SPATIAL REGRESSION MODELS

Michael D. Ward
University of Washington

Kristian Skrede Gleditsch
University of Essex

1. INTRODUCTION

1.1 Interaction and Social Science

Social scientists are interested in situations in which various types of agents—individuals, political parties, groups, countries—interact with one another. In many cases, the outcomes or incentives for actions of individual actors do not depend solely on the attributes of particular individuals but on the structure of the system, their position within it, and their interactions with other individuals. Even something as prosaic as the common flu has a social component, since it is spread through social interaction. If we want to predict the likelihood that a particular individual will come down with a rhinovirus, we would look at whether something has "been going around" lately and whether the individual has been in contact with others who have become ill with this disease. Some diseases are spread via interaction, where infected individuals transmit the disease through contact with others. Clearly, different types of interaction patterns can give rise to different disease dynamics. Although demonstrably false, the spread of the HIV retrovirus to the United States is often claimed to have originated from an index case of a single Canadian airline attendant in the late 1970s (Watts, 2003).

Oddly, the role of interactions and their structures is almost completely absent from most empirical analyses in the social sciences. Consider, for example, the case of voter turnout. Differences in turnout have typically been explained using individual characteristics, such as higher education, believed to be important for political behavior. However, interaction and ties to other individuals can be as important as personal characteristics. For example, the so-called get-out-the-vote phone calls increase voting turnout by about six percentage points (±3) on average (Imai, 2005). Similarly,

linkages to organizations such as churches and labor unions are also known to increase voter turnout. Baybeck and Huckfeldt (2002) show that even in dispersed networks individuals who are distant from one another are less likely to interact on a frequent basis. Such studies are the exception, not the rule; most studies of voter turnout still assume that all voters make independent decisions.

Clearly, treating observations as unrelated would be patently absurd for the flu example. Perhaps some people have weaker immune systems and are more likely to fall sick during an epidemic. However, we would not try to predict an individual's risk of the flu from his or her own attributes alone, independent of whether other individuals are infected. For example, parents are rarely "similar" to their children in terms of income, hours slept, and smoking. Yet if one is affected the other is typically also at risk. The social relations model grew out of an interest by psychologists to separate the independent and interactive effects of groups versus individuals, and it provides one attempt to model such dependencies (see, e.g., Kenny, 1981; Malloy & Kenny, 1986).

In this book, we examine how insights from *spatial* analysis can help researchers take dependence between observations into account and deal with spatially clustered phenomena. In particular, we focus on two important regression models with spatially dependent observations. The first of these concerns situations in which there is a spatially lagged dependent variable. The second focuses on spatially correlated errors. We recognize that there is a much larger set of interesting spatial modeling perspectives. This book is not intended to survey these but rather serves to introduce models with spatially lagged dependent variables and those with spatially correlated error terms. Many empirical undertakings in social science may benefit from these approaches, which have until very recently been widely ignored in the empirical social science literature. These types of models allow us to examine the impact that one observation has on other proximate observations. We believe this is important not only from first principles but also from the simple fact that many social phenomena are spatially "clustered." There are many forms of spatially organized data, ranging from geolocated individual locations for observations to regional data that are attributed to some geographical area. The latter are often called *areal* or *lattice* data, while the former are known as *point data*. In this book, we concentrate on regional data, which typically deal with units such as counties, states, provinces, and countries.[1]

Examples of spatially clustered phenomena are widespread in the social sciences. Regional voting clusters are often thought to be important in American political behavior. Political cleavages often overlap with

economic and ethnic cleavages. As such, models of voter turnout may need to take into account the spatial clustering of overlapping cleavages (West, 2005). Similar examples can be found from studies in comparative politics, sociology, and economics. For example, studies of the impact of the various policy choices made by central banks have been examined for their independence from the central governments as well as the preferences of the central bankers themselves. It is widely thought that central banks are constrained by a variety of local contexts apart from how independent they are from the national authorities. Thus, even if they are independent of local authorities, are central bank policies independent of each other (Adolph, 2004; Franzese, 1999)? Murdoch, Sandler, and Sargent (1997) examine interdependent decision making in the voluntary and nonvoluntary aspects of behavior regarding emissions of sulfur dioxide and nitrous oxide in Europe during the 1980s. As pollutants are spatially dispersed without regard to national boundaries, spatial analysis techniques will help in highlighting the spillover effects of pollution as well as the interdependence of compliance issues. Inequality and poverty are thought to be intertwined in cross-national studies. The most skewed wealth and income distributions are often in the poorest countries. Recent work has shown that corruption is often the consequence, as well as plausible cause, of poverty. However, it turns out that income inequality may increase the level of corruption, even more so than poverty. It may be that the distributions of wealth and corruption share a spatial clustering that complicates this effect. Spatial analysis can help untangle this conundrum. Recent work along these lines includes You and Khagram (2005). Finally, organizational forms may also spread in much the same way—policy emulation. Holmes (2006) addresses the contagion of unionism with spatial models.

In short, there are myriad studies across the gamut of the social sciences that employ data that are actually organized on a spatial template, whether the units are counties, cities, states, countries, or firms. It often turns out that the characteristics of these units are highly clustered in particular spatial regions. In many of these applications, it is plausible to assume that there may be dependencies across the observations. In practice, this clustering is generally ignored or treated as a nuisance. Ignoring these dependencies imposes a substantial price on our ability to generate meaningful inferences about the processes we study. Spatial analysis provides one way of reducing that price and taking advantage of the information we have about how social processes are interconnected. We next turn to a simple example of how this works in an important area of social science—namely, the study of the diffusion of democratic institutions.