IJSA (2021), Vol. 5, No. 1, 205-219
DOI: https://doi.org/10.29244/ijsa.v5i1p205-219

Indonesian Journal of Statistics and Its Applications
Available at https://ijsa.stats.id
e-ISSN: 2599-0802

Determining Critical Yield Index of Area Yield Insurance based on Basis Risk Constraint*

Valantino Agus Sutomo[1‡], Dian Kusumaningrum[2], Aurellia Layvieda[3], and Rahma Anisa[4]

[1,2,3]Business Mathematics Program, School of Applied STEM (Universitas Prasetiya Mulya), Indonesia
[4]Department of Statistics, Faculty of Mathematics and Natural Sciences, Indonesia
‡corresponding author: valantino.sutomo@prasetiyamulya.ac.id

Abstract

Area yield index insurance at district level faces heterogeneous basis risk due to geographical conditions which implies to obtain unprecise critical index (y_c). Clustering and zone-based area yield scheme can reduce heterogeneous basis risk that leads to determine the suitable alternative for y_c. On the previous research, we have obtained 7 clusters and 2 level of paddy productivity based on clustering assumption from primary data in Java. The suitable clustering assumption for calculating y_c is cluster based assumption, which gives the homogeneous paddy productivity under 7 clusters in Java. Therefore, our goal is to develop area yield index at district level (cluster based) with minimize basis risk at certain constraints for paddy farmer productivity in Java Indonesia. There are some methods for calculating (y_c) such as mean, median, winsor mean, one sigma, two sigma and Q_1 (first quartile) method on the basis risk constraints using confusion matrix. Furthermore, two basis risk constraints are the difference between overpayment and shortfall is not extremely far, and total basis risk does not exceed 20% of its total claim occurrence. Two sigma method has the lowest basis risk, overpayment, and shortfall, but it has lowest pure premium, small probability of claim, and low range of claim. Hence, we consider to use Q_1 (first quartile) method as alternative and suitable method to calculate y_c that satisfied two basis risk constraints. In conclusion, our research provides analytical calculation for area yield index at district level with pure premium as Rp 152,151 using $y_c = 4.67 \frac{ton}{ha}$ (Q_1 method), which is sufficient to cover the total claim and consistent with the simulation.

Keywords: area yield index insurance, basis risk constraints, bootstrap, crop insurance, group risk plan.

* Received: Jan 2021; Reviewed: Mar 2021; Published: Mar 2021

1. Introduction

Agriculture which is often faced by risks is one of the main fields occupation in Indonesia. Despite the uncertainty and changing over time, the geographical conditions such as soil fertility, climate, and natural disasters are one of the most important factors in agriculture especially in terms of crop yields. Therefore in 2012, agriculture insurance has been introduced in Indonesia by Ministry of Agriculture (MoA) to protect farmers from loss of their crops. Furthermore, for agriculture insurance, MoA appointed Jasindo as a state insurance company to conduct indemnity subsidized crop insurance policy, that was indemnity-based-crop-policy or also known as multi-peril crop insurance (MPCI). Crop insurance policy has some disadvantages such as high risk of moral hazard and adverse selection, high administrative cost, and low quality of human resources.

Sutomo et al. (2019) stated group risk plan (GRP) could be the solution or alternative crop insurance policy in Indonesia. GRP has one drawback due to land area in Indonesia which is very heterogeneous. Hence, Sutomo et al., (2019) could not obtain the precise critical yield index (Y_c) per group. As a solution, Haryastuti R. et al., (2020) proposed clustering method to obtain Y_c and zone-based area yield scheme to be the alternative policy that can improve crop insurance in Indonesia.

Haryastuti et al. (2020) stated a basis risk review is needed to determine the most suitable alternative for critical yield index. Therefore, we have conducted an analysis to determine critical yield index for farmers in Java based on basis risk constraint but only for paddy productivity. According to Haryastuti et al., (2020), several methods had been used for calculating critical yield index such as mean, median, winsor mean and two sigma method. The result is two sigma method provides the smallest estimated maximum loss (Haryastuti et al*.,* 2020). In this study, we use the same method for calculating critical yield index with additional new methods such as one sigma and Q_1 (first quartile). All of these methods will be applied based on two assumptions, namely Cluster and Level of productivity based on clustering method (Haryastuti et al., 2020). The basis risk calculation will be carried out for each method and assumption. Hence, we consider the best method and assumption for calculating critical yield index that can be applied in area yield insurance for farmers in Java.

2. Methodology

2.1 Data

In this study, we use primary data collected in the READI Project Farmer Survey 2018-2019 to calculate critical yield index for every method and assumption for farmers in Java island that we obtained from previous study (Sutomo et al*.,* 2019*)*. Primary data contains farmer's productivity (paddy productivity), planted area, *Poktan* namely by farmer group or as *kelompok tani* in Indonesia, clusters, and level of productivity. We have obtained primary data from surveys conducted in several regions in Java as shown in Table 1.

Furthermore, primary data is processed using the bootstrap method to ensure the sample that we have can represent the population. The replication was carried out as much as the number of Poktan multiplied by 25 since the Poktan members ranges from 20-25 farmers (Montgomery, 2007).

Table 1: Number of district and poktan in each cluster and level

Level of Productivity	Cluster	District	Poktan
High	DIY	5	21
	JBR2	31	1061
	JTG1	1	15
	JTM1	9	109
Middle	JBR3	1	4
	JTG2	1	2
	JTM2	9	859
	Total	57	2071

2.2 Area Yield Insurance on District Level

Area yield insurance on district level is insurance policy that has an index (area yield Y_c) as a determinant of whether claims will be accepted (paid) or not, the index will be compared with average yield in each district. Haryastuti et al. (2021) stated that area yield Insurance can be an alternative policy for MPCI that has no specific limitations. (Kusumaningrum et al., 2021) proposed scenario of area yield insurance based on district level (scenario 1) as one of the alternative policies for MPCI. In this study, we will analyze scenario 1: district level to determine the most suitable critical yield index for this scenario from both simulation and analytical calculations. Simulation is based on the calculation of basis risk with several methods and assumption to calculate the critical yield index for scenario 1. Analytical calculation is derived from the claim formula for scenario 1 (Kusumaningrum et al., 2021) to obtain pure premium formula and support the simulation. When the best method and assumption have been found from the simulation result, we can calculate the amount pure premium for scenario 1 based on analytical calculation. We need to compare the amount pure premium from simulation and analytical calculations to ensure there is consistency in both results.

2.3 Basis Risk

In scenario 1, the index will be compared with average yield in each district. However, the performance of farmers in every district will vary depending on geographical factors, farming methods, pests, or diseases. Due to this condition, there may be overpayment of claims to farmers when the yield area is low but individual farmer's productivity is high and there is a possibility that claim (shortfall claim) cannot be made due to high yield area but individual farmer's productivity is low. This type of event is called basis risk, which always appear when we are using index in insurance product.

Basis Risk is the risk that arise when the calculations of the index do not match with the actual policyholder's loss. This will cause imperfect correlation between loss measured and loss experienced by the policyholder. Basis risk cannot be eliminated but we can lower basis risk with determining which method and assumption that is suitable for calculating critical yield index. Therefore, we need to calculate basis risk for every critical yield index from each method and assumption. One method to calculate basis risk is using confusion matrix. Confusion matrix is a table that contains information about actual and predicted classifications done by a classification system and describe performance of classification system (Santra & Christy, 2012). Confusion matrix table and its term are described in Table 2 and the application of confusion matrix for calculating basis risk is described in Table 3.

Table 2: Confusion matrix

		Prediction	
		Yes	No
Actual	Yes	**True Positive**	**False Negative**
	No	**False Positive**	**True Negative**

Table 3: Terms and Condition in basis risk

Terms in Basis Risk	**Terms in Confusion Matrix**	**Condition**
True Covered	True Positive	Average yield (district)< Y_c and individual yield<Y_c
True Not Covered	True Negative	Average yield (district) >Y_c and individual yield>Y_c
Shortfall	False Negative	Average yield (district) >Y_c and individual yield<Y_c
Overpayment	False Positive	Average yield (district) <Y_c and individual yield>Y_c

We will choose the best method and assumption for calculating critical yield index (Y_c) in area yield index insurance at district level (scenario 1), depends on their basis risk performances. As a result, we will obtain the most suitable Y_c for area yield insurance in scenario 1.

2.4 Analysis Procedure

A more detailed explanation of algorithm used to find the most suitable critical yield index (Y_c) in scenario 1 based on basis risk follows:

i. Use bootstrap method to find bootstrap sample for farmer's productivity (y_{ij}) and land area with 100 repetitions and save the average of every repetitions as the result.

ii. Calculate average yield or $\bar{y}_j$ for each district from y_{ij} in bootstrap sample. The function we used can be written as:

iii.

$$\bar{y}_j = \frac{1}{n}\sum_{i=1}^{n} y_{ij}$$

With n denotes the number of farmers in j-th district; and y_{ij} denotes the individual farmer's productivity.

iv. Calculate critical yield index using average, median, average of Winsor, one sigma, two sigma and first quartile (Q_1) under assumption cluster and level of productivity.

a. Based on average

$$y_{ck} = \frac{1}{N_k}\sum_{j=1}^{d}\sum_{i=1}^{f} y_{ij}$$

With k denotes the k^{th} cluster/level of productivity assumption; y_{ck} denotes critical yield index based on cluster/level of productivity assumption; N_k denotes number of farmers based on cluster/level of productivity assumption; d denotes the number of district based on cluster/ level of productivity assumption; f denotes the number of farmer in each district based on cluster/level of productivity assumption.

b. Based on median

$$y_{ck} = \begin{cases} y_{ij\,\left[\frac{N_k+1}{2}\right]}, if\ N_k\ is\ odd \\ \dfrac{y_{ij\,\left[\frac{N_k}{2}\right]} + y_{ij\,\left[\frac{N_{k+2}}{2}\right]}}{2}, if\ N_k\ is\ even\,. \end{cases}$$

With k denotes the k^{th} cluster/level of productivity assumption; y_{ck} denotes critical yield index based on cluster/level of productivity assumption; N_k denotes number of farmers based on cluster/level of productivity assumption; y_{ij} denotes bootstrap sample for individual farmer's productivity based on cluster/level of productivity assumption which have been arranged in order; $\frac{N_k+1}{2}$, $\frac{N_k}{2}$, and $\frac{N_{k+2}}{2}$ denote value in order statistics.

c. Based on winsor mean / winsorized mean

$$y_{ck} = \frac{1}{N_k}\left[\sum_{z=w+1}^{N_k}\left(y_{ij_{(z:N_k)}} + w\, y_{ij_{(w:N_k)}}\right)\right], 1 \leq w < N_k$$

with k denotes the k^{th} cluster/level of productivity assumption; N_k denotes number of farmers based on cluster/level of productivity assumption; y_{ij} denotes bootstrap sample for individual farmer's productivity based on cluster/level of productivity assumption which have been arranged in order. Winsorized mean is a method to calculate mean/average (arithmetic mean) with replacing the smallest and largest values with the observation closest to them in array (Vasanth et al., 2015). The W^{th} winsorized mean refers to the repetition of the W smallest and largest observations with W denotes a value in order statistics.

d. Based on one sigma

$$y_{ck} = \frac{1}{N_k}\sum_{j=1}^{d}\sum_{i=1}^{f} y_{ij} - \sigma_{y_{ij}}$$

with k denotes the k^{th} cluster/level of productivity assumption; N_k denotes number of farmers based on cluster/level of productivity assumption; y_{ij} denotes bootstrap sample for individual farmer's productivity based on cluster/level of productivity assumption; d denotes the number of district based on cluster/ level of productivity assumption; f denotes the number of farmer in each district based on cluster/level of productivity assumption; $\sigma_{y_{ij}}$ denotes the standard deviation for farmer's productivity based on cluster/level of productivity assumption. One sigma allows about 68% of observation lies within one standard deviations of mean.

e. Based on two sigma

$$y_{ck} = \frac{1}{N_k}\sum_{j=1}^{d}\sum_{i=1}^{f} y_{ij} - 2\sigma_{y_{ij}}$$

with k denotes the k^{th} cluster/level of productivity assumption; N_k denotes number of farmers based on cluster/level of productivity assumption; y_{ij} denotes bootstrap sample for individual farmer's productivity based on cluster/level of productivity assumption; d denotes the number of district based on cluster/ level of productivity assumption; f denotes the number of farmer in each district based on cluster/level of productivity assumption; $\sigma_{y_{ij}}$ denotes the standard deviation for farmer's productivity based on cluster/level of productivity assumption. Two sigma allows about 95% of the population lies within two standard deviations of mean, for which data has unimodal symmetrical distribution (Klugman et al., 2012; Montgomery, 2007)

f. Based on first quartile (Q_1)

$$y_{ck} = y_{ij\left[\frac{1}{4}(N_k+1)\right]}$$

with k denotes the k^{th} cluster/level of productivity assumption; y_{ck} is critical yield index based on cluster/level of productivity assumption; y_{ij} denotes bootstrap sample for individual farmer's productivity based on cluster/level of productivity assumption which have been arranged in order; N_k denotes number of farmers based on cluster/level of productivity assumption; y_{ij} denotes bootstrap sample for individual farmer's productivity based on cluster/level of productivity assumption which have been arranged in order; $\frac{1}{4}(N_k + 1)$ denotes a value in order statistics. y_{ck} is calculated using first quartile of all farmer's productivity from any district at certain k^{th} cluster/level.

Repeat (a.), (b.), (c.), (d.), (e.), and (f.) for bootstrap sample based on Cluster which consist of seven clusters (DIY, JBR2, JBR3, JTG1, JTG2, JTM1,JTM2) and Level of Productivity which consist of two level of productivity (Middle and High productivity) that listed on Table 1.

v. Calculate claim amount or indemnity (indemnity paid and actual indemnity) and number of claim (claim occurrence and actual claim) for every assumption using equations:

- Number of Claim at certain k^{th} cluster/level

$$Claim\ Occurrence = max\left(y_{ck} - \overline{y_j}, 0\right)$$
$$Actual\ Claim\ Occurrence = max\left(y_{ck} - y_{ij}, 0\right)$$

we create dummy variables for claim occurrence and actual claim with following criteria:

1) Claim Occurrence: "Yes" means district productivity $\overline{y_j} < y_{ck}$ & "No" means $\overline{y_j} > y_{ck}$

2) Actual Claim Occurrence: "Yes" means farmer's productivity $y_{ij} < y_{ck}$ & "No" means $y_{ij} > y_{ck}$.

- Indemnity (Claim Amount) at certain k^{th} cluster/level

$$Indemnity = max(y_{ck} - \overline{y}_j, 0) \cdot SI \cdot L_{ij}, \ i = 1,2, \dots \dots$$
$$Actual\ Indemnity = max(y_{ck} - y_{ij}, 0) \cdot SI \cdot L_{ij}, i = 1,2, \dots \dots, j = 1,2, \dots \dots$$

with SI denotes sum insured in Rupiah ($\frac{Rp\ 6{,}000{,}000}{4.4\ ton}$), Kusumaningrum et al.(2021) mentioned that value of 4.4 ton per hectare comes from minimum average paddy productivity in Indonesia from 2007 up to 2018; L_{ij} denotes land area for every farmers in each cluster.

vi. Calculate basis risk with confusion matrix comparing claim occurrence and actual claim from farmer side (consist number of claim) and insurance side (consist the amount of indemnity in rupiah).

vii. Determine two constraints to find the most suitable y_c in area yield insurance at district level (scenario 1):

a. The difference between Shortfall and Overpayment is not extremely far. Since we want to consider from both insurance side and farmer side.

b. Total basis risk does not exceed 20% of its total claim occurrence (True Negative + True Positive part in confusion matrix from farmer side), since we need to set boundary how much basis risk that we can tolerate for finding the best critical yield index.

viii. Calculate performance from confusion matrix for every method and assumption with indicators:

a. Confusion matrix from farmer side

- Accuracy rate: shows how accurate the model that we use.

$$Accuracy\ Rate = \frac{(True\ Positive + True\ Negative)}{(TruePositive + TrueNegative + FalseNegative + FalsePositive)}$$

- Shortfall rate: shows the percentage of shortfall.

$$Shortfall\ Rate = \frac{(False\ Negative)}{(TruePositive + TrueNegative + FalseNegative + FalsePositive)}$$

- Overpayment rate: shows the percentage of overpayment.

$$Overpayment\ Rate = \frac{(False\ Positive)}{(TruePositive + TrueNegative + FalseNegative + FalsePositive)}$$

- Basis risk constraint: shows the proportion of total basis risk compared to total claim occurrence

$$Basis\ Risk\ Constraint = \frac{(False\ Negative + False\ Positive)}{(TruePositive + TrueNegative)} \times 100\%$$

b. Confusion matrix from insurance side

- Shortfall (Profit): shows the amount of shortfall in rupiah / profit for insurance company.
- Overpayment (Loss): shows the amount of overpayment in rupiah / loss for insurance company.
- Basis risk: shows the amount of overpayment in rupiah.

$$Basis\ Risk = Overpayment\ +\ Shortfall$$

- Pure premium: shows the amount of pure premium.

$$Pure\ Premium = (True\ Positive\ Rate \cdot True\ Positive) + (Overpayment\ Rate \cdot\ Overpayment)$$

with $True\ Positive\ Rate = \frac{True\ Positive}{(TruePositive+TrueNegative+FalseNegative+FalsePositive)}$

and overpayment rate are from confusion matrix from farmer side; True positive and overpayment are from confusion matrix from insurance side.

- |shortfall - overpayment| shows the difference amount between shortfall and overpayment.
- Total claim, total pure premium, and pure premium sufficiency (for the two best method).

ix. Determine the best assumption with comparing all performance based on assumption.
x. Determine the best method in the best assumption with comparing all performance.
xi. Find the best method and assumption to calculate y_c in scenario 1 from simulation result and find the pure premium for scenario 1.
xii. Find the pure premium formula for analytics calculation derived from claim formula for scenario 1 (Kusumaningrum, 2021).
xiii. Calculate the amount of pure premium for scenario 1 based on the analytical calculation/ claim formula (Kusumaningrum, 2021).
xiv. Compare the result from simulation and analytical calculation to make sure the result is consistent.

3. Result

The results discussed are derived from simulation and analytical calculations. The two results (simulation and analytics) were compared to ensure consistency between the simulation and analytical calculations.

3.1 Simulation Result

Simulation is based on the basis risk calculation/ confusion matrix for every method and assumption. The indicators that describe performance of confusion matrix from farmer side and confusion matrix from insurance side are applied to all the method and assumptions. All the calculation results of the indicators are shown in Table 4.

First, we need to find the best assumption for calculating critical yield index. Table 4 shows level of productivity assumption has better performance compare to cluster assumption. However, Table 4 shows the difference between performance of cluster and level of productivity assumption is not extremely far. Level of productivity in this case only gives a label to the province without clear standard regarding the labeling (high, middle, and low), while cluster is obtained from clustering method (Haryastuti et al., 2021). Moreover, level of productivity fluctuates overtime depending on geographical factors and its condition makes us cannot rely on level of productivity. Therefore, we choose cluster as proper assumption for calculating critical yield index (y_c).

Table 4: Performance of confusion matrix from farmer side and insurance side

Performance	Mean	Median	Winsor Mean	One Sigma	Two Sigma	Q_1
			Cluster			
Accuracy Rate	0.82	0.81	0.81	0.82	0.97	0.86
Shortfall Rate	0.07	0.11	0.12	0.07	0.02	0.02
Overpayment Rate	0.11	0.08	0.07	0.02	0.01	0.12
Pure Premium	Rp354,935	Rp429,848	Rp373,378	Rp560,355	Rp7,851	Rp106,880
Shortfall (Profit)	Rp521,580	Rp795,826	Rp734,817	Rp387,923	Rp156,626	Rp466,085
Overpayment (Loss)	Rp311,678	Rp300,676	Rp213,626	Rp81,421	Rp91,771	Rp197,001
|Shortfall-Overpayment|	Rp209,902	Rp495,150	Rp521,191	Rp306,502	Rp64,855	Rp269,084
Basis Risk	Rp833,259	Rp1,096,502	Rp948,443	Rp469,343	Rp248,397	Rp663,086
Basis Risk Constraint	22% of its Total Claim Occurrence	23% of its Total Claim Occurrence	24% of its Total Claim Occurrence	10% of its Total Claim Occurrence	3% of its Total Claim Occurrence	17% of its Total Claim Occurrence
			Level of Productivity			
Accuracy Rate	0.82	0.82	0.81	0.93	0.98	0.86
Shortfall Rate	0.10	0.07	0.07	0.06	0.00	0.12
Overpayment Rate	0.07	0.10	0.12	0.01	0.02	0.02
Pure Premium	Rp346,560	Rp408,986	Rp363,835	Rp68,697	Rp7,338	Rp118,444
Shortfall (Profit)	Rp552,822	Rp818,064	Rp784,709	Rp496,466	Rp259,798	Rp490,843
Overpayment (Loss)	Rp254,575	Rp259,819	Rp178,864	Rp131,384	Rp89,815	Rp324,034
|Shortfall-Overpayment|	Rp298,248	Rp558,245	Rp605,844	Rp365,082	Rp169,983	Rp166,810
Basis Risk	Rp807,397	Rp1,077,884	Rp963,573	Rp627,851	Rp349,613	Rp814,877
Basis Risk Constraint	21% of its Total Claim Occurrence	22% of its Total Claim Occurrence	24% of its Total Claim Occurrence	8% of its Total Claim Occurrence	3% of its Total Claim Occurrence	16% of its Total Claim Occurrence

**Note: Pure Premium, Shortfall, Overpayment and Basis Risk calculated in average.*

Second, we need to find the best method in cluster assumption for calculating critical yield index. We have two constraints for selecting the best method which is the difference between overpayment and shortfall is not extremely far and total basis risk does not exceed 20% of its total claim occurrence. From the first constraint, we need to calculate the difference between overpayment and shortfall for all methods in cluster assumption. The differences between overpayment and shortfall for all method are shown in Figure 1.

Haryastuti et al. (2021) stated that two sigma method provides the smallest maximum loss. Figure 1 shows two sigma method has the lowest difference between overpayment and shortfall. Table 4 shows two sigma method has the lowest basis risk, overpayment, and shortfall. Although two sigma method provides the smallest basis risk, it has lowest pure premium with extremely low amount of pure premium that listed on Table 4 part pure premium (Rp 7,851). It indicates two sigma method has a low range of claim. Hence, we need to check the number of claims for two sigma method. Number of claims for two sigma are shown in Figure 2.

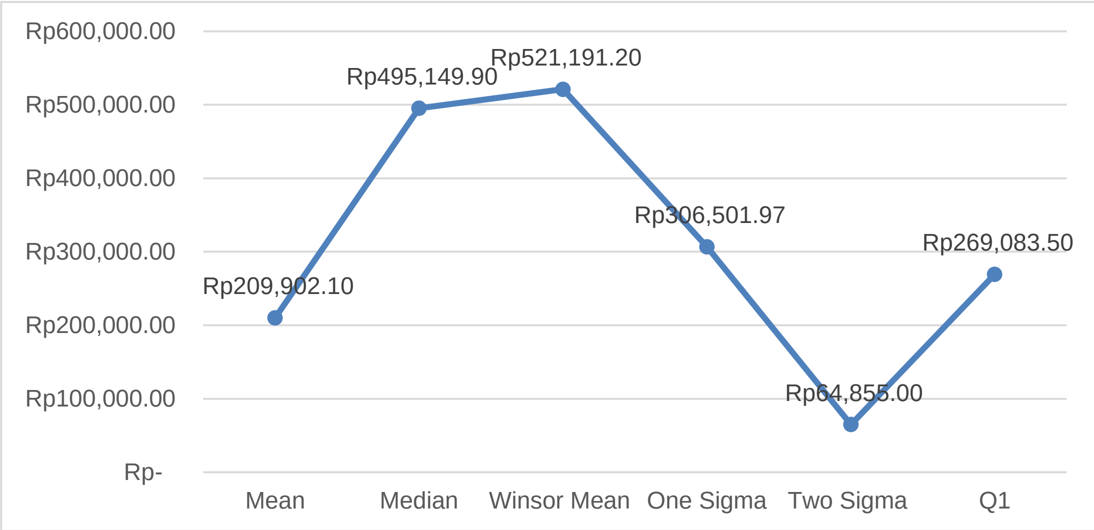

Figure1: Difference between Shortfall and Overpayment in Cluster.

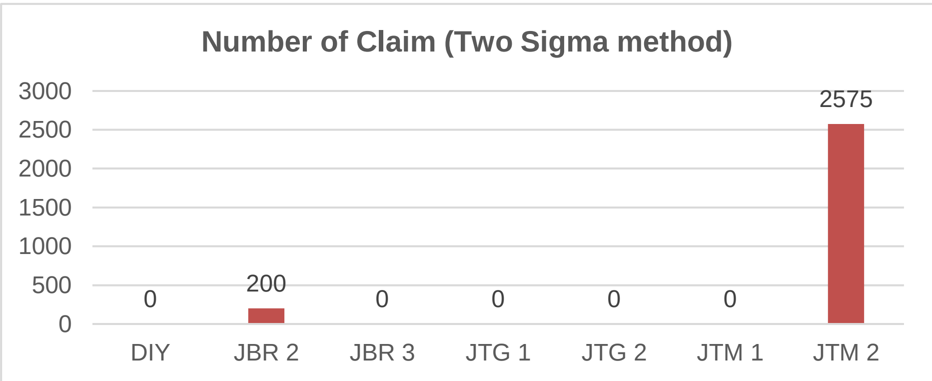

Figure 2: Number of claim for two sigma method in cluster assumption

Figure 2 shows only two clusters from seven cluster that have number of claim greater than zero. It means the probability claim is extremely small (5.36%) and this condition cause the amount of pure premium for two sigma method that listed on Table 4 is extremely low (Rp 7,851). It implies that farmers will be more disadvantage than insurance company due to low range of claim, and extremely small probability claim (5.36%). Since we want to consider both insurance side and farmer side, we cannot choose two sigma as the best method. Figure 1 shows the other methods that have small difference between overpayment and shortfall is mean and Q_1. Therefore, we will compare mean and Q_1 method to find the best method. Performance comparison between mean and Q_1 method are shown in Table 5.

We will focus on the two constraints (basis risk constraint ≤ 20%, difference between shortfall and overpayment is not far), accuracy rate, pure premium, and basis risk for comparing mean and Q_1 method that are shown in Table 5. Table 5 shows Q_1 method has a higher accuracy rate than mean and the lower pure premium, overpayment, shortfall, and basis risk. From Table 5 we know although Q_1 method has the lower pure premium than mean method, the total claim and total pure premium collected from Q_1 method are lower than mean method where pure premium for mean and Q_1 method are sufficient to cover total claim amount. Table 5 shows total basis risk from Q_1 method does not exceed 20% of its total claim occurrence. We can see from Table 5 that mean method has a lower difference between shortfall and overpayment but the differences are quite similar for mean and Q_1 method. From all of the considerations, we choose Q_1 method as the best method to calculate critical yield index y_c based on basis risk constraint and its performance for data sample that we have.

Table 5: Performance of Mean and Q_1 method.

Performance	Mean	Q_1
Accuracy rate	0.82	0.86
Shortfall rate	0.07	0.02
Overpayment rate	0.11	0.12
Pure premium	Rp 354,935	Rp 106,880
Shortfall (Profit)	Rp 521,580	Rp 466,085
Overpayment (Loss)	Rp 311,678	Rp 197,001
|Shortfall - Overpayment|	Rp 209,902	Rp 269,084
Basis Risk	Rp 833,259	Rp 663,086
Basis Risk constraint	22% of its total claim occurrence	17% of its total claim occurrence
Total claim	Rp 18,376,734,464	Rp 5,533,703,067
Total pure premium collected	Rp 18,376,759,625	Rp 5,533,712,000
Pure premium Sufficiency	Sufficient	Sufficient

3.2 Analytical Calculation

Analytical calculation is derived from claim formula for scenario 1 (Kusumaningrum et al., 2021) to obtain the amount and formula for pure premium. In scenario 1, the payment/claim will be evaluated at the district level (average yield for each district) using a critical yield index as the trigger. The claim formula and pure premium calculation of area yield index at district level (scenario 1) are given by:

a. Claim formula (Kusumaningrum et al., 2021)

$$Claim = max\left(y_c - \bar{y}_j, 0\right).SI\ ,\ \ j = 1,2,3 \dots \dots$$

with y_c denotes critical yield index to determine whether a farmer should be compensated or not; j denotes district ; $\bar{y}_j$ denotes average seasonal crop yield at the j-th district; SI denotes sum insured in Rupiah.

b. Pure premium calculation

Pure premium based on expected value of claim:

$$E(Claim) = E\left\{max\left(y_c - \bar{y}_j, 0\right)\right\}.SI \cdot Percentage\ of\ Claim$$

with $\bar{y}_j = \frac{1}{n}\sum_{i=1}^{n} y_{ij}$ and y_{ij} ~ Lognormal (μ, σ^2) because y_{ij} have large number in head portion and small number in tail portion and it is similar to lognormal distribution; Percentage of claim denotes the percentage of claim occurrence in aggregate = $\frac{Number\ of\ Cluster\ that\ Number\ of\ Claim>0}{Total\ Cluster}$, because we use cluster assumption it means we have several y_c.

We want to approximate the distribution for $\bar{y}_j$ using $\left(\prod_{i=1}^{n} y_{ij}\right)^{\frac{1}{n}}$. Under assumption $y_{ij} \in [0,1]$ (premium underrated) implies $\prod_{i=1}^{n} y_{ij} < \sum_{i=1}^{n} y_{ij}$, then from arithmetic-geometric mean inequality we know that $\left(\prod_{i=1}^{n} y_{ij}\right)^{\frac{1}{n}} < \frac{1}{n}\sum_{i=1}^{n} y_{ij}$. Therefore, we can approximate the distribution for $\bar{y}_j \approx \left(\prod_{i=1}^{n} y_{ij}\right)^{\frac{1}{n}} \sim Lognormal(e^{\left(\mu+\frac{1}{2}\frac{\sigma^2}{n}\right)}, e^{\left(2\mu+\frac{\sigma^2}{n}\right)}[e^{\left(\frac{\sigma^2}{n}\right)} - 1])$. Noted that our approximation $\bar{y}_j \approx \left(\prod_{i=1}^{n} y_{ij}\right)^{\frac{1}{n}}$ resulted with underrated pure premium. After that we find the equation for pure premium calculation of area yield index insurance at district level:

$$E(Claim) = Percentage\ of\ Claim \cdot SI \left[y_c\, \Phi\left(\frac{\ln y_c - e^{\left(\mu+\frac{1}{2}\frac{\sigma^2}{n}\right)}]}{\sqrt{e^{\left(2\mu+\frac{\sigma^2}{n}\right)}\left[e^{\left(\frac{\sigma^2}{n}\right)}-1\right]}} \right) - exp\left(e^{\left(\mu+\frac{1}{2}\frac{\sigma^2}{n}\right)} + \frac{1}{2} e^{\left(2\mu+\frac{\sigma^2}{n}\right)}\left[e^{\left(\frac{\sigma^2}{n}\right)} - 1\right]\right) \Phi\left(\frac{\ln y_c - e^{\left(\mu+\frac{1}{2}\frac{\sigma^2}{n}\right)} - e^{\left(2\mu+\frac{\sigma^2}{n}\right)}\left[e^{\left(\frac{\sigma^2}{n}\right)}-1\right]}{\sqrt{e^{\left(2\mu+\frac{\sigma^2}{n}\right)}\left[e^{\left(\frac{\sigma^2}{n}\right)}-1\right]}} \right) \right] \quad ..(1)$$

With n denotes the number of district in each cluster. Based on the equation of pure premium for scenario 1, we can try to calculate the amount of pure premium for scenario 1 analytically. In this calculation, we use bootstrap sample as the data sample (51,775 farmers in total) and Q_1 (first quartile) to determine y_c. We only select one cluster which is JBR 2 cluster for analytic calculation to obtain percentage of claim, y_c, μ, σ^2, and n since JBR 2 has a large amount of data sample based on number of district that is shown in Table 1. Note that parameters (μ and σ^2) we use in analytic calculation include productivity affected by crop failures and disasters. For sum insured (SI), we use ($\frac{Rp\ 6{,}000{,}000}{4.4\ ton}$) (Kusumaningrum et al., 2021). From bootstrap sample we can obtain $y_c = 4.66667\ (Q_1)$, $\mu = 1.65405$, $\sigma^2 = 0.316668398$, $n = 31$ districts, and percentage of claim = 0.26. Therefore, the pure premium for area yield index insurance at district level (scenario 1) analytically is:

$$E(Claim) = Rp152{,}151 \text{ (per each farmer)}$$

$$Total\ of\ Pure\ Premium\ Collected = Rp\ 7{,}877{,}618{,}025$$

Total of pure premium collected denotes total pure premium paid by all farmers or $E(Claim) \times number\ of\ farmers\ in\ bootstrap\ sample$. The amount of pure premium for scenario 1 above is appropriate and in line with percentage of claim, y_c , μ and σ^2 that are pretty low. We will compare pure premium for scenario 1 both from analytical calculation and the simulation.

3.3 Results Comparison (Simulation and Analytical Calculation)

We compare the amount of pure premium in scenario 1 by using analytic calculation and simulation. The amount of pure premium in scenario 1 from analytical calculation is Rp 152,151 with total pure premium collected Rp 7,877,618,025 and from the simulation that listed on Table 5 is Rp 106,880 with total pure premium collected Rp 5,533,712,000. There is small difference on the pure premium amount in scenario 1 between analytical calculation and the simulation under Q_1 method. Since on analytical calculation, we use only 1 cluster and independent farmer assumption, but on the simulation we use more than one cluster and non-independent farmer assumption is used for the computation. According to law of large number it is reasonable the pure premium in scenario 1 from analytical calculation is more expensive than the simulation. Moreover, the amount of pure premiums in scenario 1 for Q_1 method from both simulation and analytical calculation are proven sufficient to cover the total claim. Therefore, it proves analytical calculation for pure premium in scenario 1 is in line and consistent with the simulation and insurance company will still survive and sustain for the following year.

However, we cannot obtain the best critical yield index for every data because it depends on data and business model (the range of claim y_c at cluster level). We calculate critical yield index from bootstrap sample data, it means if we have different sample data to do a bootstrap method then the result will be different. The best critical yield index for insurance company will rely on their data and business model. Therefore, Q_1 is the best method to calculate critical yield index y_c satisfied two basis risk constraints for area yield index insurance at district level.

4. Discussion

Group risk plan (GRP) is proposed to be the alternative for crop insurance policy in Indonesia but due to land area in Indonesia is very heterogeneous, a precise critical yield index cannot be obtained (Sutomo et al*.,* 2019*)*. Clustering method could be a solution but a basis risk review is needed to calculate critical yield index for area yield index (Haryastuti et al*.,* 2021*).* After comparing all of the results from both simulation and analytical calculation, the result shows cluster as the best assumption and Q_1 as the best method to calculate critical yield index (y_c) that still satisfied two basis risk constraints of area yield index insurance in Java. Hence, Q_1 method is the best method to calculate critical yield index for our data sample since to choose the best critical yield index will depend on data and business model chosen by the insurance company. Table 6 shows the value of y_c under Q_1 method among all cluster in Java province.

Table 6: Critical Yield Index / y_c using Q_1 method.

Cluster	Level of Productivity	Number of District	$Y_c^{Q_1}$ Cluster
DIY	High	5	4.64 ton/Ha
JBR2	High	31	4.67 ton/Ha
JBR3	Middle	1	5.33 ton/Ha
JTG1	High	1	3.69 ton/Ha
JTG2	Middle	1	4.33 ton/Ha
JTM1	High	9	3.80 ton/Ha
JTM2	Middle	9	5.55 ton/Ha

Based on basis risk review and our data sample, we suggest Ministry of Agriculture (MoA) would design area yield index insurance based on cluster assumption with different critical yield index (y_c) for each cluster using its first quartile of yield productivity as y_c. However, the cluster and critical yield index in this study can only be applied for provinces in Java. It is suggested to model cluster, to calculate critical yield index for another province in Indonesia, and to analyze scenario 2: two step level with claim formula given by (Kusumaningrum et al., 2020):

$$Claim = max\left(y_c - y_{ij}, 0\right) \cdot SI \cdot 1_{(\overline{y_j} < y_c)}\ ,\ \ i = 1,2,3,4\ldots.,j = 1,2,3\ldots\ldots$$

Thus, in Scenario 2, payment or claim will be evaluated twice at the district level and individual level. It means Scenario 2 will eliminate overpayment / basis risk by itself. Therefore, it is possible that Scenario 2 will generate the lowest basis risk compare to scenario 1.

5. Conclusion

Cluster of productivity assumption was considered to have a better performance compared to level assumption when used for calculating critical yield index that satisfies two basis risk constraints. Level of productivity in this case only gives a label to the province based on the average productivity (high, middle, and low), while cluster is obtained from clustering method based on historical province productivity. Moreover, the level of productivity fluctuates overtime depending on geographical factors and causing level of productivity to switch over time. Thus, in this research, we choose cluster as a more proper assumption for calculating critical yield index (y_c). Furthermore, two sigma methods showed to have the lowest difference between overpayment and shortfall. Two sigma methods also have the lowest basis risk, overpayment, and shortfall. Although two sigma method provides the smallest basis risk, it has an extremely low pure premium (Rp 7,851). Thus, indicating that the two-sigma method has a low unreasonable range of claim. Hence, Q_1 method was considered as the best result to calculate critical yield index that satisfies the two basis risk constraints for our data sample since to choose the best critical yield index will depend on data and business model chosen by the insurance company. Finally, the amount of pure premium in scenario 1 (AYI district cluster) from analytical calculation is Rp152,151. This premium is in line and consistent with the simulation results. Moving forward, insurance companies should adjust the pure premium depending on y_c and SI as critical yield index and sum insured, which varies among clusters and external factors due to weather, pest, disease risks at certain period.

Acknowledgement. First, the authors would like to acknowledge the funding support from the Risk Management, Economic Sustainability and Actuarial Science Development in Indonesia (READI) funded by the Global Affairs Canada and Department of Statistics and Actuarial Science of University of Waterloo as appointed university lead partner. The authors are also very grateful to the Ministry of Agriculture Indonesia and Jasindo for sharing their data and knowledge. The authors also would like to thank Prof Ken Seng Tan who provided valuable feedback and comments.

References

Haryastuti R., Aidi M.N., Pasaribu S.M., Sumertajaya I.M., Sutomo V.A., Kusumaningrum D., Anisa R. (2021). Cluzster Based Area Yield Scheme for Crop Insurance Policy in Java. 6th International Conference on Mathematics: Pure, Applied and Computation (ICoMPAC 2020).

Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2012). *Loss models: from data to decisions* (Vol. 715). John Wiley & Sons.

Kusumaningrum D., Anisa R., Sutomo V.A., Tan K.S. (2021). Alternative Area Yield Index Based Crop Policies in Indonesia. Mathematical and Statistical Methods for Actuarial Sciences and Finance (eMAF2021), Natural Springer

Kusumaningrum D., Anisa R., Sutomo V.A., Robert R. I., Layvieda A., Tan K.S. (2020). Introducing a Wider Range of Area Yield Index Based Crop Insurance Policies in Indonesia. Which is Better? Manuscript in preparation.

Montgomery, D. C. (2007). *Introduction to statistical quality control*. John Wiley & Sons.

Santra, A., & Christy, C. J. (2012). Genetic algorithm and confusion matrix for document clustering. International Journal of Computer Science Issues *(IJCSI)*, *9*(1): 322.

Sutomo, V. A., Kusumaningrum, D., Anisa, R., & Paramita, A. (2019). A Bootstrap Simulation for comparison of Group Risk Plan and Multi-Peril Crop Insurance Policy. *Journal of Physics: Conference Series*, *1366*(1), 012075. IOP Publishing.

Vasanth, K., Manjunath, T., & Raj, S. N. (2015). A decision based unsymmetrical trimmed modified winsorized mean filter for the removal of high density salt and pepper noise in images and videos. *Procedia Computer Science*, *54*: 595–604.