STATISTICS IN TRANSITION new series, March 2025
Vol. 26, No. 1, pp. 1–8, https://doi.org/10.59139/stattrans-2025-001
Received – 10.07.2023; accepted – 25.11.2024

Generalised spatial autocorrelation coefficients

Janusz L. Wywiał[1]

Abstract

The article focuses on properties generalised to the multidimensional case of known coefficients of spatial correlation. The main result of the work is the decomposition of the introduced generalised autocorrelation coefficients into the sum of ordinary autocorrelation coefficients, but calculated on the basis of the principal components of the originally observed multidimensional variable. The development is illustrated with an empirical example. The coefficients provide a more detailed description of the spatial relationships of a set of variables defined in a population.

Key words: Moran coefficient, Geary coefficient, spatial autocorrelation, Mahalanobis distance, principal components.

1. Introduction

Exploration of various phenomena in natural, social, economic and other populations requires an approach involving the analysis of relationships among observations of many features defined in these populations. This applies to populations in which a distance between pairs of its members is defined. This can lead to the division of the population into a set of homogeneous subpopulations. This was the inspiration for the preparation of this work. The known Moran (1950) and Geary (1954) spatial autocorrelation coefficients described below allow for the analysis of the spatial similarity in terms of single variables. The properties of autocorrelation coefficients were considered by, among others, Getis and Ord (1992), Griffith and Chun (2022). Recently, in the works of Krzyśko et al. (2023), Krzyśko et al. (2024), the autocorrelation coefficients were significantly generalised to the multivariate case. These generalizations use advanced functional analysis to simultaneously analyze the spatio-temporal autocorrelation of time-varying vector observations. Du (2012) generalised the Geary coefficient to a random vector. It could also be adapted to the Moran coefficient.

Let $x_i, i = 1, ..., N$ be observations of x variable. Moran (1950) defined the coefficient of spatial autocorrelation in the following way:

$$I^M = \frac{1}{wv} \sum_{i=1}^{N} \sum_{j=1}^{N} (x_i - \bar{x})(x_j - \bar{x}) w_{ij} = \frac{1}{v} \sum_{i=1}^{N} \sum_{j=1}^{N} (x_i - \bar{x})(x_j - \bar{x}) q_{ij} \tag{1}$$

[1]Deaprtment of Statistics, Econometrics and Mathematics, University of Economics in Katowice, Katowice, Poland. E-mail: wywial@ue.katowice.pl. ORCID:https://orcid.org/0000-0002-3392-1688.

where $w_{ij} \geq 0$, $w_{ii} = 0$, $w = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}$, $v = \sum_{i=1}^{N} (x_i - \bar{x})^2/N$, $\bar{x} = \sum_{i=1}^{N} x_i/N$, $q_{ij} = w_{ij}/w$, $0 \leq q_{ij} \leq 1$. When neighbors are more similar (more different) than observations in general, then Moran's coefficient coefficient takes positive (negative) values. Values of this coefficient close to zero indicates absence of spatial similarity. Usually, $-1 \leq I^M \leq 1$, see Cliff and Ord (1981) or Overmars et al. (2003). However, this is not always the case. The range of coefficient variability may take into account such distribution features of the examined variable, as kurtosis or skewness.

Geary (1954) proposed the following coefficient:

$$I^G = \frac{N-1}{2Nvw} \sum_{i=1}^{N} \sum_{j=1}^{N} (x_i - x_j)^2 w_{ij} = \frac{N-1}{2Nv} \sum_{i=1}^{N} \sum_{j=1}^{N} (x_i - x_j)^2 q_{ij} \geq 0. \tag{2}$$

The value of the Geary coefficient greater (smaller) than one means large differences (similarity) of neighboring objects. The value of this coefficient close to one means the lack of substantial spatial autocorrelation in the sense described above.

Weight w_{ij}, $i \neq j = 1,...,N$ can be defined in several ways. For instance, the weights may indicate the economic relationship between sub-areas. They may, for example, indicate cooperative connections between economic regions, characterised by observations of a multidimensional variable. In particular, these connections may be financial flows between these companies. In this case, e.g. well-known input-output matrix of Leontief (1986) could be used to construct the weights. Getis and Ord (1992) suggested to set that $w_{ij} = 1$, when $|x_i - x_j| \geq d_0$ and $w_{ij} = 0$ in otherwise case, $i \neq j = 1,...,N$. For example, the constant d_0 could define the minimum flow of funds from one region to another or the maximum distance (in km) between them.

2. Generalization and decomposition of spatial autocorrelation coefficients

Let x_{it} be the i-th observation of the t-th variable, $i = 1,...,N$, $t = 1,...,k$. These data are elements of $X = [x_{it}]$ matrix of dimension $N \times k$, $k \leq N$, $X = [x_{*1}...x_{*t}...x_{*k}]$, where x_{*t} is the t-th column of X, $x_{*t}^T = [x_{1t}...x_{it}...x_{Nt}]$. The i-th row of X is denoted by $x_{i*} = [x_{i1}...x_{it}...x_{ik}]$. In particular, for $k = 1$, $X = [x_{11}\, x_{21}...x_{N1}]^T = [x_1\, x_2...x_N]^T$. The variance-covariance matrix is denoted by $V = [v_{jt}]$ where $v_{jt} = \frac{1}{N} \sum_{i=1}^{N} (x_{ij} - \bar{x}_j)(x_{it} - \bar{x}_t)$, $\bar{x}_t = \frac{1}{N} \sum_{i=1}^{N} x_{it}, t, j = 1,...,h$. We assume that V is nonsingular.

Du et al. (2012)) proposed the following generalization of Geary's coefficient:

$$I_k^G = \frac{N-1}{2kNw} \sum_{i=1}^{N} \sum_{j=1}^{N} (x_{i*} - x_{j*}) V^{-1} (x_{i*} - x_{j*})^T w_{ij} = \frac{N-1}{2Nk} \sum_{i=1}^{N} \sum_{j=1}^{N} d_{ij} q_{ij}. \tag{3}$$

where q_{ij} is explained below the expression (1) and

$$d_{ij} = (x_{i*} - x_{j*}) V^{-1} (x_{i*} - x_{j*})^T \tag{4}$$

is the Mahalanobis distance between x_i* and x_j*. Values of I_k^G close to unity indicate lack of

similarity or differences of neighboring objects due to the multidimensional variable value vectors observed in them than in the case of all the objects (not necessary neighbours).

When $I_k^G > 1$, there is a tendency that neighboring objects are more dissimilar from each other in terms of Mahanalobis distance than in the case of $I_k^G < 1$. For instance, taking into account the aforementioned suggestion of Getis and Ord (1992) we can assume that $w_{ij} = 1$ if $d_{ij} \geq d_0$ and $w_{ij} = 0$ in otherwise case, $i \neq j = 1,...,N$, $d_0 > 0$.

Let us generalize Moran's coefficient for the case when $k \geq 1$ as follows:

$$I_k^M = \frac{1}{wk} \sum_{i=1}^{N} \sum_{j=1}^{N} (x_{i*} - \bar{x}) V^{-1} (x_{j*} - \bar{x})^T w_{ij} = \frac{1}{k} \sum_{i=1}^{N} \sum_{j=1}^{N} b_{ij} q_{ij} \tag{5}$$

where

$$b_{ij} = (x_{i*} - \bar{x}) V^{-1} (x_{j*} - \bar{x})^T. \tag{6}$$

Positive values of I_k^M coefficient indicate that the observations of the vectors of the multivariate variable are similar in terms of the direction of their deviation from the vector of means. If the observation vectors of variables in neighboring objects deviate from the average vector in different directions, then we can expect that the autocorrelation coefficient is negative. Values of the autocorrelation coefficient close to zero indicate lack of similarity or dissimilarity of neighboring objects due to the multivariate variable. Just like it was in the case of I^G we can assume that $w_{ij} = 1$ if $b_{ij} \geq b_0$ and $w_{ij} = 0$ in otherwise case, $b_0 > 0$, $i \neq j = 1,...,N$.

In order to decompose the coefficients let us assume that C is such orthogonal matrix that $C^T C = U_k$ and $C^T V C = \lambda$ where U_k is $k \times k$ identity matrix, $\lambda = [\lambda_t]$ is the diagonal matrix consisting of the eigenvalues of V denoted by $\lambda_t \geq 0$, $t = 1,...,k$, see, e.g. Harville (1997) or Morrison (1976). Note that $V c_t = \lambda_t U_k$ where c_t is the t-th column of C, $c_t^T = [c_{1t}...c_{kt}]$ and it is the t-th eigenvector of V. Observations of the t-th principal component are determined by $z_t = X c_t$. The components of the vector $\lambda_t c_t$ are covariances between the t-th principal component z_t and the entire variables represented by the columns of X. The correlation coefficient between the t-th principal component and observations of the i-th original variable represented by the column x_{*i} is as follows:

$$r(z_t, x_{*i}) = c_{it} \sqrt{\frac{\lambda_t}{v_i}}, \quad i = 1,...,k. \tag{7}$$

In Appendix we show that the generalised Moran coefficient could be decomposed as follows:

$$I_k^M = \frac{1}{k} \sum_{t=1}^{k} I_{k,t}^M \tag{8}$$

where

$$I_{k,t}^M = \frac{1}{\lambda_t} \sum_{i=1}^{N} \sum_{j=1}^{N} (z_{it} - \bar{z}_t)(z_{jt} - \bar{z}_t) q_{ij} \tag{9}$$

is the ordinary Moran spatial autocorrelation coefficient calculated based for the t-th principal component of X. Hence, I_k^M is the average of the Moran autocorrelation coefficients

calculated for the principal components. If this average is equal to zero, the coefficients for the principal components may be non-zero – they happen to cancel each other out.

Similarly to (8) we derive (see Appendix) the following decomposition of Geary's coefficient:

$$I_k^G = \frac{1}{k}\sum_{t=1}^{k} I_{k,t}^G \tag{10}$$

where

$$I_{k,t}^G = \frac{N-1}{2N\lambda_t}\sum_{i=1}^{N}\sum_{j=1}^{N}(z_{it}-z_{jt})^2 q_{ij} \tag{11}$$

is the ordinary Geary's spatial autocorrelation coefficient calculated based on the t-th principal component of X. Thus, I_k^G is average of Geary's autocorrelation coefficients calculated for the principal components.

Example

We illustrate the generalised autocorrelation coefficient with an example of the following variables defined for Polish voivodships: revenues from total economic activity (x_1), sold production of industry (x_2), capital expenditures per capita (x_3), gross value of fixed assets per capita (x_4), average monthly gross salaries (x_5). Data are available at: https://bdl.stat.gov.pl/bdl/start. Variables have been scaled to have the value of each variable divided by the value assigned to the capital voivodship.

The values of the ordinary Moran autocorrelation coefficient (see expression (1)) for the listed variables $x_1,...,x_5$ are as follows: -0.2242, -0.3408, -0.2328, 0.2478, -0.2227. So, all Moran's coefficients are negative except x_4. The values of the ordinary Geary autocorrelation (given by expression (5)) for these variables are as follows: 2.7022, 2.7408, 2.4931, 1.8085, 2.5498.

Moran's and Geary's generalised coefficients take the following values -0.0332 and 1.0039, respectively. Thus, both coefficients would indicate that the spatial autocorrelation for all variables is very weak.

Now, let us consider the decomposition of the generalised coefficients. The eigenvalues (variances of principal components) of the considered $x_1,...,x_5$ are: 0.1633, 0.0318, 0.0082, 0.0060, 0.0018. The shares of these eigenvalues in their sum are as follows (%): 77.3, 15.1 3.9. 2.9, 0.8. The first two principal components explain 92.4% of the overall variation of $x_1,...,x_5$. Thus, the first two principal components explain almost all of the variability of $x_1,...,x_5$. So, the other three principal components can be ignored.

The Moran coefficient for the successive principal components are as follows: -0.2811, 0.3227, -0.0494, -0.7024 and -0.0856. The Geary coefficient for the successive principal components are as follows: 2.8019, 1.3663, 1.7604, 1.7323 and 2.3778.

The matrix of the ordinary correlation coefficients between the principal components and the original variables is as follows:

$$\begin{bmatrix} -0.9563 & -0.9167 & -0.8233 & 0.7170 & -0.8099 \\ 0.1825 & -0.3515 & -0.2944 & -0.6659 & -0.1934 \\ 0.0048 & -0.1692 & 0.4652 & 0.0167 & 0.2725 \\ 0.2270 & -0.8640 & -0.1176 & 0.2072 & 0.0776 \\ 0.0247 & -0.9779 & -0.0726 & 0.0137 & -0.4759 \end{bmatrix} \tag{12}$$

In the i-th row there are correlation coefficients between the i-th principal component and the original variable, $i = 1,...,5$. The first principal component representing the dispersion of all the original variables is strongly correlated with the original variables (see the first row of the matrix given by expression (12)). The second and last principal components are distinctly correlated with the original variables denoted by x_2, x_4 and x_2, x_5, respectively. The third and fourth principal components are rather clearly correlated with variables x_3 and x_2, respectively. The last three principal components explain less than 9% variability of the original variables. Therefore, it suffices to consider only spatial autocorrelation coefficient for the first and second component. Moran's and Geary's coefficients calculated on the basis of the first component are -0.2811 and 2.8019, respectively. Therefore, it can be concluded that neighboring Polish voivodeships differ in their observations of the first principal component. Moran's and Geary's coefficients calculated on the basis of the second component are 0.3227 and 1.3663, respectively. In this case, the coefficient indicated similarity and dissimilarity, respectively.

Note that the values of both the generalised Moran and Geary coefficients (calculated for the original vector observations) are close to zero and one, respectively. This means that there is no tendency to similarity or dissimilarity between the values of a multivariate variable observed on neighboring objects. In our case this is due to the fact that the generalised autocorrelation coefficients are the average values of the ordinary autocorrelation coefficients calculated for the individual principal components of a multivariate variable.

3. Conclusions

The results of considerations on the properties of generalised spatial autocorrelation coefficients of the population objects characterized by observation vectors of a multidimensional variable are as follows. For this purpose, a generalization of the Moran coefficient was defined in a similar way to the generalization of the Geary coefficient proposed by Du et al. (2012). Both generalised coefficients indicate the degree of similarity between neighboring objects due to the distance between the observation vectors of the multidimensional variable observed in them. The principal components of a multivariate variable allow for the presentation of each of the generalised coefficients as the arithmetic mean of the ordinary spatial autocorrelation coefficients, but calculated on the basis of the principal components. It was shown that the decomposition of the original variable into principal components can lead to a substantial simplification of the analysis of multivariate spatial autocorrelation. Moreover, it was concluded that the interpretation of the generalised autocorrelation coefficients may lead to misleading results and therefore must be carried out simultaneously with

the analysis of ordinary autocorrelation coefficients determined on the basis of individual principal components. Finally, we can say that the obtained results allow the use of principal component analysis to enrich the interpretation of generalised spatial autocorrelation coefficients.

Acknowledgements

The author would like to thank the reviewers for their valuable comments on this article.

References

Cliff, A. D., Ord, J. K., (1981). Spatial Processes: Models and Applications. Pion, London.

Du, Z., Jeong, J. S., Jeong, M. K. and Kong, S. G., (2012). Multidimensional local spatial autocorrelation measure for integrating spatial and spectral information in hyperspectral image band selection. *Applied Intelligence*, 36, pp. 542–552.

Geary, R. C., (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5 (3), pp. 115–145.

Getis, A., Ord, J. K., (1992). The analysis of spatial association by use of distance statistic. *Geographical Analysis*, 24(3), pp. 189–206.

Griffith, D. A., Chun, Y., (2022). Some useful details about Moran coefficient, Geary ratio and the joint count indices of spatial autocorrelation. *Journal of Spatial Econometric*, 3:12.

Harville, D. A., (1997). *Matrix Algebra from a Statistician's Perspective*, Springer New York, Berlin, Heidelberg, Barcelona, Hong Kong, London, Milan, Paris, Singapore, Tokyo.

Krzyśko, M., Nijkamp, P., Ratajczak, W., Wołyński, W., Wojtyła, A. and Wenerska, B., (2023). A novel spatio-temporal principal component analysis based on Geary's contiguity ratio. *Computers, Environment and Urban Systems*, 103, pp. 1–8.

Krzyśko, M., Nijkamp, P., Ratajczak, W.,Wołyński, W., Wojtyła, A. and Wenerska, B., (2024). Spatio-temporal principal component analysis. *Spatial Economic Analysis*, 19:1, pp. 8–29. doi: 10.1080/17421772.2023.2237532.

Leontief, W. W., (1986). *Input - Output Economics*, Oxford University Press, New York.

Moran, P. A. P., (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37 (1), pp. 17–23. doi:10.2307/2332142.

Morrison, D. F., (1976). *Multidimensional Statistical Methods*, McGraw-Hill New York.

Overmars, K. P., de Koning, G. H. J. and Veldkamp, A., (2003). Spatial autocorrelation in multi-scale land use models. *Ecological Modelling*, 164, pp. 257–270.

Appendix

According to notation introduced in Section 2, the equation $CVC^T = \lambda$ is transformed to the following $V = C^T \lambda C$ because $C^{-1} = C^T$. The t-the principal component is determined by $z_{*t} = Xc_t$, $t = 1,...,k$ and $Z = [z_{*1}...z_{*k}] = XC$, $C = [c_1...c_k]$.

The equation $C^{-1} = C^T$ let us write $V^{-1} = (C\lambda C^T)^{-1} = (C^T)^{-1}(C\lambda)^{-1} = C\lambda^{-1}C^T$. Moreover: $ZC^T = X$, $z_{i*}C^T = x_{i*}$, $\bar{x} = U_N^T X/N = U_N^T ZC^T/N = \bar{z}C^T$, $i = 1,...,N$. These results let us rewrite the equation (6) as follows:

$$\begin{aligned} b_{ij} &= (z_{i*}C^T - \bar{z}C^T)V^{-1}(z_{j*}C^T - \bar{z}C^T)^T = (z_{i*} - \bar{z})C^T V^{-1} C(z_{j*} - \bar{z})^T = \\ &= (z_{i*} - \bar{z})C^T C\lambda^{-1}C^T C(z_{j*} - \bar{z})^T = (z_{i*} - \bar{z})\lambda^{-1}(z_{j*} - \bar{z})^T = \\ &= [(z_{i1} - \bar{z}_1)...(z_{it} - \bar{z}_t)...(z_{ik} - \bar{z}_k)][\lambda_t^{-1}][(z_{j1} - \bar{z}_1)...(z_{jt} - \bar{z}_t)...(z_{jk} - \bar{z}_k)]^T = \\ &= [(z_{i1} - \bar{z}_1)\lambda_1^{-1}...(z_{it} - \bar{z}_t)\lambda_t^{-1}...(z_{ik} - \bar{z}_1)\lambda_k^{-1}][(z_{j1} - \bar{z}_1)...(z_{jt} - \bar{z}_t)...(z_{jk} - \bar{z}_k)]^T = \\ &= \sum_{t=1}^{k}(z_{it} - \bar{z}_t)\lambda_t^{-1}(z_{jt} - \bar{z}_t) = \frac{1}{\lambda_t}\sum_{t=1}^{k}(z_{it} - \bar{z}_t)(z_{jt} - \bar{z}_t). \end{aligned}$$

This and equations (1) and (5) lead to the following:

$$\begin{aligned} I_k^M &= \sum_{i=1}^{N}\sum_{j=1}^{N} b_{ij}q_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\lambda_t}\sum_{t=1}^{k}(z_{it} - \bar{z}_t)(z_{jt} - \bar{z}_t)q_{ij} = \\ &= \sum_{t=1}^{k}\frac{1}{\lambda_t}\sum_{i=1}^{N}\sum_{j=1}^{N}(z_{it} - \bar{z}_t)(z_{jt} - \bar{z}_t)q_{ij}. \end{aligned}$$

This directly leads to equation (8).

Similarly, equation (10) could be derived as follows:

$$\begin{aligned} d_{ij} &= (z_{i*}C^T - z_{j*}C^T)V^{-1}(z_{i*}C^T - z_{j*}C^T)^T = (z_{i*} - z_{j*})C^T V^{-1} C(z_{i*} - z_{j*})^T = \\ &= (z_{i*} - z_{j*})C^T C\lambda^{-1}C^T C(z_{i*} - z_{j*})^T = (z_{i*} - z_{j*})\lambda^{-1}(z_{i*} - z_{j*})^T = \\ &= [(z_{i1} - z_{j1})...(z_{it} - z_{jt})...(z_{ik} - z_{jk})][\lambda_t^{-1}][(z_{i1} - z_{j1})...(z_{it} - z_{jt})...(z_{ik} - z_{jk})]^T = \\ &= [(z_{i1} - z_{j1})\lambda_1^{-1}...(z_{it} - z_{jt})\lambda_t^{-1}...(z_{ik} - z_{jk})\lambda_k^{-1}][(z_{j1} - z_{j1})...(z_{jk} - z_{jk})]^T = \\ &= \frac{1}{\lambda_t}\sum_{t=1}^{k}(z_{it} - z_{jt})^2. \end{aligned}$$

This and equations (2), (3) lead to the following:

$$I_k^G = \frac{N-1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N} d_{ij}q_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{N-1}{2N\lambda_t}\sum_{t=1}^{k}(z_{it}-z_{jt})^2 q_{ij} =$$

$$= \sum_{t=1}^{k}\frac{N-1}{2N\lambda_t}\sum_{i=1}^{N}\sum_{j=1}^{N}(z_{it}-z_{jt})^2 q_{ij}.$$

This directly leads to equation (10).