for three crops separately.

On the basis of the previous steps, we can easily calculate pure premium of crop revenue insurance with rainfall index. The average of the empirical crop yield data is treated as the expected yield and the average empirical price is treated as the expected price. We can multiply them to get the expected revenue (ER) of unit area (mu) which is also set as the agreed revenue (T) in this designed insurance. In the process of premium calculation, the 10,000 sets of simulation data that we generated before are treated as the true yield, price and rainfall index data. The rainfall index standard (IND_α) is α quantile of the rainfall index distribution. It is shown in Table 1 that, for instance, the rainfall threshold in the growing period of wheat is equal to 150.27 millimeter when α=0.2. With such a standard, the insurance premium is about 8.15 per mu. For revenue insurance with different rainfall index standards, there are different indemnity, so we can get different expected loss payment. The larger the rainfall standard is, the more likely loss payments will occur. As a result, the premium will rise gradually with the increase of rainfall index standard.

5. EFFECTIVENESS

In this section, we verify the reliability and practicability of our work from two aspects. First of all, we contrast the efficiency of our proposed method to calculate premium with some more common methods and find that the combination of a density ratio model and Vine Copula does improve the accuracy of premium calculation. Secondly, by comparing the premium of revenue insurance based on rainfall index and traditional revenue insurance, the insurance we designed is proved to be more affordable.

5.1 EFFICIENCY OF DIFFERENT METHODS

We first illustrate the reliability of our method of premium calculation. We base our simulations on actual data. Specifically, we treat the estimated crop yield densities and the joint distribution of yield, price and rainfall index measured through Vine Copula as a "true" distribution and draw random samples from them to use in our simulations. The premium in this case is regarded as a "true" premium (R). We conduct experiments on all three crops with 500 samples of small sample size: T=20. For each experiment, we estimate the premium based on several approaches ($\hat{R}_i$) and then calculate the Mean Squared Error (MSE) of each estimated premium with respect to the "true" premium.

$$MSE = \frac{1}{500}\sum_{i=1}^{500}(\hat{R}_i - R)^2 \tag{14}$$

Particularly, in our proposed method, we estimate the yield densities of simulated data through a density ratio model and find the joint distribution of them using Vine Copula. In a common approach, we get the individual distributions of simulated data of each variable by separate estimation and assume that these variables are independent of each other. We consider both parametric and nonparametric methods to estimate the yield distribution. We also take into account the combination of individual yield distribution estimation with Vine Copula. The MSE results are shown in Table 2. It is seen that the premiums based on our proposed method are generally more accurate than the ones calculated by all the other ways, which shows that the combination of a density ratio model and Vine Copula improves the efficiency of premium calculation.

5.2 TRADITIONAL REVENUE INSURANCE PREMIUM

In traditional revenue insurance, a fixed proportion (γ) of the expected revenue is usually regarded as the agreed revenue. In addition, the loss payment occurs only when the actual revenue of a farmer fails to reach the value of target. The payment P^L is equal to:

$$T = \gamma * ER \tag{15}$$

$$P^L = \begin{cases} T - I_t & if\ T \geq I_t \\ 0 & if\ T < I_t \end{cases} \tag{16}$$

We also calculate the expected loss payment as pure premium. Compared with the insurance with rainfall index, the premium of a traditional full revenue insurance of wheat with the same agreed income in Table 3 is about 52.02 per mu, which is considerably unattractive. Even if the coverage level drops to γ=0.9, the premium of 20.67 per mu is still relatively high. For soybean and corn insurance, the results are clearly consistent. Thus, the high premium of traditional insurance may be responsible for the insufficient insurance demand, while this revenue insurance based on rainfall index allows farmers access to defense under which they can be covered by insurance against drought at a lower premium.

6. CONCLUSION

In this paper, we first introduce the “Insurance + Futures” mode in China, and analyze the current demand for crop insurance. Then we find that there is a lot of room for development in the research of crop insurance under such a background. Until now, most of the research on crop yield distribution is based on the actual historical yield approach, which needs a large number of historical yield data in each small area. However, the lack of long-term data reduces the efficiency of this estimation. Due to the scarcity of direct information from small areas, the realization of the precise estimation is based on the availability of accurate auxiliary information. The development of small area statistical methods has made it possible to realize a reliable estimation. Based on the above analysis, this paper focuses on the estimation of crop yield distribution of 16 cities in Shandong Province by density ratio model. With this method, we can utilize information from all the cities to estimate an individual yield distribution while retaining the simplicity of a direct estimation. By comparison, the crop yield distributions of different cities estimated by a density ratio model keep both similarities and differences.

The key to affect farmers' enthusiasm for insurance is the lack of insurance coverage and the barrier of premium. Revenue insurance can stimulate farmers' enthusiasm for insurance by providing comprehensive safeguard against price risk and yield risk. Under the background of advocating “Insurance + Futures” mode in China, crop revenue insurance is expected to become an important means to fight against agricultural production risks. So, we consider the use of agricultural futures prices in insurance design and pay attention to a functional revenue insurance with rainfall index that can protect farmers at a lower price. As the key element of insurance, premium is of great significance to both insurance suppliers and demanders. Then we calculate premium with the help of Vine Copula which can model a structure to illustrate the non-linear relationship between price, yield and rainfall index. The results for three kinds of crops prove that, compared with other premium calculation approaches, the new one shows a lot of room for improvement. Meanwhile, our newly designed insurance is more affordable for farmers than the traditional revenue insurance with a same target revenue.

Our work demonstrates the validity of the application of small area estimation in crop yield measurement and the practicability of the new insurance in agricultural risk management. The combination of insurance and futures plays an important role in agricultural trade and stability of farmers' actual income, which also can provide strong support for the development of agricultural insurance market. In view of the fact and our estimated results, we put forward some policy recommendations to provide the reference for further research and development of “insurance+ futures” mode. On the one hand, for future study, we can take into account more factors such as temperature that may affect crop yields by changing the specification of the small area estimation model. Meanwhile, a higher dimensional Vine Copula can help us to construct a related structure of more factors. On the other hand, for practical applications, the high cost may be one of the main reasons why traditional insurance is not attractive to farmers. Because the risk undertaken in crop revenue insurance is high, it needs the joint efforts of the insurance market and the futures market to promote the further development of agricultural revenue insurance. In addition, it may be more efficient to carry out targeted insurance such as revenue insurance with weather index according to local farming characteristics and specific conditions. Considering these aspects, farmers, insurers and policy makers all can benefit more from agricultural insurance in the near future.

Reference

1. Annan, F., Tack, J., Harri, A. and Coble, K. (2014). Spatial Pattern of Yield Distributions: Implications for Crop Insurance. *American Journal of Agricultural Economics*, 96(1): 253–268

2. Bedford, T. and Cooke, R. M. (2002). Vines—a new graphical model for dependent random variables. T*he Annals of Statistics*, 30(4): 1031–1068.

3. Berg, A., Quirion, P. and Sultan, B. (2009). Can weather index drought insurance benefit to least developed countries farmers? A case study on Burkina Faso. *Weather, Climate and Society*, 1: 71-84.

4. Bokusheva, R. (2011). Measuring dependence in joint distributions of yield and weather variables. *Agricultural Finance Review*, 71(1): 120-141.

5. Bokusheva, R. (2018). Using copulas for rating weather index insurance contracts. *Journal of Applied Statistics*, 45(13): 2328–2356.

6. Embrechts, P., Lindskog, F. and Mcneil A. (2003). Modelling Dependence with Copulas and Applications to Risk Management. *Handbook of Heavy Tailed Distributions in Finance*, 329–384.

7. Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74(366a): 269-277.

8. Goodwin, B. K. and Ker, A. P. (1998). Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance Contracts. *American Journal of Agricultural Economics*, 80(1): 139-153.
9. Goodwin, B. K. and Hungerford, A. (2015). Copula-Based Models of Systemic Risk in U.S. Agriculture: Implications for Crop Insurance and Reinsurance Contracts. *American Journal of Agricultural Economics*, 97(3): 879-896.
10. Harri, A., Erdem, C., Coble, K. H. and Knight, T. O. (2009). Crop Yield Distributions: A Reconciliation of Previous Research and Statistical Tests for Normality. *Review of Agricultural Economics*, 31(1): 163–182.
11. Ker, A.P. and Goodwin, B. K. (2000). Nonparametric Estimation of Crop Insurance Rates Revisited. *American Journal of Agricultural Economics*, 82(2): 463–478.
12. Ker, A. P. and Coble, K. (2003). Modeling Conditional Yield Densities. *American Journal of Agricultural Economics*, 85(2): 291-304.
13. Koundouri, P. and s Kourogenis, N. (2011). On the Distribution of Crop Yields: Does the Central Limit Theorem Apply? *American Journal of Agricultural Economics*, 93(5): 1341–1357.
14. Kriegler, B. and Berk, R. (2010). Small area estimation of the homeless in Los Angeles: An application of cost-sensitive stochastic gradient boosting. *The Annals of Applied Statistics*, 4(3): 1234-1255.
15. Patton, A.J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2): 527-556.

16. Ren, X., Li, S., Lv, C. and Zhang Z. (2014). Sequential Dependence Modeling Using Bayesian Theory and D-Vine Copula and Its Application on Chemical Process Risk Prediction. *Industrial & Engineering Chemistry Research*, 53(38): 14788–14801.
17. Richard, J. and Claassen, R. (2011). Heterogeneity and Distributional Form of Farm-Level Yields". *American Journal of Agricultural Economics*, 93(1): 144-160.
18. Sherrick, B. J., Zanini, F. C., Schnitkey, G. D. and Irwin, S. H. (2004). Crop Insurance Valuation under Alternative Yield Distributions. *American Journal of Agricultural Economics*, 86(2): 406- 419.
19. Vedenov, D. V. and Barnett, B. J. (2004). Efficiency of weather derivatives as primary crop insurance instruments. *Journal of Agricultural and Resource Economics*, 29(3): 387–403.
20. Xu, W., Filler, G. and Odening, M. (2010). On the systemic nature of weather risk. *Agricultural Finance Review*, 70(2): 267-84.
21. Zhang, Y. Y. (2017). A Density-Ratio Model of Crop Yield Distributions. *American Journal of Agricultural Economics*, 99(5): 1327–43.

Figures and Tables

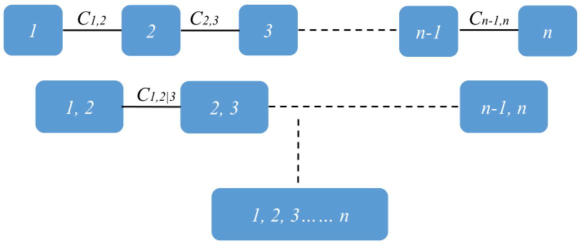


Figure 1 Vine Copula with dimension = *n*

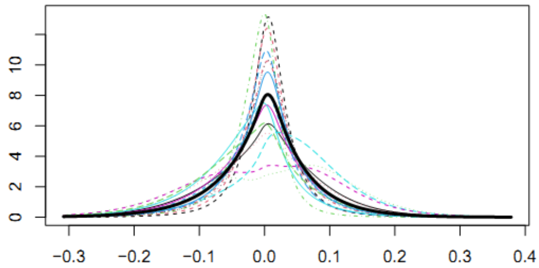


Figure 2 Soybean: DR density estimation

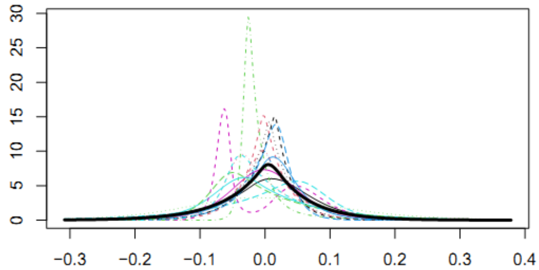


Figure 3 Soybean: Log-spline density estimation

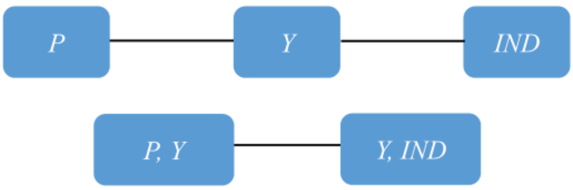
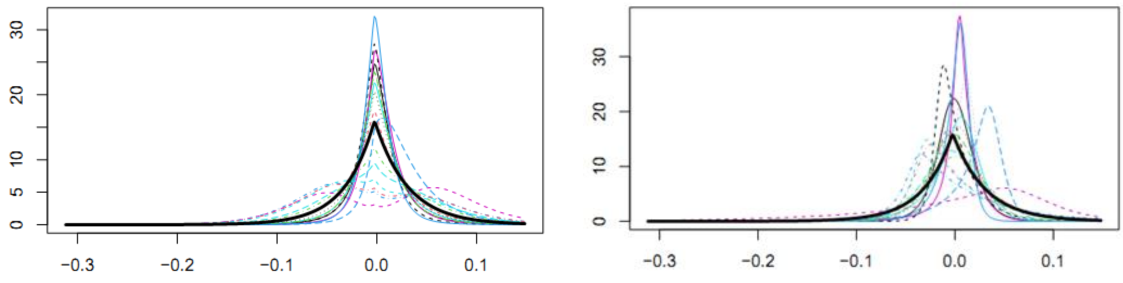


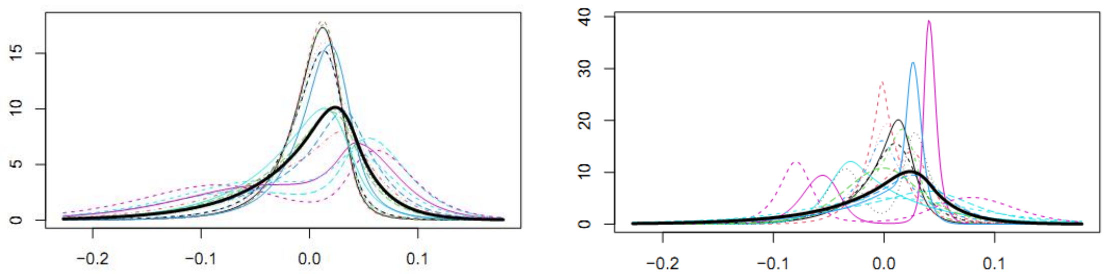
Figure 4 Copula trees



(a) (b)

Figure A1 Wheat: density estimation

Note: (a) DR density estimation; (b) Log-spline density estimation



(a) (b)

Figure A2 Corn: density estimation

Note: (a) DR density estimation; (b) Log-spline density estimation

Table 1　Pure Premium with different IND_α

		α=0.1	α=0.15	α=0.2
Soybean	Agreed Revenue	731.9928	731.9928	731.9928
	IND_α	256.9464	297.0046	330.2077
	Pure Premium	9.661363	13.00758	15.90772
Wheat	Agreed Revenue	938.8556	938.8556	938.8556
	IND_α	129.0235	140.5882	150.2735
	Pure Premium	4.49389	6.306156	8.151515
Corn	Agreed Revenue	850.2856	850.2856	850.2856
	IND_α	362.7898	386.8875	407.1661
	Pure Premium	6.36952	9.384964	12.39117

Table 2　Mean MSE of revenue insurance based on rainfall index

Method	Par	Soybean	Wheat	Corn
(a)	α=0.1	0.940545	3.006757	3.263752
	α=0.15	1.329755	7.445063	15.3775
	α=0.2	4.558302	13.13189	44.03544
(b)	α=0.1	1.087056	3.117843	3.37459
	α=0.15	1.588571	9.095507	15.77137
	α=0.2	5.79707	16.65897	45.11319
(c)	α=0.1	1.978901	4.661137	3.876745
	α=0.15	2.52015	14.80792	17.38596
	α=0.2	9.571977	33.21671	48.15548
(d)	α=0.1	3.891524	3.483988	15.20611
	α=0.15	5.957949	9.245392	30.38802
	α=0.2	7.437923	28.17604	48.20606

Note: (a) DR density estimation and Vine Copula; (b) normal yield distribution and Vine Copula;

(c) individual estimation (log-spline yield distribution); (d) individual estimation (normal yield distribution);

Table 3　Pure premium of traditional revenue insurance

		γ=0.9	γ=0.95	γ=1
Soybean	Agreed Revenue	658.7935	695.3932	731.9928
	Pure Premium	16.99601	28.68784	43.80601
Wheat	Agreed Revenue	844.97004	891.91282	938.8556
	Pure Premium	20.66837	33.48724	52.01646
Corn	Agreed Revenue	765.257	807.7713	850.2856
	Pure Premium	25.20893	40.52323	60.654

Table A1 AIC value from K=1 to 6

	AIC					
	K=1	K=2	K=3	K=4	K=5	K=6
Soybean	381.5338	371.6521	388.2469	406.5147	416.8709	439.0287
Wheat	381.1370	362.0566	384.5792	399.7736	411.6064	429.7811
Corn	404.3247	380.4258	387.1804	396.2008	401.5713	425.0892

Table A2 Kendall correlation test

	τ		
	Soybean	Wheat	Corn
(Y_t, P_t)	-0.303	-0.168	-0.121
(Y_t, IND_t)	0.303	0.303	-0.394
(IND_t, P_t)	-0.091	-0.015	-0.061

Table A3 Weibull distribution parameters and K-S test P-value

		Distribution	Parameter		P-value
Soybean	P_t	Weibull	Shape	9.377398	0.9398
			Scale	4.180464	
	IND_t	Weibull	Shape	2.991485	0.5166
			Scale	545.184940	
Wheat	P_t	Weibull	Shape	8.301697	0.7457
			Scale	2.536767	
	IND_t	Gamma	Shape	10.69168680	0.6926
			Rate	0.05249672	
Corn	P_t	Gamma	Shape	35.27334,	0.6168
			Rate	18.02637	
	IND_t	Lognormal	Meanlog	6.2300575	0.9759
			Sdlog	0.2623648	

Table A4 The best copula

		(u_1, u_2)	(u_2, u_3)	$F_{1\|2}(x_1\|x_2), F_{3\|2}(x_3\|x_2)$
Soybean	Copula	Clayton	Gaussian	Gaussian
	Par	0.06	0.09	0.99
	τ	0.03	0.06	0.93
Wheat	Copula	Gaussian	Frank	Gaussian
	Par	-0.19	-1.12	0.97
	τ	-0.12	-0.12	0.8
Corn	Copula	Gaussian	Frank	Gaussian
	Par	-0.02	-0.1	0.99
	τ	-0.02	0.01	0.98