Neighborhood Effects

Sang-Il Lee, Department of Geography Education, Seoul National University, Seoul, South Korea

Glossary

Buffer zones Areas of user-defined width surrounding geographical features.
Dummy variable A numerical variable used in regression analysis to represent categories.
Geocoding The process of positing a geographical feature on a map as a point by using a code representing its location.
Logistic regression A regression analysis used for categorical dependent variables.
Measurement scale The physical size of the spatial unit in a study, sometimes called spatial resolution.
Misspecification error A modeling error in a regression framework committed mainly by omitting relevant variables.
Operational scale The spatial extent to which a spatial process operates.
Spatial dependence A particular relationship between geographical similarity of locations and numerical similarity of values observed on them.
Spatial heterogeneity A lack of spatial uniformity in mathematical and statistical properties.

Neighborhood effects refer to sociogeographical milieu's influences on the ways in which people think and act. In other words, the neighborhood effect is a structural factor which may drive individuals in a place to possess a collective behavioral disposition, leading to similar behavioral outcomes or choices. In social sciences, neighborhood effects have been regarded as a synonym or a particular type of contextual effects. The concept of neighborhood effects has provided a conceptual foundation for those who seek to emphasize the importance of sociogeographical contexts in explaining various types of behavioral outcomes such as educational achievements, adolescent misconducts, crimes, and so on. The concept has also been utilized in housing studies, epidemiology, psychology, and so on.

It is in the field of political geography, however, that the most attention has been given to the concept. Political geographers have used the concept in order to explain why different people in the same place vote in similar ways while similar people in different places vote in different ways. It is generally accepted, in particular, by political scientists that individuals' voting decisions are largely determined by who they are. However, if the voting decisions are also associated with where they live or have been socialized even after the effects of who they are were controlled, one might argue that there exist geographical contexts' influences on the political behaviors and the influences might be conceptualized as neighborhood effects. On the contrary, if who they are explains the voting patterns or if the explanatory power of where they live is not sufficiently large in comparison with that of who they are, one might argue that neighborhood effects do not exist.

Subsequently, this article briefly discusses some theoretical underpinnings which the concept may rely on, and then moves to some methodological issues as to how to formulate and model neighborhood effects. All the discussions revolve around voting behaviors.

Theoretical Underpinnings

How could sociogeographical contexts possibly drive individuals in a place to possess a collective behavioral disposition? It has been known that the conceptual origins of neighborhood effects lie in the work of a Swede and an American political scientist, H. Tingsten and V. O. Key, in 1930s–1940s. It is by Kevin Cox, David Reynolds, and David Butler and Donald Stokes in the late 1960s that the concept emerged as a keystone for electoral studies. Kevin Cox, in his seminal 1969 paper, identified four factors (or biases), on the basis of a theory of communication and information flow, which influence people within a sociospatially structured information network. The geographical distance bias refers to the fact that people living nearby are more likely to interact than those apart. The acquaintance circle bias facilitates more frequent and intensive interaction and thus tends to lead to more opinion agreement between members of the same acquaintance circle. This bias is closely related to the first one; people who live close together are more likely to be members in the same acquaintance circle than those who live further apart. In practice, the information flows within a neighborhood are often diffusive, influential, and persuasive such that the situations may heighten the possibility that some having minority opinion are converted to the dominant majority opinion. The forced field bias indicates some situations whereby an established political culture in a local area appears to be quite coercive enough to mobilize opinions in a particular way. This is most clearly observed where place-based political parties are pervasive. The reciprocity bias refers to the fact that the chance of information being passed between two individuals is greater where there already is a tie between them. For example, the relationships between parents and children, siblings, husbands and wives, friends, and colleagues provide a greater

channel for information exchange. All these factors can be articulated in spatial terms, and collectively constitute the fundamental rationale for the concept of neighborhood effects.

In sociology and other social sciences, there are several theories about how neighborhood effects occur, some of which are congruent with Cox's notions, but others not. Contagion theory underlines the power of peer group influences to spread particular types of behaviors. Collective socialization theory emphasizes how neighborhoods provide role models and sources of social control. Institutional theory stresses the roles of various institutions located in local areas such as schools, businesses, political organizations, social service agencies, and so on. Competition theory and relative deprivation theory are less relevant here. On the basis of all these theoretical contemplations, five processes by which neighborhood effects are produced can be listed as follows (after Johnston et al., 2005):

1. Local social interaction: the classic neighborhood effect. This involves a process of "conversion through conversation" or "those who talk together vote together"; "I talk with them and vote as they do."
2. Environmental selection: according to this process people choose to live (to the extent that they can) among people they wish to associate with; "I want to be like them so I live with them."
3. Emulation: in this process, people choose to behave like their neighbors, even without interacting with them, on the basis of observed (or inferred) behavioral patterns; "I live among them and want to be like them."
4. Environmental observation: people see and hear about issues in their local area and vote with their neighbors accordingly, in order to promote local interest; "what I observe around me makes me vote with them."
5. Local pressure: political parties actively seek support through canvassing and local campaigns, and people may therefore be influenced by the intensity of that local pressure to vote in a particular way; "they want me to vote for them here."

Debates

There have been two large-scale debates on whether neighborhood effects or contextual effects really matter: first in *Political Geography Quarterly* and second in *Political Geography*. Those debates were seen as conflicts between two academic fields—geography and political science. In the 1987 debate, political scientists reported that if individual factors were controlled, contextual effects would vanish. In contrast, political geographers demonstrated that substantial contextual effects remained when individual factors were held constant. The former argues that neighborhood effects are mere myth or embellishment, while the latter contends that the alleged pure individuals have already been saturated with contextual effects.

In the 1996 debate, Gary King, a political scientist, argued "The *geographical variation* is usually quite large to begin with, but after we control for what we have learned about voters, there isn't much left for *contextual effects*." and "geographical variation yes, contextual effects no." John Agnew, a political geographer, contended "The concept of geographical context can be used to draw attention to the spatial situatedness of human action in contrast to the non-spatial sorting of people out into categories based on census and other classification schemes that inspires most conventional social science." He further argued, "[T]he hierarchical-geographical context or place channels the flow of interests, influence and identity out of which political activities emanate." and thus "political behavior is inevitably structured by a changing configuration of social-geographical influences …" (Agnew, 1996).

It is worthwhile to indicate here that the boundary between individual effects and contextual effects is often blurred; what are usually regarded as individual variables are indeed dependent upon contextual variables in many cases. For example, an individual's union membership may be significantly affected by local norms and traditions; individual income and education levels are often dependent upon the quality of schools and jobs in the local neighborhood; an individual's status in the occupational ladder is likely to be constrained and shaped by the local labor market economy.

In a methodological sense, the essence of the arguments raised by those who are against the geographical/contextual effects is that empirical models favoring the effects are committing a kind of misspecification error; that is, contextual variables turn out to be significant simply because important individual variables are omitted. Thus, what is at issue in the debates is methodological, rather than theoretical.

Methodologies

The methodologies for formulating and estimating neighborhood effects may be classified into two categories: area-based and individual-based. One could think of a typical research framework from each of the categories. A study with aggregated data may involve a regression analysis in order to explain spatial variation of voting results with various independent variables which are usually derived from census data. Although this kind of analysis could provide a formula to predict the voting outcomes at a locality on the basis of the locality's characteristics, it gives little insight into neighborhood effects. As a matter of fact, neighborhood effects reside intact in a set of residuals. On the other hand, a study with individual data may also involve a regression model, usually logistic, by positing an individual's choice in a binary format on the left-hand side and their attributes on the right-hand side which usually include:

1. socioeconomic status (occupational class, educational attainment, housing tenure, employment status, income, etc.);
2. ascribed characteristics (sex, age, race, ethnicity, etc.); and
3. family background (parents' political propensity and socioeconomic status).

This approach, however, completely ignores the neighborhood effects and has been overwhelmingly adopted in political scientists' literature.

Methods for Aggregate Data

A study based on aggregate data examines the neighborhood effects in reliance on an area's collective characteristics which are expected to explain collective outcomes under investigation. Prior to a discussion on methodological principles utilizing area data, a distinction between compositional variables and contextual variables needs to be made. In general, a compositional variable refers to an individual's socioeconomic characteristic, whereas a contextual variable relates to location which provides a milieu within which an individual lives and is socialized. When aggregate data are considered, this distinction needs to be furthered. In a conceptual sense, compositional variables are associated with whom people in a place are and what they have, whereas contextual variables describe environmental properties, physical and/or human, that a place possesses as a macroscale entity. For example, geographical variations in death rates are not only due to different age structures of areas (the compositional effect) but also due to some detrimental attributes of the areas that may have a direct or indirect effect on people's health (the contextual effect). It should be noted, however, that the two types of variables are dependent upon each other such that it is often impossible to separate one from the other.

One way of tackling neighborhood effects with aggregate data is to regress socioeconomic variables (usually compositional variables) of local areas plus "locational variables" on voting outcomes. The locational variables are mainly associated with where an area stands as a spatial object in relation to other areas or specific features. For example, they could be each area's distance from a focal point, for example, central business district (CBD), or dummy codes indicating membership in buffer zones or to particular geographical categories based on either administrative regionalization or functional classification, or an areal unit's identity itself. If the locational variables have a statistically significant explanatory power with other socioeconomic variables being held constant, the existence of neighborhood effects would be confirmed. For example, it has been acknowledged that suburban areas are more pro-conservative than their inner-city counterparts. One can follow this analytical line by incorporating a series of higher-level regional dummies along with socioeconomic characteristics of basic spatial units. A significant level of heterogeneity in the regression coefficients of the dummy variables could evidence, at least partially, the point that geographical contexts count.

A more intelligent way of tackling neighborhood effects in the regression framework using aggregate data might be to attempt to define a set of contextual variables rather than the simple, locational variables, and then to put them on the right-hand side along with a set of compositional variables. Contextual variables may include economic experiences, sociocultural practices, capital–labor relationships, and the level of social capital of a place. If the set of contextual variables has sufficient explanatory power in comparison with the set of compositional variables, one can say that there exist neighborhood effects. One problem in implementing this framework is that it is often difficult to measure genuine contextual variables and to separate them from compositional variables. A carefully designed set of contextual variables could identify neighborhood effects which would otherwise be hidden in the residuals.

There are other types of area-based analyses utilizing aggregate datasets, some variables of which are derived from individual-level survey data. We know the proportion of votes favoring a particular political party from aggregate data, but do not know what percentage of people belonging to a particular social class actually vote for the party. By possessing that kind of information, one may be able to devise a way of proving the existence of neighborhood effects. It has been found that the more dominant the class is in an area, the greater the proportion of its members who vote along the expected class lines. This echoes the finding that the more residentially segregated cities would show greater polarization between social classes.

In comparison with individual-based research, area-based studies provide indirect or "circumstantial" evidences for neighborhood effects, basically because they are never completely free from the problem of ecological fallacy, though there have been some attempts to bridge the aggregate–individual gap. They also are subject to modifiable areal unit problem (MAUP); there is no such thing as a set of "natural" spatial units; therefore all the statistical results are dependent upon the scale and configuration of spatial units engaged.

Methods for Individual Data

Purely individual-based research can be seen as a simple extension of the primitive approach; the dataset is composed of individuals' characteristics plus their locational properties. The locational properties simply refer to which regional unit the individuals' residential locations belong to; the regionalization scheme could vary according to what spatial scale is considered. In most cases, a logistic regression model for a binary dependent variable is utilized to see if the locational variables really count when the individual characteristic variables are held constant. If the individual variables as well as the locational variables are all significantly related to electoral choice, then one could conclude that contextual effects on voting are important.

Individual-based methods, in many cases, combine individual-level survey data and aggregate census data. Rather focusing on where people live, the hybrid approach pays more attention to the characteristics that an area possesses. Those areal characteristics are captured by compositional variables (rarely by contextual ones) which usually include ones related to socioeconomic performance, class or occupational structure, and demographic structure. The variables are first obtained from census in an aggregated form at a particular spatial scale, and the variables are then coded onto the individual-level data according to individuals' residential location, so that each individual is allocated a separate set of areal attributes. Once all the variables are prepared, a regression model is fitted.

One of the most advanced methodologies to implement the hybrid model is multilevel modeling. It is contended that multilevel models operate at more than one scale, so that a single model can handle the microscale of people and the macroscale of places simultaneously. Multilevel modeling is expected to resolve the analytical dilemma not only between individual scale and aggregate scale but among different aggregate scales. Here it may be sufficient to provide how a multilevel model is specified for an empirical study. One might take a three-level model where level one is the individual level, level two the local level, and level three the regional level. Independent variables at the level one may include some individual variables such as age, occupational class, housing tenure, employment status, and so on, and compositional (and/or contextual) variables at the level two may include unemployment rate, unemployment rate change, and proportion of workers in a particular industry. By conducting a multilevel modeling, one may find that the relationships of the dependent variable not only with each of the independent variables at the level one but with each of ones in the level two vary region to region (the level three). Empirical studies utilizing the multilevel modeling report that typically 10%–20% of the variance in the response is attributed to contextual effects.

Some Lessons From Spatial Data Analysis

Recent advances in spatial data analysis and geographical information Science (GISc) have increasingly provided new insights into neighborhood effect research, though their applicability still remains limited. One of the most important sources of the limitation is that there are some conceptual and methodological difficulties in applying spatial data analysis techniques to neighborhood effect modeling. In spatial data analysis terms, the neighborhood effect is associated not with first-order effects but with second-order effects. In general, first-order effects relate to a global or deterministic or large-scale trend, while second-order effects refer to local or stochastic or small-scale variations, resulting from the spatial dependence in the process, that is, the tendency of neighboring locations to be correlated with each other in terms of the deviations from the first-order effects. In short, the first-order effects are associated more with compositional effects while the second-order with contextual effect.

In the individual-based approach, some second-order point pattern analysis techniques could provide a promising room for application. When voters' locations are geocoded, Ripley's *K*-function can gauge the degree of the spatial clustering of voting patterns. A bivariate *K*-function, however, could provide more insights into the nature of neighborhood effects. For example, if the geocoded voters' locations are given a categorical value, one for supporters of a particular party, and two for supporters of the other party, then we have two sets of point patterns. By applying a bivariate *K*-function, one could gain an understating of whether they are spatially independent or not. In a regression setting, aggregate or individual, spatial linear regression models such as various types of spatial autoregressive models could edify the problem of spatial autocorrelation in residuals endemic to the ordinary least square (OLS) regression framework which has overwhelmingly been adopted in the neighborhood effect studies.

In the cases of aggregate data, various techniques of exploratory spatial data analysis (ESDA), particularly utilizing local measures of spatial association such as local Moran's *I* and Getis-Ord's G_i^*, could broaden the analytical spectrum by allowing for exploring spatial dependence and spatial heterogeneity which voting patterns usually display. Such ESDA techniques, however, may be more useful when a particular assumption on the spatial scale at which neighborhood effects operate is satisfied: the operational scale should be larger than the measurement scale, that is, the (average) size of spatial units. Only when the assumption turns out to be conceptually and practically sustainable, (positive) spatial dependence observed in voting patterns can be regarded as a circumstantial hint of the existence of "interlocal" neighborhood effects as opposed to "intralocal." However, it should be noted that many of ESDA techniques do not allow for heterogeneity in the first-order effect, that is, constant compositional effect over space, which would not be tenable in most cases under investigation. Thus, genuine (interlocal) contextual effects can be tackled only when the first-order, compositional parts are removed from the original variance of voting patterns.

Geographically weighted regression (GWR) can be seen as a method to deal with intralocal and interlocal neighborhood effects simultaneously. One of the most important rationales of GWR is that local areas possess their own peculiarities which are irreducible to the global trend. Thus, GWR allows for spatial heterogeneity of causal relationships between variables, that is, spatially drifting regression parameters. Meanwhile, GWR fulfills the task by having values at original locations referencing ones at their adjacent locations in the estimation process. In the context of neighborhood effects, GWR can show how spatially different the relationship of the dependent variable with each of not only compositional variables but contextual variables is (intralocal neighborhood effects) and how similar each of the relationships between adjacent local areas (interlocal neighborhood effects) is.

Some critical issues for future research need to be addressed. The first issue revolves around how to define a neighborhood: Is there any natural foundation on which neighborhoods are delineated? How can we be sure that our data units are congruent with the concept of neighborhood? What is the spatial extent of a typical neighborhood? Some GISc-based approaches such as automated neighborhood identification procedures open up a promising space for this issue. Second, there is another scale issue, that is, the multiscale nature of neighborhood effects. To gain understanding of the whole picture of neighborhood effects, it may be desirable to investigate the hierarchical and/or interlocking nature of the effects occurring at different spatial scales. Recent advances in multilevel modeling could provide a viable starting point to address this issue. The third issue is about how to specify neighborhood characteristics, either compositional or contextual. In order to adequately model neighborhood effects, it is crucial to possess explanatory variables which properly measure socioeconomic and environmental conditions of a place. Some efforts have recently been made to devise composite measures of neighborhood characteristics such as measures of areal deprivation by exploiting a variety of data sources in a GIS environment.

See Also: Cartography, Electoral; Ecological Fallacy; Exploratory Spatial Data Analysis; Geographically-Weighted Regression; Geographic Information Science and Systems; Modifiable Areal Unit Problem; Point Pattern Analysis; Regression: Linear and Nonlinear; Spatial Autocorrelation; Spatial Clustering, Detection and Analysis of; Spatially Autoregressive Models; Statistics, Overview; Statistics, Spatial.

Further Reading

Agnew, J.A., 1996. Mapping politics: how context counts in electoral geography. Polit. Geogr. 15, 129–146.
Anselin, L., Cho, W.K.T., 2002. Spatial effects and ecological inference. Polit. Anal. 10, 276–297.
Bailey, T., Gatrell, A.C., 1995. Interactive Spatial Data Analysis. Longman, Harlow, Essex.
Cox, K.R., 1969. The voting decision in a spatial context. Prog. Geogr. 1, 81–117.
Dietz, R.D., 2002. The estimation of neighborhood effects in the social sciences: an interdisciplinary approach. Soc. Sci. Res. 31, 539–575.
Eagles, M. (Ed.), 1995. Spatial and Contextual Models in Political Research. Taylor & Francis, London.
Haining, R., 2003. Spatial Data Analysis: Theory and Practice. Cambridge University Press, Cambridge.
Johnston, R.J., Pattie, C.J., 2006. Putting Voters in Their Place: Geography and Elections in Great Britain. Oxford University Press, New York.
Johnston, R.J., Propper, C., Sarker, R., Jones, K., Bloster, A., Burgess, S., 2005. Neighbourhood social capital and neighbourhood effects. Environ. Plan. 37, 1443–1459.
Jones, K., Duncan, C., 1996. People and places: the multilevel model as a general framework for the quantitative analysis of geographical data. In: Longley, P., Batty, M. (Eds.), Spatial Analysis: Modelling in a GIS Environment. GeoInformation International, Cambridge, pp. 79–104.
King, G., 1996. Why context should not count. Polit. Geogr. 15, 159–164.
O'Loughlin, J., 2003. Spatial analysis in political geography. In: Agnew, J., Mitchell, K., Toal, G. (Eds.), A Companion to Political Geography. Blackwell Publishing, Oxford, pp. 30–46.
Sykes, B., Kuyper, H., 2009. Neighbourhood effects on youth educational achievement in The Netherlands: can effects be identified and do they vary by student background characteristics? Environ. Plan. 41, 2417–2436.