Chapter 8
At the Frontier Between Local and Global Interactions in Regional Sciences

Gary Cornwall, Changjoo Kim, and Olivier Parent

8.1 Introduction

Regional scientists have long stressed the importance of spatial spillover effects on local economic outcomes. In his seminal work, Marshall (1890) emphasizes that when economic agents locate in close proximity, they can take advantage of market interactions, knowledge spillovers, and linkages between intermediate and final goods producers. Due to such conveniences, people tend to cluster at specific locations and benefit from the subsequent agglomeration of economies. This clustering not only ends up providing conveniences in markets and economic activity but also fosters, at some level, local growth and development. Measuring the extent to which spillovers are localized remains a key challenge to empirical work in the field. By considering the role of geographic proximity in evaluating spillover effects, LeSage (2014) illustrates the fundamental role of appropriate model specification.

A spatial spillover arises when the decision or outcome of an agent is influenced by a corresponding decision or characteristic of some neighboring agent. Feedback effects are observed when this influence is projected back upon the original agent via a first order reaction to the neighbor's new decision. Spillovers are said to be global when endogenous feedback effects are present.

With the emergence of social network models (Manski 1993; Brock and Durlauf 2001; Bramoullé et al. 2014), researchers have been interested in new forms of

G. Cornwall • O. Parent (✉)
Department of Economics, Carl H. Lindner College of Business, University of Cincinnati, Cincinnati, OH, USA
e-mail: cornwagj@mail.uc.edu; olivier.parent@uc.edu

C. Kim
Department of Geography, University of Cincinnati, Cincinnati, OH, USA
e-mail: changjoo.kim@uc.edu

R. Jackson, P. Schaeffer (eds.), *Regional Research Frontiers - Vol. 2*, Advances in Spatial Science, DOI 10.1007/978-3-319-50590-9_8

local or group interactions based on spillovers and social distance. Economic agents belonging to the same cluster tend to behave similarly. New spatial econometrics models have been developed to incorporate intragroup interaction (Lee 2007). Similar to the local spatial spillover effect, those models assume that interaction is limited and does not spread across clusters. Interactions between agents do not spill across cluster boundaries, and within a cluster the same weight is often attributed to all individuals leaving aside geographical or social group-wise variations. A clear distinction is made with local spatial spillovers which do not involve endogenous feedback effects. LeSage (2014) discusses in detail the distinction between global and local specifications, advocating respectively for the implementation of the Spatial Durbin (SDM) and the Spatial Durbin Error Models (SDEM).

One of the primary challenges in analyzing interactions amongst economic agents is the inherent complexity in their connectivity structure or network. In standard peer effects models, the local interaction effects represent strategic complementarity in effort across neighboring agents. An agent's incentive to make a particular decision increases as the number of neighboring agents making a similar decision increases. Strategic complementarities correspond to positive partial cross-derivatives. In addition to local complementarities, global interactions across all agents have recently been introduced by Ballester et al. (2006) to reflect strategic substitutability.

Interdependencies can take a variety of forms and little is currently known about their structure. As researchers become more skilled at leveraging geographic information system (GIS) technologies, new types of data will improve the understanding of spatial interactions. Defining a suitable topological structure for network modeling can present a number of GIS challenges and, in general, empirical work has yet to really analyze the transmission of interactions among economic agents. Future research in regional science will greatly benefit from properly specifying the endogenous process that makes economic agents connected. Assuming that connections between agents are mainly explained by exogenous geographical proximity is overly restrictive and could cast serious doubt on causal interpretations of spillover effects. To evaluate the magnitude of local spillover effects, empirical studies in regional science have been exclusively implementing either an SDEM or the so-called SLX model containing exogenous interaction effects. Future research will acknowledge that feedback effects could play an important role in explaining local spillovers effects while being restricted to a limited set of observations or neighborhoods. Moreover, new models will accommodate the possibility that local externalities do not conform to administrative boundaries and will allow for more heterogeneity in the level of spatial dependence.

The remainder of this chapter addresses these challenges as follows. The following section presents modeling issues related to spatial network analysis specifically oriented to GIS. Section 8.3 discusses the limit of a spatial interaction model when regions or groups of society are well delineated. Section 8.4 questions the central issue of endogeneity in the interaction structure. Section 8.5 proposes new spatial mixture models allowing for parameters to be heterogeneous across clusters, and cluster membership is not known to the econometrician. Section 8.6 concludes and points at future work.

8.2 Identifying Networks Using GIS

Regional scientists have long been paying attention to whether agents in close geographical, social, or virtual proximity interact with each other. Their interactions create a conduit by which information is transmitted, and form the fabric of regional development, all of which demands the attention of researchers. The combination of mobile technology and comprehensive datasets have changed how agents interact across space, and new approaches to both local and global interactions will be developed in future regional science research. Today, data is available in exceptional volume and easily accessed over current communication networks more than ever before and has created a new dimension in the study of regional science. In addition to the extended network, GIS has now advanced into new spheres, such as the modeling and analysis of spatio-temporal networks facilitating the understanding of decision making. Despite the great potential, Brugere et al. (2014) consider the intersecting research between spatial networks in GIS and temporal networks in related fields still in its infancy.

Mobile communication tools allow interactive data publishing, which tracks how agents interact with each other and records under what dimensions they are connected. No longer is this data restricted to geographic boundaries and often is contextualized in network structures through social media (i.e. Twitter, Facebook, LinkedIn, etc.). These platforms diminish the importance of traditional measures of distance and, instead, create relationships that may be tangential to those same measures but nevertheless of great importance. Geo-demographics generated in these virtual environments have a great deal of potential when measuring spatial spillover effects. It is now convenient to analyze populations based on who and where under a less restrictive spatial paradigm.

Mobile telecommunications technologies are contributing significantly to the voluminous amount of data being generated by daily online activities. Cameras, phones, and cars have been, and are being, infused with location-aware software designed in some capacity to give producers insights into consumer activities. These devices have, in effect, begun to sense and communicate their absolute and relative positions with locational tags providing a significant medium for organizing, browsing, and retrieving interactions across space. Location-based services have begun to make use of geographic position by identifying the local (global) network of related devices and people across the world.

GIS can also generate social or virtual proximity that could help to detect spatial dependence among individuals beyond physical boundaries as well as geographical proximity. GIS has been playing a significant role in identifying and generating a realistic network of spatial interaction of social processes. With the help of GIS, networks can be developed at the resolution of individual people by their connections. This often requires that large amounts of interaction data are managed and manipulated across scales. Identifying and building a network of massive and hidden connections using GIS is potentially of great value in regional science in providing new tools for advanced model building and in adding spatial dimension and spatial thinking into regional science. Modeling interaction data in both physical and virtual environments will be future challenges in dealing with local and global interactions in regional science.

8.3 Groupwise Spatial Dependence and Spatial Fixed Effects

Researchers have recently recognized the importance of spatial econometric models in identifying and estimating social interaction models. In the empirical literature of regional science, a region, district, or a group of society can be considered a spatial unit whose neighboring units could be defined in terms of a certain socio-economic or physical distance.

One key challenge is to identify the main determinants of the correlation between outcomes of those spatial units who interact with each other. In a seminal work, Manski (1993) points out the difference between endogenous effects capturing the influence of peer behavior and the contextual effects measuring the influence of exogenous peer characteristics. He also mentions the importance of unobserved, correlated effects capturing the likelihood of units to behave similarly due to the similarity of characteristics and/or environment.

Consider some population of n spatial unit for which y_i is the outcome of individual $i = (1 \ldots, n)$. To model how individual units exert some influence on each other, we assume that this influence could be mediated by a network of peer relationships or any socio-economic or physical distances. To constrain those influences, each spatial unit belongs to a group. The interaction between units may occur within a group but not across. For each group $r = (1, \ldots, R)$, we observe n_r units, where $n = \sum_{r=1}^{R} n_r$. As explained in Lee (2007), a group interaction model based on a block diagonal matrix $W = diag(W_1, \ldots, W_R)$ for which each element $w_{ij,r} = 1$ if i and j are direct neighbors or friends, and $W_{ij} = 0$, otherwise.

Lee (2007) and Bramoullé et al. (2009) have rewritten the generic neighborhood effects model described by Manski (1993) as the following Spatial Durbin Autoregressive specification for each group r as:

$$Y_r = \rho W_r Y_r + X_r \beta + W_r X_r \gamma + \iota_{n_r} \alpha_r + \epsilon_r \tag{8.1}$$

where ϵ_r is a n_r-dimensional vector consisting of i.i.d. disturbances with zero mean and a variance σ_2. X_r is an $n_r \times k$ matrix of explanatory variables and Y_r is the n_r-dimensional vector of observation in the rth group.

The spatial weight matrix reflects in principle the structure of the interaction process, and ignoring this process when one is present will induce a misspecified model. The consequence of such a misspecification is that estimates will be biased and inferences will be misleading. To better understand the issue, the reduced form of the spatial lag model can be rewritten as:

$$Y_r = (I_{n_r} - \rho W_r)^{-1}(X_r\beta + W_r X_r \gamma + \iota_{n_r}\alpha_r) + (I_{n_r} - \rho W_r)^{-1}\epsilon_r \tag{8.2}$$

where $(I_{n_r} - \rho W_r)^{-1}\epsilon_r$, is now a spatially correlated and heteroskedastic error term. By using the Taylor's series for the inverse matrix,

$$(I_{n_r} - \rho W_r)^{-1} = I_{n_r} + \rho W_r + \rho^2 W_r^2 + \ldots + \rho^n W_r^n \tag{8.3}$$

Magnitude and significance of spillover effects are assessed via the partial derivatives of the expectation of y_r. LeSage and Pace (2009) show that direct effects are based on the diagonal elements of (8.3), while the off-diagonal elements contain the indirect or spillover effects. An important characteristic of these models is that spillovers only spread within each group or neighborhood r. Unlike a traditional model, they are not global anymore and do not spread across all neighborhoods.

One way to define the neighborhood structure is to assume that all individuals in the same group are neighbors of each other. Each element $w_{ij,r}$ of the spatial weight matrix W is now equal to $1/(n_r - 1)$, and each $n_r \times n_r$-dimensional block matrix W_r can be rewritten as

$$W_r = [1/(n_r - 1)]J_{n_r} - [1/(n_r - 1)]I_{n_r} \tag{8.4}$$

where $J_{n_r} = \iota_{n_r}\iota'_{n_r}$, ι_{n_r} is an $n_r \times 1$ dimensional vector of ones, and I_{n_r} is an identity matrix of dimension n_r. The reduced form of Eq. (8.1) would involve the following inverted matrix for each block r:

$$(I_{n_r} - \rho W_r)^{-1} = \delta_{1,n_r} J_{n_r} + \delta_{2,n_r} I_{n_r}, \tag{8.5}$$

where $\delta_{1,n_r} = \rho/((n_r - 1 + \rho)(1 - \rho))$ and $\delta_{2,n_r} = (n_r - 1)/(n_r - 1 + \rho)$. This model has received substantial attention in the spatial econometric literature for social interaction (Lee 2007). It is important to note that the spatially lagged dependent variable Wy asymptotically becomes proportional to the unit vector. In this case, a spatial fixed effects model is asymptotically equivalent to the SDM with group-wise weights. Spatial correlation should disappear by removing the fixed effects.

A spatial fixed effects specification seems appropriate when individual observations are distributed across well-defined groups for which some characteristics α_r are unobserved. However, there are two main issues that are associated with the use of spatial fixed effects. First, the fixed effects are influencing in an identical fashion all observations within a group. If the data were to exhibit heterogeneity or spatial interaction across neighboring individuals within a group, the result would produce correlation in the error term. In this case, the spatial fixed effects would not correct for the presence of spatial correlation, and the model would be misspecified. Second, and more importantly, the spatial delineation of groups or neighborhood is often ambiguous. There is no reason why administrative districts should be used to delineate spatial areas, except as a matter of convenience. Incorrect delineation might exacerbate spatially correlated and heteroskedastic error terms and create additional model misspecification. In other words, unless the structure of the model results in a set of group-wise constants equivalent to the fixed effects, the inclusion of spatial fixed effects will not be robust to the model misspecification.

8.4 Endogeneity in Dependence Within Groups

A key issue with the causal interpretation of estimates in the peer effect Eq. (8.1) is that the connectivity structure between agents may be endogenous. Spatial econometrics has typically been relying on the ad hoc assumption of exogeneity for the spatial weight matrix. This very strong assumption might not be reasonable when assessing the influence of decisions from neighboring agents. In assessing fiscal policy interdependence and budget spillovers across states, Case and Rosen (1993) underline that economic similarities between regions are more likely to exert influence on each other rather that simply sharing a common border. Several subsequent studies have questioned the narrowly defined connectivity structure that relies exclusively on geographical proximity (see Kelejian and Piras 2014). The main concern has become that estimates of regression that do not account for the endogeneity of the spatial weight matrix should suffer from bias, casting doubt on causal interpretations of the peer effects (Qu and Lee 2015).

By modeling group formation, Jackson (2008) makes the assumption that the decision between two agents to form a link is the outcome of two choices. The net utility stemming from the agreement to form a link can be seen as positive. The utility for agent i to form a link with agent j can be defined as $U_i(j)$ and, therefore, the interaction between both agents can be expressed as

$$D_{ij} = \mathbb{1}_{U_i(j)>0} \times \mathbb{1}_{U_j(i)>0} \tag{8.6}$$

In this framework, each potential pair of neighboring agents evaluates the utility of a link between them at the same point in time. The important implication is that those individual utilities depend on the characteristics of the two individuals, conditional on the network at the beginning of the period. Goldsmith-Pinkham and Imbens (2013) propose a Bayesian estimation procedure that separates the likelihood function of the network formation from the likelihood function of the outcome. They find that indirect effects coming at least from the second order neighbors (friends-of-friends) are hard to assess and largely driven by the functional form assumption that ties these indirect effects. The main issue in developing models that allow for endogeneity in the interaction structures between individuals is to define a rule that keeps them separate from each other. As explained by Qu and Lee (2015), estimating a connectivity structure that relies purely on economic distance might be challenging. He underlines the importance of imposing restrictions on the spatial weights, which depend not only on the ad hoc geographical distance but also the magnitude of neighboring effects through socio-economic distance. Interesting extensions would include an examination of how endogeneity over time might change the interaction structure. For all of those situations, the task of properly estimating direct and indirect effects remains daunting.

8.5 Unobserved Dependence Across Groups

It is a common practice in regional science to adopt administrative boundaries for convenience (e.g., census tract or census block boundaries). There is no reason, however, to believe that social interactions will remain within such boundaries. In fact, it is likely that generic neighborhood effects (such as crime, air quality, employment search, etc.) will not conform to such boundaries and will have heterogeneous areas.

As explained in Autant-Bernard et al. (2007), spatial spillovers may occur through collaborative networks (social, scientific, technological, etc.) giving rise to myriad forms of spatial interaction. The geographical dimension of spillover effects appears to be closely related to other mechanisms that are barely measurable. Clusters of individuals should not only rely on geographical proximity. We often observe that across neighboring observations, two individuals might exhibit different patterns or, more specifically, if we consider those patterns to be probabilistic in nature, different distributions. In fact, an aspect that is often overlooked is the considerable heterogeneity of behavior across individuals whether they belong to the same neighborhood or not. Though unobserved heterogeneity across clusters is more difficult to take into account, there is a rapidly growing literature in econometrics using mixture models (see Keane and Wasi (2013) for a review). These models account for unobserved heterogeneity by assuming the data are drawn not from a single distribution but from a finite number of distributions. In fact, they assume, different agents in the population have varying preferences and estimate the proportion of each type.

Cornwall and Parent (2016) consider estimation of spatial data models when the parameters are heterogeneous across groups, and group membership is not known to the econometrician. Thus, they allow parameters to be homogeneous within a group but heterogeneous across groups. This is a form of model-based clustering which partitions a set of data, y_i into G groups according to how near they are to one another. This is easily distinguishable from the aforementioned analysis in which the objective is to understand how the delineated groups differ. It is also important to note that they are allowing the parameters to vary across groups rather than confining themselves to marginal effects, which differ through splitting the sample based on the values of regressors.

Model-based clustering takes as a starting point that a set of data with a group structure is generated by a mixture of distributions such that an observation drawn from sub-population g has density $f_g(y_i|\beta_g, \sigma_g^2)$. If z_i is the identifying label, i.e., $z_i = g$ if unit i belongs to group g, then one can define the dependent variable y_i as being drawn from g different normal distributions with probability $p(z_i = g) = w_g$ and $\sum_{g=1}^{G} w_g = 1$. The normal mixture distribution has means and variances that are different for each group g:

$$P(y_i|\beta, \sigma^2, p_g) = \sum_{g=1}^{G} w_g N(X_i\beta_g, \sigma_g^2). \tag{8.7}$$

We define by $I_g = \{i : z_i = g\}$ the set of agent belonging to the mixture component g and whether an individual belongs to a mixture component g is not known. Cornwall and Parent (2016) develop a spatial extension for which a new dependent variable is defined as $\tilde{y}_{i,r} = y_{i,r} - \rho \sum_{j=1}^{n_r} w_{ij,r} y_{j,r}$, where $w_{ij,r}$ represents the neighborhood structure as defined in (8.4) that is typically based on geographical proximity. This spatial model could be easily extended to the SDM presented in (8.1). In fact, the spatial mixture would then take the following expression:

$$P(\tilde{y}_{i,r}|\beta, \sigma^2, p_g) = \sum_{g=1}^{G} w_g N(\alpha_{r,g} + X_{i,r}\beta_g + \sum_{j=1}^{n_r} w_{ij,r} X_{j,r}\gamma_g, \sigma_g^2). \tag{8.8}$$

Bayesian estimation procedures can be adopted to estimate this model. The introduction of spatial mixtures of distributions relaxes the assumption of independence between observations whether they belong to the same mixture or not. Geographical proximity generates spatial dependence across neighboring individuals even if they exert different behavior and are not part of the same mixture.

8.6 Conclusion

With the increased interest in social interaction, research in regional science has gradually moved from a pure spatial definition of neighboring effects toward a multidimensional measure relying on a different form of socio-economic distances. The emergence of social networking tend to show that agents belonging to a network might not be in close geographical proximity. Moreover, there is no reason why neighborhood effects should be delineated across well-defined groups. It is possible for neighborhood effects to spill over administrative boundaries, and this possibility must be accommodated when modeling such processes. The difficulty in detecting and measuring spillover effects call for a stronger theoretical basis of the interaction structure. Simple weight matrix based on geographical distance might not be enough. Future work will need to rely on the endogeneity of those interactions along with the heterogeneity of behavior that is influenced by physical and socio-economic distance. Promising future direction in regional science will utilize GIS to incorporate data-rich sources from physical and virtual networks to better assess the magnitude of spillover effects.

References

Autant-Bernard C, Mairesse J, Massard N (2007) Spatial knowledge diffusion through collaborative networks. Pap Reg Sci 86:341–350
Ballester C, Calvo-Armengol A, Zenou Y (2006) Who's who in networks. Wanted: the key player. Econometrica 74:1403–1417
Bramoullé Y, Djebbari H, Fortin B (2009) Identification of peer effects through social networks. J Econometrics 150:41–55
Bramoullé Y, Kranton R, D'amours M (2014) Strategic interaction and networks. Am Econ Rev 104:898–930
Brock W, Durlauf S (2001) Discrete choice with social interactions. Rev Econ Stud 68:235–60
Brugere I, Gunturi V, Shekhar S (2014) Modeling and analysis of spatio-temporal social networks. In: Encyclopedia of social network analysis and mining. Springer, New York, pp 950–960
Case AC, Rosen HS (1993) Budget spillovers and fiscal policy interdependence: evidence from the states. J Public Econ 52:285–307
Cornwall GJ, Parent O (2016) Mixture models with spatial dependence. Working paper, University of Cincinnati
Goldsmith-Pinkham P, Imbens GW (2013) Social networks and the identification of peer effects. J Bus Econ Stat 31:253–264
Jackson M (2008) Social and economic networks. Princeton University Press, Princeton
Keane MP, Wasi N (2013) Comparing alternative models of heterogeneity in consumer choice behavior. J Appl Econometrics 28:1018–1045
Kelejian H, Piras G (2014) Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. Reg Sci Urban Econ 46:140–149
Lee LF (2007) Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. J Econometrics 140:333–374
LeSage JP (2014) What regional scientists need to know about spatial econometrics. Rev Reg Stud 44:13–32
LeSage JP, Pace, RK (2009) An introduction to spatial econometrics. Chapman Hall/CRC Press, Boca Raton, FL
Manski C (1993) Identification of endogenous social effects: the reflection problem. Rev Econ Stud 60:531–542
Marshall A (1890) Principles of economics. Macmillan, London
Qu X, Lee LF (2015) Estimating a spatial autoregressive model with an endogenous spatial weight matrix. J Econometrics 184:209–232

Gary Cornwall is a Ph.D. candidate at the University of Cincinnati. His primary research interests are spatial econometrics, income dynamics/inequality, and Bayesian econometric methods. He received his M.A. in Applied Economics from the University of Cincinnati in 2012 and is expected to complete his Ph.D. in 2017.

Changjoo (CJ) Kim is an associate professor, Department of Geography, University of Cincinnati. His research and teaching interests are in geographic information science, transportation, and spatial modeling. His research addresses theoretical and practical questions in urban and economic geography through the application of GIS methods. He investigates a range of urban and economic concerns including travel activity, urban sprawl, commuting, airline industry, retailing, etc. Dr. Kim earned the Ph.D. in geography from The Ohio State University in 2004.

Olivier Parent is associate professor, Department of Economics, at the University of Cincinnati (UC). His primary research interests are spatial econometrics; regional economics; and Bayesian econometrics. Dr. Parent earned the Ph.D. in economics from the University of Saint-Etienne, France, in 2005.