





ARTICLE

Spatial autocorrelation analysis of multivariate rural insurance data

 Walef Machado de Mendonça* and  Patrícia de Siqueira Ramos*

Instituto de Ciências Sociais Aplicadas, Universidade Federal de Alfenas, Varginha-MG, Brazil

*Corresponding author. Email: walefm2@gmail.com; patricia.amos@unifal-mg.edu.br

(Received: October 25, 2022; Revised: March 22, 2023; Accepted: May 02, 2023; Published: December 01, 2023)

Abstract

The environment in which agricultural activities are developed presents high risk and great uncertainty. Several factors related to the agricultural sector can generate fluctuations in the income of producers. These fluctuations must be faced through risk management support policies such as, for example, the hiring of rural insurance. This type of insurance enables the recovery of the financial capacity of the producer in the occurrence of adverse events that cause economic damage. Considering the relevance of rural insurance in the agricultural sector, this study aims to evaluate the spatial distribution of the variables of this type of insurance in Brazilian municipalities from 2006 to 2019. The data used were obtained from rural insurance censuses compiled by the Ministry of Agriculture, Livestock, and Supply. Principal Component Analysis (PCA) was used to reduce data dimensionality and Exploratory Spatial Data Analysis (ESDA) using the scores of the first PC was used to investigate the presence of spatial distribution patterns of rural insurance. By using PC scores, it was found that the highest concentrations of rural insurance policies are located in the South and Midwest regions of Brazil, and there is a tendency for an increase in the spatial dependence of rural insurance throughout the analyzed period. The identification of these areas shows how rural insurance is heterogeneously distributed in Brazil. This result suggests that some strategies can be adopted by policy makers and insurers in order to serve areas that have demand and are not yet covered by rural insurance.

Keywords: Rural insurance. Agricultural policy. Spatial statistics. Spatial autocorrelation. Moran's *I*.

1. Introduction

Agricultural activities are of great relevance in the Brazilian economy. According to a study carried out by the Center for Advanced Studies in Applied Economics, the agribusiness's share in the Brazilian GDP was 20.5% in 2019 and, in 2020, this percentage reached 26.6% (CEPEA, 2021). According to the latest Agricultural Census, between 2006 and 2017, there was an increase of approx-

imately 5.8% in the total area of agricultural establishments, and both the total area and agricultural and livestock production showed considerable growth (IBGE, 2019).

However, the environment in which agricultural activities are carried out presents high risk and relevant uncertainty. This insecurity is mainly due to climatic instabilities and sanitary threats, which can affect production, or market reasons, such as exchange and interest rates, or conditions related to the business environment, such as changes in regulatory frameworks and in public policies. All these variables related to agricultural markets generate variations in the sector's income, which are commonly addressed through policies to support risk management (BRASIL, 2018).

Risk management in agriculture can occur in various ways. However, the purchase of rural insurance is one of the most common forms. This type of insurance works to mitigate losses and enable the producer to recover's financial recovery in the event of losses. Rural insurance provides a more favorable environment for the development of agricultural activities, as it guarantees tincome flow, favors an increase in the planted area and facilitates obtaining financing. In addition, rural insurance is an instrument that enables the sharing of agricultural risk with other economic agents and sectors (BRASIL, 2019).

It is important to highlight that the rural insurance market does not consolidate without the participation of the State. Problems such as high investments and administrative costs, the possibility of catastrophic risks, the strong influence of moral hazard and adverse selection in the formation of portfolios, formation limit the efficiency of private initiatives in offering products. In this sense, the government is called upon to interfere in the market, either by acting directly as an insurer or by creating programs that stimulate the supply and demand for insurance products (ibid.).

Considering the relevance of rural insurance in the agricultural sector, this study aims to analyze the spatial distribution of multivariate data on rural insurance in Brazil municipalities between 2006 and 2019. Furthermore, the study seeks to investigate the presence of spatial distribution patterns in the data, more specifically, whether there is a presence of spatial dependence or heterogeneity. Finally, this work intends to provide information that can contribute to the debate around the improvement of the rural insurance system in Brazil. To achieve these objectives, the work makes use of Principal Component Analysis (PCA) to reduce the dimensionality of the data and Exploratory Spatial Data Analysis (ESDA) to investigate the presence of spatial distribution patterns. The data used come from the Rural Insurance Census, compiled by the Ministry of Agriculture, Livestock and Supply (BRASIL, 2021).

This paper is structured as follows: section 2 describes the methodology used in this study, the data source and the computational resources. Section 3 discusses the results obtained. Finally, section 4 brings the final considerations.

2. Materials and Methods

2.1 Source of data and computational resources used

The data used are annual, referring to rural insurance policies of Brazilian municipalities from the year 2006 to the year 2019. The variables used in the analysis are presented in Table 1. Regarding spatial aggregation, for all variables, the municipal aggregation level was used. In cases where the available information referred to other localities, such as districts, farms and villages, such information was assigned as the corresponding municipalities.

The data on rural insurance come from the website of the Ministry of Agriculture, Livestock and Supply (MAPA). Data containing geographic attributes, such as the position and shape of the Brazilian territory, were obtained from the website of the Brazilian Institute of Geography and Statistics (IBGE, 2021).

Before performing the analysis itself, some descriptive statistics of the variables were obtained for all years to provide a general sense of what the original data looks like.

Table 1. Description of the variables used.

Variable	Acronym
Total of policies contracted	TPC
Sum of insured amount (R\$ million)	SIA
Sum of premiums (R\$ million)	SPR
Total subsidy (R\$ million)	TSB
Sum of indemnities paid (R\$ million)	SIP
Average rate applied to policies	ARP
Number of indemnified policies	NIP

This study was carried out using the programming language *Python* (Van Rossum *et al.*, 2007), using the interface *Jupyter* (Kluyver *et al.*, 2016). In addition, the following libraries were used: *Pandas* (McKinney *et al.*, 2010), for data manipulation, *NumPy* (Van Der Walt *et al.*, 2011), which enables numerical computation with *Python*, *Matplotlib* (Hunter, 2007) and *Seaborn* (Waskom, 2021), which are libraries to create graphics and *Jenkspy* for the use of the Fisher-Jenks algorithm (Jenks, 1977). PCA was performed using the library *Scikit-learn* and the libraries *GeoPandas* (Jordahl, 2014) and *PySAL* (Rey & Anselin, 2010) enabled the spatial analysis.

2.2 Principal component analysis

First, the correlation matrix between the variables was obtained, and then the principal component analysis (PCA) was used to explain the covariance structure of the p variables, $\mathbf{X}^T = [X_1 \ X_2 \ \dots \ X_p]$, through linear combinations of the original variables. The principal components (PCs) obtained are the p linear combinations $\mathbf{Y}^T = [Y_1 \ Y_2 \ \dots \ Y_p]$, not being correlated with each other (Everitt & Hothorn, 2011; Mingoti, 2007).

Furthermore, the PCs are ordered in such a way that the first ones already explain most of the variance of the original variables. Therefore, they can be used to provide a reduction in the dimensionality of the data (Everitt & Hothorn, 2011; Mingoti, 2007). According to Mingoti (2007), the j -th principal component is defined by:

$$Y_j = a_{j1}X_1 + a_{j2}X_2 + \dots + a_{jp}X_p. \quad (1)$$

The vector of coefficients defining the j -th principal component, \mathbf{a}_j , is the eigenvector of the covariance matrix of the data (\mathbf{S}), associated with its j -th largest eigenvalue. The variance of the j -th principal component is given by λ_j , where $\lambda_1, \lambda_2, \dots, \lambda_p$ are the eigenvalues of \mathbf{S} subject to the constraint $\mathbf{a}_j^T \mathbf{a}_j = 1$ (Everitt & Hothorn, 2011).

As PCA was applied to the correlation matrix (\mathbf{R}), the j -th principal component shown in (1) is defined by:

$$Y_j = a_{j1}Z_1 + a_{j2}Z_2 + \dots + a_{jp}Z_p, \quad (2)$$

where Z_i is the standardized variable, that is, subtracted from the mean and divided by the standard deviation. The vector of coefficients defining the j -th principal component, \mathbf{a}_j , is the eigenvector of the data correlation matrix (\mathbf{R}), associated with its j -th largest eigenvalue. The variance of the j -th principal component is given by λ_j , where $\lambda_1, \lambda_2, \dots, \lambda_p$ are the eigenvalues of \mathbf{R} . The proportion of the total variance of Z explained by the j -th principal component is defined by

$$\frac{\text{Var}(Y_j)}{\text{Total variance of } Z} = \frac{\lambda_j}{\text{trace}(\mathbf{R})} = \frac{\lambda_j}{\sum_{i=1}^p \lambda_i}.$$

The scores, that is, the numerical values of the components, are obtained by replacing the values of the standardized variables (Z_i) in each of the components Y_j of (2). The scores of the first principal component (CP1) were used as the variable to be used in the exploratory analysis of spatial data.

2.3 Exploratory Spatial Data Analysis

In this work, the Exploratory Spatial Data Analysis (ESDA) was used with the objective of evaluating the spatial distribution of rural insurance in Brazilian municipalities in the period from 2006 to 2019. The ESDA seeks to capture patterns of spatial association of variables and verify the existence of spatial clusters. In addition, the ESDA seeks to analyze the presence of different spatial regimes or the identification of atypical observations, the spatial outliers (Almeida, 2012).

To perform ESDA, it is necessary to use the matrix of spatial weights in which each entry represents the connection between two municipalities and it is called spatial weight. The spatial weighting matrix, with dimension $n \times n$, is denoted by \mathbf{W} .

The continuity spatial weight matrix with the queen convention was adopted, considering the first-order neighbors. In addition, the municipalities of Fernando de Noronha and Ilhabela, belonging to the states of Pernambuco and São Paulo, respectively, were removed. These municipalities were disregarded because, in addition to being islands, they do not have any rural insurance policy taken out during the years under analysis.

In the binary contiguity weighting matrix used, the municipalities that have physical boundaries in common are considered neighbors and the value 1 is assigned to the corresponding element in the matrix, otherwise it is assigned the value 0. Moreover, by convention, it is assumed that $w_{ij} = 0$, for all $i = j$, that is, a region is not considered a neighbor of itself (ibid.).

Spatial autocorrelation was used in order to measure the relationship of a certain variable in a region with the values of that same variable in neighboring regions. One of the ways to calculate the autocorrelation is using Moran's I . This indicator measures the ratio of the standard deviation of a variable z in an area I to the standard deviation of neighboring areas of the same variable z . Formally, the statistic is expressed by:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} z_i z_j}{\sum_{i=1}^n z_i^2}, \quad (3)$$

where z_i denotes the value of the variable of interest standardized in region i , n represents the number of regions and S_0 is the sum of all weights matrix \mathbf{W} (ibid.). The global Moran's I was calculated and its significance obtained through the pseudo- p , obtained from 999 random permutations. In this study, a significance level of $\alpha = 0.05$ was considered.

The local spatial autocorrelation indicator used to detect local patterns of autocorrelation was the local indicator of spatial association (*LISA*), also called Local Moran's I . While Moran's I provides a global statistic (for the entire distribution), *LISA* provides a local statistic (for each observation), thus allowing to verify if there are spatial clusters statistically significant. For Anselin (1995), a local indicator of spatial association must satisfy two criteria: a) it must indicate spatial clusters statistically

significant; and b) the sum of the local indicators should lead to the global indicator. The local Moran coefficient I can be expressed as

$$I_i = z_i \sum_{j=1}^n w_{ij} z_j.$$

The calculation of I_i is performed considering only the neighboring regions of i , defined through the spatial weight matrix. As the calculation of the I_i is performed for each of the n observations, it generates a large amount of information, a more efficient way to visualize the set of generated statistics is to present them in a cluster (Almeida, 2012).

3. Results and discussion

Table 2 presents the mean, median, standard deviation and maximum values of the rural insurance variables in all the years analyzed. It is possible to observe that the median value of the variables is equal to 0, that is, more than half of the Brazilian municipalities present rural insurance data in the years studied. Furthermore, it can be seen that the averages of the variables referring to the premium and subsidy values present growth during the analyzed period. Analyzing Table 2 one can also see that, with the exception of the variables related to indemnity, the averages of the variables showed a drop between the years 2014 and 2015. The drop in the mean values of the variables in 2015 may be caused by the containment undertaken by the federal government, which released only R\$400 million of the R\$700 million expected to subsidize rural insurance that year (Andrade, 2017).

It can be seen in Table 2 that the average ARP rises between 2006 and 2019. According to Santos & Silva (2017), it was expected that, as the rural insurance system consolidates, policy prices would decrease due to agricultural productivity gains and the reduction of risk factors. These reductions could occur due to the adoption of agricultural zoning guidelines, a greater knowledge of the history of climatic events and claims, as well as the use of technologies such as irrigation, for example. However, it was found that this reduction in rates did not occur during the analyzed period.

Figure 1 presents the values of Pearson's correlations between each of the pairs of rural insurance variables. The lighter the color of the rectangle, the closer to 1 is the value of the correlation coefficient and, therefore, the greater is the degree of positive association between the variables. On the other hand, the darker the color of the rectangle, the lower is the value of the correlation coefficient and, therefore, the lower is the degree of association between the variables. Although the correlation values may be in the range of -1 to 1 , in the case of the data in this work, no value was negative.

When analyzing Figure 1, it is possible to observe that there are groups of variables that present a greater correlation with each other. The variable that represents the total amount of subsidy granted (TSB) presents, in almost all years, a high correlation with variables such as the number of policies contracted (TPC), sum of insured amount (SIA) and the sum of premiums (SPR). In particular, the correlation value of the total amount of subsidy with the sum of the premiums is equal to 0.94 in the year 2006 and, approximately, equal to 1 in the other analyzed years. The positive correlation between the total subsidy granted and the variables related to the total insured and the premium indicates that, as pointed out by Ferreira & da Rocha Ferreira (2009), the government participation, through the subsidy granted by the PSR, is fundamental and must guarantee the sustainability of rural insurance in order to ensure the stability of farmers and other participants in the system.

In addition, there is a high correlation between the variable number of policies contracted (TPC) and total insured (SIA), as well as between the total insured (SIA) and the premium amount (SPR). The variables that showed lowest correlation, although still showing a positive correlation, are the variables that are related to the occurrence of claims, such as the number of indemnities (NIP) and

Table 2. Statistical summary of rural insurance variables. Brazil 2006 – 2019

Year		Variable						
		TPC	SIA	SPR	TSB	SIP	ARP	NIP
2006	mean	3.90	513,139.68	12,425.23	5,546.29	3,717.63	0.01	0.07
	median	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	std	19.87	19,026,436.38	103,890.51	39,040.28	134,395.00	0.08	44958
	max	380.00	1,415,464,375.40	4,154,993.94	1,659,547.01	9,350,283.85	3.72	44.00
2007	mean	5.59	469,420.48	22,213.60	10,772.33	7,452.96	0.07	0.34
	median	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	std	27.85	3,563,113.91	167,293.06	75,687.45	120,700.20	2.56	4.61
	max	693.00	204,837,891.17	7,183,889.39	3,559,961.08	7,148,143.24	190.96	226.00
2008	mean	10.72	1,286,463.25	57,122.13	28,003.68	39,374.12	0.08	1.55
	median	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	std	40.15	5,694,151.72	262,279.75	124,862.29	297,909.03	0.23	12.13
	max	1,030.00	229,181,750.40	10,123,979.60	5,018,884.76	12,331,088.53	45111	469.00
2009	mean	12.38	1,582,182.36	79,811.71	43,981.50	37,197.58	0.11	1.29
	median	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	std	48.40	6,539,617.53	373,034.05	214,089.25	300,949.34	0.33	9.95
	max	1,206.00	151,916,401.52	14,493,978.15	8,691,945.76	15,868,083.32	4.50	417.00
2010	mean	9.53	1,227,298.15	66,583.42	35,860.17	26,245.61	0.10	0.69
	median	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	std	39.80	5,367,333.12	348,127.13	195,933.90	376,059.15	0.34	7.54
	max	1,054.00	123,799,481.96	13,982,021.38	8,340,637.55	19,494,606.91	45051	419.00
2011	mean	10.84	1,370,505.60	85,400.49	46,805.44	81,562.26	0.12	2.43
	median	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	std	46.19	5,785,824.87	462,511.56	267,927.91	638,564.73	0.37	18.57
	max	1,183.00	130,168,446.33	20,276,057.12	12,137,894.94	29,254,409.29	5.26	653.00
2012	mean	11.15	1,533,742.30	99,493.97	56,170.41	37,119.41	0.11	45170
	median	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	std	47.53	6,153,350.71	497,293.14	292,702.26	305,581.63	0.30	7.85
	max	1,275.00	125,894,452.33	19,011,091.57	11,394,874.34	12,383,907.06	4.96	274.00
2013	mean	18.40	3,041,795.25	180,811.06	100,654.96	106,750.72	0.18	2.39
	median	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	std	60.03	10,461,154.95	739,099.44	425,365.88	568,479.26	0.46	45000
	max	1,115.00	188,329,966.51	20,466,874.27	12,262,535.86	21,878,374.82	8.94	624.00
2014	mean	21.10	3,315,890.75	220,444.19	123,610.50	128,776.50	0.23	2.32
	median	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	std	65.92	10,716,152.24	849,241.98	488,032.22	723,734.14	0.56	11.95
	max	1,065.00	192,496,634.61	28,850,484.01	17,171,901.10	23,519,898.83	7.70	374.00
2015	mean	7.15	969,564.55	82,878.79	49,636.62	57,038.15	0.14	1.31
	median	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	std	31.75	4,102,739.52	413,582.75	251,064.07	329,787.73	0.43	8.30
	max	649.00	83,681,426.11	16,710,515.68	9,987,996.89	7,399,237.48	7.37	251.00
2016	mean	13.33	2,301,323.78	162,518.17	69,198.86	83,000.11	0.20	1.32
	median	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	std	50.15	8,364,211.14	721,854.18	294,880.01	633,735.72	0.52	6.38
	max	1,064.00	172,381,269.16	29,156,880.45	10,929,039.13	33,615,986.22	8.42	116.00
2017	mean	11.91	2,152,306.48	152,355.58	65,126.64	33,627.19	0.18	0.80
	median	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	std	44.92	7,693,927.36	692,632.56	297,856.99	221,923.03	0.47	5.15
	max	949.00	179,356,410.10	25,701,862.75	11,553,888.46	6,463,084.06	8.60	237.00
2018	mean	11.33	2,241,032.91	153,339.39	65,859.70	166,125.76	0.20	3.17
	median	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	std	44.15	8,122,254.75	666,529.21	288,098.75	955,950.08	0.55	45157
	max	926.00	180,919,573.67	23,153,925.46	10,380,635.01	37,201,500.70	8.25	777.00
2019	mean	45033	3,605,376.90	227,034.98	78,838.51	60,790.78	0.25	1.63
	median	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	std	59.41	11,867,132.91	833,176.45	288,837.39	277,397.42	0.64	7.36
	max	1,141.00	203,834,786.15	19,587,142.59	6,824,214.77	6,888,474.65	44967	185.00

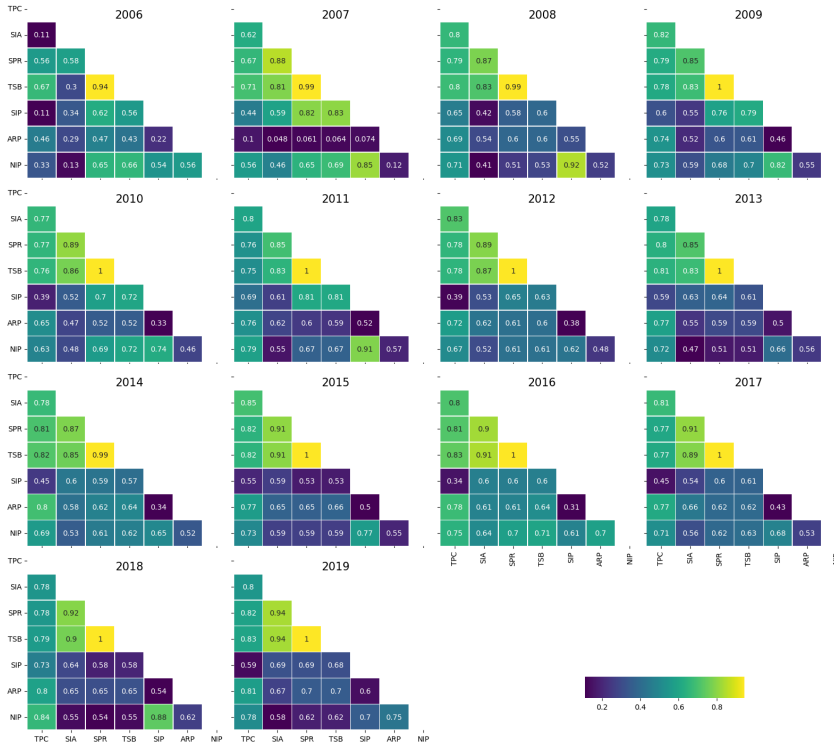


Figure 1. Correlation between rural insurance variables. Brazil 2006 – 2019

total indemnities paid (SIP). These variables present, in all years, lower correlations with the other variables of the data set.

The graph in Figure 2 presents the proportion of total variance explained by the principal components for each year analyzed. In this graph it is possible to observe that the first principal component is responsible for most of the explained variance in all the years analyzed. That is, only the first principal component is able to provide information on the variance of the entire set of rural insurance variables. For example, in 2006, the first PC explained 53% of the total variance of the data, and in 2019, the first PC explained 78% of the total variance.

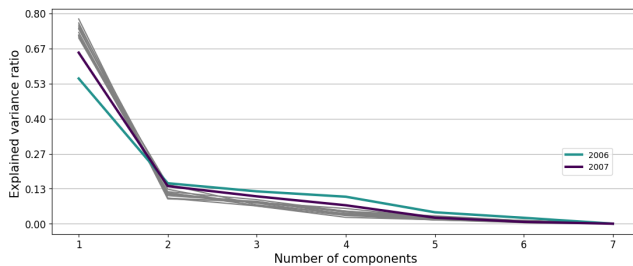


Figure 2. Proportion of total variance explained by the principal components. Brazil 2006 - 2019

It is necessary to emphasize that, in the years 2006 and 2007 (highlighted in Figure 2), the variance explained by the first component was less than 0.72, being equal to 0.55 in 2006 and 0.65 in 2007. However, it is possible to consider that the use of the scores of the first principal component

in the exploratory analysis of spatial data is adequate, since the first component is able to represent, on average, 0.72 or 72% of the variability of the original data (Everitt & Hothorn, 2011).

Figure 3 shows the effect of rural insurance variables on each of the principal components. The lighter the color of the rectangle, the closer to 1 is the coefficient value, therefore, the greater the degree of positive association between the principal components and the variables. On the other hand, the darker the color of the rectangle, the lower the value of the coefficient, therefore, the lower the degree of association between the principal components and the variables.

It is possible to see that, in all years between 2006 and 2019, the first principal component is positively correlated with all variables in the dataset. This makes it easier to interpret the contribution of each variable to the value of the PC. The higher the value of each variable, the higher the value of the PC. The distribution of the effect of the variables in the first component is relatively homogeneous among the variables analyzed, since the values of the coefficients have similar values. Despite SPR and TSB variables present higher weight values, the effects of all the variables in the first component are close to each other.

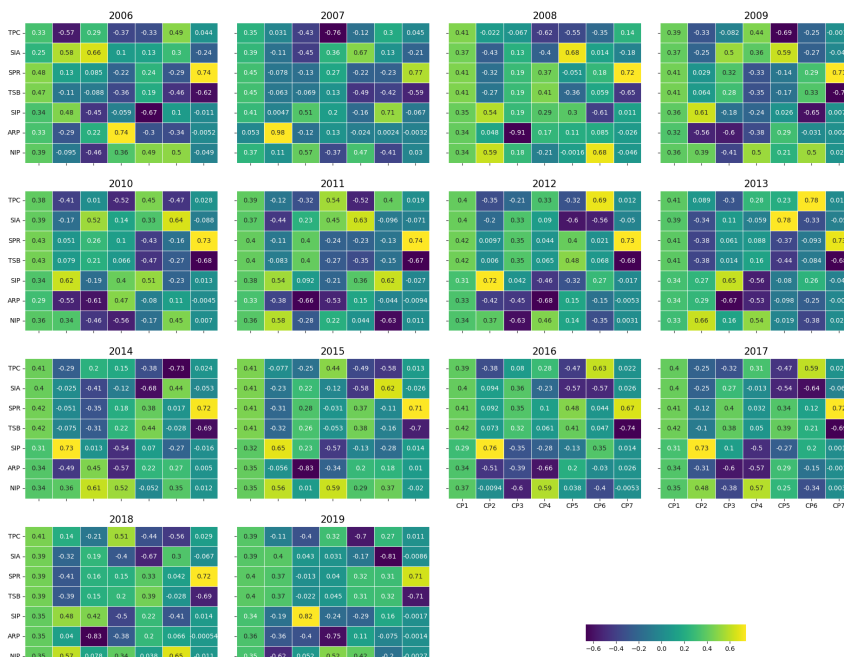


Figure 3. Effect of variables on each component. Brazil 2006 – 2019.

Figure 4 presents the biplots between the first two PCs. In this graph, each point represents a municipality and the axes show the PC scores. All vectors start at the origin and their values show the weight of the variable in the PC under which the vector is projected. Furthermore, the correlation between pairs of variables is represented by the angle between the vectors, in this case, the greater the angle, the lower the correlation between the variables.

By analyzing the Figure 4, it is possible to observe that all variables have a strong positive relationship with the first PC. The variables NIP and SIP, with the exception of 2007, have greater weights related to the second PC, as well as NIP and ARP for most of the years analyzed. In addition, it appears that the SPR and TSB variables are correlated in all years and the SIP and NIP variables are positively correlated. In any case, as shown in the first column of Figure 3, the correlations between all variables and the first PC are expressive.

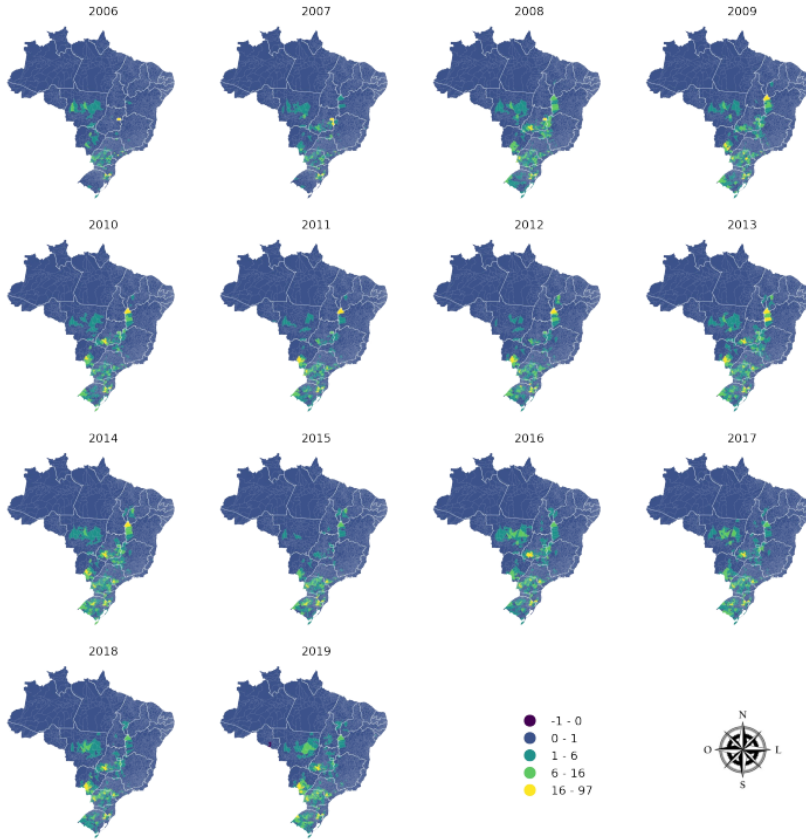


Figure 5. Spatial distribution of the first principal component (PC1). Brazil 2006 – 2019.

The value of Moran's I was calculated using the scores of the first component for all years between 2006 and 2019. The results obtained are shown in Figure 6. When analyzing the graph in Figure 6, it is possible to observe that all values I were positive and significant (the average Moran's I was 0.46 in the analyzed period). This indicates the presence of positive spatial autocorrelation of rural insurance in Brazilian municipalities. That is, the fact that a municipality has above-average values for variables related to rural insurance is influenced, in addition to other factors, by the value of this set of variables in the region neighboring this municipality. Thus, the fact that a municipality has a high value of the sum insured is influenced, among other factors, by the sum of the sum insured in the municipalities in its surroundings. The same applies to the variables sum of premiums, total subsidy, sum of indemnities paid, average rate applied to policies and number of indemnified policies.

In addition to the presence of positive autocorrelation in all years, the analysis of Figure 6 indicates that the spatial dependence grew over the analyzed period. The lowest observed spatial autocorrelation value is equal to 0.216 in 2007 and the highest value occurred in 2019 and was equal to 0.606. Between 2006 and 2019, the value of spatial autocorrelation showed a growth of about 2.62 times the value observed in 2006. Thus, it is possible to conclude that, during the analyzed period, there was an increase in spatial autocorrelation in all the rural insurance variables analyzed.

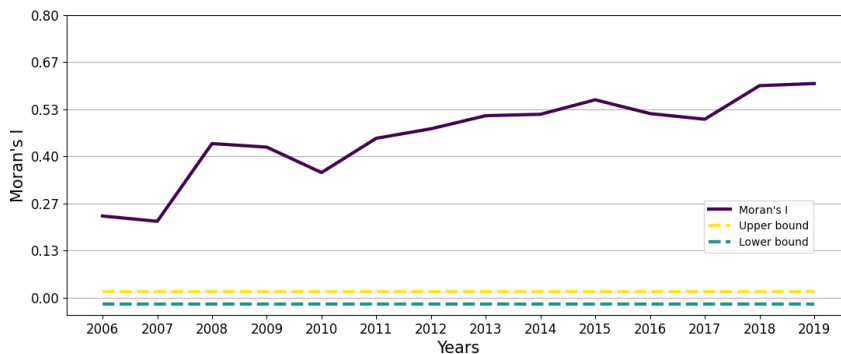


Figure 6. Spatial autocorrelation (Moran's I) Brazil 2006 – 2019

3.1 Local spatial autocorrelation

The *LISA* maps presented in Figure 7 contain four groups with different statistically significant spatial association characteristics for each of the years analyzed. In 2006, HH-type clusters, that is, groups formed by municipalities whose variables presented high values and that had neighbors also with high values of rural insurance variables, were concentrated exclusively in the Midwest, Southeast and South regions. In 2006, the South region had 74.56% of the municipalities in the HH group, 64.32% of which belonged to the state of Paraná, 6.43% to Rio Grande do Sul and 3.8% to Santa Catarina.

In the Southeast region, in 2006, about 50 municipalities belonged to the state of São Paulo, a percentage corresponding to 14.6% of the municipalities in the HH group. In addition, the Midwest region had 10.82% of the municipalities belonging to HH cluster HH. The state of Mato Grosso had 7.6% of the municipalities of HH group and the state of Mato Grosso do Sul had 2.63% of the HH municipalities.

In 2019, the municipalities whose belonging to the HH group were significant, were also concentrated in the Midwest, Southeast and South regions. More specifically, 8.10% of the municipalities in the HH group were located in the state of Mato Grosso do Sul, 5.56% were located in the state of Goiás and 2.78% in Mato Grosso. In the Southeast, significant HH municipalities were located mainly in São Paulo (11.57% of the total) and Minas Gerais had only 0.46% of HH municipalities.

The South Region, in 2019, concentrated about 79.86% of significant HH type municipalities. Most of these municipalities (48.61%) were located in the state of Paraná, 19.67% of the type HH municipalities belonged, in 2019, to the state of Rio Grande do Sul and, finally, 3.24% belonged to Santa Catarina state. In addition, it is worth noting that the years 2007 and 2015, even with positive and significant spatial autocorrelation, had the lowest numbers of municipalities belonging to the BB group, that is, groups formed by municipalities whose variables presented low values and that had neighbors also with low values of rural insurance variables.

According to dos Santos & da Silva (2017), historically, the policies contracted for rural insurance, the insured amount and, consequently, the subsidy amount through the PSR are regionally concentrated. This is due to the presence of higher weather risks in the Southern Region states, in São Paulo and Minas Gerais. States such as Mato Grosso do Sul, Mato Grosso, Goiás, and the region formed by the states of Maranhão, Tocantins, Piauí and Bahia, more recently, have also been adhering to rural insurance contracts due to climatic factors.

In addition, several factors can be pointed out as determinants of the concentration of the rural insurance market in a few regions, with few crops, with reduced participation of insurance companies. According to dos Santos *et al.* (2013), some factors that should be considered are: the small number and discrepancy in the size of insurers that offer this type of insurance; difficulties by bank-

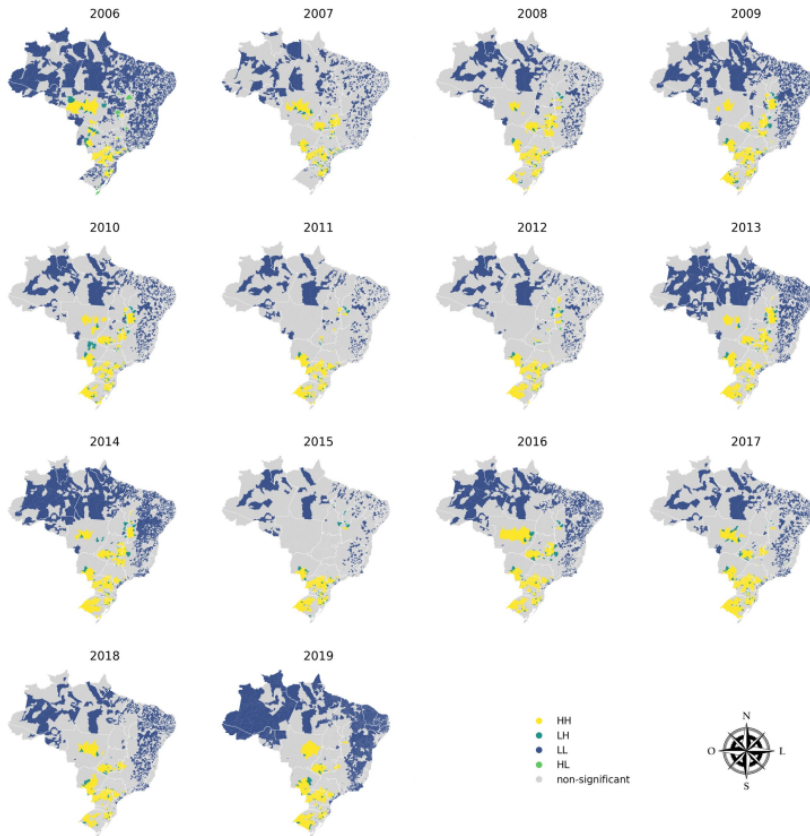


Figure 7. LISA for the first principal component. Brazil 2006 – 2019.

ing institutions with operations in the rural environment, leading to inaccurate information and the elevation of risks and prices; government partnerships with operators, such as official credit programs operated by Banco do Brasil, which is also the controller of the largest agricultural insurance company; the degree of opportunity assessed as small by insurers, and finally; the small weight of partnerships and assessment of opportunities involving the brokerage segment.

4. Conclusions

With the use of PCA, it was possible to verify that one variable was able to represent the set of rural insurance variables efficiently. Instead of using the original seven variables, it was possible to perform the spatial analysis with only one, the first principal component. For almost all the years analyzed, except 2006 and 2007, the first principal component was able to explain more than 70% of the variation in the original data.

The ESDA results indicate that there is spatial dependence in all the years analyzed, that is, there are statistically significant spatial association patterns in the rural insurance data. It was also possible to identify the presence of significant spatial clusters. In general, it was identified that the largest concentrations of rural insurance policies are located in the South, Midwest and Southeast regions, in the southern part of the State of São Paulo. The identification of these areas shows how rural insurance is heterogeneously distributed in Brazil. This result suggests that some strategies can be adopted by policy makers and insurers in order to serve areas that have demand and are not yet

covered by rural insurance.

In addition, it was found that there is an increase in the spatial dependence of rural insurance over the analyzed period. That is, municipalities that have greater adherence to rural insurance tend to be geographically close to municipalities that also have greater adherence to rural insurance. However, the analysis of the *LISA* indicates that the spatial concentration of type HH of rural insurance has changed little over the years.

Since it was found that there is an increase in the spatial dependence of rural insurance over the analyzed period, there is a market to be explored. A cooperation between insurers and policy makers can be established in order to serve regions with lower contracting, especially in the North and Northeast regions of Brazil. Insurers can use this information to identify areas with high demand and low supply and focus their marketing efforts to serve them.

A limitation of the study was the lack of incorporation of the temporal dimension in the analysis, thus some aspects of the dynamics of the variables involved may not have been captured, which could be addressed in future work. Another suggestion for future work would be to identify producing areas in Brazil that still do not contract rural insurance and cross-reference the results of this research. Thus, it would be possible to identify potential regions to be served by this type of insurance that do not yet do so.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. We would like to thank reviewer and editor for their valuable comments that improved the paper.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

1. Almeida, E. Econometria espacial. *Campinas-SP. Alínea* (2012).
2. Andrade, L. Os Limites e Alternativas do Seguro Rural no Brasil. *Globo Rural*. <https://globorural.globo.com/limites-e-alternativas-do-seguro-rural-no-brasil.html> (2017).
3. Anselin, L. Local indicators of spatial association—LISA. *Geographical analysis* 27, 93–115 (1995).
4. BRASIL. *Programa de Subvenção ao Prêmio do Seguro Rural: Relatório de Resultado 2017* tech. rep. (Brasília, DF: Ministério da Agricultura, Pecuária e Abastecimento (MAPA), 2018).
5. BRASIL. *Programa de Subvenção ao Prêmio do Seguro Rural: Relatório de Resultado 2020* (Brasília, DF: Ministério da Agricultura, Pecuária e Abastecimento (MAPA), 2021).
6. BRASIL. *Relatório de Resultado 2018* (Brasília, DF: Ministério da Agricultura, Pecuária e Abastecimento (MAPA), 2019).
7. CEPEA. *PIB do Agronegócio Brasileiro 2020* tech. rep. (Centro de Estudos Avançados em Economia Aplicada (CEPEA), Piracicaba, SP, 2021).
8. Dos Santos, G. R. & da Silva, F. C. *Dez anos do programa de subvenção ao prêmio de seguro agrícola: proposta de índice técnico para análise do gasto público e ampliação do seguro* tech. rep. (Texto para Discussão, 2017).
9. Dos Santos, G. R., de Sousa, A. G. & Alvarenga, G. *Seguro agrícola no Brasil eo desenvolvimento do programa de subvenção ao prêmio* tech. rep. (Texto para Discussão, 2013).

10. Everitt, B. & Hothorn, T. *An introduction to applied multivariate analysis with R* (Springer Science & Business Media, 2011).
11. Ferreira, A. L. C. J. & da Rocha Ferreira, L. Experiências internacionais de seguro rural: as novas perspectivas de política agrícola para o Brasil. *Econômica* **11** (2009).
12. Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in science & engineering* **9**, 90–95 (2007).
13. IBGE. *Censo Agropecuário 2017* Rio de Janeiro, RJ: IBGE. 2019.
14. IBGE. *Portal de mapas* Disponível em: <https://portaldemapas.ibge.gov.br/>. Acesso em: 04 mar. 2021. Rio de Janeiro, 2021.
15. Jenks, G. F. Optimal data classification for choropleth maps. *Department of Geography, University of Kansas Occasional Paper* (1977).
16. Jordahl, K. *GeoPandas: Python tools for geographic data* <https://github.com/geopandas/geopandas>. 2014.
17. Kluyver, T. *et al. Jupyter Notebooks—a publishing format for reproducible computational workflows*. (2016).
18. McKinney, W. *et al. Data structures for statistical computing in python* in *Proceedings of the 9th Python in Science Conference* **445** (2010), 51–56.
19. Mingoti, S. A. in *Análise de dados através de métodos estatística multivariada: uma abordagem aplicada* 295–295 (2007).
20. Rey, S. J. & Anselin, L. in *Handbook of applied spatial analysis* 175–193 (Springer, 2010).
21. Santos, G. R. & Silva, F. C. *Dez anos do Programa de Subvenção ao Prêmio de Seguro Agrícola: proposta de índice técnico para análise do gasto público e ampliação do seguro* tech. rep. (Ipea, Rio de Janeiro, 2017). http://repositorio.ipea.gov.br/bitstream/11058/7718/1/td_2290.pdf (2021).
22. Van Der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy array: a structure for efficient numerical computation. *Computing in science & engineering* **13**, 22–30 (2011).
23. Van Rossum, G. *et al. Python Programming Language*. in *USENIX annual technical conference* **41** (2007), 1–36.
24. Waskom, M. L. Seaborn: statistical data visualization. *Journal of Open Source Software* **6**, 3021 (2021).