CHAPTER 3

Spatial Methods

Steve Gibbons*, Henry G. Overman*, Eleonora Patacchini[†]
*London School of Economics, London, UK
[†]Cornell University, Ithaca, NY, USA

Contents

Abstract

This chapter is concerned with methods for analyzing spatial data. After initial discussion of the nature of spatial data, including the concept of randomness, we focus most of our attention on linear regression models that involve interactions between agents across space. The introduction of spatial variables into standard linear regression provides a flexible way of characterizing these interactions, but complicates both interpretation and estimation of parameters of interest. The estimation of these models leads to three fundamental challenges: the "reflection problem," the presence of omitted variables, and problems caused by sorting. We consider possible solutions to these problems, with a particular focus on restrictions on the nature of interactions. We show that similar assumptions are implicit in the

Handbook of Regional and Urban Economics, Volume 5A
ISSN 1574-0080, http://dx.doi.org/10.1016/B978-0-444-59517-1.00003-9

empirical strategies—fixed effects or spatial differencing—used to address these problems in reduced form estimation. These general lessons carry over to the policy evaluation literature.

Keywords

Spatial analysis, Spatial econometrics, Neighborhood effects, Agglomeration, Weights matrix

JEL Classification Codes

R, C1, C5

3.1. INTRODUCTION

This chapter is concerned with methods for analyzing spatial data. When location is simply a source of additional information on each unit of observation, it adds little to the complexity of analyzing and understanding the causes of spatial phenomena. However, in situations where agents are able to interact, relative locations may play a role in determining the nature of those interactions. In these situations of spatial interdependence, analysis is significantly more complicated and the subject of ongoing epistemological and methodological debate. It is these issues that are the focus of this chapter.

Even when units of observation can be located in some space, it is possible that location is irrelevant for understanding data pertaining to those units. In such circumstances it makes sense to think of the spatial dimension as random—a concept that can be made precise using notions from spatial statistics (Cressie, 1993; Diggle, 2003). In contrast, when location matters, the spatial dimension is nonrandom and our understanding of the data will be increased if we can allow for and explain this nonrandomness. Such nonrandomness is pervasive in areas of interest to urban economics. Why do individuals and firms concentrate geographically in dense (urban) areas? How does concentration affect outcomes and how does this explain why some cities perform better than others? To what extent do firms in particular industrial sectors cluster geographically? Why does this clustering happen and how does it influence outcomes for firms? Is the spatial concentration of poverty within cities a manifestation or a determinant of individual outcomes? Does location determine how individuals, firms, and other organizations, including government, interact and if so, how does this help us understand socioeconomic outcomes?

Answering such questions about nonrandomness is clearly central to increasing our understanding of how urban economies function. Unfortunately, as we explain in detail below, detecting departures from nonrandomness is not always straightforward. Distinguishing between the causes of nonrandom spatial outcomes is exceptionally difficult, because it requires us to distinguish between common influences and interaction effects that might explain the observed nonrandomness. For example, all individuals that live in New York City may be affected by the density of the city, its cost of living, or many other shared environmental factors. As a consequence, their outcomes—such as wages, health, behavior, and well-being—change together as these factors change. However, this correlation of outcomes across individuals need not imply that these individuals directly influence each other. If, in contrast, individual New Yorkers' behavior is directly influenced by (expectations of) the behavior of other New Yorkers, then the correlation across individuals is the result of social interactions.

Consideration of these issues is further complicated by the fact that the terminology used to talk about these effects is often imprecise and dependent on the disciplinary background. For example, "spatial interactions," "social interactions," "neighborhood effects," "social capital," "network effects," and "peer effects" are all terms that are often used synonymously but may have different connotations (Ioannides, 2013). These differences in terminology may also reflect important differences in the theoretical models that underlie empirical specifications. For example, in the network effects literature, the definition of an interaction effect is often based on interdependent objective functions (utility, profit, etc.). If my utility (and choice) is based on yours and vice versa, the equilibrium outcomes observed in the data are a complex function of both utility functions. Common influences do not imply such interdependency. However, social interactions defined more broadly need not involve such direct interdependency in objective functions (Manski, 2000). Social interactions may involve the availability of information, for example, about the value of education, job opportunities, or one's own ability (Banerjee and Besley, 1991). Or they may arise because of the effect that one person's actions have on another owing to the constraints they both face, for example, when one child's misbehavior diverts a teacher's attention from another child, allowing them to misbehave (which is a standard explanation of educational peer effects). In contrast, in the spatial econometrics literature, spatial interactions in outcomes may be posited for individual-level or area-level outcomes with no reference made to any underlying objective function or any other economic microfoundations. Of course, this begs the question whether one could microfound such models without recourse to interdependent objective functions. Many models within the new economic geography tradition show that this is indeed possible. In the Krugman (1991b) core-periphery model, for example, firms are sufficiently small that they ignore their impact on other firms (and hence ignore reactions from those firms), while workers' utility functions depend only on consumption of a continuum of manufacturing sector varieties and an agricultural good (not directly on the utility of other workers). Yet in these models the location of both firms and workers is interdependent in equilibrium.[1] Similarly, in the urban peer effects literature, Benabou (1993) shows how segregation can arise when the skill of neighborhood peers affects the costs of acquiring skills (in schools), and how this in turn can affect the incentives to

[1] Similarly, a range of search models can also be used to provide microfoundations for spatial interactions without the need for interdependent objective functions. See, for example, Patacchini and Zenou (2007) and Zenou (2009).

acquire skills. Epple and Romano (2011) review a range of other theoretical models that explain social interactions without directly interdependent objective functions.

Regardless of the terminology, recent research on spatial econometrics (and the related literature on network effects) has shown that the nature of the interconnection between individuals, firms, or places is crucial when it comes to identifying parameters or causal effects in spatial models that involve interactions. This literature has given us a far better understanding of the kind of data-generating processes where we can, in principle, distinguish between the different causes of nonrandomness and the information that is then needed to do so in practice. In particular, it is important to distinguish between two broad types of interaction structure. On the one hand, there is the context where a group of individuals or firms may influence one another jointly. For example, all firms in a cluster, or individuals in a neighborhood, may jointly impact each other. Estimation in this case would look to determine, for example, whether cluster-level R&D spending determines firm-level R&D spending[2] or if the local crime rate is relevant to explain the individual propensity to commit crime.[3] In this case the interaction scheme is complete because all agents in a given group are connected to all others in the group.

Distinguishing between a common influence and an interaction effect in this setting is particularly challenging, because when one estimates the propensity of a firm or individual to make a decision as a function of the average behavior of its group, a unique type of endogeneity arises. In particular, if outcomes are modeled as a linear function of group outcomes (e.g., R&D), and exogenous individual and group characteristics (e.g., firm age and average firm age), it becomes difficult to distinguish between the influence of the group outcome and other group-level characteristics. Econometrically, problems arise because group-averaged outcomes are perfectly collinear, or nearly collinear, with the group-averaged exogenous variables unless specific types of restrictions are imposed on the structure of interactions, or on other aspects of the specification. Conceptually, the issue is that the average outcome for the group is an aggregation of outcomes or behaviors over other group members, and hence is an aggregation of individual characteristics over other group members. This problem is known as the "reflection problem" (Manski, 1993). It is an often misunderstood problem, which frequently results in the inappropriate interpretation of neighborhood and peer effects. Specifically, positive significant coefficients on group averages are often misinterpreted as identifying endogenous social interactions even in situations where the full set of exogenous characteristics that determine behavior are not available. This problem is pervasive even in cases when assignment to groups is random as, for example, in Sacerdote (2001).

The alternative to complete interactions occurs in contexts where some, but not all, individuals or firms in a group influence one another: that is, the interaction scheme is

[2] See, for example, the extensive knowledge production function literature initiated by Jaffe (1989).

[3] Case and Katz (1991) provide an early example.

"incomplete." For example, firm-level R&D may be influenced by interaction with specific peers, rather than a cluster (or industry) as a whole.[4] If firm A interacts with firm B, firm B interacts with both firm A and firm C but firm C does not interact with firm A, the interaction scheme is not complete. In this case the influence of the group outcome and the influence of other group-level characteristics can, in principle, be separately identified. In a similar vein, individuals may be influenced by only some (rather than all) neighbors when taking decisions. If one can specify the details of such an incomplete interaction scheme, then this avoids the reflection problem. Indeed, this is the "solution" to the identification problem that has traditionally been (implicitly and artificially) imposed in the spatial econometrics literature through the use of standard, *ad hoc* spatial weight matrices (e.g., rook or queen contiguity). We discuss these issues in much more depth below.

Unfortunately, in practice, the number of situations where we have detailed information on the true structure of interactions is limited—especially in terms of common spatial interactions that may be of interest. The problems of distinguishing between different causes become even more pronounced in situations where we do not know all of the relevant individual factors or common influences that explain outcomes, and do not know the structure of interactions or whether the structure of interactions is endogenously determined (i.e., decisions of individual agents determine who is influenced, not just how they are influenced). In these situations, Gibbons and Overman (2012) propose adopting a reduced form approach, focusing on finding credibly exogenous sources of variation to allow the identification of causal processes at work. Again, we discuss these issues further below.

This chapter is organized as follows. We lay out some of the basic intuitions regarding the modeling of spatial data in Section 3.2 and provide more formal consideration in Section 3.3, focusing our attention on the linear regression model with spatial effects. This section also considers the distinction between spatial and social interactions. In Section 3.4 we consider issues relating to identification and estimation with observational data, with a particular focus on how the existence of spatial interactions might complicate the reduced form approach to identification. An alternative to focusing on the reduced form in quasi-experimental settings is to adopt an experimental approach where the researcher uses randomization to provide an exogenous source of variation. Such an approach is particularly associated with the estimation of treatment effects.

[4] The importance of networks has long been recognized in the literature on research productivity (broadly defined). However, empirical papers have tended to focus on the construction of summary statistics (i.e., social network analysis measures) for use as additional explanatory variables in knowledge production function specifications. See, for example, Abbasi et al. (2011) and Harhoff et al. (2013). A second literature uses shocks to networks as an exogenous source of variation in the composition of peers. See, for example, Borjas and Doran (2012). Only recently has the focus shifted toward network structure as a source of identification, as we discuss further in Section 3.4.

We devote Section 3.5 to the estimation of treatment effects in the presence of spatial interactions. Section 3.6 concludes the chapter.

3.2. NONRANDOMNESS IN SPATIAL DATA

Underlying all spatial data are units of observation that can be located in some space. Locational information provides us with the position of one observation relative to others (distance and direction) and can be recorded in a number of ways. In many examples we will be interested in physical locations, but the methods we discuss can be applied more broadly (e.g., to location within a nonphysical network). Figure 3.1 presents a stylized set of spatial data that allow us to introduce the basic identification problem. Each panel in this figure maps location for two groups of observations. Group membership is identified through the use of different symbols—hollow points to represent membership of group 1, solid points to represent membership of group 2. In the left-hand panel the location of all observations is randomly determined, while in the right-hand panel it is nonrandomly determined (with solid points over represented toward the South and West and hollow points over represented toward the North and East).

The precise meaning of randomness for this kind of spatial data can be formalized using concepts developed for the analysis of spatial point patterns (Cressie, 1993; Diggle, 2003). Traditionally, that literature has focused on the null hypothesis of complete spatial randomness, which assumes that space is homogeneous, so that points are equally likely to be located anywhere. As argued in Duranton and Overman (2005), this hypothesis is unlikely to be particularly useful in many economic situations where location choices are constrained by a range of factors. To address this problem, those authors propose comparing the distribution of the sample of interest with some reference distribution. In their specific application, the groups of interest are specific industry sectors, while the reference distribution is the location of UK manufacturing as a whole. Comparison to this distribution allows one to test for geographical clustering of specific sectors—in terms of both the extent of clustering and its statistical significance.

For given spatial data, randomness can be uniquely defined (either using the assumption of homogeneous space or relative to some reference distribution) but deviations

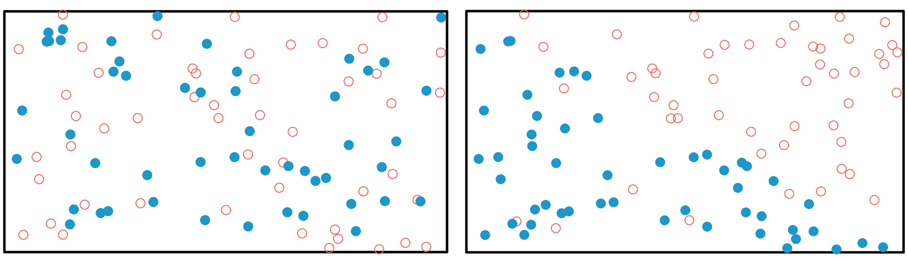

Figure 3.1 Randomness versus nonrandomness.

from randomness can happen along many dimensions. For example, in their study of segregation in the United States, Massey and Denton (1987) characterize racial segregation along five dimensions: evenness, concentration, exposure, clustering, and centralization. In contrast to these multiple causes of nonrandomness, tests for departures from randomness must be based on the calculation of index numbers that characterize the underlying distribution. A given index will have a unique distribution under the null hypothesis, but the power of the test will often depend on the causes of nonrandomness. In many cases, the distribution under the null cannot be derived analytically, leaving tests to rely on bootstrapping to determine appropriate test values. In short, while it may be conceptually simple to define randomness, detecting departures from randomness is more complicated in practice.

Until relatively recently, the mainstream economics literature largely ignored these problems and focused on the use of indices calculated using areal data (e.g., district, region) and constructed to characterize certain features of the data. For example, in the segregation literature, Cutler et al. (1999) use two indices of segregation. The first is a measure of dissimilarity which captures "what share of the black population would need to change areas for the races to be evenly distributed within a city." The second is a measure of isolation which captures the exposure of blacks to whites. Changes in both these indices over a long time period are then used to characterize the "rise and decline of the American Ghetto." In the international trade literature, similar indices such as the spatial Gini index and the Krugman specialization/concentration index (which is just two times the dissimilarity index) have been used to describe patterns of specialization and geographical concentration. Again, the focus has usually been on changes over time or on comparisons across geographical areas or industries rather than on the statistical significance of any departure from randomness. Ellison and Glaeser (1997) moved the literature closer to the statistical point pattern literature by worrying about the appropriate definition of randomness (specifically, the extent to which any index of spatial concentration should adjust for industrial concentration). But their criteria for high and moderate spatial concentration relied on the use of arbitrary cutoff points, defined with respect to the observed distribution of index values across industries rather than the underlying distribution of the index conditional on the assumption of randomness. Combes and Overman (2004) provide an overview and assessment of different measures.

Using ideas from the spatial point pattern literature, a number of authors have subsequently developed a new generation of tests for nonrandomness that can be applied to nonaggregated data with detailed location information. All of these tests use information on some moment of the bilateral distribution of distances between points to allow comparison of the sample with the reference distribution. Duranton and Overman (2005) make the case for comparison to be based on the density function for the full set of bilateral distances. In contrast, Marcon and Puech (2003) develop more traditional measures based on the use of cumulative distribution functions (Ripley's K and L; Ripley, 1976).

Subsequent contributions to this literature have developed alternative tests which differ in terms of the way in which the moments of the distribution of distances are used to assess for nonrandomness. Some of these alternative tests (e.g., those focusing on distances to the *k*-nearest neighbors) simplify calculations for large distributions—remembering that the number of bilateral distance calculations increases with the square of the number of sample points. Other authors (e.g., Klier and McMillen, 2008; Vitali et al., 2009; Ellison et al., 2010; Kosfeld et al., 2011) have suggested approximations or algorithmic improvements for tests based on the complete distribution of bilateral distances that similarly reduce computational complexity. Scholl and Brenner (2012) provide a relatively recent overview of different measures, while Scholl and Brenner (2013) provide discussion of computational issues. Debate still continues as to the "best" method for detecting departures from randomness. Our own view is that in situations where we wish to test for nonrandomness, the choice of the method is a second-order consideration relative to the first-order decision of whether or not to treat space as continuous. If the data allow it, using insights from the spatial point pattern literature and treating space as continuous, rather than discrete, allows for more powerful tests of nonrandomness.

Unfortunately, in many circumstances, researchers have access to only spatial aggregates for units of observations that correspond to areas rather than the individual units of observation. Duranton and Overman (2005) refer to this process of aggregation as moving from "points on a map to units in a box." Any such discretization and corresponding aggregation implies a loss of information and makes it harder to test for departures from randomness. Still, such areal data are often all that researchers have available to them. In these cases, tests for nonrandomness can be based on the concentration/segregation indices, discussed above, that have traditionally been used in the population and industrial location literature (such as the Herfindahl–Hirschman index, Krugman/dissimilarity index, and Ellison and Glaeser index; see, respectively, Herfindahl, 1959; Hirschman, 1964; Krugman, 1991a; Ellison and Glaeser, 1997) or on "global indicators of spatial association" developed in the spatial statistics and econometrics literature (such as Moran's I or Getis–Ord statistics; see, respectively, Moran, 1950; Getis and Ord, 1992).

Once we have applied one or more of these tests and rejected the null hypothesis of randomness, we may want to find out where within our geographical study area this nonrandomness occurs. For example, once we have established that crime is nonrandom across space in New York, we may want to visualize where in New York the crime hot spots occur. A range of spatial methods exist for doing just that, facilitated today by the integrated data analysis and mapping capabilities of geographical information systems (GIS) and related spatial software. Standard kernel density and spatial interpolation methods can be easily implemented in a modern GIS to visualize these patterns using point pattern data. For more aggregated data "local indicators of spatial association" (Anselin, 1995) such as the local Moran's I and Getis–Ord $Gi^\star$ statistics (which are simply the spatially disaggregated components of their global counterparts) are also readily available in standard GIS software to statistically test for and visualize these local spatial departures from randomness (see Felkner and Townsend, 2011, for one example). All these methods are, however, purely descriptive and say nothing about the causes (or consequences) of the departure from randomness. It is these questions which are the main motivation behind the development and application of the spatial methods that are discussed in detail in the remainder of this chapter. Thinking about the possible causes of nonrandom location and the way in which the consequence of nonrandom location feeds back into location decisions gives us some idea about the difficulties that lie ahead. For example, assume that the points in Figure 3.1 represent either firms or workers and the color represents different types of economic activity. There are several ways in which the nonrandom pattern in the right-hand panel in Figure 3.1 can emerge. First, firms may be randomly allocated across space but some characteristic of locations varies across space and influences outcomes. We might think of farmers who are randomly distributed across space, with the type of crops they produce driven by locational differences in underlying soil type and fertility.[5] Second, location may have no causal effect on outcomes, but outcomes may be correlated across space because heterogenous individuals or firms are nonrandomly allocated across space. We might think of highly educated workers producing R&D in one area, while less educated workers assemble manufactured goods in another area.[6] Third, individuals or firms may be randomly allocated across space but they interact, and so a decision by one agent affects outcomes of other agents. We might think of students choosing among different college majors, where the choice of each student influences the choices of their fellow students.[7] Similarly, in R&D, knowledge might spill over beneficially between nearby scientists, so the decision to undertake research in a specific field, or the registration of patents by inventors, varies systematically across space (as indicated by the color of the dots). Fourth, individuals or firms may be nonrandomly allocated across space and the characteristics of others nearby directly influence individual outcomes. For example, growing up among educated, employed, and successful neighbors might be beneficial in raising children's expectations about their life chances, and this may directly influence their own educational outcomes and through that their employment outcomes.[8]

[5] See, for example, Holmes and Lee (2012), who attempt to distinguish whether soil characteristics (explanation number 1 in our list) or economies of density (explanation number 3) explain crop choice in North Dakota.

[6] See, for example, Ellison and Glaeser (1997), who consider the role of "natural advantages" in explaining geographical concentration of industrial activity. Their broad definition of natural advantages allows a role for resources (e.g., coal), factor endowments (e.g., skilled workers), and density to influence geographical concentration. That is, they assess the role of the first, second, and fourth factors (in our list) in determining sector of economic activity.

[7] See, for example, Sacerdote (2001) and De Giorgi et al. (2010).

[8] A vast literature on childhood neighborhood effects considers this possibility; for example, Aaronson (1998), Patacchini and Zenou (2012), and Gibbons et al. (2013).

Understanding the causes of nonrandomness requires us to discriminate between these four different causes of nonrandomness in situations where one or more of them may explain departures from randomness. In empirical settings, the situation is further complicated because we may not observe all individual factors that determine outcomes. This makes it even harder to distinguish between different causes of nonrandomness. This adds a further potential explanation for nonrandomness—that individuals appear to be randomly located, in terms of observables, but they are in fact nonrandomly located in terms of unobserved characteristics that determine outcomes. The next section formalizes a number of these issues and considers what information is required to enable us to distinguish between different causes of nonrandomness.

3.3. SPATIAL MODELS

This section sets up a very general framework for linear regression models that involve interactions between agents across space. We show how the standard regression approach can accommodate spatial factors by the addition of "spatial variables." These allow the outcomes for an individual to be influenced by the choices, outcomes, and characteristics of other individuals who interact with the individual, and by other characteristics of the location of the individual. In practice, these spatial variables are typically constructed as linear combinations of the observations in neighboring locations, aggregated with a sequence of scalar spatial or group weights. Traditionally, the literature has summarized this information in a (spatial) weights matrix (G in the network literature, W in the spatial econometrics literature), constructed on the basis of the definition of reference groups—the set of individuals or firms that may impact other agents' outcomes. We provide a number of examples below. Both the nature of the reference group and the way in which individual outcomes depend on group membership have fundamental implications for the interpretation, estimation, and identification of spatial models. We deal with questions of interpretation in this section, and also consider the implication for estimation if spatial factors are present, but ignored. The next section then shows how the nature of the reference group, as captured in the structure of the weights matrix, is essential in determining whether the parameters on spatial variables are identified, or can be estimated (and if so, what is the appropriate identification strategy).

3.3.1 Specification of linear spatial models

We start with the standard linear regression model of a variable y relating to some unit of observation i such as a firm, individual, or household (or an areal aggregate of these, e.g., a zip code). For convenience in what follows, we often refer to these units of observation as "individuals." We suppress the constant term and assume that all variables are in deviations from means, allowing us to write the standard linear regression model as

$$y_i = x_i'\gamma + \varepsilon_i, \tag{3.1}$$

where y_i is some outcome, such as output (for a firm) or income (for an individual), and x_i is a vector of characteristics, such as capital, labor, and material inputs (for a firm), or education, age, gender, etc. (for an individual), which determine outcomes and are observed in the data available. Unobserved characteristics that affect outcomes are represented by ε_i. In what follows we assume that ε_i is random and set aside the potential problems that arise if ε_i is not random and correlated with x_i, since the econometric issues involved in this case are well known and we will not address them here.[9] This is a completely nonspatial model, in that there is no explicit reference to where individuals are located in space, to any of the characteristics of the space in which they are located, or to any interconnections between individuals. Suppose we have additional information about the geographical locations s of the individuals whose behavior we want to model. This information is what makes data spatial. Variable s_i might be a point in space referenced by coordinates, or a geographical zone, or some other locational identifier (school, position in a network, etc.).

Let us now modify Equation (3.1) by adding new terms that reflect the fact that the individual choice or outcome y_i may be influenced not only by the characteristics of the individual i, but also by the choices, outcomes, and characteristics of other individuals who interact with the individual i and by other characteristics of the location s_i of individual i. Individuals may interact with each other for a number of reasons, but the important point here is that their interaction is based on some relationship in terms of their spatial location s—for example, they are neighbors or belong to some common group. We will say more about how this "neighborliness" or grouping can be defined below. As we have outlined already, spatial patterns arise through two primary channels: (1) the influence of area characteristics on individuals, both in determining the characteristics acquired by individuals, and through the sorting of already heterogenous individuals across space; and (2) the interaction of neighboring individuals with each other. A framework that captures almost anything researchers try to do with linear regressions when investigating the importance of these spatial factors—both how spatial characteristics affect individuals in the economy, and how neighboring individuals affect each other—is based around the following generalization of Equation (3.1):

$$y_i = x_i'\gamma + m_y(y,s)_i\beta + m_x(x,s)_i'\theta + m_z(z,s)_i'\delta + m_v(v,s)_i\lambda + \varepsilon_i. \tag{3.2}$$

Here, as before, y_i is the outcome for an individual at location s_i, and x_i is the vector of characteristics of i. The expressions $m_{.}(.,s)_i$ are a general representation of "spatial

[9] A general, textbook-level treatment can be found in Angrist and Pischke (2009). Chapter 1 considers how insights from the experimentalist paradigm advocated by Angrist and Pischke (2009) can be applied to questions of causal inference in urban economics. This chapter complements the chapter by Baum-Snow and Ferreira by specifically considering the complications introduced by spatial or social interactions.

variables," the interpretation of which we come to in more detail below. These are functions that generate linear, or sometimes nonlinear, aggregations of variables that are spatially connected with location s_i using information on the vector of locations s. We consider four kinds of spatial variables relating to outcomes (y_i), a vector of individual characteristics (x_i), a vector of characteristics (z_i) of other entities or objects (other than individuals i), and a variable that captures all characteristics of either individuals or entities and objects that are unobservable to the econometrician (v_i). We are keeping things very general at this stage, so we allow the form of $m(.,s)_i$ to be different for y, x, z, and v, and indeed for x and z, possibly different for different elements of these vectors, so that each variable could have its own aggregating or averaging function.

The spatial connections between locations, which form the basis for aggregation, can be defined through absolute or relative positions in geographical space, the position within networks, or other methods. In general, these functions $m_{.}(.,s)_i$ can be thought of in a number of ways, as forming estimates of the means of the variables or expectations at location s_i, as spatial smoothing functions that estimate how the variables vary over locations s, or as structural representations of the connections between locations s. Depending on the setting, these functions may capture interpersonal effects that are passive or deliberate (which might be distinguished as "externalities" vs. "interactions"). These effects may also occur directly or may instead by mediated through the market (leading, for example, to the distinction between pure/technological externalities and pecuniary externalities).

To give a specific example, the outcome under consideration might be earnings, for individuals, and the aim is to estimate Equation (3.2) on a sample of individuals. If y_i is individual earnings, $m_y(y, s)_i$ allows for the possibility that some spatial aggregation of individual outcomes—for example, the mean earnings for individuals living in the same city—may affect individual earnings. The vector x_i might include individual years of education, so $m_x(x, s)_i$ might be defined to capture the mean years of education in some interconnected group—for example, individuals working in the same city. Vector z_i might include indicators of firm industrial classification in an auxiliary sample of firms, so one component of $m_z(z, s)_i$ could be defined to capture the proportion of firms or the total number of firms in each industry category in i's city. Vector z_i might also include average yearly temperature readings from weather stations, such that a second component of $m_z(z, s)_i$ yields mean city temperature. In this example, the share of educated workers (a component of $m_x(x, s)_i$) and the number of firms by sector (a component of $m_z(z, s)_i$) may have a direct effect on earnings or a pecuniary effect (if the share of educated workers is also a measure of labor supply, while the number of firms is also a measure of labor demand).[10] Importantly, Equation (3.2) allows spatial aggregates of the unobservables

[10] This distinction has received some consideration in the literature on human capital externalities (Ciccone and Peri, 2006) but has largely been ignored in the agglomeration literature looking at productivity effects or urban wage premium.

$m_v(v, s)_i$ to influence y_i, to allow for the possibility either that individuals interact with each other across space on unobserved dimensions, or that there are spatially correlated shocks from other sources that affect spatially interconnected individuals simultaneously. To continue the example above, v_i might include individual abilities that are not represented in x, or unobserved productive advantages of the places s in which individuals are located, but which are not represented by variables in z. Again, the spatial aggregate $m_v(v, s)_i$ might then be defined as the mean of these unobserved factors. It is, of course, possible to add a time dimension to this specification, for estimation on a panel or repeated cross sections of individuals, but for now we focus on the cross-sectional case only.

For a set of observations on variables at locations s_j, the "spatial" variables $m_{.}(.,s)_i$ are typically linear combinations of the observations in neighboring locations, aggregated with a sequence of scalar spatial or group weights $g_{ik}(s_i, s_j)$ that depend on the distance (or some other measure of the degree of interconnection) between observations at the corresponding locations s_i and s_j. Let us define

$$m_x(x, s_i) = \sum_{j=1}^{M} g_{ij}(s_i, s_j) \cdot x_j = G_{xi}x, \tag{3.3}$$

where G_{xi} is a $1 \times M$ row vector of the set of weights relating to location s_i, and x is an $M \times 1$ column vector of x for locations $s_1, s_2, \ldots, s_M$. Sometimes it is more convenient to work with matrix notation for all observations i, where G is an $N \times M$ matrix, so

$$m_x(x, s) = G_x x, \tag{3.4}$$

and similarly for z, y, and v. Note that in cases where spatial variables are created by aggregating over the N individuals for whom Equation (3.2) is to be estimated, $N = M$. With use of Equation (3.4) and similar expressions for y, x, and v, Equation (3.2) becomes

$$y = X\gamma + G_y y\beta + G_x X\theta + G_z Z\delta + G_v v\lambda + \varepsilon. \tag{3.5}$$

This notation is favored in the spatial econometrics literature, where the weights matrix is usually designated using W instead of G, assumed common across variables (so $W_y = W_x = W_z = W_v$), and Wy, WX, WZ, and Wv are called "spatial lags." Restrictions on Equation (3.5) yield a typology of spatial econometrics models—for example, the spatially autoregressive (SAR) model ($\delta = 0$, $\lambda = 0$, $\theta = 0$), the spatially lagged x model[11] ($\beta = 0$, $\lambda = 0$), the spatial Durbin model ($\lambda = 0$), and the spatial error model ($\beta = 0$, $\delta = 0$). In what follows, we use the notation G in preference to W, because W has become associated with a set of spatial weights which specify *ad hoc* connections between

[11] The distinction between Z and X is often irrelevant in much applied spatial econometrics research, which usually works with aggregated spatial data units. In this case the data for individuals (x) and for other spatial entities (z) have already implicitly been through a first stage of aggregation. Hence, the standard terminology refers simply to the spatially lagged x model without distinguishing between x and z.

neighboring places, and with a spatial econometrics literature that seeks to distinguish between competing models through statistical testing of model fit. Instead, we wish to focus attention on the fact that the nature of interactions within social and spatial groups is central to theoretical interpretation, identification, and estimation.

In contrast, the social interactions literature favors an alternative notation, where Equations (3.2) and (3.5) are typically written out in terms of expected values of the variables in the groups to which i belongs. Here, the expected values are taken to imply the mean characteristics (observed or unobserved) of the group, or expectations about behaviors or characteristics which are unobserved by individuals or not yet realized. The structural specification analogous to Equation (3.2) in the social interactions literature is thus

$$y_i = x_i'\gamma + E(y|G_i)\beta + E(x|G_i)'\theta + E(z|G_i)'\delta + E(v|G_i)'\lambda + \varepsilon_i. \tag{3.6}$$

In practice, in empirical implementations, the expectations are replaced by empirical counterparts with the estimates $\hat{E}(y|G_i) = G_y y$, $\hat{E}(x|G_i) = G_x x$, and $\hat{E}(z|G_i) = G_z z$ so the spatial models and social interactions models are for the most part isomorphous. Manski (1993) introduced a useful and popular typology of interaction terms in this kind of specification. In this typology, β represents "endogenous" effects, whereby individuals' behavior, outcome, or choices respond to the anticipated behavior outcome or choices of the other members in their reference group. In contrast, θ represents "contextual" or "exogenous" interactions in which individuals respond to observable exogenous or predetermined characteristics of their group (e.g., age and gender). Manski refers to λ as "correlated" effects, in which peer-group-specific unobservable factors affect both individual and peer behavior. For example, children in a school class may be exposed to common factors such as having unobservably good teachers, which can lead to correlation between individuals and peers which look like interactions, but are not. Of course, some of these peer-group-specific factors may also be observable (e.g., teacher qualifications or salaries), and the effects of these observable characteristics are captured in our notation by δ.

3.3.2 Specifying the interconnections

We now turn to the various ways that are used in the literature to define reference groups—the set of agents that impact other agents' outcomes. Both the nature of the reference group and the way in which individual outcomes depend on group membership have fundamental implications for the interpretation, estimation, and identification of spatial models.

The most basic structure for G, and one that is implicitly used in many regression applications that are not ostensibly "spatial," is a block grouping structure. Assume that there are N individuals (or firms, households, areas, etc.; although we continue to focus on individuals for ease of exposition) divided into $k = 1,\ldots,K$ groups, each with n_k members, $i = 1,\ldots,n_k, \sum_{k=1}^{K} n_k = N$. The interaction scheme can be represented by a matrix $G = \{g_{ij}\}$ whose generic element g_{ij} would be 1 if i is connected to j (i.e., interacts with j) and 0 otherwise. Usually, such matrices are row normalized, such that premultiplying an $N \times 1$ vector x by the $N \times N$ matrix G generates an $N \times 1$ vector of spatial averages.[12] For example, consider seven individuals, from each of two neighborhoods: $k = 1,2$. Individuals $i = \{1,2,3\}$ belong to neighborhood $k = 1$ and individuals $i = \{4,5,6,7\}$ belong to neighborhood $k = 2$. The associated G matrix is shown below:

$$G = \begin{array}{c} \\ 1\\2\\3\\4\\5\\6\\7 \end{array}\!\!\begin{array}{c} \begin{array}{ccccccc} 1&2&3&4&5&6&7 \end{array} \\ \left[\begin{array}{ccccccc} \frac{1}{3}&\frac{1}{3}&\frac{1}{3}&0&0&0&0\\ \frac{1}{3}&\frac{1}{3}&\frac{1}{3}&0&0&0&0\\ \frac{1}{3}&\frac{1}{3}&\frac{1}{3}&0&0&0&0\\ 0&0&0&\frac{1}{4}&\frac{1}{4}&\frac{1}{4}&\frac{1}{4}\\ 0&0&0&\frac{1}{4}&\frac{1}{4}&\frac{1}{4}&\frac{1}{4}\\ 0&0&0&\frac{1}{4}&\frac{1}{4}&\frac{1}{4}&\frac{1}{4}\\ 0&0&0&\frac{1}{4}&\frac{1}{4}&\frac{1}{4}&\frac{1}{4} \end{array}\right] \end{array}, \quad GG = \begin{array}{c} \\ 1\\2\\3\\4\\5\\6\\7 \end{array}\!\!\begin{array}{c} \begin{array}{ccccccc} 1&2&3&4&5&6&7 \end{array} \\ \left[\begin{array}{ccccccc} \frac{1}{3}&\frac{1}{3}&\frac{1}{3}&0&0&0&0\\ \frac{1}{3}&\frac{1}{3}&\frac{1}{3}&0&0&0&0\\ \frac{1}{3}&\frac{1}{3}&\frac{1}{3}&0&0&0&0\\ 0&0&0&\frac{1}{4}&\frac{1}{4}&\frac{1}{4}&\frac{1}{4}\\ 0&0&0&\frac{1}{4}&\frac{1}{4}&\frac{1}{4}&\frac{1}{4}\\ 0&0&0&\frac{1}{4}&\frac{1}{4}&\frac{1}{4}&\frac{1}{4}\\ 0&0&0&\frac{1}{4}&\frac{1}{4}&\frac{1}{4}&\frac{1}{4} \end{array}\right] \end{array}. \tag{3.7}$$

Notice that in this example, the weights are set to $1/n_k$, where n_k is the number of neighbors in group k, to achieve row normalization. More importantly, this matrix has two important properties. First, it is block diagonal, and transitive such that the neighbors of i's neighbors are simply i's neighbors. Second, it is symmetric-idempotent, and as a result $GG = G$. This feature will be both useful for interpretation and harmful to estimation. The interpretation is clear: all individuals from 1 to 3 and from 4 to 7 are in a given neighborhood and therefore the spatial influence is constrained to that neighborhood. Indeed, in this case, the values that populate the matrix indicate both group membership and the extent of the influence of any one individual on other individuals. This will not be the case with other specifications of G.

A simple modification that is commonly used in practice is to exclude i from being his or her own neighbor, by putting zeros on the diagonal. This maintains the transitive property, although the matrix is no longer idempotent, for example,

[12] We discuss averaging versus aggregating in more detail below.

$$G=\begin{array}{cc} & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{array}{c} 1\\ 2\\ 3\\ 4\\ 5\\ 6\\ 7 \end{array} & \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix} \end{array}, \quad GG=\begin{array}{cc} & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{array}{c} 1\\ 2\\ 3\\ 4\\ 5\\ 6\\ 7 \end{array} & \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{2}{9} & \frac{2}{9} & \frac{2}{9} \\ 0 & 0 & 0 & \frac{2}{9} & \frac{1}{3} & \frac{2}{9} & \frac{2}{9} \\ 0 & 0 & 0 & \frac{2}{9} & \frac{2}{9} & \frac{1}{3} & \frac{2}{9} \\ 0 & 0 & 0 & \frac{2}{9} & \frac{2}{9} & \frac{2}{9} & \frac{1}{3} \end{bmatrix} \end{array}. \tag{3.8}$$

A simple structure for G that breaks both the transitivity property and the idempotent property could be based on the two nearest neighbors, where 1 is nearest to 2 and 7, 2 is nearest to 1 and 3, 3 is nearest to 2 and 4, 4 is nearest to 3 and 5, 5 is nearest to 4 and 6, and 6 is nearest to 5 and 1. The associated G matrix is shown below, and it is clear in this case that $GG \neq G$—that is, the neighbors of i's neighbors are not simply i's neighbors:

$$G=\begin{array}{cc} & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{array}{c} 1\\ 2\\ 3\\ 4\\ 5\\ 6\\ 7 \end{array} & \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \end{array}, \quad GG=\begin{array}{cc} & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{array}{c} 1\\ 2\\ 3\\ 4\\ 5\\ 6\\ 7 \end{array} & \begin{bmatrix} \frac{1}{3} & \frac{2}{9} & \frac{1}{9} & 0 & 0 & \frac{1}{9} & \frac{2}{9} \\ \frac{2}{9} & \frac{1}{3} & \frac{2}{9} & \frac{1}{9} & 0 & 0 & \frac{1}{9} \\ \frac{1}{9} & \frac{2}{9} & \frac{1}{3} & \frac{2}{9} & \frac{1}{9} & 0 & 0 \\ 0 & \frac{1}{9} & \frac{2}{9} & \frac{1}{3} & \frac{2}{9} & \frac{1}{9} & 0 \\ 0 & 0 & \frac{1}{9} & \frac{2}{9} & \frac{1}{3} & \frac{2}{9} & \frac{1}{9} \\ \frac{1}{9} & 0 & 0 & \frac{1}{9} & \frac{2}{9} & \frac{1}{3} & \frac{2}{9} \\ \frac{2}{9} & \frac{1}{9} & 0 & 0 & \frac{1}{9} & \frac{2}{9} & \frac{1}{3} \end{bmatrix} \end{array}. \tag{3.9}$$

Similar matrices would summarize the pattern of influence in a situation where individuals are asked to name their two closest friends.[13] Of course, the number of neighbors need not be the same for all i. Allowing for varying numbers of bordering neighbors, this

[13] See, for example, the National Longitudinal Study of Adolescent Health, which asks adolescents in grades 7–12 to name up to five male and five female friends. Fryer and Torelli (2010), Calvó-Armengol et al. (2009), Weinberg (2007), and Ioannides (2013) provide other examples.

form of the G matrix gives a contiguity matrix that is commonly used in the spatial econometrics literature for regressions involving areas (districts, regions, etc., rather than individuals) in which the weights are constructed to indicate whether areas share a border. The previous example would correspond to the contiguity matrix for seven areas located sequentially around a circle, with area 1 contiguous to areas 2 and 7, area 2 contiguous to areas 1 and 3, etc.

As should be clear from these three examples, different specifications of G provide a fairly flexible way of constructing spatially weighted variables. A nonexhaustive list of other common structures includes constructing G on the basis of

- "buffers" based on the choice of a fixed distance threshold within which interaction occurs;
- queen or rook contiguity (for geographies with two or higher dimensions), the distinction between the two being whether to regard areas touching at a vertex as contiguous or only those sharing a common border;
- inverse distance weighting;
- connectivity measures along some network.

Observe that the matrix G could be symmetric or asymmetric, depending on the nature of the interactions. It is symmetric in case of bilateral influences between any two units, and—in the case of row normalization—when each unit has the same number of neighbors. It will be asymmetric if interactions are assumed to flow one way, or if units have different numbers of neighbors. The appropriate definition will, of course, depend on the specific application. Note also that the spatial grouping or weights matrix can be defined so that it generates either spatial averages or spatial aggregates of neighboring observations. To produce averages, the G matrix must be row normalized as in the examples above, so that the weights in any row sum to 1. That is, for the spatial weights corresponding to an observation at location s, the weighting vector is

$$G_i = 1/\sum_{j=1}^{M} g_{ij}(s_i, s_j) \times [g_{i1}(s_i, s_1) \quad g_{i2}(s_i, s_2) \quad \ldots \quad g_{iN}(s_i, s_N)],$$

while for aggregation, the weighting vector is simply

$$G_i = [g_{i1}(s_i, s_1) \quad g_{i2}(s_i, s_2) \quad \ldots \quad g_{iN}(s_i, s_N)].$$

The distinction between these two operations could be important, since aggregation adds up the effects of neighboring individuals, firms, or places, thus taking into account the number of these within the appropriate group as specified by the weighting structure. In contrast, averaging takes out any influence from the number of individuals, firms, or places that are close by. Which of these schemes is appropriate is essentially a theoretical consideration. Averaging has been the standard approach in most fields, including those on neighbor and peer effects (Epple and Romano, 2011). Aggregating is more appropriate, and is usually applied, in work on agglomeration, or transport accessibility where the focus is on economic mass or "market potential" (Graham, 2007; Melo et al., 2009), although the literature on human capital externalities in cities has generally favored averaging (see Chapter 5). In cases where there is no guidance from economic considerations, it may be possible to use statistical tests to choose between the different specifications. In regression specifications such as (3.2) it is in principle straightforward to test whether to use aggregation or averaging, since both versions are nested within the expression $n_{ki}m_x(x,s)_i'\theta_1 + m_x(x,s)_i'\theta_2 + n_{ki}\theta_3$, in which n_{ki} is the group size for person i, $m_x(x,s)_i$ is a row-normalized (averaging) aggregator, and $n_{ki}m_x(x,s)_i$ is the interaction of the two, which gives non-row-normalized (aggregating) specification. Including all these terms in a regression specification and testing for restrictions on the parameters would provide one way to distinguish these cases statistically, with $\theta_2 = \theta_3 = 0$, $\theta_1 \neq 0$ implying aggregation, and $\theta_1 = 0$, $\theta_2 \neq 0$, $\theta_3 \neq 0$ implying that separate mean and group size effects are more relevant. There may, of course, be practical collinearity problems when implementing such a test. Liu et al. (2014) provide another test procedure to discriminate between the local-average and local-aggregate models with network data.

Another potentially important consideration is whether or not the number of individuals in the groups over which variables are averaged increases as the sample size increases ("infill" asymptotics). The number of cases over which the averages are constructed increases with sample size for inverse distance weighting or fixed distance buffer groups, and may also do so with block diagonal structures (e.g., if the block specifies different cities, and the cases are individuals). In contrast, this is not necessarily the case with contiguity matrices based on a fixed geographical structure of areas (unless sample size is increased by adding more observations of the same areas over time), or with a fixed number of nearest neighbors or friends. Sample size increases in this case require obtaining more groups ("increasing domain" asymptotics). This issue is important because it affects the way the variance of the spatial means $m_x(x,s)_i$, $m_v(x,s)_i$ behaves as the sample size increases, which will naturally matter when we come to consider questions of identification and estimation of these spatial models.

3.3.3 Interpretation

A vast range of empirical studies on urban, regional, and neighborhood questions, plus research on peer groups and other social interactions, have been based on some version of Equation (3.2). Usually in such studies, the primary focus is on estimating one or more elements of δ or θ, the effect of spatially aggregated observed characteristics for individuals (x_i) or other entities (z_i) on individual outcomes y; or sometimes on estimating β, the effect of neighboring individual outcomes (y_i) on the outcome of an individual entity.

For example, in a typical study of neighborhood effects on the education of children, y would be a child's educational attainment, $G_y y$ (using matrix notation) would be the mean of the attainment of neighboring children, x could include child prior achievement, age, gender, and family background, $G_x x$ might include the mean of these characteristics among neighboring children, and $G_z z$ might include attributes of the child's home location (average local school quality, number of libraries, or average distance to nearest schools). Potentially unobserved factors in $G_v v$ include the quality of teaching in the local school, motivation and aspirations of neighbors, other local resources that facilitate education, etc. This literature is discussed in Chapter 9. To take a second example, studies of agglomeration effects on firm productivity typically specify y_i as firm output, restrict the coefficient on $G_y y$, $\beta = 0$, and define $G_x x$ as a measure of employment density based on aggregating neighboring firm employment or $G_z z$ as a measure of market potential based on aggregating population or income in an auxiliary population sample or census. Firm characteristics such as capital, labor, and material inputs appear in x. Unobservables in $G_v v$ probably include climate, terrain, and other local productive advantages. Depending on whether the specification was in terms of $G_x x$ or $G_z z$, the coefficient θ or δ would then be interpreted as an estimate of the impact of agglomeration economies on total factor productivity. Chapter 5 provides a summary of this literature.

The aim of researchers employing a specification such as Equation (3.2) for these kinds of applications is usually to estimate the "causal" relationship between changes in one or more of the right-hand-side variables and changes in y_i. A good definition of causality is the subject of much debate, and there are a number of interpretations.[14] One definition of a causal estimate is the expected change in y in response to an exogenous manipulation of some particular right-hand-side variable, including any indirect effects that operate through other determinants of y that may also be influenced by the exogenous manipulation of the right-hand-side variable in question. Another definition is the expected change in y for a change in x, with all other factors being held constant. We do not worry too much about these definitions here, except to note that neither looks particularly satisfactory in terms of understanding the parameter β on $G_y y$. Since $G_y y$ is an aggregate of the dependent variable, there is no sense in which it can be directly, exogenously manipulated within the population or sample to which Equation (3.2) relates. Nor can it be changed while holding other factors constant, since if other factors are constant, then y is constant and so is $G_y y$. To return to the education example, it is impossible to think of a hypothetical experiment that would directly manipulate average neighborhood educational outcomes. Instead, one would have to manipulate some other determinant of educational outcomes (e.g., teacher quality in $G_z z$, or neighborhood composition $G_x x$ or the unobserved determinants of $G_v v$) that in turn change average educational outcomes. But in this case this implies a change in

[14] See, for example, the "Con out of Economics" symposium in the *Journal of Economic Perspectives*, 24 (2) (spring 2010). See also Heckman (2005).

$G_z z$, $G_x x$ or $G_v v$, and $G_y y$. As we shall see below, there are structures of G for which we could think of (3.2) applying to one subgroup of the population, while we causally manipulate $G_y y$ by changing $G_z z$ or $G_x x$ for some other subgroup of the population to which they are connected. We return to this issue in Section 3.5. Given these conceptual problems, an alternative is to approach Equation (3.2) as a structural, law-like relationship that determines the process generating y, with the goal of estimating the parameters characterizing this process, setting aside questions over the causal interpretation of β. In this case, the specification to be estimated will need to be derived from some underlying theoretical model. Chapter 2 provides further discussion.

3.3.3.1 Spatial versus social interactions

A particular class of the spatial models described above, which adopt a structural interpretation of the parameter β on $G_y y$, are so-called social interactions models. Social interactions models, as a class, are concerned with modeling these interactions between agents at the microlevel. More specifically, social interactions models are concerned with estimating the parameters that describe the way individuals behave given what they can observe about the group to which they belong, and especially how they expect other individuals in their group to behave. These models and their behavioral foundations have been the focus of much recent attention in the research literature, and are discussed in greater detail in Chapter 9. They provide two crucial insights in the context of the spatial methods considered here. First, as a result of this research, considerable progress has been made in our understanding of the importance of the structure of G in achieving identification of the class of models that involve endogenous interactions in outcomes $G_y y$. We discuss this in the next section. Second, and perhaps less widely recognized, is that the social interactions literature clarifies the circumstances in which the structural equation for y will involve terms in $G_y y$.

In fact, there is a sense in which these social interaction models in which individuals make simultaneous decisions about some action are the only class of models for which the structural equation for y will involve terms in $G_y y$. To see this, note that in any situation where there is no direct interaction in decisions, we should be able to explain the outcome for individual i as a function of own characteristics and group characteristics without needing to know $G_y y$. A concrete example may help clarify this. Imagine a situation where an individual is deciding on the price at which he or she will sell his or her house. We might think that one piece of information the individual will use to set prices is the price of any neighboring houses that have been sold recently. In such situations, it may be convenient to model individual house prices as a function of neighborhood house prices $G_y y$. But this cannot be the structural form, because the timing of sales means that the prices for earlier houses are not determined by the future sales prices of neighboring houses (ignoring any expectation effects that may influence the demand for housing). With information on both prices and the timing of sales, the appropriate structural form involves no term in $G_y y$ because the sales prices of neighboring houses are predetermined from the point of view of any individual price and should thus be treated as an element of X.[15] In contrast, the structural equation for y will involve $G_y y$ in situations of social interaction where decisions are simultaneous. For example, a teenager's decision to start smoking may be dependent on the simultaneous decisions of his or her friends ($G_y y$)—which implies a joint decision based on what each expects the other to do—although even here, an individual's decision to start smoking may be more affected by what that individual observe his or her friends already doing (in which case timing matters and $G_y y$ does not enter the structural form for y).[16]

Another way of putting this is that the scope for including spatial lags in y is more limited than would seem to be implied by the applied spatial econometrics literature. Indeed, in that literature, terms in $G_y y$ are often included without any consideration of whether decisions that determine y are truly simultaneous. In some circumstances, this assumption may be justified. For example, in the tax competition literature, local tax rates are a function of neighboring government tax rates if governments simultaneously set taxes in response to (expectations of) taxes in contiguous neighboring jurisdictions. More generally, however, many spatial models simply assume that any interaction (between individuals in neighborhoods or schools, between neighboring or otherwise interconnected firms, between inventors and other agents of innovation, between neighboring governments and other institutions, etc.) can be used to justify the inclusion of terms in $G_y y$.

3.3.3.2 Pecuniary versus technological externalities

Another important distinction, but one that has received relatively little attention in the literature, is whether spatial interactions arise as a result of pecuniary or technological externalities. As we discussed above, in the general spatial model terms in Gy, GX, and GZ can capture interactions that either occur directly or are mediated though the market (i.e., may capture either technological or pecuniary externalities, respectively). We have provided several examples where either may arise. For example, models in the new economic geography tradition can motivate empirical specifications that model employment in area i as a function of employment in nearby areas Gy. As we explained in Section 3.1, in these models firms are sufficiently small that they ignore their impact on other firms (and hence ignore reactions from those firms), while workers' utility functions depend only on

[15] For an empirical example, see Eerola and Lyytikainen (2012), who use the partial release of public information on past house sales to examine the impact of information on past transactions on current house prices. Ioannides and Zabel (2008), Kiel and Zabel (2008), and Ioannides (2013) provide a more general discussion of neighborhood effects on housing demand and the use of neighborhood information in hedonic regressions.

[16] See, for example, Krauth (2005) and Nakajima (2007). Simons-Morton and Farhat (2010) provide a review of the literature on peer group influences on adolescent smoking.

consumption of a continuum of manufacturing sector varieties and an agricultural good (not directly on the utility of other workers). Given that, at least in the general spatial form, these two kinds of externalities are observationally equivalent, it is likely that theory will need to provide additional structure if applied work is going to distinguish between these different sources of interaction. Chapter 2 provides further discussion.

3.4. IDENTIFICATION

All researchers working with spatial data have to confront fundamental challenges that render the identification and estimation of Equation (3.2) a difficult empirical exercise. These challenges are (a) the so-called reflection problem, (b) the presence of correlated unobservables or common shocks, and (c) sorting—that is, the presence of omitted variables which are correlated with location decisions and outcomes. Problem (a) occurs when the aim is to estimate β (i.e., the effect of group outcomes or behavior on individual outcomes) as distinct from θ (i.e., the effect of group characteristics), while problems (b) and (c) may arise regardless of whether we are estimating models with or without endogenous interactions. We consider these problems in turn and discuss the solutions proposed in the existing literature.

3.4.1 Spatially autocorrelated unobservables, when these are uncorrelated with the observables

Even in the simplest setting where we know the structure of group membership and the individual and group variables that determine outcomes, the reflection problem can prevent the estimation of all coefficients of interest. The problem arises when the aim is to separately estimate β (the effect of group outcomes or behavior on individual outcomes) and θ (the effect of group characteristics) in situations where there are unobservable factors that also vary at the group level. The presence of these variables means that estimation must rely on recovering the structural parameters from parameters on the exogenous variables in the reduced form. This is usually not possible without imposing further restrictions.

To focus on this specific issue, let us initially assume that group membership is exogenous and that these unobservables are uncorrelated with the observable characteristics. This spatial autocorrelation in unobservables could occur because individuals are interacting on unobserved dimensions. For example, in a model of neighborhood effects on school grades, individual effort (unobserved by the researcher) may influence other individuals' effort within the neighborhood, even before the outcomes of that effort—school grades (y)—are observed. Or it could occur because the group members are exposed to similar unobservables. For example, in a model of the effect of cluster employment on firm employment, different clusters could be subjected to area shocks that are not directly related to the performance of the cluster. Both these processes show up as autocorrelated unobservables, so are observationally equivalent from the researcher's perspective.

As mentioned above, Manski (1993) refers to these as "correlated effects," the presence of group-specific unobservable factors, uncorrelated with individual observables, but affecting both individual and group behavior. Spatial econometricians refer to models containing these spatially autocorrelated unobservables as spatial error models. Applied economists in many other fields generally refer to these as "common shocks" to capture the idea that individuals in spatial or peer groups are subject to unobserved influences in common. These group-specific differences in unobservables are almost inevitable in situations where estimation is based on observational survey, census, or administrative data, and there is no explicit manipulation of the data by experimentation or policy. In situations where we are not interested in the estimation of β, the presence of these unobservable factors that are uncorrelated with x and z requires no more than adjustment to standard errors. Standard approaches to correcting the standard errors in the case of intragroup correlation and groupwise heteroscedasticity can be applied in this case (Cameron and Miller, 2015). However, these methods require discrete spatial groups, with no intergroup correlation, and can seem *ad hoc* in settings where space is best thought of as continuous. Conley (1999) provides analogous methods for continuous space. For a deeper discussion of these issues, see Barrios et al. (2012). Alternatively, researchers could resort to Monte Carlo methods in which the null distribution is simulated by random assignment across space, an approach that is common in spatial statistics.[17] Unfortunately, in models involving $G_y y$ the implications are more serious.

For models involving $G_y y$, the presence of unobserved effects, even if uncorrelated with the included variables, leads to a basic estimation problem because the ordinary least squares (OLS) estimate of β—the endogenous effect or SAR parameter—is biased and inconsistent. The intuition behind this is simply that the model is a simultaneous equation model. For any individual i, group outcomes $G_y y$ are partly determined by the outcome for individual i. Therefore, group outcomes for individual i, $G_y y$, are explicitly correlated with individual i's own unobservables. In other words, the spatial lag term contains the dependent variable for "neighbors" (i.e., members of the same group), which in turn contains the spatial lag for their neighbors, and so on, leading to a nonzero correlation between the spatial lag $G_y y$ and the error terms—that is,[18]

$$\underset{n\to\infty}{p\lim} = n^{-1}\left(G_y y' \varepsilon\right) = 0. \tag{3.10}$$

[17] Tests for spatial autocorrelation in the residuals from a regression analysis can also be helpful in establishing whether such corrections to the standard errors are justified. These tests can be based on Moran's I or other statistics that measure spatial autocorrelation, as outlined in Section 3.2.

[18] More technically, the pure SAR model $y = G_y y\beta + \varepsilon$ has the following reduced form: $y = (I - G_y\beta)^{-1}\varepsilon$. Hence, $G_y y = G_y(I - G_y\beta)^{-1}\varepsilon$. Let us define $S = G_y(I - G_y\beta)^{-1}$, then $E(G_y y', \varepsilon) = E(\varepsilon'^{-1} G_y', \varepsilon) = E(\mathrm{tr}(S\varepsilon'), \varepsilon) = \mathrm{tr}(S)E(\varepsilon'\varepsilon) \neq 0$. There is no reason to believe that $\mathrm{tr}(S) = 0$.

As a consequence, OLS estimates of parameters in a specification such as Equation (3.5) are inherently biased, unless $\beta = 0$. This is a mechanical endogeneity problem generated by the two-way feedback between individuals in a spatial setting. Much spatial econometrics, since Anselin (1988), is concerned specifically with this problem and adopts maximum likelihood methods or instrumental variables estimators (in the case where there are exogenous variables in the model).[19] While this basic estimation problem is pervasive, solutions to it are well understood. The biases that arise in situations where $G_y y$ determines y but is omitted from the estimating equation are also well understood and are discussed in Appendix A. The much more substantive problem concerns the question of whether the underlying parameters are identified (or, equivalently, whether valid instruments are available). It is to this issue that we now turn.

3.4.1.1 The reflection problem

To focus on this specific issue, let us define these unobservables as $u = G_v v\lambda + \varepsilon$. We assume these are uncorrelated with the observable characteristics x and z—that is, there is no sorting and no omitted spatial variables (we return to this problem in Section 3.4.3). Using this definition of u, we can write Equation (3.5) as

$$y = X\gamma + G_y y\beta + G_x X\theta + G_z Z\delta + u. \tag{3.11}$$

Premultiplying by $G_y y$ gives

$$G_y y = G_y X\gamma + G_y G_y y\beta + G_y G_x X\theta + G_y G_z Z\delta + G_y u. \tag{3.12}$$

Now, the spatial aggregate or average y, $G_y y$ is explicitly correlated with u by virtue of the model structure, even if $E[u|X, Z] = 0$. Evidently then $E[u|G_y y] \neq 0$, and least squares estimates of Equation (3.11) are biased. Given this dependence of the spatial average y on the remaining spatially averaged unobservables (the common unobserved interactions/shocks/correlated effects), methods for estimating β in Equation (3.11) must rely on being able to recover the parameters β, θ, and δ from parameters on the exogenous observables X and Z in the reduced form. The reduced form is obtained by substituting out $G_y y$ in Equation (3.11) to obtain an expression that contains only the exogenous variables and their spatial lags. Unfortunately, in general, it is not easy to recover these parameters from the reduced form without imposing further restrictions.

The fundamental issue which makes it difficult to recover the parameters in Equation (3.11) from its reduced form is that, in this linear specification, the spatially averaged outcomes $G_y y$ are likely to be perfectly collinear with the spatially averaged

[19] See Lee (2004) for details of the maximum likelihood approach and Kelejian and Prucha (1998, 1999, 2004, 2010) for details of the instrumental variables approach. A basic review of the estimation methods for linear spatial models can be found in Anselin (1988).

exogenous variables $G_x X$ and $G_x Z$, except in so far as $G_y y$ is determined by the spatial unobservables u. This holds unless specific types of restrictions are imposed on the structure of G, or on other aspects of the specification, as we discuss in detail below. In other words, $m_y(y, s)_i$ is an aggregation of outcomes or behaviors over "neighbors" (i.e., members of the relevant group) at location s_i, and hence is an aggregation of $m_x(x, s)_i$, $m_z(z, s)_i$ (and u) over neighbors at s_i.

This is easiest to see if we choose the very simple mean-creating, block diagonal, idempotent, and transitive grouping structure as in Equation (3.7), and define a common $G = G_y = G_x = G_z$. In this case,

$$y = X\gamma + Gy\beta + GX\theta + GZ\delta + u, \tag{3.13}$$

$$\begin{aligned} Gy &= GX\gamma + Gy\beta + GX\theta + GZ\delta + Gu \\ &= GX(\gamma+\theta)/(1-\beta) + GZ\delta/(1-\beta) + Gu/(1-\beta). \end{aligned} \tag{3.14}$$

Plugging the expression for Gy in Equation (3.14) into the expression for y yields a reduced form:

$$y = X\gamma/(1-\beta) + GX(\gamma\beta+\theta)/(1-\beta) + GZ\delta/(1-\beta) + u + Gu\beta/(1-\beta), \tag{3.15}$$

$$y = X\tilde{\gamma} + GX\tilde{\theta} + GZ\tilde{\delta} + \tilde{u}. \tag{3.16}$$

The parameters β, θ, and δ cannot be separately identified from the composite parameters $\tilde{\theta} = (\gamma\beta+\theta)/(1-\beta)$ and $\tilde{\delta} = \delta/(1-\beta)$ in this reduced form. This is the Manski (1993) "reflection problem," which Manski originally discussed in the context of social interactions, where we are trying to infer whether individual behavior is influenced by the average behavior of the group to which the individual belongs. Although our exposition above assumes an idempotent G matrix, the problem is not limited to only that case. For example, the problem still arises if, as is common practice in spatial econometrics, we exclude the influence of an individual i on itself in defining G—that is, we set the diagonals to zero to render G nonidempotent as in Equation (3.8). To see this, define $G^\star$ and G as zero-diagonal and non–zero-diagonal matrices for the same grouping structure, with equal-size groups with M members. It follows that

$$G^* = \frac{M}{M-1}G - \frac{1}{M-1}I.$$

It is evident from this that there is no additional information in $G^\star$ that could be used for identification, since it only differs from G in subtracting the contribution made to each group by individual i. To see this more formally, define $a = \frac{M}{M-1}$ and $b = \frac{1}{M-1}$. Now, using the zero-diagonal grouping matrix in Equation (3.13) and disregarding $G_z z$, for which the concept of zero diagonals is irrelevant since the z come from entities other than the individuals under investigation,

$$\begin{aligned} y &= X\gamma + G^{*}y\beta + G^{*}X\theta + u \\ &= X\gamma + Gy\beta b + GX\theta b - ay\beta - aX\theta + u \\ &= Gy\beta b + X(\gamma - a\theta)/(1 + a\beta) + GX\theta b/(1 + a\beta) + u/(1 + a\beta). \end{aligned} \tag{3.17}$$

Evidently, comparing Equation (3.17) with Equation (3.13), we see there is no gain from using zero diagonals in terms of identification, when group sizes are equal, because we have no additional exogenous variables. A similar argument holds when group sizes are large, because $M \xrightarrow{\lim} \infty \; a = 1$ and $M \xrightarrow{\lim} \infty \; b = 0$, so $M \xrightarrow{\lim} \infty \; G^{*} = G$. The reflection problem carries through in general to any case where Gy, GX, GZ forms the averages or expectations of y, X, and Z conditional on the groups defined by G.[20]

To summarize, to be able to estimate an equation such as (3.5) or (3.6), the researcher must be able to observe differences between the spatial means defined by $G_y y$, $G_x X$, $G_z Z$ in the data, otherwise there is insufficient variation to allow estimation. But if group-specific differences lead to variation in $G_y y$, $G_x X$, $G_z Z$, then they almost certainly lead to differences between groups in terms of unobservables. In large groups of individuals (e.g., census data from cities), these differences can arise only because there is nonrandom sorting of individuals across space. In smaller groups (e.g., samples based on friendship networks), the process of assignment to these groups must also be nonrandom, or else the groups must be sufficiently small that the researcher can make an estimation from the random sampling variation in the group means. Of course, if the researcher is conducting an experiment or is investigating the consequences of a specific policy intervention, then that researcher may have much greater control over assignment of individuals to groups and manipulation of the variables of interest, $G_x X$ and $G_z Z$. We return to discuss these issues in Section 3.5. But for observational data, the reflection problem is very likely to occur unless we are able to impose further restrictions.

3.4.1.2 Solutions to the reflection problem

There are a number of possible solutions to the identification challenges arising from the reflection problem.

First, since the issue originates in the fact that individual outcomes are linear in group-mean outcomes, and group-mean outcomes are, in turn, linear in group-mean characteristics, the use of nonlinear functional forms provides one parametric solution

[20] In cases where the group size is small and varies across groups, it is technically possible to identify the parameters in Equation (3.13), with a zero-diagonal block diagonal matrix, as discussed in, for example, Lee (2007) and Bramoullé et al. (2009). This identification comes from the fact that the neighborhood or peer effect for individuals in a given group is a weighted average of the simple mean in the group (from which we have shown that β is not identified) and their own contribution to the mean. These weights vary with group size. The relationship between the simple mean generated by G and the mean generated by $G^{\star}$ is, for a given individual, $G_i^{*}y = \frac{M_k}{M_k - 1} G_i y - \frac{y_i}{M_k - 1}$. Technically, identification can come from the weights $\frac{M_k}{M_k - 1}$. This is clearly a tenuous source of identification, particularly if there are separate group size impacts (i.e., direct effects) of M_k on the outcome. In addition, in practice, problems may arise because as the group sizes become similar, $\mathrm{Var}(M_k) \to 0$, and as the group sizes become large, $\frac{M_k}{M_k - 1} \to 1$ and $\frac{1}{M_k - 1} \to 0$.

(e.g., Brock and Durlauf, 2001). For instance, if an outcome is binary (e.g., either to smoke or not to smoke) and thus the probability of smoking is nonlinear in individual characteristics, then identification could come from the assumed functional form of the relationship between covariates and the probability of smoking. However, these kinds of structural assumptions clearly assume that the theoretical structure is known a priori. Further discussion can be found in Chapter 9 and Ioannides (2013). Empirical examples can be found in Sirakaya (2006), Soetevant and Kooreman (2007), Li and Lee (2009), Krauth (2005), and Nakajima (2007).

A second strategy would be to impose restrictions on the parameters on the basis of theoretical reasoning. Obviously, as discussed above, setting $\beta = 0$ and assuming away endogenous effects would be one solution, but would not be very helpful if the aim is to estimate β or we are interested in a structural estimate of γ. Restrictions on some or all of the coefficients on group-means GX are another possibility. That is, if there is some x_r that affects outcomes whose group-mean does not affect outcomes, then the group-average can be used as an instrument for Gy in Equation (3.13). These assumptions are quite difficult to defend, and the exclusion restrictions on θ can appear arbitrary. Goux and Maurin (2007), for example, experiment with using neighbors' age as an instrument for neighbors' educational achievement in their study of neighborhood effects in France, but recognize that neighbors' age may have direct effects. Gaviria and Raphael (2001) simply assume away all contextual effects from GX completely.

The third strategy builds on our discussion of the interaction matrix G in Section 3.3.2. It relies on imposing a specific structure for the interaction matrix G that is not block diagonal or transitive, and has the property that $GG \neq G$. This approach to identification has long been proposed in the spatial econometrics literature (Kelejian and Prucha, 1998). Recently, this same approach has been the focus of a number of papers dealing with the identification and estimation of peer effects with network data (e.g., Bramoullé et al., 2009; Calvó-Armengol et al., 2009; Lee et al., 2010; Lin, 2010; Liu and Lee, 2010; Liu et al., 2012).

In the general spatial model in Equation (3.11), if G is characterized by a known nonoverlapping group structure, such that $G_yG_y \neq G_y$, $G_yG_x \neq G_x$, or $G_yG_z \neq G_z$, then the parameters β, θ, and δ can be separately identified. More explicitly, suppose $G_y = G_x = G_z = G$, but $GG \neq G$. As before we can get an expression for Gy by multiplying through by G:

$$y = X\gamma + Gy\beta + GX\theta + GZ\delta + u, \tag{3.18}$$

$$\begin{aligned} Gy &= GX\gamma + Gy\beta + GX\theta + GZ\delta + Gu \\ &= GX(\gamma + \theta)/(1-\beta) + GZ\delta/(1-\beta) + Gu/(1-\beta). \end{aligned} \tag{3.19}$$

Now, however, when we plug Gy back into the estimating equation, the fact that $GG \neq G$ means we end up with additional terms in G^2X, G^2Z, and G^2y (using the notation that $GG = G^2$). Repeated substitution for Gy gives the reduced form of Equation (3.11) as

$$\begin{aligned} y &= X\gamma + GX(\gamma\beta+\theta) + G^2X(\gamma\beta^2+\theta\beta) + G^3X(\gamma\beta^3+\theta\beta^2) \\ &\quad + \cdots + GZ\delta + G^2Z\delta\beta + G^3Z\delta\beta^2 + \cdots + u + Gu\beta + G^2u\beta + \cdots . \end{aligned} \tag{3.20}$$

In this case, in comparison with Equation (3.15), there are additional exogenous variables which are the spatially double-lagged and spatially multiply lagged observables G^2X, $G^3X,\ldots$ and G^2Z, $G^3Z,\ldots$ which affect y only via their influence on $G_y y$. There are at least as many reduced form parameters as structural parameters, so technically, the structural parameters are identified. For example, the ratio of the coefficients on the corresponding elements of the vectors GZ and G^2Z provides an estimate of β. That estimate, combined with the estimate of γ (the coefficient on X) can then be used to back out θ from the coefficient on GX. Alternatively, we could use terms in G^2X, $G^3X,\ldots$ and G^2Z, $G^3Z,\ldots$ as an instrument directly for $G_y y$ using two-stage least squares. The intuition behind this result is simple: when the interaction structure is incomplete, we can find "neighbors of my neighbors" whose behavior influences me only via the influence that they have on my neighbor. The characteristics of these second-degree neighbors are thus correlated with my neighbors' behavior, but have no direct influence on my behavior, satisfying the relevance and excludability criterion for a valid instrument.

In principle, these results are widely applicable, because in many real-world contexts, an individual or firm may not necessarily be influenced by all the others in a given group. For example, firms in an industry may not be in contact with all the others in the industry, but may be in contact only with those firms from which they buy inputs. Or a child may not be affected by all children in its school, but may be affected only by those children with whom that child is friends on Facebook. These cases are examples of an incomplete network—that is, everybody is not connected with everybody else. Rather, each individual has its own group of contacts, which differ from individual to individual. When this occurs, $GG \neq G$, and this solves the reflection problem as just discussed. The network structure provides a good context to summarize the intuition for the formal result. Consider a simple network with three individuals A, B, and C as illustrated in Figure 3.2.

A and B play piano together and B and C swim together, but A and C have never met. Then, the only way C could influence A's behavior is through B. The characteristics of C are thus a good instrument for the effect of the behavior of B on A because they certainly influence the behavior of B but they do not influence directly the behavior of A.

To identify network effects, one needs only one such intransitivity; however, in most real-world networks, there are a very large number of them.

While in principle this solution to the reflection problem might apply in a large number of situations, its application in many spatial settings is problematic. The identification

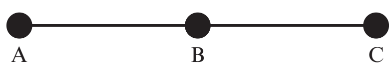

Figure 3.2 A simple network.

strategy relies on having detailed and accurate data on the interactions between agents (i.e., one needs to know exactly who interacts with whom). In particular, it hinges upon nonlinearities in group membership (i.e., on the presence of intransitive triads). If links are incorrectly specified, then the exclusion restrictions are violated. Going back to our example in Figure 3.2, if C in fact knows A but we assume that she does not, then identification fails. In the network literature, restrictions on the interaction scheme are often imposed on the basis of data that specifically seek to identify relevant linkages (Bramoullé et al., 2009; Calvó-Armengol et al., 2009; Lee et al., 2010; Lin, 2010; Liu and Lee, 2010; Liu et al., 2012) or are explicitly derived from theory.

In contrast, in the spatial econometrics literature, the requirement that $GG \neq G$ has been largely met through the use of *ad hoc* spatial weight matrices pulled from a pick-list of popular forms—for example, constructed on the basis of rook or queen contiguity, or inverse distance weighting, which are non-block diagonal and nonidempotent as discussed in Section 3.3.2. In our view, while $GG \neq G$ provides a solution to the reflection problem, any such restrictions require careful justification on the basis of institutions, policy, or theory, or (as in the network literature) need to be imposed on the basis of data that specifically seek to identify relevant linkages. This is something which is very hard to achieve when simply imposing many of the popular spatial weight matrices.

Unfortunately, identification fails if these restrictions (whether carefully justified, based on data, or imposed *ad hoc*) are invalid. The network literature suggests that the problems of missing data (on nodes, but not on links) may be less severe. Helmers and Patnam (2014), Liu et al. (2012), and Liu et al. (2013) present Monte Carlo evidence on the bias of the estimator when misspecification of the social network structure is due to data for individuals missing at random because of sampling (but where all links are observed). Liu et al. (2013) develop a nonlinear estimator designed to address sampling issues over networks. The common finding seems to be that random sampling with known network structure induces a consistent downward bias in the estimates at all sample sizes and at all spatial parameter values. That is to say, as in more standard settings, nonsystematic measurement error causes attenuation bias on the parameters of interest. This implies that, in the presence of a known network structure but random measurement error for nodes, estimated coefficients are likely to provide a lower bound for the importance of social interactions. There is little chance, however, that random measurement errors are inducing us to detect the presence of peer effects when they are not existent (see Conley and Molinari, 2007; Kelejian and Prucha, 2007 for studies showing the robustness of variance–covariance estimators to location misspecification). In other words, if G is known and the only source of measurement error is random missing data for specific nodes, point estimates of peer effects are likely to be higher and standard errors remain roughly unchanged. Note, however, that these results do not provide much reassurance in situations where missing data are nonrandom or where there are errors on the interaction structure (e.g., due to the endogeneity of the interaction structure, missing links in the network, or the fact that the restriction $GG \neq G$ has been arbitrarily imposed by choosing one of the popular spatial weight matrices).

Even when G is known and the network is incomplete, so that G^2X, G^3X, G^2Z, G^3Z (and so on) provide valid instruments, the weakness of the instruments may prove a serious threat to identification and estimation.[21] This weak instruments problem arises if the instruments G^2X, G^3X,G^2Z, G^3Z (and so on) are highly correlated with the explanatory variables GX and GZ, so that, conditional on GX and GZ, there is little variation in the instruments. Therefore, while identification is technically possible, there may be little variation in the instruments to allow estimation. This is potentially a serious problem when G represents spatial connections between neighboring agents or places, when G is row normalized so that it creates the means of the neighbors (as G is commonly specified), and where there is strong spatial autocorrelation in X and Z (usually the case empirically). In this case Gx, for example, estimates the mean of a variable x at each location on the basis of the values of x at neighboring locations, G^2x estimates the means at each location on the basis of the means of the means of x at each location, and so on. So, Gx, G^2x, and G^3x are all just estimates of the mean of x at each location using different weighting schemes. Indeed, this use of neighbors to estimate location-specific means underpins nonparametric kernel regression methods, and spatial interpolation methods in GIS applications. In practice, in cases where the groups formed by G are small (e.g., three nearest neighbors, or contiguous districts), there may be enough sampling variation in these means to ensure that Gx, G^2x, G^3x, and higher-order spatial lags are not perfectly collinear, so estimation may be possible. The problem is, however, potentially especially serious in the situations, noted at the end of Section 3.3, where the numbers of observations in a group becomes very large. The means estimated by Gx, G^2x, and G^3x converge to the population mean of x at each location as the group size goes to infinity, implying the spatial lags are all perfectly collinear and so identification fails.[22]

This weak instruments problem is potentially less pervasive in peer group network applications with individual data (see Chapter 9) when the information on social connections is rich and if individuals make diverse and idiosyncratic choices about their friends. In this case, unlike the spatial setting with spatial autocorrelation, the characteristics of an individual's friends provide little or no information about the individual's own characteristics. However, in cases where peer groups are formed by strongly assortative or

[21] As discussed in Bound et al. (1995), weak instruments lead to a number of problems. The two-stage least squares estimator with weak instruments is biased for small samples. Any inconsistency from a small violation of the exclusion restriction is magnified by weak instruments. Finally, estimated standard errors may be too small. Stock et al. (2002) propose a first-stage F test that can be used to guide instrument choice when there are concerns about weak instruments.

[22] For example, the mean of a variable x among the 1000 nearest neighbors of an individual will not be very different from the mean among the 1000 nearest neighbors of that individual's nearest neighbor, so Gx, G^2x, G^3x, and so on will be almost perfectly collinear.

disassortative matching processes, the weak instruments issue may still create a potential threat to estimation and identification.[23]

We have considered three possible solutions to the reflection problem—the use of functional form, the imposition of exclusion restrictions, and the use of an incomplete interactions matrix such that $GG \neq G$. The last of these, in particular, has received considerable attention in the recent social interactions literature focusing on the identification and estimation of peer effects with network data. These methods may be applicable in a broader set of spatial settings. However, any such restrictions require careful justification on the basis of institutions, policy, or theory, or need to be imposed on the basis of data that specifically seek to identify relevant linkages. While these issues have received careful consideration in both the networks literature and the theoretical spatial econometrics literature, much applied work continues to rely on *ad hoc* restrictions implicitly imposed through the choice of popular spatial weight matrices.

3.4.2 Spatially autocorrelated unobservables, when these are correlated with the observables

So far we have set aside the possibility, explicit in Equation (3.2) or (3.5), that there are spatial or group-specific unobservables, $m_v(v, s)_i$ or $G_v v$ using the matrix form, which are correlated with the explanatory variables. The second challenge arises once we drop this assumption and allow for the possibility that unobservables $u = G_v v\lambda + \varepsilon$ are correlated with the observable characteristics x and z. In many situations observable individual, location, and neighbor characteristics x, $G_x x$, and $G_z z$ are very likely related to the unobservable location and neighbor characteristics $G_v v$. We can identify two mechanisms. First, group membership is exogenous and the correlation arises because of spatially omitted variables that are correlated for individuals in the same group. These omitted variables may directly affect y, or they may determine x or z and hence indirectly affect y. Second, group membership is endogenous and the correlation arises because of the sorting of individuals with different characteristics x into locations with different $G_v v$. For example, in the agglomeration literature the link between urban wages and urban education may arise because cities that offer high returns to education have unobserved characteristics that encourage individuals to acquire more schooling (as in the literature on human capital externalities, reviewed in Moretti, 2004), or highly educated workers may move into cities that offer high returns to their education (as in the urban wage premium literature; e.g., Combes et al., 2008). In either case, if the factors that determine city-specific returns to education are not all observable, x and spatial aggregates of x (i.e., $G_x x$) or variables that are included in $G_z z$ are correlated with $G_v v$.

[23] Lee and Liu (2010) propose a generalized method of moments with additional instruments to try to circumvent the weak instrument problem.

It is important to note that while the urban economics literature has traditionally recognized these two mechanisms through which $G_x x$ and $G_z z$ may be correlated with $G_\nu \nu$, it has tended to treat these symmetrically. However, in most cases "sorting" is better thought of as the situation where group membership is endogenous. That is, the correlation between $G_x x$ or $G_z z$ and $G_\nu \nu$ arises because G_x, G_z, and G_ν are endogenous. In this subsection, we set aside this possibility to consider the situation where group membership is exogenous (although not necessarily fixed over time) and correlation arises because of spatially omitted variables that are correlated for individuals in the same group.

Suppose that the aim is to estimate a specification without endogenous interactions, either because endogenous interactions are being ruled out, or because this is viewed as the reduced form of a model with endogenous specifications. Restricting our attention to spatial interactions that can be represented by a set of spatial weight matrices implies

$$y = X\gamma + G_x X\theta + G_z Z\delta + G_\nu \nu\lambda + \varepsilon. \tag{3.21}$$

Standard nonexperimental approaches to estimating Equation (3.21) all involve, in some way, transforming the estimating equation in a way that "partials" out $G_\nu \nu$ so that it no longer enters the estimating equation. For example, an increasingly common way to partial out $G_\nu \nu$ is to apply "spatial differencing," which transforms all variables by subtracting some appropriately constructed spatial mean (Holmes, 1998). Assume, for the moment, that we know G_ν, then spatial differencing is equivalent to premultiplying Equation (3.21) by a transformation matrix $[I - G_\nu]$ to give (where ζ is another random error term)

$$y - G_\nu y = (X - G_\nu X)\gamma + (G_\nu - G_\nu G_x)X\theta + (G_z - G_\nu G_z)Z\delta + (G_\nu - G_\nu G_\nu)\nu\lambda + \zeta. \tag{3.22}$$

If $\operatorname{plim}(G_\nu - G_\nu G_\nu)\nu = 0$, this transformation eliminates spatial unobservables $G_\nu \nu$, allowing consistent estimation of Equation (3.22) by OLS. Clearly, from the above, this condition will hold when we know G_ν and where G_ν has an idempotent structure (e.g., block group structures similar to the example in Equation (3.7)), in which case $G_\nu - G_\nu G_\nu = 0$, so

$$y - G_\nu y = (X - G_\nu X)\gamma + (G_\nu - G_\nu G_x)X\theta + (G_z - G_\nu G_z)Z\delta + \zeta. \tag{3.23}$$

This is just a standard fixed effects estimator, in which variables have been differenced from some group mean (where the groups are defined by G_ν) or where the regression includes a set of dummy variables for the groups defined by G_ν.

Indeed, if we have panel data providing multiple observations for individuals over time and define G_ν to have a block group structure for each individual, this is just the standard fixed effects estimator. The transformation matrix $[I - G_\nu]$ eliminates the individual-level mean and allows us to consistently estimate Equation (3.21) providing that group-level characteristics are correlated only with time-invariant individual-level unobservables. Individual-level time-varying shocks will still lead to inconsistent estimates if they are correlated with group-level characteristics. This is the approach adopted in the standard mincerian wage regression approach to estimating city-level productivity or wage differences (Combes et al., 2008; Di Addario and Patacchini, 2008; Mion and Naticchioni, 2009; De la Roca and Puga, 2014; Gibbons et al., 2014; and many others). In that literature, the identifying assumption is that city location (i.e., group membership) can be correlated with time-invariant individual characteristics (such as ability), but not with time-varying shocks (e.g., to an individual's income).

Just as with the standard individual fixed effects approach, there are evidently further limitations to the application of spatial differencing. Suppose in the absence of any other information, we simply assume that the spatial weighting/grouping functions $m(.,s)$ are the same for all variables—that is, $G_x = G_z = G_\nu = G$. In this case, Equation (3.23) reduces to

$$y - Gy = (X - GX)\gamma + \zeta. \tag{3.24}$$

Note that spatial differencing removes both $GX\theta$ and $GZ\delta$, so while the parameters γ on X are identified, the parameters on the spatial variables GX or GZ are not. This is, of course, just the standard problem that the parameters on variables that are collinear with group fixed effects cannot be estimated. Clearly, if one is willing to assume that the structure of connections in terms of unobservables G_ν is different from the ones in terms of observables (G_x and G_z), then demeaning the variables using the spatial means of G_ν would not eliminate GX and GZ and allow estimation of θ and δ.[24] However, imposing a different structure of connections for the observables and unobservables is a strong assumption. This discussion illustrates a crucial point: even in the most basic strategy for eliminating spatial unobservables, researchers are making fairly strong assumptions about the structure of the implied interconnections between observations, and the structure of the (implicit) G matrices that link different observations together on observable and unobservable dimensions.

There are cases where this assumption may serve as a reasonable approximation. For example, a study of neighborhood effects on labor market outcomes might be prepared to assume that the observable variables of interest—for example, neighborhood unemployment rates—are linked at the neighborhood level (defined by G_x), but that unobservable labor market demand factors (G_ν) operate at a large labor market level. A good research design should ground this identifying assumption on sound theoretical reasoning or on supporting evidence (e.g., about institutional arrangements).

One increasingly popular approach in spatial settings, "boundary-discontinuity" design (which is a particular spatial case of regression discontinuity design), provides an explicit justification for having a distinct set of weights for observables and unobservables. In this setup, the researcher cites institutional and policy-related rules as a justification for assuming that the spatial connections between places in terms of the

[24] Estimation of γ does not require this assumption as shown above.

characteristics of interest are very different from those that affect unobservables ν. This difference may arise because, for example, administrative boundaries create discontinuities in the way $G_z Z$ varies over space but (so it is assumed) do not create discontinuities in the way $G_\nu \nu$ varies over space. Typical applications include studies of the effects of school quality on house prices (Black 1999), the effect of local taxes on firm employment (Duranton et al., 2011), and the evaluation of area-based initiatives (Mayer et al., 2012; Einio and Overman, 2014). This boundary-discontinuity design amounts to defining G_ν to be a block diagonal matrix, in which pairs of places that share the same nearest boundary and are close to the boundary (e.g., within some distance threshold) are assigned equal nonzero (row-normalized) weights. G_z, on the other hand, is structured such that a row for an individual i, located at s_i, assigns nonzero weights to places on the same side of the administrative boundary, and zero weights (or much smaller weights) to places in different administrative districts to location s_i. Restricting G_ν in this way implicitly assumes that observations close to an administrative boundary share the same spatial unobservables, but that area-level determinants are at work at the administrative district or sub-administrative district level. The main threat to identification in this boundary-discontinuity regression discontinuity design is that this assumption may not hold. For example, individuals may sort across the boundary in response to cross-boundary differences in $G_z Z$, so unobserved individual characteristics will differ across the boundary, leading to a change in $G_\nu \nu$ across the boundary. Again, note that it is the assumptions on the structure of $G_\nu \nu$ that have failed in this example.

There are also extensions to the spatial differencing/fixed effects idea in which G is not idempotent, but $\text{plim}[G_\nu G_\nu] = \text{plim}[G_\nu]$. This would be true for any case in which G_ν forms an estimate of the mean of ν at each location s, because $E[E[\nu|s]|s] = E[\nu|s]$. This is the case if each row of G, $g(s)$ is structured such that it comprises a sequence of weights $[gi1 \;\; gi2 \;\; gi3 \;\; \ldots]$ which decline with the distance of locations 1,2,3,… from location s, and sum to 1, which yields a standard kernel weighting structure. Applications of this approach are given in Gibbons and Machin (2003) and Gibbons (2004). However, the basic problem remains that the spatial weights used to aggregate spatial variables of interest $G_x X\theta$ and $G_z Z\delta$ must be different from the spatial weights used in the transformation to sweep out the unobservables ν.

As with the reflection problem, if $G_y = G_x = G_z = G_\nu = G$ is known and the network is incomplete, then G^2X, G^3X, G^2Z, G^3Z,… continue to provide valid instruments for Gy, although not for Gx or Gz. That is, an incomplete structure for G can solve the reflection problem and allow estimation of the coefficient on endogenous effects ($G_y y$) in the presence of peer-group-specific effects that are correlated with observables. But this cannot provide us with an estimate of the coefficients on either Gx or Gz. More generally, the other way to think about these spatial models with sorting and correlated spatial shocks is in terms of the class of general problems where x and z may be correlated with the error term and to look for ways of instrumenting using variables that are exogenous but correlated with the included variables. This approach requires theoretical reasoning about appropriate instruments. However, even then, the instruments must be orthogonal to the spatial unobservables, so it is often necessary to apply instrumental variables combined with spatial-differencing-based methods (see, e.g., Duranton et al., 2011).

In a nutshell, when group membership is exogenous and there are unobservable variables that are correlated with observables, our ability to estimate coefficients of interest depends on the structure of the spatial interactions. If we are willing to assume that the interconnections between individuals on these unobserved dimensions are best described by a matrix of interconnections G_ν that is symmetric and idempotent, then these unobservables can be partialled out using standard differencing/fixed effects methods. If we wish to estimate the coefficients on the spatial explanatory variables G_xX, G_zZ, we must further assume that the interconnections between individuals that form the group-level or spatial averages of the explanatory variables (i.e., G_x and G_z) must be different from G_ν. If this assumption holds, the spatial differencing/fixed effects design eliminates the spatially correlated unobservables, but does not eliminate the spatial explanatory variables. Neither of these assumptions is sufficient to allow the estimation of $G_y y$. If we wish to estimate the coefficient on $G_y y$, then we must assume a known incomplete interaction matrix. This solves the reflection problem and allows the estimation of the coefficient on $G_y y$ but not on G_xX or G_zZ (in either the structural or the reduced form).

Note that the issues and solutions discussed in this section are essentially the same as those for standard omitted variables, but where the correlation between unobservables and observables arises through channels that may not be immediately obvious without thinking about the spatial relationships at work. A subtler consequence of omitted spatial variables is the so-called modifiable areal unit problem (see, e.g., Openshaw, 1983; Wong, 2009; Briant et al., 2010) in which estimates of parameters can change as the spatial aggregation of the units of analysis changes. We say more about this issue in Appendix A.

3.4.3 Sorting and spatial unobservables

In the previous section we considered the possibility, explicit in Equation (3.2) or Equation (3.5), that there are spatial or group-specific unobservables, $m_\nu(\nu,s)_i$ or $G_\nu \nu$ using the matrix form, which are correlated with the explanatory variables. Our discussion there assumed that group membership was exogenous. In this section we allow for the possibility that group membership is endogenous so that the correlation between $G_x x$ and $G_z z$ with $u = G_\nu \nu\lambda + \varepsilon$ stems from individual-level decisions about group membership. As discussed above, while the urban economics literature has traditionally recognized these two mechanisms through which $G_x x$ and $G_z z$ may be correlated with $G_\nu \nu$, it has tended to treat these symmetrically. However, when group membership is endogenous, the correlation between $G_x x$ or $G_z z$ and $G_\nu \nu$ arises because G_x, G_z, and G_ν are endogenous.

If the individual-level variables that affect location also affect outcomes, then a fixed effects approach can do little to alleviate this problem as the individual-level unobservables would not be eliminated when subtracting a group-mean. To return to the urban wage premium example, including individual-level and city-level fixed effects does not consistently identify the urban wage premium if unobserved shocks (e.g., a change in labor market circumstances) affect both wages and location.

In much of the urban economics literature, the response to this problem has been to suggest that this is the best that can be achieved in the absence of random allocation across locations (we consider this further in the next section). An alternative is to impose more structure on the location problem. Ioannides and Zabel (2008), for example, use factors influencing neighborhood choice as instruments for neighbors' housing structure demand when estimating neighborhood effects in housing structure demand. The literature on equilibrium sorting models and hedonics may lead to further theoretical insights into identification of neighborhood effects when the researcher is prepared to impose more structure on the neighborhood choice process (Kuminoff et al., 2013).

Various estimation techniques have recently been developed in the econometrics of network literature to address the issue of endogenous group membership. These have not yet been applied in spatial settings although they may be helpful (particularly for researchers taking a more structured approach). There are three main methodological approaches. In the first approach, parametric modeling assumptions and Bayesian inferential methods are employed to integrate a network formation model with the model of behavior over the formed networks. The selection equation is based on individual decisions and considers all the possible couple-specific correlations between unobservables. This is a computationally intense method where the network formation and the outcome equation are estimated jointly (Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2013; Mele, 2013; Del Bello et al., 2014; Patacchini and Rainone, 2014). The alternative approach is the frequentist approach, where a selection equation based on individual decisions is added as a first step prior to modeling outcome decisions. An individual-level selection correction term is then added in the outcome equation. The properties of the estimators are analytically derived. Observe that, while the idea is similar to a Heckman-type estimation, inference is more difficult because of the complex cross-sectional interaction scheme. This approach is considered in Liu et al. (2012). Finally, another strategy is to deal with possible network endogeneity by using a group-level selection correction term. The group-level selection correction term can be treated as a group fixed effect or can be estimated directly. Estimation can follow a parametric approach as in Lee (1983) or a semiparametric approach as in Dahl (2002). This method is considered in Horrace et al. (2013).

In the peer groups/social interactions literature that employs the network structure as a source for identification, network or "component" fixed effects can sometimes be used to control for sorting into self-contained networks or subsets of the networks (Bramoullé et al., 2009; Calvó-Armengol et al., 2009; Lee et al., 2010; Lin, 2010; Liu and Lee, 2010). For example, children whose parents have a low level of education or whose level of education is worse than average in unmeasured ways are more likely to sort into groups with low human capital peers. If the variables that drive this process of selection are not fully observable, potential correlations between (unobserved) group-specific factors and the target regressors are major sources of bias. The richness of social network data (where we observe individuals over networks) provides a possible way out through the use of network fixed effects, for groups of individuals who are connected together, assuming individuals fall into naturally disconnected subgroups, or some cutoff in terms of connectivity can be used for partitioning into subgroups. Network fixed effects are a potential remedy for selection bias that originates from the possible sorting of individuals with similar unobserved characteristics into a network. The underlying assumption is that such unobserved characteristics are common to the individuals within each network partition.[25] This may be a reasonable assumption where the networks are quite small—for example, a network of school students. When networks contain instead a large number of agents who are not necessarily drawn together by anything much in common—for example, a network of LinkedIn connections—this is no longer a viable strategy as it is not reasonable to think that the unobserved factors are variables which are common to all members. As another example, networks of transactions in the housing market that involve a large number of properties may contain different types of unobservables for different properties, even though all the properties belong to the same network of buyers and sellers. In this case, the use of network fixed effects would not eliminate endogeneity problems. A similar context is provide by trading networks with financial data. Also in this case, when the number of transactions is high, the use of network fixed effect is not a valid strategy, although network topology can still contain valuable information (see Cohen-Cole et al., 2014). Obviously, it must also be feasible to partition individuals into mutually exclusive sets of individuals (or units) who are not directly or indirectly related in the network in order to define the fixed effects, so this is not a solution in networks where all individuals are indirectly related to each other.

3.4.4 Spatial methods and identification

To summarize, all researchers working with spatial data face fundamental identification and estimation challenges. Spatial methods can provide a partial solution to these challenges. Restrictions on functional form, on the exogenous variables that directly determine outcomes, and on the nature of interactions may solve the reflection problem and allow identification of interaction effects. But identification fails if these restrictions

[25] Testable implications of this assumption can be verified using the recent approach proposed by Goldsmith-Pinkham and Imbens (2013). Patacchini and Venanzoni (2014) apply this approach to an urban topic.

are invalid. Further challenges to identification arise if there are omitted variables that are correlated with observables. These challenges arise when estimating models with or without endogenous interactions. Standard solutions to these problems (e.g., fixed effects, spatial differencing) imply restrictions on the nature of spatial interactions. Reformulating these approaches within a spatial econometrics framework makes these restrictions explicit. If the omitted variables problem arises because of sorting across space (i.e., location is endogenous), this raises further identification problems. Again, reformulating sorting within the spatial econometrics framework, specifically as giving rise to an endogenous interaction matrix, helps clarify these issues. The network literature and the spatial econometrics literature suggest some solutions to the sorting problem although all of these require further assumptions and restrictions on the model that determines location. In situations where researchers are unwilling to impose these restrictions, it is often suggested that the use of standard spatial methods (e.g., fixed effects or spatial differencing) provides the best estimates that we can hope for in the absence of random allocation across locations. Unfortunately, recent literature questions the extent to which even random allocation may help. It is to this question that we now turn.

3.5. TREATMENT EFFECTS WHEN INDIVIDUAL OUTCOMES ARE (SPATIALLY) DEPENDENT

In this section, we recast the discussion so far in terms of the framework used in the policy evaluation literature, where the aim is to estimate the treatment (causal) effect of some policy intervention.[26] We consider the extent to which explicit experiments—for example, randomized controlled trials (RCTs)—can be designed to overcome the basic identification problems discussed above. Doing so helps reinforce the intuition provided above by considering the issues within a different conceptual framework, as well as providing a link to the evaluation literature that applies RCTs in settings where spatial or network dependence may be important.

3.5.1 (Cluster) randomization does not solve the reflection problem

As discussed above, the reflection problem can prevent estimation of β (the effect of neighbor outcomes or behavior on individual outcomes) separately from θ (the effect of neighbor characteristics) in situations where there are unobservable factors that also vary at the group level. Unfortunately as this section shows, without the imposition of further restrictions, randomization does not generally solve the reflection problem.

[26] A burgeoning literature considers the application of treatment effect analysis to economic problems. Early surveys include those of Angrist and Krueger (1999) and Heckman et al. (1999), while Lee (2005) provides a book-level treatment. Angrist and Pischke (2011), among a number of others, provide further discussion.

To think this through, consider the design of an experiment that would identify the parameters from a standard linear (spatial) interactions model where outcome y is determined by both individual characteristics and the outcome, observed and unobserved characteristics of some reference group (for simplicity we disregard Z or assume it is subsumed in X, and we suppress the constant):

$$y = X\gamma + G_y y\beta + G_x X\theta + u. \tag{3.25}$$

If each individual is a member of at most one reference group (i.e., G is block diagonal), then an RCT could use the existing reference groups (summarized by G) as the basis for the random allocation of treatment. That is, the group, rather than the individuals, can be randomized into treatment. This is the approach taken by cluster randomized trials, which have seen widespread application in the public health literature (see, e.g., Campbell et al., 2004). Note that, although G may be endogenously determined, randomization of groups into treatment ensures that u is uncorrelated with treatment status (at least when there are a large number of available groups). We can model treatment as changing some element of x_i for all members of treated groups while holding everything else constant. Given that there is complete interaction within each group (and assuming G is row normalized), $G_y y$ and $G_x X$ form the sample mean within each group. Thus, treatment affects individuals directly through x_i, and indirectly via both $G_y y$ and $G_x X$. As highlighted by Manski (2013), and discussed further below, these assumptions imply restrictions on the treatment response functions (which characterize the way in which outcomes change with treatment) that are not trivial.

Suppose we have just two groups, group 0 and group 1, with random assignment of treatment to all members of group 1 rather than to members of group 0. We have

$$\text{Treatment group:} \quad E[y|1] = E[x|1](\gamma+\theta)/(1-\beta) + E[u|1]/(1-\beta), \tag{3.26}$$

$$\text{Control group:} \quad E[y|0] = E[x|0](\gamma+\theta)/(1-\beta) + E[u|0]/(1-\beta), \tag{3.27}$$

where random assignment implies $E[y|1] - E[y|0] = 0$, given that $E[x|1] - E[x|0] = 0$, $E[u|1] - E[u|0] = 0$. Now we expose all members of the treatment group to some known treatment, by changing some element of x_i for all members of the treatment group (group 1) while holding everything else constant, to give $E[x|1] - E[x|0] = x^\star$. This gives the reduced form, causal effect of the treatment:

$$\begin{aligned} E[y|1] - E[y|0] &= (E[x|1] - E[x|0])(\gamma+\theta)/(1-\beta) \\ &= x^*(\gamma+\theta)/(1-\beta). \end{aligned} \tag{3.28}$$

For many policy evaluation purposes this is sufficient, but it is clear that cluster randomization does not solve the reflection problem and allow the separate estimation of γ, θ, and $(1-\beta)$. With control over within-cluster assignment to treatment it is possible to go further (under the assumptions imposed so far) and separately identify the direct effect of the intervention (γ) from the effects due to social interactions. We show an example in Appendix B. Note, however, that control over group membership when individuals are members of only one group (i.e., G is block diagonal) does not provide a solution to the reflection problem or allow us to separately identify θ or $(1-\beta)$.

In addition, note that applying cluster randomization to existing reference groups raises issues with respect to inference when (a) group membership is endogenous, or (b) there are omitted group-specific variables that affect outcomes. Both situations imply that the characteristics of individuals are correlated with the characteristics of others in their group. This within-group correlation in terms of either observable or unobservable characteristics (often referred to as intracluster correlation) reduces the effective sample size in a way that depends on both the size of the within-group correlation and the average group size relative to the total sample size. When within-group correlation equals 1 (so that individuals are identical within groups in terms of characteristics which determine γ), the effective sample size is equal to the number of groups. When within-group correlation in the characteristics that determine γ is 0, the effective sample size is equal to the total number of individuals in the two groups. For intermediate situations, basing inference only on the number of groups will result in standard errors that are too large, while using the total number of individuals will result in standard errors that are too small. Using conservative standard errors (based on group size) will exacerbate concerns over power (i.e., the probability of correctly rejecting the null hypothesis of no treatment effect when the null is false) in situations where the number of groups is small and the within-group correlation is large.

In situations where the researcher has control over group membership, random assignment of individuals to treatment and control groups, rather than random assignment of treatment to all members of existing groups, helps address these concerns over inference. This is because individual-level randomization reduces this within-group correlation in terms of both observable and unobservable characteristics, given that group membership is no longer endogenously determined. It also ensures that u is uncorrelated with treatment status in situations where unobservable characteristics are correlated within groups (as will usually be the case when group membership is endogenous). However, even if we randomly allocate individuals to treatment and control groups, if we want these individuals in the treated group to interact, then they have to be colocated somewhere and if they are colocated, then they will be subject to place-specific unobservables. Therefore, even this form of randomization does not completely eliminate the problems for inference induced by treating people in groups.

In practice, it is perhaps difficult to think of situations where we would have such strong control over both group membership and treatment assignment within groups. But thinking about the appropriate RCT helps clarify intuition about the kind of quasi-random variation needed to achieve identification of the direct effect γ separately from the effects of interaction between agents. Conditional on the assumption about the treatment response function,[27] an RCT with control over both group membership and individual assignment into treatment allows us to eliminate biases due to selection on unobservables into the two groups, and to estimate the reduced form effect of changes in x and group average x. The quasi-experimental methods for causal analysis on non-experimental data discussed in Chapter 1 are therefore perfectly applicable to this problem providing they can use two sources of quasi-random variation: the first to determine assignment into treatment, the second to determine assignment into the reference group. Note, however, that simple treatment/control randomization does not solve the "reflection" problem of separate identification of β and θ, so clearly methods based on quasi-random variation will also fail in this respect.

Is there an experiment that separately identifies β and θ? As before, we must impose more structure on the problem to achieve identification. It should be clear from Section 3.4 that an appropriate identification strategy must rely on overlapping but incomplete network structures (i.e., a nonidempotent G matrix with intransitive network relationships). Appendix B provides an example of a simply hypothetical experiment that fulfills these criteria.

As can be seen, the requirements for a successful RCT to identify the separate causal parameters in the general spatial model of Equation (1) are rather stringent. Two key components are required: (a) randomization into different groups; (b) a known and enforceable "incomplete" network structure that defines the permissible interactions between agents in these groups. Even then there are evidently problems when trying to design such a hypothetical experiment to answer questions that are specifically spatial, such as questions about neighborhood effects or geographical spillovers. For example, in the hypothetical experiment discussed in Appendix B, individuals are assigned into a control group and three treatment groups (groups 1–3). The crucial restriction for identification is that individuals in group 1 are connected to individuals in group 2 and individuals in group 2 are connected to individuals in group 3, but individuals in groups 1 and 3 are not connected. If the connections are spatial, then ensuring compliance is not so straightforward, since group 1 must overlap with group 2 in space and group 2 must overlap with group 3 in space, so it is very hard to ensure that group 3 does not overlap with group 1 in geographical space. Given the difficulties of designing a hypothetical experiment to recover these parameters, it becomes clear that recovering them from observational data when there is no explicit randomization and/or the true network structure of G is unknown is going to be difficult.

The situation is further complicated once we relax the assumption on the treatment response function that we have imposed so far (i.e., that treatment affects individuals directly through x_i, and indirectly via both $G_y y$ and $G_x X$). As emphasized by Manksi (2013), once we allow for the possibility of social interaction, it is hard to maintain

[27] That is, that treatment affects individuals directly through x_i, and indirectly via both $G_y y$ and $G_x X$.

the assumption that individual outcomes only vary with own treatment, and not with treatment of other members of the population. That is, the stable unit treatment value assumption (Rubin, 1978) that underpins much of the treatment effects literature is unlikely to hold. As Manski (2013) makes clear, the stable unit treatment value assumption, or "individualistic treatment response" assumption (as he calls it) is quite restrictive in situations that allow for social interaction. Indeed, in the examples above, we dropped this assumption to allow the treatment effect to depend on both the individual treatment and the average level of treatment in the group (as captured by $G_y y$ and $G_x X$). Manski (2013) defines this as a functional interaction response (the interaction occurs only through some function of the distribution of treatments across the groups—in this case the mean). Relaxing this assumption would give us what Manski calls distributional interactions (where individual treatment response depends on the distribution of treatments across others in the group but not on the size of the group or the identity of those treated). A further relaxation gives anonymous interactions (the outcome of person j is invariant with respect to permutations of the treatments received by other members of his group, but the size of the group could matter). Progressively weaker assumptions on the treatment response function make identification more difficult. The situation is further complicated if we allow reinforcing or opposing interactions (two examples of "semimonotone treatment response functions"). Treatment could also influence group structure if, for example, treatment is observable and individuals sort on the basis of treatment. In short, even in situations where G is known and structured such that $GG \neq G$, further assumptions on the nature of the treatment response function are required to identify treatment effects of interest. The literature that considers these issues is in its infancy.

3.5.2 Randomization and identification

It is increasingly common for the applied urban economics literature to suggest that the application of spatial methods (e.g., fixed effects, spatial differencing) represents the "best we can do" in the absence of explicit randomization. While this may be true, this section showed that randomization itself may be insufficient to solve fundamental identification problems, especially where the aim is to identify endogenous neighborhood effects or spillovers of the SAR variety in spatial econometrics. Even in situations where the researcher has control over group structure and treatment, identification of β (the effect of neighbor outcomes or behavior on individual outcomes) separately from θ (the effect of neighbor characteristics) is not straightforward. Uncertainty about treatment response (i.e., the appropriate functional form) or the endogeneity of group membership (especially to treatment) further complicates the problem, as well as providing an additional set of challenges to researchers interested in identifying reduced form treatment effects. The nascent literature considering this latter issue is yet to receive widespread consideration in the applied treatment effects literature. However, this emerging literature makes it clear that much applied work relies on restrictions on the treatment response function, in particular the individual treatment response assumption, which may not hold in practice. Dealing with these issues is one of the key challenges facing those who wish to develop and apply the treatment effects approach in spatial settings.

3.6. CONCLUSIONS

This chapter has been concerned with methods for analyzing spatial data. After initial discussion of the nature of spatial data and measuring and testing for departures from randomness, we focused most of our attention on linear regression models that involve interactions between agents across space. The introduction of spatial variables—functions that generate (usually linear) aggregations of variables that are spatially connected with a specific location using information on all locations—into standard linear regression provides a flexible way of characterizing these interactions. The introduction of these spatial variables complicates both interpretation and estimation of model parameters of interest. This raises the question of whether one could ignore these spatial variables and still correctly determine the impact of some specific variable x on some outcome y? As is usually the case, however, model misspecification—in this case ignoring interactions between individuals when they are relevant—means that OLS results may be misleading. In some circumstances—for example, when we are interested in the impact of some policy intervention x on some outcome y—the OLS bias may not be problematic. In other cases, this bias will be a problem. This is one reason to consider how to estimate models which allow for spatial interactions. A second, more substantive, reason is that the spatial interactions themselves may be objects of interest.

Once we switch focus to the estimation of models including spatial variables, we face three fundamental challenges which are particularly important in the spatial setting: the so-called reflection problem, the presence of omitted variables that imply correlated effects (or common shocks), and problems caused by sorting.

In most settings using observational data, the reflection problem is very likely to occur unless we are able to impose further restrictions. We consider three possible solutions involving restrictions on the functional form, (exclusion) restrictions on the exogenous variables that directly determine outcomes, and restrictions on the nature of interactions. This last solution has been widely applied in the spatial econometrics literature through the use of *ad hoc* spatial weight matrices that assume interactions are incomplete, so have the property that $GG \neq G$. This strategy has been more recently applied in the social interaction literature, which exploits the architecture of network contacts to construct valid instrumental variables for the endogenous effect (i.e., by using the characteristics of indirect friends). However, in our view, these restrictions require careful justification on the basis of institutions, policy, or theory (or need to be imposed on the basis of data that identify relevant linkages). These issues have received careful consideration in the networks and theoretical spatial econometrics literature, but much applied work continues to rely on *ad hoc* restrictions imposed through the choice of popular spatial weight matrices. Unfortunately, identification fails if these restrictions (whether carefully justified or imposed *ad hoc*) are invalid.

For some, especially those working within the experimentalist paradigm, the information requirements associated with these techniques are sufficiently profound that they may favor estimation of the reduced form with a specific focus on addressing problems created by sorting and omitted spatial variables. However, as we have shown, similar assumptions on the structure of G are implicit in the frequently applied empirical strategies—fixed effects or spatial differencing—used to address these problems. Our discussion above makes these assumptions explicit, which suggests that there may be an argument for greater use of the general spatial form in structuring applied microeconometric studies. Unfortunately, when the source of the omitted variables is due to endogenous sorting, it is very difficult to make progress without imposing further assumptions on the process that determines location. We show that these general lessons carry over to the policy evaluation literature, where the aim is to estimate the causal effect of some policy intervention. In particular, the requirements for a successful RCT to identify the separate causal parameters in the general spatial model are stringent. The difficulties inherent in designing the hypothetical experiment serve to emphasize the challenges for studies using observational data as well as pointing out the limits of RCTs in addressing these problems.

If there is one overarching message to emerge from this chapter, it is that while the use of spatial statistics and econometrics techniques to answer relevant questions in urban economics is certainly a promising avenue of research, the use of these techniques cannot be mechanical. As we discussed in this chapter, there are a variety of challenges and various possible solutions. Ultimately, the choice of the most appropriate model, identification, and estimation strategy depends on the mechanism underlying the presence of spatial effects and cannot be based only on statistical considerations.

APPENDIX A: BIASES WITH OMITTED SPATIAL VARIABLES

Even when estimation of spatial or social interactions is not the main goal, omission of salient spatial variables and variables capturing social interactions can obviously have important consequences for the estimates of other parameters. This is just a standard omitted variables problem. In the main text, we show that interactions between individuals may stem from the effects of (1) group-level individual characteristics, (2) group-level characteristics of other entities or objects, or (3) the outcomes for other individuals in the reference group. Omitting any of these sources of interaction leads to biases on the estimates of the effects of the other variables, although the importance of these biases in practice depends to some extent on the intended purpose of the estimation.

Suppose interactions really occur only through group-level characteristics—that is, contextual effects—so Equation (3.5) becomes (using matrix notation)

$$y = X\gamma + G_x X\theta + \varepsilon.$$

Now suppose we try to estimate γ using a (misspecified) standard regression model in which individual outcomes depend only on own characteristics:

$$y = X\gamma + \varepsilon. \tag{A.1}$$

There is now a standard omitted variables bias due to omission of $G_x X\theta$, given that $G_x X$ is correlated with X by construction. The bias in the OLS estimate of γ is increasing in the importance of neighbors' or peers' characteristics in determining individual outcomes, θ:

$$\hat{\gamma}_{\mathrm{OLS}} = \gamma + (X'X)^{-1} X' G_x X\theta. \tag{A.2}$$

An analogous argument holds for omission of external attributes of the group $G_z Z$, when the correct specification is

$$y = X\gamma + G_z Z\delta + \varepsilon,$$

although clearly the magnitude of the bias will depend on the extent to which $G_z Z$ and X are correlated.

Suppose instead that interactions genuinely occur as a result of individuals' responses to other individuals' outcomes—that is, endogenous effects—so Equation (3.5) becomes

$$y = X\gamma + G_y y\beta + \varepsilon.$$

If we mistakenly estimate γ using Equation (A.1), the OLS estimator is

$$\begin{aligned}
\hat{\gamma}_{\mathrm{OLS}} &= \gamma + (X'X)^{-1} X' G_y y\beta \\
&= \gamma + (X'X)^{-1} X' G_y X\gamma\beta + (X'X)^{-1} X' G_y^2 y\beta^2 \\
&= \gamma + (X'X)^{-1} X' G_y X\gamma\beta + (X'X)^{-1} X' G_y^2 X\gamma\beta^2 \\
&\quad + (X'X)^{-1} X' G_y^3 X\gamma\beta^3 + \cdots
\end{aligned} \tag{A.3}$$

by repeated substitution, implying an infinite polynomial series of bias terms. OLS will be biased if $\beta > 0$. The bias goes to infinity when β approaches 1 (where the estimator is not defined) and it goes to 0 as β goes to 0. The intuitive reason for this bias is simply that the effect of X operating through γ is amplified through feedback between neighbors or peers, with the effect of X on one individual having an effect on its neighbor, and vice versa. In the case where G_y is a simple symmetric block diagonal, mean-creating matrix such as Equation (3.7), this bias expression simplifies to

$$\hat{\gamma}_{\mathrm{OLS}} = \gamma + (X'X)^{-1} X' G_y X\gamma\beta/(1-\beta). \tag{A.4}$$

Finally, let us consider the case where interactions occur in terms of both group-level characteristics and outcomes—that is, the real relationship is

$$y = X\gamma + G_y y\beta + G_x X\theta + \varepsilon.$$

If we estimate γ using model (A.1)—that is, omitting both endogenous effects, $G_y y$, and contextual effects, $G_x x$—the OLS estimator is

$$\begin{aligned}\hat{\gamma}_{\text{OLS}} &= \gamma + (X'X)^{-1}X'G_xX\theta + (X'X)^{-1}X'G_y y\beta \\ &= \gamma + (X'X)^{-1}X'G_xX\theta + (X'X)^{-1}X'G_yX\gamma\beta \\ &\quad + (X'X)^{-1}X'G_yG_xX\theta\beta + (X'X)^{-1}X'G_y^2 y\beta^2 \\ &= \gamma + (X'X)^{-1}X'G_xX\theta + (X'X)^{-1}X'G_yX\gamma\beta \\ &\quad + (X'X)^{-1}X'G_yG_xX\theta\beta + (X'X)^{-1}X'G_y^2X\gamma\beta^2 \\ &\quad + (X'X)^{-1}X'G_y^2G_xX\theta\beta^2 + \cdots,\end{aligned} \tag{A.5}$$

and again if $G_y = G_x = G$ is a simple block diagonal mean-creating idempotent matrix, this simplifies to

$$\hat{\gamma}_{\text{OLS}} = \gamma + (X'X)^{-1}X'GX(\gamma\beta + \theta)/(1-\beta). \tag{A.6}$$

If we disregard the pathological case where $\beta\gamma = -\delta$, OLS will be baised, with the bias depending on both β and θ. The bias goes to infinity when β goes to 1 or θ goes to infinity and it goes to 0 if both β and θ go to 0. Again the bias is intuitive and includes effects due to omitted contextual interactions working through θ and the individual impacts γ, both amplified by the feedback effect between neighbors β.

Of course, for a policy maker interested in the effect of some treatment X, this “biased” parameter is exactly what that policy maker is interested in: the reduced form effect of the policy, taking into account the amplifying effects of the spatial interactions between agents—both in the sense that individuals are affected by their own treatment γ and the treatment of their neighbors δ, and because there is feedback via the outcomes that the treatments induced (the multiplicative factor $1/(1-\beta)$). Whether this estimate should be considered the “causal” effect of treatment depends on the definition of causality as discussed in the main text, although in the usual interpretation in the program effects literature this biased parameter is indeed a causal parameter. Regardless, this reduced form interpretation of the OLS coefficient is the fundamental reason why researchers interested in policy treatment effects may care more about other threats to identification than about carefully delineating the various types of spatial or social interaction. We discussed these issues further in Section 3.5.

In some situations, where researchers are interested in trying to understand the structure of spatial and social interactions out of curiosity, rather than for any instrumental policy purpose, this reduced form interpretation is not very helpful. A researcher may be interested specifically in the identification of the structural parameter γ, or the interaction terms θ and β may be of substantive interest. If simply disregarding the interaction effects is not an attractive option, the researcher needs to adopt methods for estimation which allow for the inclusion of these interactions, although as we have shown in Section 3.4, identification of these parameters is not easy.

Omitting spatial variables can also lead to a lot of confusion, because it gives rise to the problem usually called the modifiable areal unit problem (see, e.g., Openshaw, 1983; Wong, 2009; Briant et al., 2010). This refers to the empirical observation that estimates of parameters can change substantially as the researcher changes the level of spatial aggregation of the data on which the analysis is conducted (moving, for example, from individual microdata, to districts to regions, or even abstract regular geometric aggregations as shown in Briant et al., 2010). The reasons for this problem in regression applications are clear from the above discussion, in that changing the level of aggregation changes the relative weights of the individual effects γ and the effects arising from spatial interactions (or other spatial variables). For example, suppose the underlying relationship at the individual level is

$$y = X\gamma + G_x X\theta + \varepsilon$$

as in the first example above, and we estimate a regression of y on X using individual data, omitting the spatial variable $G_x X$. Then as shown above, the OLS estimate is $\hat{\gamma}_{\text{OLS}} = \gamma + (X'X)^{-1} X' G_x X\theta$. This is a weighted average of γ and θ which depends on the sample covariance between $G_x X$ and X and the sample variance of X. As we perform aggregation up from the individual level to higher geographical levels of aggregation, the weight on θ increases, until, if we perform estimation at the level of aggregation defined by G_x—that is, we estimate $G_x y = G_x X\gamma + G_x X\theta + \varepsilon$—we obtain $\hat{\gamma}_{\text{OLS}} = \gamma + \theta$. Similar issues arise if the omitted variable is not $G_x X$, but is any other spatial variable that is correlated with X.

APPENDIX B: HYPOTHETICAL RCT EXPERIMENTS FOR IDENTIFYING PARAMETERS IN THE PRESENCE OF INTERACTIONS WITHIN SPATIAL CLUSTERS

In Section 3.5 we noted that standard clustered RCT designs can identify only a composite parameter characterizing a combination of the direct effects of an intervention plus the social multiplier effects from contextual and endogenous interactions between treated individuals in spatial clusters. However, we noted that experiments could potentially be designed to recover some or all of these parameters. Here, we provide some simple examples, which we hope further elucidate the more general problems of identifying the parameters in models with spatial and social interaction.

The standard clustered RCT experiment described around Equation (3.26) allowed us to estimate the overall effect of a policy intervention $x^\star$ in the presence of interactions within the randomly treated spatial clusters: $E[y|1] - E[y|0] = x^\star(\gamma + \theta)/(1 - \beta)$.

Suppose now, rather than randomly treating some clusters (treatment) and not others (control), we have control over the share of individuals who are randomly treated within each cluster. We use s to denote the share of individuals who are treated within a cluster, such that for those individuals $E[x|1] - E[x|0] = x^{\star}$, but for the cluster we have $E[x|s] = x^{\star}s$.

From this experiment we could estimate the means of the outcomes for the treated individuals in each cluster, the nontreated individuals in each cluster, and the mean outcome in each cluster, which would vary with the share s treated.[28] Mean outcome in cluster is:

$$\begin{aligned} E[y|s] &= \beta E[y|s] + x^* s(\gamma + \theta) \\ &= x^* s(\gamma + \theta)/(1 - \beta). \end{aligned} \tag{B.1}$$

Individual treated directly in cluster with share s treated

$$\begin{aligned} E[y|1,s] &= \beta E[y|s] + x^*(\gamma + s\theta) \\ &= x^* s[\beta(\gamma + \theta)/(1 - \beta) + \theta] + \gamma x^*. \end{aligned} \tag{B.2}$$

Individual not treated directly, in cluster with share s treated

$$\begin{aligned} E[y|0,s] &= \beta E[y|s] + x^* s\theta \\ &= x^* s[\beta(\gamma + \theta)/(1 - \beta) + \theta]. \end{aligned} \tag{B.3}$$

And subtracting the mean for those not treated from the mean of those treated recovers the direct effect of the treatment:

$$E[y|1,s] - E[y|0,s] = x^* \gamma. \tag{B.4}$$

Hence, with two or more clusters available, with different shares treated, we can identify γ and a composite parameter representing the strength of social interactions $\beta(\gamma + \theta)/(1 - \beta) + \theta$. However, this still does not provide a solution to the reflection problem and allow the separate estimation of θ and $(1 - \beta)$.[29]

Attempting to separately identify the endogenous interactions β is more complex, and requires that the experimental structure mimics the intransitive network grouping structure discussed as a prerequisite for identification in Section 3.4. The idea is to create some groups of individuals who are treated directly, some groups of individuals who are treated indirectly through interaction with the individuals treated directly (endogenous and contextual effects), and some individuals who are treated only indirectly through interaction with others who are treated only indirectly (endogenous effects).

We create four groups of individuals (groups 0, 1, 2, and 3), in which group 0 is a control group. Individuals are randomly assigned to equal-size groups 1, 2, and 3 in triads

[28] Here we are assuming the standard linear in means expression for individual outcomes as in (3.6).

[29] We could also use group assignment to identify γ and $\theta/(1 - \beta)$ by completely isolating some agents. For isolated agents, the difference in expected outcomes between treated and untreated individuals is $E[y|1] - E[y|0] = (E[x|1] - E[x|0])\gamma = x^{\star}\gamma$, which provides estimates of the direct effect γ.

in which an individual in group 1 interacts with an individual in group 2 and this individual in group 2 also interacts with an individual in group 3, but the individual in group 1 does not interact with an individual in group 3. Also, for simplicity of notation, we assume that individuals in a given group cannot interact with other individuals in that group. Again, we set aside practical considerations about how this system of interactions might be enforced. Agents are randomized across all three groups, so $E[y|j] - E[y|k] = E[x|j] - E[x|k] = E[u|j] - E[u|k] = 0$ for all j and k. Group 1 is subject to an intervention $x^\star$

For a simple example of only two agents in each group, the structure of the G matrix is, by design,

$$G = \begin{array}{cc} & \begin{array}{cccccccc} a & b & c & d & e & f & g & h \end{array} \\ \begin{array}{c} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \end{array} & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array},$$

where a and b belong to group 0, c and d belong to group 1, e and f belong to group 2, and g and h belong to group 3. Clearly $GG \neq G$, so we could simply apply the results from Section 3.4. Once again, however, we think it is instructive to work through this specific example within the case–control RCT paradigm to further develop understanding of how identification is achieved and what this tells us about how difficult this might be in nonexperimental settings.

Following the standard structure of linear interactions and using the notation $DE[x_i|j] = E[x_i|j] - E[x_i|0]]$ and so on (i.e., differences from control group means), we find the expressions for individuals in each group are as follows:

$$E[y|0] = E[x|0]\gamma + E[u|0], \tag{B.5}$$

$$E[y|1] = E[y|2]\beta + E[x|1]\gamma + E[u|1], \tag{B.6}$$

$$E[y|2] = (E[y|1] + E[y|3])\beta/2 + (E[x|1] + E[x|3])\theta/2 + E[x|2]\gamma + E[u|2], \tag{B.7}$$

$$E[y|3] = E[y|2]\beta + E[x|2]\theta + E[x|3]\gamma + E[u|3]. \tag{B.8}$$

With randomization and intervention in group 1,

$$DE[y|1] = DE[y|2]\beta + x^*\gamma, \tag{B.9}$$

$$DE[y|2] = (DE[y|1] + DE[y|3])\beta/2 + x^*\theta/2, \tag{B.10}$$

$$DE[y|3] = DE[y|2]\beta. \tag{B.11}$$

We get the reduced form for $DE[y|2]$ by substituting $DE[y|1]$ and $DE[y|3]$ in Equation (B.10):

$$\begin{aligned} DE[y|2] &= DE[y|2]\beta^2 + x^*(\gamma\beta+\theta)/2 \\ &= x(\gamma\beta+\theta)/2(1-\beta^2) \\ &= x^*\pi, \end{aligned} \tag{B.12}$$

where π is the composite parameter $(\gamma\beta/2 + \theta)/2(1 - \beta^2)$

Since $DE[y|3] = x^\star\pi\beta$ and $DE[y_i|2] = x^\star\pi, \beta = DE[y|3]/DE[y|2]$. In other words, an estimate of the endogenous interaction coefficient β could be obtained from this experiment by taking the difference between means outcomes of group 3 and group 0, and dividing by the difference in means between group 2 and group 0. This is equivalent to an instrumental variables estimate, using the intervention $x^\star$ as an instrument for $DE[y|2]$ in the regression of $DE[y|3]$ on $DE[y|2]$ (with obvious parallels to the way identification is achieved in the network literature as described in Section 3.4).

REFERENCES

Aaronson, D., 1998. Using sibling data to estimate the impact of neighborhoods on children's educational outcomes. J. Hum. Resour. 33 (4), 915–946.

Abbasi, A., Altmann, J., Hossain, L., 2011. Identifying the effects of co-authorship networks on the performance of scholars: a correlation and regression analysis of performance measures and social network analysis measures. J. Informetr. 5 (4), 594–607.

Angrist, J., Krueger, A., 1999. Empirical strategies in labor economics. In: Ashenfelter, A., Card, D. (Eds.), Handbook of Labor Economics 3A. North-Holland, Amsterdam.

Angrist, J., Pischke, J.S., 2009. Mostly harmless econometrics. Princeton University Press, Princeton.

Angrist, J., Pischke, J.S., 2011. The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. J. Econ. Perspect. 24, 3–30.

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

Anselin, L., 1995. Local indicators of spatial association. Geogr. Anal. 27 (2), 93–115.

Banerjee, A., Besley, T., 1991. Peer Group Externalities and Learning Incentives: A Theory of Nerd Behavior. Princeton University, Mimeo.

Barrios, T., Diamond, R., Imbens, G.W., Kolesar, M., 2012. Clustering, spatial correlations, and randomization inference. J. Am. Stat. Assoc. 107 (498), 578–591.

Benabou, R., 1993. Workings of a city: location, education, and production quarterly. J. Econ. 108, 619–652.

Black, S.E., 1999. Do better schools matter? Parental valuation of elementary education. Q. J. Econ. 577–599.

Borjas, G., Doran, K., 2012. The collapse of the Soviet Union and the productivity of American mathematicians. Q. J. Econ. 127 (3), 1143–1203.

Bound, J., Jaeger, D., Baker, R., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. J. Am. Stat. Assoc. 90 (430), 443–450.

Bramoullé, Y., Djebbari, H., Fortin, B., 2009. Identification of peer effects through social networks. J. Econom. 150, 41–55.

Briant, A., Combes, P.P., Lafourcade, M., 2010. Dots to boxes: do the size and shape of spatial units jeopardize economic geography estimations? J. Urban Econ. 67 (3), 287–302.

Brock, W.A., Durlauf, S.N., 2001. Interactions-based models. In: Heckman, J.J., Leamer, E.E. (Eds.), Handbook of Econometrics, first ed., vol. 5. Elsevier, pp. 3297–3380 (Chapter 54).

Calvó-Armengol, A., Patacchini, E., Zenou, Y., 2009. Peer effects and social networks in education. Rev. Econ. Stud. 76, 1239–1267.
Cameron, A.C., Miller, D.L., 2015. A practitioner's guide to cluster-robust inference. J. Hum. Resour. forthcoming.
Campbell, M.K., Elbourne, D.R., Altman, D.G., 2004. CONSORT statement: extension to cluster randomised trials. BMJ 328, 702.
Case, A., Katz, L., 1991. The company you keep: the effects of family and neighborhood on disadvantaged youths. National Bureau of Economic Research, Inc, NBER Working papers 3705.
Ciccone, A., Peri, G., 2006. Identifying human-capital externalities: theory with applications. Rev. Econ. Stud. 73 (2), 381–412, Oxford University Press.
Cohen-Cole, E., Kirilenko, A., Patacchini, E., 2014. Trading networks and liquidity provision. J. Financ. Econ. 113 (2), 235–251.
Combes, P.P., Overman, H.G., 2004. The spatial distribution of economic activities in the European Union. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics. Cities and Geography, vol. 4. Elsevier, Amsterdam.
Combes, P.P., Duranton, G., Gobillon, L., 2008. Spatial wage disparities: sorting matters!. J. Urban Econ. 63 (2), 723–742.
Conley, T.G., 1999. GMM estimation with cross sectional dependence. J. Econom. 92 (1), 1–45, Elsevier.
Conley, T.G., Molinari, F., 2007. Spatial correlation robust inference with errors in location or distance. J. Econom. 140, 76–96.
Cressie, N.A.C., 1993. Statistics for Spatial Data. John Wiley, New York.
Cutler, D.M., Glaeser, E.L., Vigdor, J.L., 1999. The rise and decline of the American Ghetto. J. Polit. Econ. 107 (3), 455–506.
Dahl, G.B., 2002. Mobility and the returns to education: testing a Roy model with multiple markets. Econometrica 70, 2367–2420.
De Giorgi, G., Pellizzari, M., Redaelli, S., 2010. Identification of social interactions through partially overlapping peer groups. Am. Econ. J. Appl. Econ. 2 (2), 241–275.
De la Roca, J., Puga, D., 2014. Learning by working in big cities. CEMFI.
Del Bello, C., Patacchini, E., Zenou, Y., 2014. Peer effects: social or geographical distance? Working paper.
Di Addario, S., Patacchini, E., 2008. Wages and the city. Evidence from Italy. Labour Econ. 15 (5), 1040–1061.
Diggle, P.J., 2003. Statistical Analysis of Spatial Point Patterns. Oxford University Press, New York.
Duranton, G., Overman, H.G., 2005. Testing for localisation using micro geographic data. Rev. Econ. Stud. 72, 1077–1106.
Duranton, G., Gobillon, L., Overman, H.G., 2011. Assessing the effects of local taxation using microgeographic data. Econ. J. 121, 1017–1046.
Eerola, E., Lyytikainen, T., 2012. On the role of public price information in housing markets. Government Institute for Economic Research, VATT Working papers 30/2012.
Einio, E., Overman, H.G., 2014. The effects of spatially targeted enterprise initiatives: evidence from UK LEGI. LSE.
Ellison, G., Glaeser, E.L., 1997. Geographic concentration in U.S. manufacturing industries: a dartboard approach. J. Polit. Econ. 105 (5), 889–927, University of Chicago Press.
Ellison, G., Glaeser, E.L., Kerr, W., 2010. What causes industry agglomeration? Evidence from coagglomeration patterns. Am. Econ. Rev. 100, 1195–1213.
Epple, D., Romano, R.E., 2011. Peer effects in education: a survey of the theory and evidence. In: Benhabib, J., Bisin, A., Jackson, M.O. (Eds.), Handbook of Social Economics, vol. 1B. Elsevier, Amsterdam (Chapter 20).
Felkner, J.S., Townsend, R.M., 2011. The geographic concentration of enterprise in developing countries. Q. J. Econ. 126 (4), 2005–2061.
Fryer, R., Torelli, P., 2010. An empirical analysis of 'Acting White'. J. Public Econ. 94 (5–6), 380–396.
Gaviria, A., Raphael, S., 2001. School-based peer effects and juvenile behavior. Rev. Econ. Stat. 83 (2), 257–268, MIT Press.
Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. Geogr. Anal. 24, 189–206.
Gibbons, S., 2004. The costs of urban property crime. Econ. J. 114 (498), F441–F463.

Gibbons, S., Machin, S., 2003. Valuing English primary schools. J. Urban Econ. 53 (2), 197–219.
Gibbons, S., Overman, H.G., 2012. Mostly pointless spatial econometrics. J. Reg. Sci. 52 (2), 172–191.
Gibbons, S., Silva, O., Weinhardt, F., 2013. Everybody needs good neighbours? Evidence from students' outcomes in England. Econ. J. 123 (571), 831–874.
Gibbons, S., Overman, H.G., Pelkonen, P., 2014. Area disparities in Britain: understanding the contribution of people versus place through variance decompositions. Oxf. Bull. Econ. Stat. 76 (5), 745–763.
Goldsmith-Pinkham, P., Imbens, G.W., 2013. Social networks and the identification of peer effects. J. Bus. Econ. Stat. 31, 253–264.
Goux, D., Maurin, E., 2007. Close neighbours matter: neighbourhood effects on early performance at school. Econ. J. 117 (523), 1193–1215, Royal Economic Society.
Graham, D.J., 2007. Agglomeration, productivity and transport investment. J. Transp. Econ. Policy 41 (3), 317–343.
Harhoff, D., Hiebel, M., Hoisl, K., 2013. The impact of network structure and network behavior on inventor productivity. Munich Center for Innovation and Entrepreneurship Research (MCIER). Max Planck Institute.
Heckman, J., 2005. The scientific model of causality. Sociol. Method. 35 (1), 1–97.
Heckman, J., Lalonde, R., Smith, J., 1999. The economics and econometrics of active labour market programs. In: Ashenfelter, A., Card, D. (Eds.), Handbook of Labor Economics, vol. 3A, North-Holland, Amsterdam.
Helmers, C., Patnam, M., 2014. Does the rotten child spoil his companion? Spatial peer effects among children in rural India. Quant. Econ. 5 (1), 67–121.
Herfindahl, O.C., 1959. Copper Costs and Prices: 1870–1957. The John Hopkins Press, Baltimore, MD.
Hirschman, A.O., 1964. The paternity of an index. Am. Econ. Rev. 54 (5), 761.
Holmes, T., 1998. The effect of state policies on the location of manufacturing: evidence from state borders. J. Polit. Econ. 106, 667–705.
Holmes, T.J., Lee, S., 2012. Economies of density versus natural advantage: crop choice on the back forty. Rev. Econ. Stat. 94 (1), 1–19, MIT Press.
Horrace, C.W., Liu, X., Patacchini, E., 2013. Endogenous network production function with selectivity. Syracuse University, Working paper.
Hsieh, C.S., Lee, L.F., 2013. A social interaction model with endogenous friendship formation and selectivity. Ohio State University, Working paper.
Ioannides, Y., 2013. From Neighborhoods to Nations: The Economics of Social Interactions. Princeton University Press, Amsterdam.
Ioannides, Y., Zabel, J., 2008. Interactions, neighbourhood selection and housing demand. J. Urban Econ. 63, 229–252.
Jaffe, A., 1989. Real effects of academic research. Am. Econ. Rev. 79 (5), 957–970.
Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance. J. Real Estate Financ. Econ. 17, 99–121.
Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. Int. Econ. Rev. 40, 509–533.
Kelejian, H.H., Prucha, I.R., 2004. Estimation of simultaneous systems of spatially interrelated cross sectional equations. J. Econom. 118, 27–50.
Kelejian, H., Prucha, I.R., 2007. HAC estimation in a spatial framework. J. Econom. 140, 131–154.
Kelejian, H.H., Prucha, I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. J. Econom. 157, 53–67.
Kiel, K., Zabel, J., 2008. Location, location, location: the 3L approach to house price determination. J. Hous. Econ. 17, 175–190.
Klier, T., McMillen, D.P., 2008. Evolving agglomeration in the U.S. auto supplier industry. J. Reg. Sci. 48 (1), 245–267.
Kosfeld, R., Eckey, H.-F., Lauridsen, J., 2011. Spatial point pattern analysis and industry concentration. Ann. Reg. Sci. 47, 311–328.
Krauth, B., 2005. Peer effects and selection effects on smoking among Canadian youth. Can. J. Econ. 38 (3), 414–433.

Krugman, P., 1991a. Geography and Trade. MIT Press, Cambridge, MA.
Krugman, P., 1991b. Increasing returns and economic geography. J. Polit. Econ. 99 (3), 483–499.
Kuminoff, N., Kerry Smith, V., Timmins, C., 2013. The new economics of equilibrium sorting and policy evaluation using housing markets. J. Econ. Lit. 51 (4), 1007–1062.
Lee, L.-F., 1983. Generalized econometric models with selectivity. Econometrica 51, 507–512.
Lee, L.-F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. Econometrica 72, 1899–1926.
Lee, M.-J., 2005. Micro-Econometrics for Policy, Program and Treatment Effects. Oxford University Press, Oxford.
Lee, L.-F., 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. J. Econom. 140, 333–374.
Lee, L.-F., Liu, X., 2010. Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. Econ. Theory 26, 187–230.
Lee, L.-F., Liu, X., Lin, X., 2010. Specification and estimation of social interaction models with network structures. Econom. J. 13, 145–176.
Li, J., Lee, L., 2009. Binary choice under social interactions: an empirical study with and without subjective data on expectations. J. Appl. Econ. 24, 257–281.
Lin, X., 2010. Identifying peer effects in student academic achievement by a spatial autoregressive model with group unobservables. J. Urban Econ. 28, 825–860.
Liu, X., Lee, L.-F., 2010. GMM estimation of social interaction models with centrality. J. Econom. 159, 99–115.
Liu, X., Patacchini, E., Zenou, Y., Lee, L.-F., 2012. Criminal networks: who is the key player? CEPR Discussion Paper No. 8772.
Liu, X., Patacchini, E., Rainone, E., 2013. The allocation of time in sleep: a social network model with sampled data. CEPR Discussion Paper No. 9752.
Liu, X., Patacchini, E., Zenou, Y., 2014. Endogenous peer effects: local aggregate or local average? J. Econ. Behav. Organ. 103, 39–59.
Manski, C.F., 1993. Identification of endogenous effects: the reflection problem. Rev. Econ. Stud. 60, 531–542, 84, 600–616.
Manski, C.F., 2000. Economic analysis of social interactions. J. Econ. Perspect. 14 (3), 115–136.
Manski, C.F., 2013. Identification of treatment response with social interactions. Econom. J. 16 (1), S1–S23.
Marcon, E., Puech, F., 2003. Evaluating the geographic concentration of industries using distance-based methods. J. Econ. Geogr. 4 (3), 409–428.
Massey, D.S., Denton, N.A., 1987. Trends in the residential segregation of Blacks, Hispanics, and Asians: 1970–1980. Am. Sociol. Rev. 94, 802–825.
Mayer, T., Mayneris, F., Py, L., 2012. The impact of urban enterprise zones on establishments location decisions: evidence from French ZFUs. PSE.
Mele, A., 2013. Approximate variational inference for a model of social interactions. Working papers 13–16, NET Institute.
Melo, P.C., Graham, D.J., Noland, R.B., 2009. A meta-analysis of estimates of urban agglomeration economies. Reg. Sci. Urban Econ. 39, 332–342.
Mion, G., Naticchioni, P., 2009. The spatial sorting and matching of skills and firms. Can. J. Econ. 42, 28–55 [Revue canadienne d'économique].
Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37 (1), 17–23.
Moretti, E., 2004. Human capital externalities in cities. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics. Cities and Geography, vol. 4. Elsevier, Amsterdam.
Nakajima, R., 2007. Measuring peer effects on youth smoking behaviour. Rev. Econ. Stud. 74, 897–935.
Openshaw, S., 1983. The Modifiable Areal Unit Problem. Geo Books, Norwich.
Patacchini, E., Rainone, E., 2014. The word on banking—social ties, trust, and the adoption of financial products, EIEF Discussion Paper No. 1404.
Patacchini, E., Venanzoni, G., 2014. Peer effects in the demand for housing quality. J. Urban Econ. 83, 6–17.
Patacchini, E., Zenou, Y., 2007. Spatial dependence in local unemployment rates. J. Econ. Geogr. 7, 169–191.

Patacchini, E., Zenou, Y., 2012. Neighborhood effects and parental involvement in the intergenerational transmission of education. J. Reg. Sci. 51 (5), 987–1013.
Ripley, B.D., 1976. The second-order analysis of stationary point processes. J. Appl. Probab. 13, 255–266.
Rubin, D.B., 1978. Bayesian inference for causal effects: the role of randomization. Ann. Stat. 6 (1), 34–58.
Sacerdote, B., 2001. Peer effects with random assignment: results for Dartmouth roommates. Q. J. Econ. 116, 681–704.
Scholl, T., Brenner, T., 2012. Detecting spatial clustering using a firm-level cluster index. Working papers on Innovation and Space 02.12: 1-29.
Scholl, T., Brenner, T., 2013. Optimizing distance-based methods for big data analysis. Philipps-Universität Marburg, Working papers on Innovation and Space.
Simons-Morton, B., Farhat, T., 2010. Recent findings on peer group influences on adolescent smoking. J. Prim. Prev. 31 (4), 191–208.
Sirakaya, S., 2006. Recidivism and social interactions. J. Am. Stat. Assoc. 101 (475), 863–875.
Soetevant, A., Kooreman, P., 2007. A discrete choice model with social interactions: with an application to high school teen behaviour. J. Appl. Econ. 22, 599–624.
Stock, J., Wright, J., Yogo, M., 2002. A survey of weak instruments and weak identification in generalized method of moments. J. Bus. Econ. Stat. 20 (4), 518–529.
Vitali, S., Mauro, N., Fagiolo, G., 2009. Spatial localization in manufacturing: a cross-country analysis. LEM Working paper Series 4, 1–37.
Weinberg, R., 2007. Social interactions with endogenous associations. NBER Working paper No. 13038.
Wong, D., 2009. The modifiable areal unit problem (MAUP). In: Fotheringham, A.S., Rogerson, P. (Eds.), The SAGE Handbook of Spatial Analysis. Sage Publications Ltd, London, pp. 105–124.
Zenou, Y., 2009. Urban Labour Markets. Cambridge University Press, Cambridge.