Elżbieta Antczak
The Chair of Spatial Econometrics
University of Lodz

Building W Matrices Using Selected Geostatistics Tools. Empirical Examination and Application

Since the early 90s it has been generally known that the construction of a spatial weight matrix (**W**) is an important problem of spatial econometrics (Kooijman, 1976). That matrix takes into account and expresses potential for interactions between pairs of observations in different locations (Anselin, 1988). The **W** matrix could be set a priori (**W** specified exogenously) by the researcher, which is not always satisfactory (Angulo et al. (2017)). On the other hand, many scientists claim that the **W** matrix may be estimated from data (Harris, 2011). Kooijman (1976) was one of the first to explicitly tackle the question of estimating the **W** matrix. He suggested weights be built by maximising the value of Moran's *I*. His procedure aroused many doubts, but provoked looking for new ways of solving the problem. For example, Lee (1982) showed that the **W** *k*-nearest neighbour problem and other seemingly unrelated problems can be solved efficiently with the Voronoi diagram; Griffith (1996) proposed to find a **W** absorbing spatial effects from data; Fernández et al. (2009) suggested a specification of **W** based on the measure of entropy; Mur and Paelinck (2010) focused on the maximisation of the Complete Correlation Coefficient; Getis and Aldstadt (2004) used the local statistical model and AMOEBA algorithm; Stewart and Zhukov (2010) indicted neighbours from the visualisation of spatial effects; Hondroyianis et al. (2012) and Keleijan, Piras (2014) assumed that elements of **W** are an unknown function of two sets of exogenous variables; or Benjanuvatra (2015) presented the QML to estimate weights in **W** directly from data.

Recalling Tobler’s first law of geography, the distance and neighbouring relations between different areas can in particular indicate to what degree spatial dependence exists and “how close places need to be” in order to be related, or spatially autocorrelated. This law makes clear that spatial relations are not static but evolve over distance.

In the structure of the spatial weights matrix based on the geographical distance, it can be difficult to determine the maximum distance to which units are interrelated (show similarity in terms of a studied feature resulting from mutual spatial relations). One of the main assumptions of the article is to describe and apply spatial statistics and geostatistics methods (spatial measures of central tendency and [semi]variograms) based on which the spatial continuity (variation, degree of spatial correlation) of specified phenomena can be effectively characterised depending on the distance. However, according to, e.g. Matheron (1963), Krige (1996), Zawadzki (2002) or Robinson, Dietrich (2016), the variogram provides a description of the overall spatial pattern and of how data are related (autocorrelated) not only with the distance but also with the distance and direction (directional variogram). In most analyses, the spatial structure is quantified as being direction independent (isotropic). However, for a lot of spatial data the direction (anisotropy) relates to processes i.e. pollution, migrations, politics, travel etc. The anisotropy of phenomena is a property of a spatial process or data in which spatial dependence (autocorrelation) changes with both the distance and direction between two locations. The idea that the direction can be important in determining spatial structure is the second of the main concerns of this paper. In order to describe the variation of a phenomenon depending on the distance and direction, the standard deviation ellipsis (allowing one to see if the distribution of features is elongated and hence has a particular orientation), directional variogram (defined above) and surface trend models (being a mathematical function, or polynomial that describes the variation in data) were used.

Through the empirical application of the above tools to construct a range of spatial weights matrices, an attempt was made to answer the following research questions: how neighbours influence each other: cumulatively, equally, or proportionally to their proximity or some other measure of decay? Should the spatial weights matrix contain information about the anisotropy of the phenomenon (identical weights without taking into account the directional character – dispersion, diffusion – of phenomena in different directions in geographical space and with different intensity)? How to determine the distance of spatial correlations and the degree of the mutual influence of units in space? Does spatial autocorrelation change solely depending on the distance or does it also depend on the direction of the courses of phenomena? How to determine values of weights in **W** matrices? Do varied values of spatial weights lead to significant differences in results of analyses? Should weights matrices be different for different years? What results of analyses will we receive if we introduce a weights matrix built without considering the nature of phenomena?

The study was carried out on a sample of about 300 Polish towns and selected years in the time span 2005-2015, as well as for averaged data for the whole period. Variables were related to the quantity of produced waste and economic development. Both ESDA analyses and estimations of spatial panel models were performed by including particular spatial weights matrices in the study (exogenous matrices, distance matrices and directional matrices constructed based on them). The research was conducted in ArcMap, RCran, SAGA and GeoDa. Received results indicated that geostatistics tools can be effectively used to build **W** matrices. Substantially, results of carried out analyses applying different spatial weights matrices did not exclude but rather supplemented (enriched, extended) one another. The most precise picture of spatial (global and local) dependences was received by including directional matrices in the analysis. What is more, models with different weights matrices (directional matrix, selected distance matrix or just an exogenous one) were sensitive to a change in that matrix in such a way that values of the assessed parameter at the regressors did not significantly change (considering the direction of the influence on the endogenous variable, in few cases, determined a slight increase of the values). There was, however, a change in the assessed value of the spatial autoregression or autocorrelation parameter (strength of influence). Nevertheless, modelling results were still substantially accurate and the application of different (endogenous, with "dedicated" weights) matrices was justified. A problem appeared to be the inclusion of matrices whose spatial weights change recurrently (from period to period) in panel econometric modelling and the fact that the creation of such matrices is time-consuming.

Key words: **W** matrixes, engogenity, geostatistics, directional matrix, (semi)variograms, TSA;
JEL: C30, C33, C46, Q01

Litearture:

1. Aldstadt, J., Getis A. (2006), Using AMOEBA *to Create a Spatial Weights Matrix and Identify Spatial Clusters*. *Geographical Analysis* 38 327-343.
2. Angulo A., Burridge P., Mur J., (2017), *Testing for Breaks in the Weighting Matrix*, Documento de Trabajo 2017-01 Facultad de Economía y Empresa Universidad de Zaragoza Depósito Legal Z-1411-2010.
3. Anselin L. (1988), *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer.
4. Benjanuvatra, S. (2015), *QML estimation of the spatial weight matrix in the MR-SAR model*. York, DERS University of York Working Paper.
   Contained in Spatial Statistical Models. In *Practical Handbook* of *Spatial Statistics,* edited by S. L. Arlinghaus. Boca Raton: CRC.
5. Elhorst, J. (2014), *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*; Springer: New York, NY, USA, 2014.
6. Fernández E., Mayor M., Rodríguez J., (2009), *Estimating spatial autoregressive models by GME-GCE techniques*. *International Regional Science Review*, 32 148-172.
7. Getis A, Aldstadt J. (2004), *Constructing the Spatial Weights Matrix Using a Local Statistic Spatial*. Geographical Analysis, 36 90-104.
8. Griffith, D. A. (1996), *Some Guidelines for Specifying the Geographic Weights Matrix*
9. Harris, R. Moffat, J. Kravtsova, V. (2011), *In search of 'W'*. Spat. Econ. Anal. 2011, 6, 249–270.
10. Hondroyianis, G., Kelejian, H. H., Mukerji, P., Tavlas, G. S. (2012), *Government bond spreads contagion among euro area countries*. Manuscript.
11. Kelejian H. Piras G. (2014), *Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes*. Regional Science and Urban Economics, 46 140.149.

12. Kooijman, S. (1976), Some *Remarks on the Statistical Analysis of Grids Especially with Respect to Ecology*. Annals of Systems Research 5.
13. Krige, D.G. (1996), A practical analysis of the effects of spatial structure and of data available and accessed, on conditional biases in ordinary kriging. International Geostatistics Congress, Wollongong, Australia 22–27 September 1996. Vol. 2. Baafi, E.Y. and Schofield, N.A. (eds). Springer. pp. 799–810.
14. Lee DER-TSAI, (1982), *On k-Nearest Neighbor Voronoi Diagrams in the Plane*, IEEE TRANSACTIONS ON COMPUTERS, VOL. C-31, NO. 6, JUNE 1982.
15. Matheron, G. (1963). *Principles of Geostatistics*. Economic Geology 58, 1246–1266.
16. Mur, J. Paelinck J. (2010), *Deriving the W-matrix via p-median complete correlation analysis of residuals*. The Annals of Regional Science, DOI: 10.1007/s00168-010-0379-3.
17. Robinson M., Dietrich S. (2016), *An Introduction to Spatial Autocorrelation and Kriging*,https://sites.ualberta.ca/~lkgray/uploads/7/3/6/2/7362679/slides-autocorrelation_kriging.pdf.
18. Stewart M., Zhukov Y. (2010), *Choosing Your Neighbors: The Sensitivity of Geographical Di_usion in International Relations*, http://isites.harvard.edu/fs/docs/icb.topic668784.files/Stewart_Zhukov_ISA.pdf
19. Tobler W. (1970), *A computer movie simulating urban growth in the Detroit region*. *Economic Geography*, 46 234-240.
20. Zawadzki J. (2002), Badania korelacji przestrzennych zawartości wybranych pierwiastków śladowych w glebach Warszawy i okolic. Ochrona Środowiska, 24 (4), p. 17-26.