PME Publication no. ICARNEH-ML-EPUB-2022-43

Spatial Dependency

*N. Uttam Singh, ***Kishore K. Das, *Kamni P Biam, *Pampi Paul, *C. Gowda H R, *Debasish Chakraborty, **Puran Chandra & *Aniruddha Roy
* ICAR Research Complex for NEH Region, Umiam, Meghalaya
**NBPGR, New Delhi
***Gauhati University, Guwahati
uttamba@gmail.com

Correct Citation: N. Uttam Singh, Kishore K. Das, Kamni P Biam, Pampi Paul, C. Gowda H R, Debasish Chakraborty, Puran Chandra & Aniruddha Roy , 2021. Spatial Dependency. ICAR Research Complex for NEH Region, Umiam, Meghalaya.

Introduction

Spatial dependence occurs when an observation recorded at one point in space is dependent on an observation(s) at another point(s) in space (Cassandra et al. 2000). Spatial autocorrelation statistic measure and analyze the degree of dependency among observations in a geographic space. Observations made at different locations may not be independent. Measurements made at nearby locations may be closer in value than measurements made at locations farther apart. This phenomenon is called spatial autocorrelation. Positive spatial autocorrelation occurs when similar values occur near one another. Negative spatial autocorrelation occurs when dissimilar values occur near one another. According to Ciccone 1996, the aggregate level of technology in each country may not only rely on externalities originated by capital accumulation within the country, but also on the aggregate level of technology of its neighbors.

Similarly, Spatial effects are important in explaining agricultural growth. States can interact strongly with each other through channels such as trade, technological diffusion, capital inflows, and common political, economic and social policies. Present study focuses only on detecting the spatial relationship among the states. Spatial autocorrelation is the correlation between the values of a single variable that is strictly due to the proximity of these values in geographical space by introducing a deviation from the assumption of independent observations of classical statistics (Griffith 2003). The most common spatial autocorrelation indicators in the literature are: the Moran's *I* and the Geary's *c*. These tests test the null hypothesis of no spatial dependence against the alternative hypothesis of spatial dependence.

Measures of Spatial Autocorrelation

Weight Matrix: To assess spatial autocorrelation, one first needs to define what is meant by two observations being close together, i.e., a distance measure must be determined. These distances are presented in weight matrix, which defines the relationships between locations where measurements were made. If data are collected at n locations, then the weight matrix will be *nxn* with zeroes on the diagonal.
The weight matrix can be specified in many ways:

- The weight for any two different locations is a constant.
- All observations within a specified distance have a fixed weight.
- K nearest neighbors have a fixed weight, and all others are zero.
- Weight is proportional to inverse distance, inverse distance squared, or inverse distance up to a specified distance.

Other weight matrices are possible. The weight matrix is often row-standardized, i.e., all the weights in a row sum to one. Note that the actual values in the weight matrix are up to the researcher.

Moran's *I*

Moran's *I* is one of the oldest statistics used to examine spatial autocorrelation. This global statistic was first proposed by Moran 1948 & 1950. Later, Cliff and Ord 1973 & 1981 present a comprehensive work on spatial autocorrelation and suggested a formula to calculate the *I* which is now used in most textbooks and software. Moran's *I* (Moran 1950) tests for global spatial autocorrelation for continuous data. It is based on cross-products of the deviations from the mean and is calculated for *n* observations on a variable *x* at locations *i, j* as:

$$I = \frac{n}{S_0}\frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

where $\bar{x}$ is the mean of the x variable, w_{ij} are the elements of the weight matrix, and S_0 is the sum of the elements of the weight matrix $S_0 = \sum_i \sum_j w_{ij}$.

Moran's *I* is similar but not equivalent to a correlation coefficient. It varies from -1 to +1. In the absence of autocorrelation and regardless of the specified weight matrix, the expectation of Moran's *I* statistic is *-1/(n-1)*, which tends to zero as the sample size increases. For a row-standardized spatial weight matrix, the normalizing fact or S_0 equals *n* (since each row sums to 1), and the statistic simplifies to a ratio of a spatial cross product to a variance. A Moran's *I* coefficient larger than *-1/(n-1)* indicates positive spatial autocorrelation, and a Moran's *I* less than *-1/(n-1)* indicates negative spatial autocorrelation. The variance is:

$$Var(I) = \frac{n\{(n^2 - 3n + 3)S_1 - nS_2 + 3S_0^2\} - k\{n(n-1)S_1 - 2nS_2 + 6S_0^2\}}{(n-1)(n-2)(n-3)S_0^2} - \frac{1}{(n-1)^2}$$

where $S_1 = \frac{1}{2}\sum\sum_{i\neq j}(W_{ij} + W_{ij})^2 = 2S_0\, for\ symmetric\ W\ containing\ 0's\ and\ 1's$

$$S_2 = \sum_i (W_{i0} + W_{0i})^2\ where\ W_{i0} = \sum_j W_{ij}\ and\ W_{0i} = \sum_j W_{ji}$$

Geary's *c*

Geary's *c* statistic (Geary 1954) is based on the deviations in responses of each observation with one another:

$$c = \frac{n-1}{2S_0}\frac{\sum_i \sum_j W_{ij}(x_i - x_j)^2}{\sum_i (x_i - \bar{x})^2}$$

Geary's *c* ranges from 0 (maximal positive autocorrelation) to a positive value for high negative autocorrelation. Its expectation is 1 in the absence of autocorrelation and regardless of the specified weight matrix (Sokal and Oden 1978a & 1978b). If the value of Geary's *c* is less than 1, it indicates positive spatial autocorrelation. The variance is:

$$Var\ (c) = \frac{1}{n(n-2)(n-3)S_0^2}\Big\{S_0^2[(n^2-3) - k(n-1)^2] + S_1(n-1)[n^2 - 3n + 3 - k(n-1)] + \frac{1}{4}S_2(n-1)[k(n^2 - n + 2) - (n^2 + 3n - 6]\Big\}$$

where S_0, S_1 and S_2 are the same as in Moran's *I*.

Comparison of Moran's I and Geary's C

Moran's *I* is a more global measurement and sensitive to extreme values of *x*, whereas Geary's *c* is more sensitive to differences in small neighborhoods. In general, Moran's *I* and Geary's *c* result in similar conclusions. However, Moran's *I* is preferred in most cases since Cliff and Ord 1973 & 1981 have shown that Moran's *I* is consistently more powerful than Geary's *c*. These tests test the null hypothesis of no spatial dependence against the alternative hypothesis of spatial dependence.

Determination of spatial dependency

Foodgrain-yield of eight states of North Eastern Region (NER) of India over the years from 1967 to 2017 has been analyzed for spatial dependency among the states. There are 408 observations in total. Each state has 51 observations for each corresponding 51 years. Out of 408 observations 19 were unobserved - first four observations in Meghalaya and Mizoram; and first 15 observation in Sikkim. The location of each observation is expressed in the corresponding latitude and longitude coordinates.

Autocorrelation analysis features of PROC VARIOGRAM of SAS 9.3 was used to compute Moran's *I* and Geary's *c* statistic. It reports the two-sided p-values for Moran's *I* and Geary's *c* coefficients under the null hypothesis that the sample values are not autocorrelated. That is the probability that the observed coefficient lies farther away from $|Z|$ on either side of the coefficient's expected value, that is, lower than - *Z* or higher than + *Z*. Smaller p-values correspond to stronger autocorrelation for both coefficients. However, the p-values do not tell us whether the autocorrelation is positive or negative. Positive autocorrelation is indicated when the *Z* score for Moran's *I* is positive ($Z_I>0$) or the *Z* score for Geary's c is negative ($Z_c<0$). Negative autocorrelation is indicated when $Z_I<0$ or $Z_c>0$. The resulting autocorrelation statistic containing Moran's I and Geary's c coefficients with distance weight is shown in table 1. Based on the small *p*-values of the reported Moran's I and Geary's c coefficients, we reject the null hypothesis of zero spatial autocorrelation in the values of foodgrain-yield. The sign of *Z* for both Moran's *I* and Geary's *c* coefficients indicates positive autocorrelation in the data values.

The distance matrix used in the above calculations is a 385×385 matrix where each off-diagonal entry (i,j) in the matrix is equal to $1/(1+h)$, where h is the distance between point i and point j. Further Moran's *I* and Geary's *c* coefficients were also estimated with binary weight by assigning a threshold distance of 1 such that pairs with distances less than 1 are considered *connected* or *close* and pairs with distances greater than 1 are not.

From table 1 it is also seen that the change in distance measure does not change the interpretation from the previous discussion with distance weight. Based on the small *p*-value of the reported Moran's *I* and Geary's *c* coefficients under binary weight, we reject the null hypothesis of zero spatial autocorrelation in the values of foodgrain-yield. The *Z* scores again indicate a positive autocorrelation.

Table 1: Autocorrelation statistic of Moran's *I* and Geary's *c* for spatial values of foodgrain-yield in the states of NER.

Distance measure	Coefficient	Observed	Expected	Std Dev	Z	Pr > |Z|
Distance Weight	Moran's I	0.053	-0.003	0.002	24.030	<.0001
	Geary's c	0.966	1.000	0.010	-3.410	0.001
Binary Weight	Moran's I	0.094	-0.003	0.007	13.380	<.0001
	Geary's c	0.908	1.000	0.028	-3.270	0.001

From table 2 it is seen that all the pair-distances are accommodated within lag 9 with a lag distance of 0.67. Highest number of pairs are found in lag 3 (26.25%) and lowest number of pairs are found in lag 7 (2.84%). The same is depicted in figure 1.

Table 2: Pairwise distance intervals for spatial analysis of values of foodgrain-yield in the states of NER.

Lag Class	Bounds		Number of pairs	Percentage of pairs
0	0	0.33	0	0
1	0.33	1	5202	0.0803
2	1	1.67	9996	0.1544
3	1.67	2.33	16999	0.2625
4	2.33	3	7395	0.1142
5	3	3.67	7395	0.1142
6	3.67	4.34	6894	0.1065
7	4.34	5	1836	0.0284
8	5	5.67	5364	0.0828
9	5.67	6.34	3672	0.0567
10	6.34	7	0	0

Figure 1: Pairwise distance intervals for spatial analysis of values of foodgrain-yield.

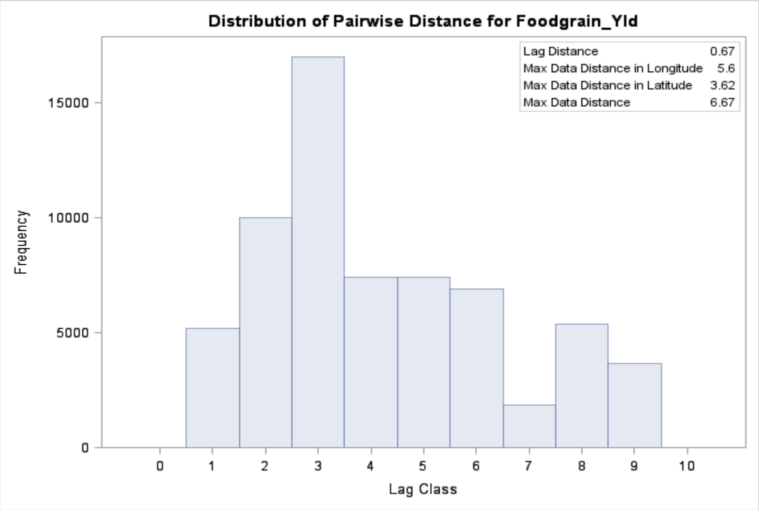
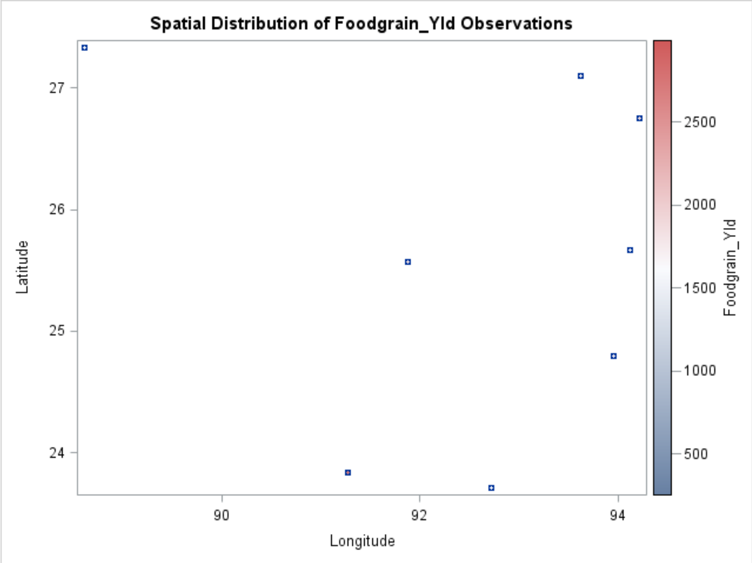


Figure 2: Spatial distribution of foodgrain-yield of states of NER based on centroid coordinates.



Conclusion

Foodgrain yield of neighboring states had closer values than that of the non-neighboring states. The detected spatial dependency was based on the centroid geographical coordinates of the states through replication of 51 time periods. However, this part of analysis in the study was just to detect whether spatial dependency in the values of foodgrain-yield of the north eastern states exist or not and from the above discussion, spatial dependency in the region was witnessed for the values of foodgrain-yield. Further analysis based on this spatial dependency can be carried out incorporating spatial panel data analysis as future scope of this study. A plot of spatial distribution of foodgrain yield is given in figure 2.

References

- Cassandra Di Rienzo, Paul Fackler and Barry K Goodwin, Modeling Spatial Dependence and Spatial Heterogeneity in Country Yield Forecasting Model, AAEA Conference, Tampa, Session 8, August 1 (2000).
- Ciccone Antonio, *Externalities and the Interdependent Growth: Theory and Evedence* (University of Califonia at Berkeley and Univeritat Pompeu Fabra, Social Science Electronic Publishing, Social Science Research Network, 1996).
- Cliff, A. D. and Ord, J. K., *Spatial Processes - Models and Applications* (London: Pion, 1981).
- Cliff, A.D. and Ord, J.K., *Spatial Autocorrelation* (London: Pion, 1973).
- Geary, R., The Contiguity Ratio and Statistical Mapping, The Incorporated Statistician **5**, 115-45 (1954).
- Griffith, D.A., *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization* (Berlin: Springer-Verlag, 2003).
- Moran, P.A.P., Notes on Continuous Stochastic Phenomena, Biometrika **37**, 17-23 (1950).
- Moran, P.A.P., The Interpretation of Statistical Maps, Journal of the Royal Statistics Society, Series B (Methodological) **10(2)**, 243 - 251 (1948).
- Sokal, R. R. and Oden, N. L., Spatial Autocorrelation in Biology 1. Methodology, Biological Journal of Linnean Society **10**, 199-228 (1978a).
- Sokal, R. R. and Oden, N. L., Spatial Autocorrelation in Biology 2. Some Biological Implications and Four Applications of Evolutionary and Ecological Interest, Biological Journal of Linnean Society **10**, 229-249 (1978b).

Acknowledgement

Published under the project "Task Force for Sustainable Hill Agriculture-National Mission for Sustaining the Himalayan Ecosystem (Phase-II)", PIMS Code: 0XX5387. The authors express their gratitude to Department of Science and Technology (DST), Climate Change Programme (CCP), New Delhi for funding the project.