چهارمین سمینار آمار فضایی و کاربردهای آن
۵–۶ آبان ماه ۱۴۰۰
27-28 Oct. 2021
4th Seminar on Spatial Statistics and Its Applications

Spatial Autocorrelation of Smuggling Crimes in Iran Provinces Using Weight Matrices

Mohadeseh Alsadat Farzammehr*
Iranian Judiciary Research Institute, Tehran, Iran

Abstract:

Statistical autocorrelation describes spatial connections between observations/regions. As an application of the theory, this paper examines the smuggling of goods and foreign exchange in 31 provinces of Iran between 2011 and 2017. Coordinates are specified between spatial units using spatial weights matrices, sometimes referred to as $\boldsymbol{W}$. The local structure and influences of the spatial relationships are described by these matrices. The choice of a weight function for the spatial weights matrix plays a significant role in the results of the spatial analysis. Here, we compare the statistical performances of different types of spatial weighting matrices based on Queen contiguity (first-order and second-order neighbor weights). The first-order contiguous Queen with the neighbors of 148 shows a significant correlation in goods smuggling and foreign exchange.

Keywords: Spatial autocorrelation, Spatial weight matrix, Crimes of goods and foreign exchange smuggling.

Mathematics Subject Classification (2010): 60Gxx, 60Hxx, 60Bxx.

---

1 Introduction

In spatial analysis, regions and spatial relationships are directly integrated into their mathematics, which is defined as analytic techniques that are associated with analyzing geographical phenomenon locations along with their spatial dimensions and features Esri (2001). Crimes involving goods smuggling and smuggling of foreign exchange are social

*Speaker: ma.farzammehr@gmail.com

phenomena that are constantly shaped by cultural and geographical details. Interdependence of measured variables is the main reason for spatial analysis, as expressed in Tobler's first law of geography: near things are more related to one another than far things Tobler (1970).

The events at a particular location are highly influenced by the events at it's neighboring locations in many spatial data applications. In order to measure spatial dependence, spatial autocorrelation is used. The spatial autocorrelation can be defined as the existence of a functional relationship between what happens at one point and what happens at another Anaselin (1988).
Each spatial unit is correlated in a unique manner using spatial weights matrices, sometimes referred to as $\boldsymbol{W}$. In evaluating spatial autocorrelation, it is necessary to take into account the spatial interactions between items and their values. Weights are expressed in an $n \times n$ matrix $\boldsymbol{W}$ where each element w_{ij} is a spatial weight. $\boldsymbol{W}$ is a non-stochastic symmetric weight matrix which generates the spatial neighborhood with zeros in the main diagonals and ones in the neighboring diagonals. By assuming a positive constant as a limit, we assume the matrices are bounded by some constant away from zero. In order to limit dependence, matrix $\boldsymbol{W}$ can also be normalized so that each row is equal to one.
The study by Griffith (1995) has shown that it will be preferable to specify the relationship between observations parsimoniously instead of assuming something about distance decay. A detailed description incorporating assumptions about, for instance, distance decay is preferable to a sparse description of the observations' relationships Wang et al. (2013). It is discussed in this paper how two functions can be used to compute the weight for a contiguity spatial autocorrelation, using the Queen contiguity by first-order neighbour and the Queen contiguity by second-order neighbour functions. As defined by Queen contiguousness a neighbour is a pair of polygons that have a common boundary or vertex. Additionally to determining the direction of contiguity, we can also determine how far a neighbouring polygon is from a given polygon. This is known as the contiguity order. First-order neighbour weights are discussed as a form of spatial weight. Higher-order neighbour regression, based on spatial neighbour relationships across locations, was also employed to construct a spatial weight matrix of order n. Comparatively, second-order contiguity counts the polygons containing the polygon in question (first-order contiguous polygons), as well as their contiguous neighbors.
Selecting a weight matrix for spatial autocorrelation is one of the most important steps in it's estimation Anselin and Griffith (1998). The result and effect of spatial analysis will be unsatisfactory if the weight function is not selected properly, or even the calculation will be distorted Chen (2009). It is possible to store $n \times n$ spatial relationships in the spatial weights matrix file, however, in most cases, each feature should only be related to

a handful of others. Sparse methodology takes advantage of this by storing only non-zero relationships. The spatial relationship values are utilized in the mathematics of several spatial statistics tools including spatial sutocorrelation (global Moran's I and Getis-Ord general G), hot spot analysis (Getis-Ord G_i^*), and cluster and outlier analysis (Getis-Ord G_i and Anselin local Moran's I). In essence, local indicators of spatial autocorrelation (LISA) allow the decomposition of global indicators, such as Moran's I, into the contribution of each individual observation. This class of indicators can have two interpretations: the assessment of significant local spatial clustering around an individual location, similar to the interpretation of the G_i and G_i^* statistics of Getis and Ord (1992); and the indication of pockets of spatial non-stationarity, or the suggestion of outliers or spatial regimes, similar to the use of the Moran scatterplot of Anselin (1996).

In this paper, we present three methods of spatial correlation: Global Moran's I, Getis-Ord General G, and LISA and we examine the impact of these different weight matrices on analytical choices. Following this, a Moran scatterplot is used to visualize conflict between Iran provinces between 2011 and 2017.

1.1 Spatial Autocorrelation

Global statistics tools, such as global Moran's I and the high/low clustering (Getis-Ord general G) tool, can assess the overall pattern and trend of a dataset. Typically, they are most effective when mapped across an area with a consistent spatial pattern. As a measure of global clustering, Moran's I will be used. An analysis of the null hypothesis is also performed here. The null hypothesis implies that the values of the variable are randomly distributed in space, meaning that we cannot predict the values of neighboring observations based on the value of the center location. A negative value of I and with the null hypothesis of no spatial autocorrelation rejected indicates a negative spatial autocorrelation meaning the clustering of dissimilar values among neighboring observations, while a positive value of I and with the null hypothesis of no spatial autocorrelation rejected indicates positive spatial autocorrelation. The null hypothesis is rejected if $|Z| > Z_{\alpha/2}$, where $Z = (I - E(I))/\sqrt{(Var(I))}$ and $Z_{\alpha/2}$ derived from standard normal distribution.

The General G statistic for overall spatial association is the second method that is commonly used when calculating spatial autocorrelation. In this case the only difference between the numerator and denominator is the weighting (w_{ij}). High/Low clustering will only work with positive values. Consequently, if your weights are binary (0/1) or are always less than 1, the range for general G will be between 0 and 1. A high/low clustering (Getis-Ord general G) tool returns four values: observed general G, expected general G, $z - score$, and p-value, with the null hypothesis that there is no spatial autocorrelation among locations. The null hypothesis is rejected if $|Z| > Z_{\alpha/2}$, where

$Z = (G - E(G))/\sqrt{(}Var(G))$.

Identification of local patterns of spatial association is an important aspect of this study. Using local indicators of spatial autocorrelation or LISA tools (like hot spot analysis), one can assess each feature within the context of neighboring features and compare the local situation to the global situation. The purpose of LISA was to decompose global autocorrelation into the contribution of individual observations Anselin (1996). By calculating local Moran's I by I_i, the LISA statistic evaluates spatial autocorrelation or clustering in the individual units. When an indicator of global spatial association is significant, it may be more appropriate to use a Moran scatterplot than interpretation of locations as hot spots to identify outliers and leverage points. Hence, it is important to note that Getis-Ord G_i have been suggested to detect significant spatial clustering at a local level when global statistics do not provide evidence of association Getis and Ord (1992).

2 Results and Discussion

The results presented in this paper describe characteristics of the smuggling of goods and foreign exchange for Iran's provinces over the years, with an average from 2011 to 2017. Creating the spatial weights matrix involved analyzing autocorrelation test results in three methods (Moran's I, Getis's Ord G, and LISA) that were applied to two types of spatial weights (neighbourhood weights and distance weights). The data set is used in a data frame and in shapefiles (.shp). The data frame contains the main variable in row and column to be tested, and the shapefile contains the geospatial data formats used to get the neighboring and coordinate data. In this data, for instance, Kurdistan province on average had 239.67 accused subject input ratio, and Khorasan Razavi had 28.78 accused subjects with the one hundred thousand population, they are ranked highest and lowest in the country, for the years 2011-2017. Note that, functions described here are available as part of a package named *spdep* which is written directly in *R*. The figures

```
Neighbour list object:
Number of regions: 31
Number of nonzero links: 148
Percentage nonzero weights: 15.40062
Average number of links: 4.774194
Weights style: W
Weights constants summary:
   n  nn S0        S1       S2
W 31 961 31 13.74246 127.7108
```

```
Neighbour list object:
Number of regions: 31
Number of nonzero links: 382
Percentage nonzero weights: 39.75026
Average number of links: 12.32258
Weights style: W
Weights constants summary:
   n  nn S0       S1       S2
W 31 961 31 5.281461 127.2631
```

Figure 1: Characteristics of weights list object

in reference 1 illustrate the Queen contigency based on the first-order and second-order neighbor weights with 148 and 382 neighbors, respectively. Connectivity graphs display a line from each point to its nearest neighbor. At the moment, we're using polygons based on the 31 regions here (provinces).In order to make our connectivity graphs, we need to get points. Typically, polygon centroids are used for this. These connectivity graphs illustrate the way the sets of contiguous neighbors of each zone are constructed based on distance and contiguity (See Figure 2). The province with the highest number of Queen

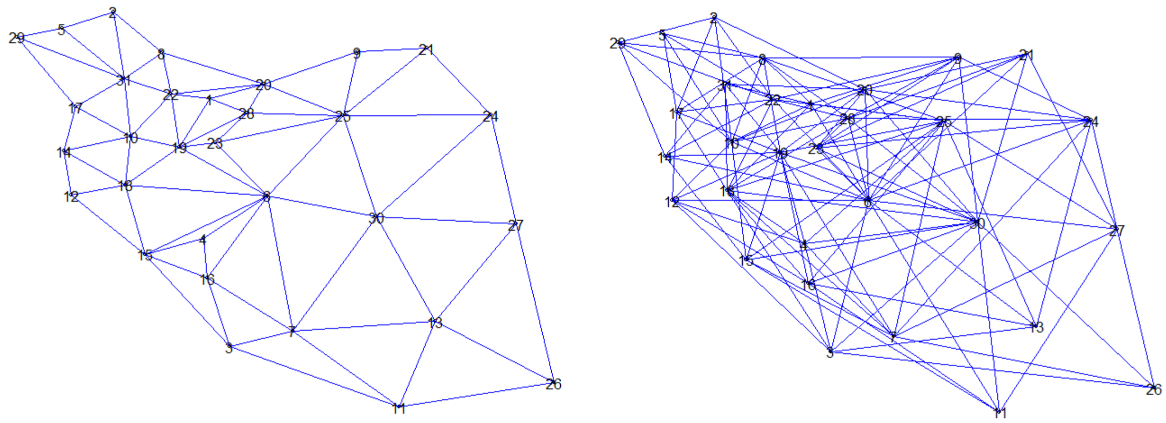

Figure 2: Connectivity graphs

contiguity based on first-order (second-order) neighbor weights is Esfahan with nine (23) neighbors, which indicates that the incidence of smuggling crimes in this location is or has been influenced by nine (23) neighbors.

Note that, Queen contiguity based on first-order (and second-order) neighbor weights produces different neighborhoods. In 31 regions, there were 148 (with 382) neighbouring regions. Through weighting Queen contiguity first-order, Golestan has only three neighboring list (Mazandaran, North Khorasan and Semnan). While, using Queen contiguity second-order, Ardabil has 8 neighboring list (Alborz, Isfahan, Gilan, Mazandaran, North Khorasan, Qazvin, Qom, Razavi Khorasan, Semnan, Tehran, Yazd).

According to first-order weighting of Queen contiguities, Golestan only has three neighboring lists (Mazandaran, North Khorasan, and Semnan). By examining the correlation between weighted Queen contiguity first-order and Getis-Ord general function, it was shown that there is a significant spatial autocorrelation. This is indicated by the value $|Z| > 1.96$ or by $p - value < 0.05$.

The existence of spatial autocorrelation suggests that there was a global dependence in smuggling crimes among Iran's provinces. The pattern of spatial autocorrelation in Moran's I statistic indicates that the crimes of goods and foreign exchange smuggling are spatially clustered (auto correlated) and the G statistic indicates there are clusters of high or low values. Figure 3 shows the Moran scatter plot. As the slope of the regression

Table 1: Spatial autocorrelation test (Global Moran's I and Getis-Ord General G tests).

Index Expectation and Z	Queen contiguity first-order	Queen contiguity second-order
Global Moran's I		
I	0.36059894	0.202480796
E[I]	-0.03333333	-0.033333333
—Z—	3.660823	4.222477
p-value	0.0001257*	1.208e-05*
Getis-Ord General G		
G	3.728891e-02	3.499129e-02
E[G]	3.333333e-02	3.333333e-02
—Z—	2.017113	1.146879
p-value	0.02184*	0.1257

Note: *) Significance at $\alpha = 5\%$.

line corresponds to the value of the Moran I, the Moran scatter plot can be used to visualize the type and strength of spatial autocorrelation. On the $x-$axis of the scatter plot is displayed the standardized value of the variable while on the $y-$axis is displayed the standardized spatial lag of the same variable. This scatter plot is sectioned in four quadrants each indicating four types of spatial association: Quadrant I is high values of observatian surrounded by high values (High-High), Quadrant II is low values of observatian surrounded by high values (Low-High), Quadrant III is low values of observatian surrounded by low values (Low-Low), and Quadrant IV is high values of observatian surrounded by low values (High-Low).

As example the crimes of the goods and currency smuggling in overall average of 2011-2017, there are 12 provinces located in the High-High group (Bushehr, Chaharmahal and Bakhtiari, Fars, Hormozgan, Ilam, Kermanshah, Khuzestan, Kohgiluyeh and Boyer-Ahmad, Kohkiluyeh and Boyer-Ahmad, Kurdistan, Lorestan, Sistan and Baluchestan, West Azerbaijan). The locations in this group has a high goods and currency smuggling which surrounded by locations that have high goods and currency smuggling crimes. The second group is Low-Low group which contains 14 sub districts (Alborz, Ardabil, Gilan, Golestan, Markazi, Mazandaran, North Khorasan, Qazvin, Qom, Razavi Khorasan, Semnan, South Khorasan, Tehran, Yazd). The location in this group has a lower amount of goods and currency smuggling crimes surrounded by areas with lower amounts of goods and currency smuggling crimes. Low-High group consists of five provinces (East Azerbaijan, Isfahan, Hamadan, Kerman, Zanjan). This group of locations has high goods and currency smuggling crimes and is surrounded by areas with lower crimes. Note that there are no High-Low provinces plotted because they are not significant. Additionally, local spatial clusters, sometimes referred to as hot spots, can be defined as contiguous locations or sets of contiguous locations in which the LISA is significant. Table 2 contains the final LISA paper results.

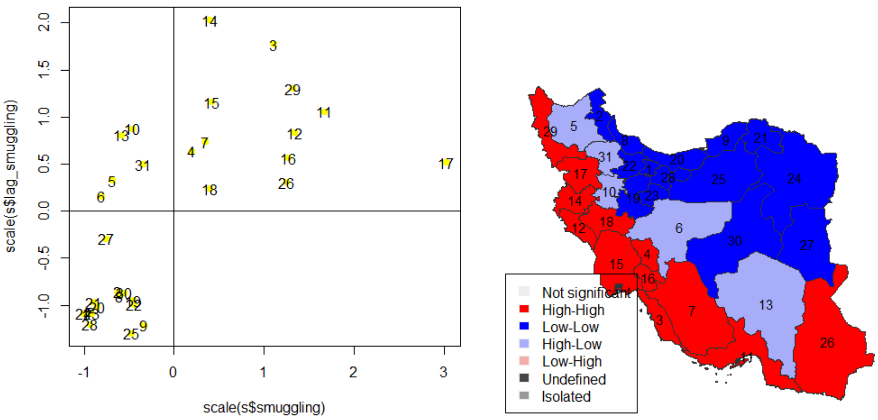

Figure 3: Moran scatter plot and mapping for the crimes of goods and currency smuggling.

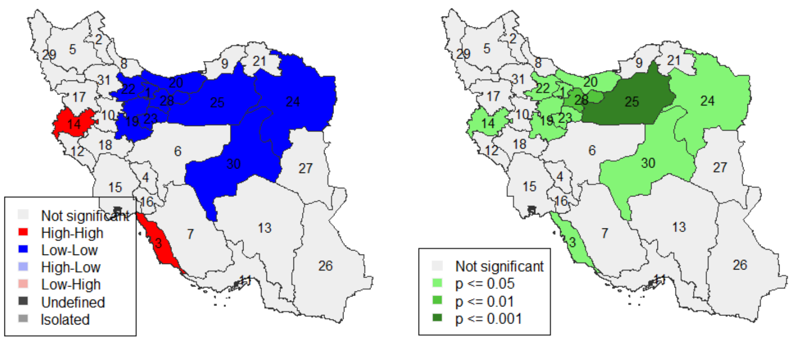

Figure 4: Map of hot-spots / cold-spots and p-value for the crimes of goods and foreign exchange smuggling.

Table 2: Measures of local indicators spatial association.

cluster analysis	pattern analysis	spatial autocorrelation	provinces	hot-spots	cold-spots
High-High	cluster	positive	(3, 4, 7, 11, 12, 14, 15, 16, 17, 18, 26, 29)	(3, 14)	-
Low-High	no-cluster	negative	(5, 6, 10, 13, 31)	-	-
High-Low	no-cluster	negative	-	-	-
Low-Low	cluster	positive	(1, 2, 8, 9, 19, 20, 21, 22, 23, 24, 25, 27, 28, 30)	-	(1, 19, 20, 22, 23, 24, 25, 30)

Conclusion

Spatial weights, $\boldsymbol{W}$, have element w_{ij} at i , j for n locations. It describes the type of structure among neighbouring or adjacent locations and displays how a particular location affects another or vice versa. There are several types of weighting based on the shape of the area and its spatial structure. Different types of weighting will result in different spatial autocorrelations. The study computes 148 neighbour lists based on the first-order Queen contiguity method and shows how spatial autocorrelation can affect smuggling of goods and forex. Queen contiguity based on second-order neighbour weights gives different results. It compute 382 neighbour list and shows that there is no significance spatial autocorrelation on the method of Getis-Ord G statistic.

References

Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Springer Science & Business Media,11.

Anselin, L. (1996), The Moran scatterplot as an exploratory spatial data analysis tool to assess local instability in spatial association,*In M. M. Fischer, H. J. Scholten and D. Unwin (eds) Spatial analytical perspectives on GIS*, (London: Taylor & Francis), 111125.

Anselin, L. and Griffith, D. A. (1988), Do spatial effects really matter in regression analysis?, **65(1)**, 1134.

Chen, Y. G. (2009), Reconstructing the mathematical process of spatial autocorrelation based on the Moran statistic, *Geographical Research*, **28 (6)**, 1449-1463.

Esri, A. (2001), *AdvancedGIS spatial analysis using raster and vector data*, NewYork: Environmental System Research Institute.

Getis, A., and Ord K. (1992), The analysis of spatial association by use of distance statistics, *Geographical analysis*, **24**, 189-206.

Griffith, D. A. (1995), Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. *Practical handbook of spatial statistics*, 6582.

Waldo Tobler, A. (1970), Computer movie simulating urban growth in the Detroit region, *Economic geography*, **46 (2)**, 234-240.

Wang, Z., Liang, H. Y., and Zhou, Q. (2013), Model for Effect of Spatial Weighted Matrix on Spatial Autocorrelation, *International Journal of Applied Mathematics and Statistics*, **44 (14)**, 431445.