MODELLING REGIONAL ECONOMIC CHANGE WITH THE AID OF GEOGRAPHIC INFORMATION SYSTEMS: AUTHENTIC REGIONAL SCIENCE?
(Presidential Address, April 12, 1991)

Mark S. Henry*

I. INTRODUCTION

A memory I have from one of the first Southern Regional Science Association meetings I attended is a conversation with an economist employed by one of the national labs. We were talking about a presentation made by Walter Isard. He had discussed a model of the region as a complex set of social, economic, political, and environmental forces at work. I was impressed. How could anyone model all these forces in a meaningful way and suggest that there were ways to make consistent estimates of key parameters in all these models? My colleague was less impressed. He said that regional science was still nothing but analysts from separate disciplines looking at the same problems with different conceptual frameworks. At best, you might have a team of analysts with different disciplinary backgrounds working together on a problem. But he disavowed any faith in the notion of regional science with "space or the region" as the focal point for developing problem solving tools.

I would guess that the multidisciplinary view of regional science is held by most attendees at this meeting of the SRSA. Yes, often it is a region or set of regions that provide the data base or funding opportunity to do research within the confines of a traditional discipline. However, the idea that regional science is different from traditional disciplines is not widely held. But regional science is different from economics, for example. Again, Isard tells us that in contrast to "the economist, who at most gives peripheral attention to space," there is . . . "the regional scientist " who "asks how the spatial setting and dimensions significantly affect the behavior of enterprises, consumers and institutions." (Isard 1975, 3)

My purpose today is to argue that recent developments in geographical informations systems (GIS) open renewed opportunities for "authentic" regional science. While GIS is not new, less expensive computer workstations, improved software for GIS, and spatial data like the census TIGER files have combined

*Professor, Department of Agricultural and Applied Economics, Clemson University. Thanks are due to my Clemson colleagues, Kerry Brooks, Jim London, Loretta Singletary and Martin Roche for collaboration on the GIS project discussed in my remarks. Funding for the project was from the Appalachian Regional Commission and Clemson University. I am thankful for their support.
over the past several years to expose the GIS tool to a wider audience of practitioners and researchers.

A GIS may be described as a combination of hardware and software tools to maintain, retrieve, and manipulate both spatial and nonspatial data about places. The chief advantage of a GIS is that it allows interactive access, manipulation or transformation of data for the purpose of replicating a real world model. It is possible to use the system for the construction of hypothetical scenarios that might result from planning decisions, for instance. In this context then, a GIS serves as an experimental laboratory for analyzing the results of planning policies (Burroughs, 1986).

I suggest that GIS will help regional scientists address several important problems. First, better definitions and delineation of regions are possible. Second, better measures of spatial relationships are possible. Third, improved procedures for developing estimators in spatial econometrics are possible. Each of these issues in regional science has a focus on space or the region as the critical element in problem definition and problem solving.

Before extolling the virtues of the GIS for doing regional science, the economist in me says that the rather substantial costs of using GIS—from hardware and software to the immense data requirements—are important constraints to its use. And in many cases the value of more accurate spatial measures may have little benefit to the study of a problem at hand. For example, a student at Clemson used the GIS system in a study of land values in a coastal area of South Carolina. GIS was used to compute distance to the beach from lots, distances to an Air Force base, and distances to selected infrastructure features thought to be important in determining lot prices. The objective of the hedonic land value models developed was to estimate the contribution of lot characteristics to lot price (Veillet 1991). The GIS system was not needed for this small case study. Here, manual estimation of the required distances would suffice, though it would have been tedious work.

However, in very applied work the usefulness of GIS for planners seems obvious—though cost considerations tend to be underestimated and can result in the abandonment of systems that consume too many scarce staff and operating resources. The need to train or hire staff, as well as the costs of data entry and maintenance, can be substantial and can limit GIS use to the best-endowed governmental units.

On the other hand, one of the most useful aspects of GIS for local planners is the capability of GIS to simulate times and distances over space under alternative networks of roads, traffic conditions, and an array of potential impediments—speed limits, intersection control, bridge crossings, etc. Obvious applications of these procedures exist in the areas of routing for school buses and emergency medical
services. Other local uses that are popular include updating tax maps and identifying solid waste disposal sites.

However, I want to focus my remarks on the potential of GIS to contribute to a better understanding of regional phenomena from a broader research perspective. Viewing GIS as an addition to our kit of research tools for applied regional science work means that it will be difficult to estimate the benefits of the tool. Still, there are benefits. GIS can identify regions of interest, improve measurement of spatial relationships, and help resolve some problems in spatial econometrics.

II. GEOGRAPHIC INFORMATION SYSTEMS AND REGIONAL DELINEATION

Because GIS is able to create maps (and related spatial data sets) based on physical, political and/or socioeconomic criteria, it obviously can be used to define regions based on the criteria deemed important to the research problem. For example, GIS efforts in South Carolina have created a statewide system for key infrastructure, social, economic, political, and land use coverages (Henry et al., 1990). Each coverage is a layer of information that may be overlaid with any other layer(s) for computer mapping. One could with little effort display trade areas around cities that conform to Reilly's law of retail gravitation. Loschian market areas and central place systems can be visualized in theory and compared with actual distributions. Functional economic areas can be identified in a more spatially sensitive manner than as county level groupings.

In South Carolina, we used GIS to create regions as overlays of county census divisions and zip code regions and to pick out regions that were touched by five kilometer buffers around four-lane highway projects. And, manufacturing plants were located in another layer of information.

We plotted distance decay functions for the distribution of firm locations from four-lane highways by year that the highway was completed and by year the firms were established (see Figure 1). These firm and highway distributions over space gave us guidelines on how to define our primary impact regions. In turn, this helped us to use both census and microdata that were geocoded to evaluate regional economic impacts of infrastructure investment.

But the importance of this computer mapping for regional scientists goes beyond the benefits of visualization of relations between social, economic, political, and environmental phenomena. A second benefit is the ability to provide continuous measures of spatial relationships.
FIGURE 1
Count of Manufacturing Firms Located Near
Four-Lane Highways Completed Between 1971-1980

* Firms established by 1960
☐ Firms established by 1970
# Firms established by 1980
△ Firms established by 1990

Distance from Highway in Kilometers
III. GIS AND THE MEASUREMENT OF SPATIAL ECONOMIC RELATIONSHIPS

The GIS allows the user to think about variables in spatial terms—in essentially three dimensions. First, we can think in terms of location in some larger system, ranging from the actual world coordinate system defining points, lines or areas, to location within some spatial hierarchy (i.e., an economic region). Second, we can think in terms of distances: one entity’s distance from another. We can describe this distance in continuous terms (e.g., meters) or in varyingly fuzzy categories: "next to," "above," "nearby." With the GIS system, ARCINFO, and its network module, as well as the proper data set, we can convert distances to time as we think of commuter flows over highways, subject to various constraints. Finally, we can add a third dimension (in addition to the actual physical topographic dimension)—we can think of an "overlay" coincidence of spatial variables. The power of a GIS is the ability to work within these multiple dimensions, creating new variables from various combinations (and, or, not, greater than, less than, etc.) in both horizontal—across one layer—and vertical dimensions—between layers of spatial data.

For example, consider the problem of slow growth in rural America. It used to be that the notions of rural and urban were defined by size of the place or by the characteristics of a county that would deem it to be a metropolitan or nonmetropolitan county. And most rural/urban relationships in space were confined to county-level analyses. Concern was expressed in the 1980s that nonmetro counties were lagging behind metro counties (see Drabenstott and Gibson 1988, or Garnick 1989, for example). Yet, there may have been areas in nonmetro counties that were outpacing the average metro county. A continuous measure of rurality and how this measure changes over time could overcome some of these problems in generalizing about how "rural" places are doing.

In ARCINFO, a network could be used to develop a continuous measure of rurality. One can compute the time that it would take to go from any point in a nonmetro county or metro county to the nearest large urban center. And these computations could be repeated for different years to reflect changes in investment in new four-lane expressways, speed limits, and the expansion of urban centers into the old hinterlands. Continuous measures of rurality over space and time could be central to testing hypotheses about the spatial distribution of firm locations between rural and urban regions.

For example, GIS could be useful in examining the effects of the growth of rural places as a consequence of a trend in industrial organization toward vertical disintegration. Immediately, one would attempt to identify industries that have undergone vertical disintegration and to document their location patterns over
time. But the temptation would be to use county-level data on metro and non-metro counties. Ideally, the data on firms in the industry would reflect location in places with strong urban center influences or in isolated rural areas—say, areas outside reasonable commuting times to the nearest urban center. GIS could play a role in identifying where these isolated regions are (and how they have changed over time) and to identify firm location patterns in rural and urban places.

GIS, of course, has to have a spatial data base to manipulate. And this kind of spatial data base for firms is expensive (Dun and Bradstreet, for example) or nonexistent. Still, some agencies like the Small Business Administration and the U.S. Department of Labor can be expected to improve the spatial context of their data bases over time. More use of tools like GIS that make use of point data should provide added incentive to agencies to improve the spatial content of the data they make available.

Another example of using GIS to measure spatial relationships is the CCD/ZIP regions we created (Henry et al., 1990) that were touched by a five-kilometer buffer around four-lane highway projects. Using quasi-experimental techniques (Isserman et al., 1982 and 1989), we compared growth in regions that were "treated" with four-lane highways vs. control regions—regions that had similar initial conditions. (See Figure 2 for an example of the control and treatment regions.) Significant post-treatment era employment growth rates differentials favoring treated over control regions suggested that the highway investments do matter.

In summary, we have found GIS to be very helpful in the creation of spatial variables that are used to test developmental policy hypotheses and to estimate parameters for more formal econometric models. And the ability to simulate time and distance responses to changes in transportation networks is adding a degree of data reality that should improve our understanding of the importance of the friction of distance for decisions made by businesses, households, and governmental units.

IV. GIS AND SPATIAL ECONOMETRIC PROBLEMS

The use of GIS to define regions and cross-sectional observations may aid in the development of regional observations that mitigate the problems of spatial dependence and heterogeneity that Cliff and Ord (1973) and Anselin and Griffith (1988) discuss. Despite the recognition of these problems by specialists in spatial statistics and econometrics, most applied research using spatial observations either ignores the problems or relegates them to the "back-burner." The increased availability of GIS systems to applied researchers can help make tests and adjust-
FIGURE 2
Control and Treatment Regions

Never Treated
1960s Only
1970s Only
1980s Only
1960s & 1970s
1960s & 1980s
1970s & 1980s
1960s, 1970s, 1980s

SCIP Project:
Cluster Group 1: Medium-High
ments for these problems more likely. In fact, if we are to accept the advice to practitioners from Anselin and Griffith, we should observe the following rules whenever we analyze spatial data sets:

1. "Do not take diagnostic and test statistics at their face value."

2. "The first step in analyzing a spatial data series should be to assess the sources, nature and degree of prevailing spatial effects." If effects seem small, then output from standard packages like SAS or SPSS or SHAZAM are a reasonable first approximation. If not, more sophisticated modelling of spatial effect is needed (Anselin and Griffith 1988, 29).

Consider the two major spatial econometric problems identified by Anselin and Griffith—spatial dependence and spatial heterogeneity. Looking first at spatial dependence, Anselin and Griffith note the following:

In many instances where regression analysis is applied to cross-sectional data, various measurement problems, spatial externalities, and other spillover effects may lead to dependent error terms. In addition, in the empirical implementation of models of spatial processes, spatial structure, or spatial interaction, the dependent variable at one point in space may be functionally related to its values at some or all other locations in the system (Anselin and Griffith 1988, 15).

Spatial dependence in the error term with the OLS estimators is shown by Anselin and Griffith to cause misleading tests of significance and measures of model fit. So much for the standard t and F tests we all know so well.

Looking next at the problem of spatial heterogeneity, Anselin and Griffith note that:

Many phenomena studied in regional science could also lead to structural instability over space, in the form of systematically varying parameters or different response functions. In addition, . . . the measurement errors that result from the use of ad hoc spatial units of analysis are likely to be nonhomogeneous and can be expected to vary with location, area, or other characteristics of the spatial unit (Anselin and Griffith 1988, 16).
Major problems include structural instability (functional form or parameter values vary over space) and heteroscedasticity (resulting from missing variables or other forms of model misspecification). Unfortunately, parameter estimates may be biased (except when one has heteroscedastic errors only), significance tests misleading, and poor prediction performance likely with spatial heterogeneity (Anselin and Griffith 1988, 17).

Can GIS help? GIS might ameliorate some of these statistical problems in several ways. First, new regions constructed through spatial grouping in GIS might be designed to avoid serious violation of error term assumptions if they are revealed in the tests that Anselin and Griffith suggest.

Second, the question of the trade-offs between degrees of freedom and spatial autocorrelation can be approached in new and productive ways. Why not use county data rather than state level data; or use County Census Division (CCD) data rather than county data; or use CCD/ZIP overlay regions rather than CCD data? Each reduction in region size increases degrees of freedom, assuming data are available for each size region. The limitation is in the increased probabilities of spatial dependence that limit the usefulness of statistical tests, despite the increase in degrees of freedom.

A key issue in developing estimates of the spatial autocorrelation parameter is the construction of a contiguity matrix, W. Most often this matrix is square and consists of ones (for regions that are contiguous) and zeros (for regions that have no common borders). In turn, W may be normalized so that row sums are equal to one and autocorrelation parameter estimates such as those in Equation (2) are made.

Equation (1) is the standard linear regression model. In Equation (2) the autocorrelation parameter is \( \rho \) which can be estimated as shown by Benirschke (1990, 8).

\[
\begin{align*}
Y &= Xb + u \\
(1-\rho W)Y &= (1-\rho W)Xb + e \quad \text{with} \quad e = (1-\rho W)u \sim N(0, \sigma^2 I)
\end{align*}
\]

It is the W matrix of contiguity that seems most amenable to manipulation by GIS techniques. In GIS, W matrices are easily constructed for alternative regional specifications.

Use of ARCINFO, for example, would greatly simplify the construction of the required contiguity matrix for almost any reasonable number of regions. In the South Carolina case, we constructed some 477 CCD/ZIP overlay regions—a tenfold increase from the 46 South Carolina counties and nearly double the number of CCD regions. The W matrix is easy to form in ARCINFO, and the ability
to carry out sensitivity analysis of alternative forms of the W matrix on the autocorrelation parameter is greatly enhanced in GIS. We might use the length of common boundaries or the time to move from the center of each region to its contiguous regions as alternatives to the 0/1 common boundary approach. In short, the GIS provides a very effective tool for managing large spatial data bases. As such, it provides the applied researcher with an improved capability to evaluate the effects of space on econometric estimators.

V. CONCLUSIONS

After working with a large GIS data base for about a year, I have found it to be an excellent resource for creating regions and measuring spatial relationships. It is probably the closest thing to an experimental laboratory that most regional scientists can access. Both its potential and cost are large for the research community. It is certainly too early to tell if substantial investments in GIS are warranted by research results in the regional science area. GIS will not produce behavioral models, but can aid in the testing of hypotheses from models and improve their empirical content.

In addition to improvements in data quality and parameter estimates, the ability to visualize large batches of spatial and temporal data using GIS helps to convey the results of research to policymakers. Much of the work done by members of the SRSA is oriented toward state and local policy analysis. In this regard, GIS may be of particular interest to many in this audience.

Finally, let me reiterate that regional science takes space and the region as its central concern in the explanation of business, household, and institutional behavior. GIS can help regional scientists keep good accounts of spatial change, to be creative in the definition of the region for the problem at hand, and to believe the parameters and tests in spatial econometric estimation. As regional scientists, we should welcome these GIS contributions to the development of the data bases and analytical tools needed to understand how space and the region matter to decisionmakers.

REFERENCES


