Available online at www.sciencedirect.com
ScienceDirect
Regional Science and Urban Economics 37 (2007) 491–496
regional SCIENCE & urban ECONOMICS
www.elsevier.com/locate/regec

Reflections on spatial autocorrelation

Arthur Getis *

Department of Geography, San Diego State University, San Diego, CA 92182, United States

Available online 12 April 2007

Abstract

The concept, spatial autocorrelation, is central to many concerns expressed in Regional Science and Urban Economics. The term, first used in 1967, had precursors embodied in such ideas as distance-decay and spatial interaction. Today, spatial autocorrelation is the mainstay of the burgeoning field of spatial econometrics. Its study has many advantages such as providing tests on model misspecification. The Moran's I statistic is one of many ways spatial autocorrelation can be represented. Interestingly, the concept has yet to make a significant appearance in the mainstream econometrics literature.

JEL classification: C21; C51; C88; R12; R15
Keywords: Spatial autocorrelation; Spatial effects; Spatial econometrics

Although all regional scientists study socio-economic aspects of earth regions, we approach it in many different ways. A typical article in RSUE, for example, will contain at least one model of regional economic activities, economic theory for an explanation of those activities, and a sophisticated quantitative analysis of the model using census or related data. The activities chosen for study, the methods of analysis, the types of models, and the array of data employed all vary considerably.

As a geographer, my point of view is not mainstream RSUE. For this essay, I thought it might be interesting for readers to learn what one geographer considers central to a better understanding of RSUE concerns. I have selected *spatial autocorrelation*, a basic concept in modern spatial analysis and discussed at length in the geographic literature. While the spatial autocorrelation is not completely missing from the pages of RSUE, it is not in evidence in mainstream econometrics, a literature familiar to the majority of those who publish in RSUE. As a result, the concept is not given enough attention in RSUE. In my judgment, recognition of the importance of spatial autocorrelation is beginning to make its way into RSUE, but it will be at least 5 or 10 years

* Tel.: +1 619 594 6639; fax: +1 619 594 4938.
E-mail address: arthur.getis@sdsu.edu.

0166-0462/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.regsciurbeco.2007.04.005

before spatial autocorrelation will appear in the mainstream econometrics literature. Spatial autocorrelation is not mentioned *at all* of in 12 currently used econometrics textbooks.

1. The origin of the term

Spatial dependence, spatial association, spatial interaction, and spatial interdependence have been among the terms used to describe a phenomenon like no other. For 200 years, since the early 1800s, scholars had identified a distance decay effect that George Kingsley Zipf (1949) would elevate to the status of a law in his landmark book, *Human behavior and the principle of least effort: An introduction to human ecology* and that Edward L. Ullman (1956) would concisely conceptualize as the basis for human spatial interaction. But it was not until the regional science conference paper by Andrew D. Cliff and J. Keith Ord in 1967 (published in 1969) that the term spatial autocorrelation made its way into the lexicon of phrases used to capture an important segment of the realm of spatial and regional understanding (Cliff and Ord, 1969). Like others before them, Cliff and Ord realized that the special case of correlation in space, when there is a relationship between nearby spatial units of the same variable, needed to be identified if research questions were to be answered accurately. One such question is fundamental to the entire discipline of geography — “Is the spatial pattern displayed by the phenomenon significant in some sense and therefore worth interpreting?” – and another question is becoming an integral part of a host of disciplines, including urban economics — “Can we obtain any information on the processes which have produced the observed pattern from an analysis of the mapped distribution of the phenomenon?” (Cliff and Ord, 1973, 1981) Cliff and Ord recognized what is known as the misspecification problem in spatial analysis, that is, models that required traditional statistics for their evaluation were misspecified if spatial autocorrelation was not taken into account when data were georeferenced. To this day, many economists are unaware that for an OLS model to be properly specified, residuals must be spatially independent in the mapped region of study.

An interesting sidelight to this is that simultaneously with the introduction of the term “spatial autocorrelation” came Jean Paelinck’s discussion (1967) of a new field that would address problems in regional and urban econometric modeling by blending economic theory, mathematical formalization, and mathematical statistics. In 1974 Paelinck named the field “spatial econometrics” (Paelinck and Klaassen, 1979). As it has evolved, the discipline of spatial econometrics is held together by various tests on the spatial autocorrelation that might exist in spatial economic and related data.

Cliff and Ord’s predecessors are many and varied. Direct inspiration came from one of Cliff’s teachers at Northwestern University, Michael F. Dacey. Dacey (1965) had already created a join count statistic that described spatial autocorrelation. Dacey had expanded on the work of the statisticians P.A.P. Moran (1948), P.V.A. Krishna Iyer (1949), and R.C. Geary (1954), who had developed spatial autocorrelation statistics without naming them thus.

2. Pre-conditions

The spatial twist on correlation came mainly from a need to compare maps of the same region and the realization that georeferenced observations generally are not independent of one another, especially if they are close in distance to one another. Such well-known spatial concepts as distance-decay, spatial interaction, and spatial randomness have a rich history, but not one that can be associated directly with the concept of spatial autocorrelation. Spatial autocorrelation is more closely linked to statistical theory than to spatial theory. This is not to say that the very same scholars who concerned themselves with spatial theory were unaware of the idea of spatial correlation. Being sensitive to the peculiarities of the spatial point of view, they adopted and expanded the concept.

Dacey realized that the contributions to the study of spatial structure failed to create a "well-developed conceptual and methodological framework that organizes and structures the description, classification and analysis of spatial distributions." (Dacey, 1973, p. 301). Because the study of and specification for spatial autocorrelation are so closely tied to the geometry of the spatial units in question, spatial autocorrelation theory evolved from the study of patterns of variables on maps. This recognition of the possible effect on the results of analysis of correlation peculiarities of the shapes, sizes, and boundaries of regions produced the spatial autocorrelation movement.

At a time when Cliff and Ord were just explicating the concept of spatial autocorrelation, Hubert, Golledge, and Costanzo looked for generalizations of spatial structure (1981). By the late 1980s, the field of spatial structure through the study of spatial autocorrelation had advanced significantly. More than any other regional scientist, however, it was Luc Anselin (1988) in his book *Spatial Econometrics: Methods and Models* who called attention to the responsibility of spatial scientists—geographers, urban planners, but especially economists — to take into account spatial effects when faced with regional-type georeferenced models and data. In a paper published by Anselin and Griffith (1988), they pinpointed the central role of what they called the spatial effects when working with theory that required georeferenced data for its evaluation.

3. Not any correlation

Although it may be viewed as a special case of correlation, spatial autocorrelation has a meaning all its own. Whereas correlation statistics were designed to show relationships between variables, autocorrelation statistics are designed to show correlations within variables, and spatial autocorrelation shows the correlation within variables across space. A cursory view might regard a test on the existence of spatial autocorrelation as an extension of the Durbin–Watson test for temporal autocorrelation. It is well-known that a fundamental aspect of time series analysis is to either assess, filter, or model temporal autocorrelation. The special effort to discern spatial autocorrelation, however, makes identifying temporal autocorrelation a relatively simple job compared to the same task for spatial autocorrelation. The multidimensional nature of spatial autocorrelation, i.e., the need to search in all directions as opposed to the one-way temporal direction, is one complicating factor. Another is the wide variety of units that are eligible for use in spatial autocorrelation (distance, neighbors, links, etc.) as compared to temporal autocorrelation (time).

Hubert et al. (1981) say that "Given a set S containing n geographical units, spatial autocorrelation refers to the relationship between some variable observed in each of the n localities and a measure of geographical proximity defined for all $n(n-1)$ pairs chosen from n."(p. 224) Cliff and Ord (1973) give it a process-like definition by saying, "If the presence of some quality in a county of a country makes its presence in neighboring counties more or less likely, we say that the phenomenon exhibits spatial autocorrelation."(p. 1) The point here is that spatial autocorrelation is a term, geometrical in character, that says that with regard to a single variable certain spatial units are related (positively) or unrelated (negatively) to other spatial units and that there is dependence between spatial units. As most researchers are aware, the vast majority of standard statistical tests are based on independent observations.

The advantages of the study of spatial autocorrelation are manifold:

- Provides tests on model misspecification;
- Determines the strength of the spatial effects on any variable in the model;

- Allows for tests on assumptions of spatial stationarity and spatial heterogeneity;
- Finds the possible dependent relationship that a realization of a variable may have on other realizations;
- Identifies the role that distance decay or spatial interaction might have on any spatial autoregressive model;
- Helps to recognize the influence that the geometry of spatial units under study might have on the realizations of a variable;
- Allows us to identify the strength of associations among realizations of a variable between spatial units.
- Gives us the means to test hypotheses about spatial relationships;
- Gives us the opportunity to weigh the importance of temporal effects;
- Provides a focus on a spatial unit to better understand the effect that it might have on other units and vice versa ("local spatial autocorrelation");
- Helps in the study of outliers.

4. Representations

Whereas temporal autocorrelation can be modeled easily in a Durbin–Watson framework, one must choose carefully from one or more of the following to represent spatial autocorrelation (this is not an exhaustive list):

- Moran's I (a global covariance representation)
- Geary's c (a global differences representation)
- Γ (a cross product representation)
- Getis and Ord's G (a global multiplicative representation)
- Ripley's K (a cumulative pairs over distance representation)
- ρ, λ (autoregressive coefficients in various regression representations)
- Getis and Ord's G_i and G_i^* (local cluster representations)
- Anselin's I_i and c_i (local indicators of spatial association (LISA) statistics)
- Ord and Getis's O (a local representation taking into account global autocorrelation)
- $1/\gamma$ (the inverse of the semivariogram; i.e., the correlogram).

All of these representations of spatial autocorrelation include some measure of the configuration of the spatial units under study, very often called the spatial weights matrix.

For examples, an element of that matrix might be:

- 1 or 0 (1 for some defined aspect of the geometry of the region being studied and 0 otherwise)
- $1/d^{-\alpha}$ (a distance decay representation)
- s/p (a common boundary representation — s is length of side in common, p is perimeter)
- d_2-d_1 (a difference in distance representation)
- G_i, I_i (a local statistics representation)

5. Recognition of spatial effects in the literature

In Anselin et al. (2004) the trends in the recognition of spatial effects in relevant economics journals is documented. There is no question of the increasing impact spatial concerns are having, even in journals in which spatial data had not heretofore been analyzed. The reasons for this are many, including the "new" requirements for the analysis of spatial data, the recent increase in the availability of spatial data, new and more powerful computers, and the introduction of now widely used software in which spatial autocorrelation is explicitly analyzed. *SpaceStat* (1995) and then *GeoDa* (2005) by Anselin, then *Spatial Econometrics* (2006) by LeSage, and *STARS* (2006) by Rey, and the inclusion of some spatial autocorrelation functions in ArcView 9.1 (2005) (a heavily used geographic information systems package) have all had a considerable influence in increasing the use of spatial analysis. Many edited volumes and textbooks detail the tools available for spatial analysis. Anselin's listserve ("Openspace") has opened wide the door for those inclined to become involved with the spatial aspects of analysis.

Interesting enough, however, is that spatial econometrics has not yet appeared in any serious way in the econometrics literature. The recent textbooks in econometrics mentioned earlier contain virtually nothing on ways to handle spatial data. This implies that all economists being educated today must learn about dealing with spatial models and data outside of their curricula. At the same time, however, economists are more concerned than ever about environmental and urban problems. No wonder the literature on such subjects is only lightly sprinkled with analyses of spatial effects and the use of the spatial autocorrelation concept. Even in the field of geostatistics, a field intimately involved in what may be called spatial analysis of continuous data, the term is rarely used.[1]

There is no doubt in my mind that spatial autocorrelation should be and become a prominent subject for study in all of the social sciences.

References

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer, Dordrecht.

Anselin, L., Griffith, D.A., 1988. Do spatial effects really matter in regression analysis. Papers of the Regional Science Association 65, 11–34.

Anselin, L., Florax, R., Rey, S., 2004. Econometrics for spatial models, recent advances. In: Anselin, L., Florax, R., Rey, S. (Eds.), Advances in Spatial Econometrics: Methodology, Tools and Applications. Springer-Verlag, Berlin, pp. 1–25.

ArcView, v 9.1, 2005, Environmental Systems Research Institute: Redlands, CA.

Cliff, A.D., Ord, J.K., 1969. The problem of spatial autocorrelation. In: Scott, A.J. (Ed.), London Papers in Regional Science. Studies in Regional Science, vol. 1. Pion, London, pp. 25–55.

Cliff, A.D., Ord, J.K., 1973. Spatial Autocorrelation. Pion, London.

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

Cressie, N., 1991. Statistics for Spatial Data. John Wiley, Chichester.

Dacey, M.F., 1965. A review of measures of contiguity for two and k-color maps. In: Berry, B.J.L., Marble, D.F. (Eds.), Spatial Analysis: A Reader in Statistical Geography. Prentice Hall, Englewood Cliffs, pp. 479–495.

Dacey, M.F., 1973. Some questions about spatial distributions. In: Chorley, R.J. (Ed.), Directions in Geography. Methuen, London.

Geary, R., 1954. The contiguity ratio and statistical mapping. The Incorporated Statistician 5, 115–145.

GeoDa, 2005. Spatial Analysis Laboratory, University of Illinois at Urbana–Champaign. see web.

[1] Noel Cressie (1991), in his 800-page discussion of geostatistics, *Statistics for Spatial Data*, a defining text for the field of geostatistics, does not mention the term spatial autocorrelation. The term correlogram, used in a few instances, represents spatial autocorrelation. Interestingly, Cressie says "I do not advocate cross-validation for confirmatory data analysis (e.g., hypothesis testing, standard error estimation, etc.) at this time, because the correlations between data give rise to complicated distribution theory that needs further study." In essence, the field of geostatistics assumes spatial autocorrelation to exist by adopting for spatial data the concept of intrinsic stationarity, a term that implies that spatial effects exist and can be described by the variogram.

Hubert, L.J., Golledge, R.G., Costanzo, C.M., 1981. Generalized procedures for evaluating spatial autocorrelation. Geographical Analysis 13 (3), 224–233.

Krishna Iyer, P.V.A, 1949. The first and second moments of some probability distributions arising from points on a lattice, and their applications. Biometrika 36, 135–141.

Le Sage, J., 2006. Econometrics Toolbox. see web.

Moran, P.A.P., 1948. The interpretation of statistical maps. Journal of the Royal Statistical Society. Series B 10, 243–251.

Paelinck, J., 1967. L'efficacite de la politique economique regionale. Namur, Faculte des Sciences Economiques 58.

Paelinck, J., Klaassen, L., 1979. Spatial Econometrics. Saxon House, Farnborough.

Rey, S.J., Janikas, M.V., 2006. STARS: space-time analysis of regional systems. Geographical Analysis 38, 67–86.

SpaceStat: A software program for the analysis of spatial data., 1995, v 1.80 Regional Research Institute, West Virginia University.

Ullman, E.L., 1956. The role of transportation and the bases for interaction. Man's Role in Changing the Face of the Earth. University of Chicago Press, pp. 862–880.

Zipf, G.K., 1949. Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology. Addison-Wesley, Reading, Mass.