JOURNAL OF HOUSING ECONOMICS 6, 368–386 (1997)
ARTICLE NO. HE970216

REVIEW ESSAY

How GIS Can Put Urban Economic Analysis on the Map*

John M. Clapp†

School of Business Administration, University of Connecticut, 368 Fairfield Road, U-41RE, Storrs, Connecticut 06269-2041

Mauricio Rodriguez‡

Texas Christian University, P.O. Box 298530, Fort Worth, Texas 76129

and

Grant Thrall§

Department of Geography, University of Florida, Gainesville, Florida 32611

Received May 9, 1997

Geographic information systems (GIS) provide powerful tools for storing and manipulating large amounts of information on spatial relationships. However, applications of GIS in urban economics and real estate are infrequent. As a symptom of this, there is often little relation between the extensive literature on residential mortgages and the housing markets literature. The purpose of this paper is to discuss the potential contributions GIS could make to urban economic and real estate economic research and open debate on GIS as applied to these fields. Beginning with theory, we determine how GIS can be useful to researchers and what is needed from future GIS. A GIS is a database management system with a spatial reference. It is capable of providing researchers with excellent control over spatial relationships. For example, one can "geocode" the housing transactions used for hedonic pricing studies. The process of geocoding assigns a latitude and longitude coordinate for each transaction. This enables much more elaborate spatial analysis than with traditional techniques such as using a ruler on a paper map. For instance, GIS allows the hedonic pricing literature to be reworked using spatial autoregressive

* The authors gratefully acknowledge comments on earlier drafts of this paper by Ayse Can, Robert Cromley, Pat McGlamery, and Susan Elshaw Thrall.

† E-mail: johnc@sbaserv.sba.uconn.edu.
‡ E-mail: mrodriguez@zeta.is.tcu.edu.
§ E-mail: thrall@geog.ufl.edu or thrall@afn.org. Website: http://www.afn.org/~thrall/.

1051-1377/97 $25.00
Copyright © 1997 by Academic Press
All rights of reproduction in any form reserved.

analysis (SAR). SAR models need to repeatedly calculate large numbers of distances from each observation to every other observation within a given radius. In this context, GIS software is primarily useful for assigning a latitude and longitude location to each observation based on the best of street, ZIP, or ZIP+4. These geocoded observations may then be exported to other software for batch processing of SAR statistics. Future development of GIS software should eliminate the need for multiple software packages. Likewise, road distances with flexible weights (e.g., depending on type of road or on construction delays) awaits further software development. © 1997 Academic Press

1. INTRODUCTION

Urban economics differs from neoclassical microeconomics in that the former discipline rigorously considers spatial economic relations.[1] Geographic information systems (GIS) provide powerful tools for storing and manipulating large amounts of information on spatial relationships.[2] However, applications of GIS in urban economics and real estate are infrequent.[3] As a symptom of this, there is often little relation between the extensive literature on residential mortgages and the housing markets literature.[4]

An emormous amount of data, including most Census data, is currently being made available to GIS software. Greater data availability together with improvements in GIS software can provide better integration of market analysis with financial analysis. For example, supply and demand in the local housing market determines net operating income to an apartment project; the value of the mortgage default option on that project will clearly depend on changes in local market conditions.

A GIS is a database management system with a spatial reference. It is capable of providing researchers with excellent control over spatial relationships. For example, one can "geocode" the housing transactions used for hedonic pricing studies. The process of geocoding assigns a latitude and longitude coordinate for each transaction. This allows for much more elaborate spatial analysis than with traditional techniques such as using a ruler on a paper map.

[1] Turnbull (1995) explains the link between basic microeconomic theory of the consumer and urban economics.

[2] For an introduction to GIS, see Huxhold (1991) and Star and Estes (1990). Thrall and Marks (1993) and Marks and Thrall (1994) provide a discussion of GIS in real estate research.

[3] Explanations include: (1) GIS software is not user friendly, (2) it is difficult to learn any major new technology, and (3) GIS software is not specifically designed for many issues addressed by urban economics and real estate. Most software is generic, not adapted for vertical applications.

[4] Greater availability of local house price indices is beginning to change this. For example, a paper by Quigley *et al.* (1994) uses house price indices to estimate current loan-to-value ratio as a determinant of default and prepayment options.

The purpose of this paper is to discuss the potential contributions GIS could make to urban economic and real estate economic research and open debate on GIS as applied to these fields. This paper adopts the viewpoint of a marine biologist rather than of a fisherman. Therefore, we will not provide a comprehensive catalog of the strengths and weaknesses of various GIS software packages. Instead, we will compare the needs of academic researchers with the capabilities of a range of typical GIS software; our treatment of software will be oriented toward researchers rather than professionals. The purpose of this is twofold:

1. to alert researchers to the pros and cons of using GIS; and
2. to open up debate on what academic researchers need from future software development.[5]

We will use theory and statistical analysis from these fields to evaluate the possible role of GIS. Beginning with theory, we determine how GIS can be useful to researchers and what is needed from future GIS. Thus, urban economic GIS applications can complement GIS development.

The next section discusses the potential significance of GIS in urban economic and real estate economic research. Section 3 evaluates the strengths and weaknesses of GIS from the perspective of academic research. Section 4 contains some concluding remarks.

2. THE SIGNIFICANCE OF GIS IN URBAN ECONOMIC AND REAL ESTATE RESEARCH

A fundamental purpose of urban economics and real estate is to analyze spatial relationships. In urban economics, this typically takes the form of costs to transport consumers, employees, goods, or services between different points in space. Real estate often seeks to quantify the relationships between supply and demand for a particular property type in a given geographic area.

Empirical studies where GIS might be applied can be divided into two groups. The first group consists of studies dealing with explicit distance to proxy for transportation costs. For example, Do *et al.* (1994) and Thorson (1994) use straight-line distance to the CBD and other significant locations.

The second group of studies make use of spatial area characteristics. For example, Sivitanidou (1996) uses location in school districts, prime reporting areas, and zoning districts to associate neighborhood characteristics with office rents. Case and Mayer (1996) regress house price appreciation on census variables, school spending, crime rates, and the like. This is done

[5] The academic market is a small but influential one. Academic research can suggest new techniques to software developers.

for 168 towns; GIS would allow assignment of house transactions to tracts or precincts. Dubin and Sung (1990) assign houses to census blocks, school districts, and prime reporting areas.

The quantification of spatial relationships by these two types of studies is relatively simple. Is GIS really necessary in order to implement these sorts of studies? The next two subsections attempt to answer this question.

2.1. *The Role of GIS with Studies Employing Explicit Distance*

The simple role of distance in this type of study is an artifact of a technology consisting of paper maps and rulers. Spatial economic theory indicates that *transportation costs* are the most important factors determining accessibility. A GIS can approximate transportation costs in a superior way by calculating distances along roads (see Rodriquez *et al.*, 1995), by weighting interstate roads less than secondary roads, and by adding delay factors for construction sites, bridges, or other obstacles. However, much of the current GIS software is limited as described in Section 3.

More importantly, a GIS opens up new areas of statistical analysis. Given that accessibility is a crucial factor in theory, it is not surprising that studies of spatial autocorrelation typically suggest omitted spatial variables. Recent tests of spatial autoregressive processes indicate that substantial statistical gains can be obtained by explicitly modeling these processes (Pace and Gilley, 1997; Can and Megbolugbe, 1995; Dubin, 1988; Ripley, 1981). Thus, as with any new information technology, GIS opens up new and better ways of doing things that were done in a rather slow and clumsy way before. As with the telephone and the computer, the new technology also allows new areas to be explored and developed.[6]

Pace and Gilley (1997) develop a simple but flexible model for weighting neighboring geographical areas when estimating a standard hedonic equation. They show that coefficients change by large amounts and the R^2 increases substantially after using their spatial autoregressive model. Can and Megbolugbe (1995) show that prediction errors are dramatically reduced by including spatial autoregression (e.g., average errors within 3 km) in the hedonic regression.

Their results are consistent with those of Dubin (1988) who argues that:

> The model excluding neighborhood variables illustrates the consequences of ignoring spatial autocorrelation: the OLS estimates are inefficient, the OLS estimates of the standard errors are biased and the predicted values may be inaccurate. (p. 474)

[6] GIS facilitate the creation of variables based on spatial relationships. For example, a traditional database may contain town names, but it may be a difficult task to create a variable based on all observations adjacent to each town. GIS technology can be used to create variables that meet particular spatial and non-spatial criteria.

Goodman and Thibodeau (1995) reach similar conclusions in the context of heteroskedastic hedonic regressions. Pace and Gilley (1997) use spatial autoregressive methods; "Spatial techniques can improve both estimated coefficients relative to our priors and greatly increase the goodness-of-fit. For example, the AGE variable becomes negative and significant and the median absolute estimation error drops by 39.4% (pp. 1–2)."

These results are plausible and consistent with conclusions from the time series literature. They suggest that all hedonic models should test for, and probably correct for, spatial autocorrelation; most of the existing hedonic literature should be reworked. GIS software removes the most important impediment to this advance in econometric technique by allowing measurement of relationships among spatial units. However, the ease of these computations is subject to specific limitations of GIS software (see Section 3.2).

2.2. *Using GIS to Analyze Spatial Area Characteristics*

Another class of studies assigns individual economic units—such as business establishments or housing transactions (Waddell *et al.* 1993; Rosenthal and Helsley, 1994)—to bounded spatial areas (such as neighborhoods, police precincts, towns, or counties). Neighborhood characteristics are assumed to be equally available to all economic units within its boundary, and available *only* to these units. This may be true of local public goods such as schools, where a short bus ride is the only factor that differentiates access to the school for households within the boundaries of the school district. But, assigning the same characteristic to all economic units within a geographic area may be a poor approximation for jobs or population where units near the center have substantially better access; those near the boundary may be strongly related to the adjacent geographic areas.

The typical approach essentially aggregates all economic units within the spatial area and assumes that all activity takes place at the center of the area. Thus, zero distance is assumed to separate all units within the geographic area.

GIS allows the more general assumption that jobs, population, and the like are spread out within the geographic area.[7] For example, with GIS we can assume that economic units are uniformly distributed within the area, or that they are uniformly distributed along roads in the geographic area. This is particularly important because a lot of data available for urban economic and real estate analysis are for spatially aggregated areas provided

[7] The geography literature provides guidance on how to aggregate address-level data into larger spatial units (Arbia, 1989). Likewise, census blocks, block groups, or tracts can be aggregated into regions that are homogenous with respect to important characteristics (see Can and Megbolugbe, 1996).

by the Bureau of the Census, the Department of Labor, or other data collection agencies.

The remainder of this paper will discuss strengths and weaknesses of GIS software from the point of view of real estate and urban economic researchers. The section on weaknesses is addressed primarily to the developers of GIS software and databases. It will also serve to alert researchers to the pitfalls of using current GIS software and databases.

3. THE RESEARCHER'S PERSPECTIVE ON GIS SOFTWARE AND DATABASES

A powerful advantage of a GIS system is the ability to geocode data; i.e., data can be assigned a latitude and longitude coordinate based on street address, zip code, or other information.[8] Second, GIS can assist in calculating great arc distances between any two points in space. Finally, GIS allows visual inspection of spatial linkages on a map and attractive presentation of results.

Each of these advantages is offset by limitations of the software and data. Table I summarizes the availability of specific GIS functions for a range of GIS programs. The programs were chosen to represent a broad mid-range of GIS software that would be particularly appropriate for academic researchers. Thus, we avoided the high-end "industrial strength" GIS systems that require the investment of substantial amounts of human capital. At the low end, we did include a thematic mapping program, Surfer, because we judged it particularly useful to researchers. The next subsections explore specific advantages and limitations of selected GIS software.

3.1. *Advantages of GIS Software and Data*

Most GIS software is adept at geocoding large quantities of data (see Point Georeferencing in Table I). This is usually done by matching street addresses to the street map in the GIS. Since street addresses are often subject to ambiguities, it may be expedient to use the zip code to assign a latitude and longitude for each address; the center of the zip code area will be the assigned location for each address within the zip code. In fact, GIS can use the best of street, ZIP+4, or ZIP. Normally, if you have a ZIP, you can geocode more accurately at the same effort and expense for ZIP+4. ZIP+4 should get you located on your street block. However, ZIP+4 is not the center of a polygon as is ZIP. ZIP+4 is a spatial average of the

[8] Projection systems other than the standard latitude and longitude are also available in most GIS software.

TABLE I
GIS Software Programs and Their Functions

GIS functions	ArcView	MapInfo professional	Maptitude	Surfer
Program target market				
General purpose GIS	Yes	Yes	Yes	No
Thematic mapping	Yes	Yes	Yes	No
Terrain mapping	Yes	No	No (available with GIS+)	Yes
Distance calculations				
Single mode	Yes	Yes	Yes	No
Batch mode	Yes	No	Yes (accessed via their Edit-Fill command, can fill a field with distances between a feature and the closest map feature in another layer)	No
Measurement types				
Road distance	Yes (available for purchase from third party vendors)	No (available for purchase from third party vendors)	Yes	No
"As the crow flies" distance	Yes	Yes	Yes	No
Manhattan distance	No	No	No	No
Database import/export				
Point				
dBase	Yes	Yes	Yes	No
Lotus or Excel	Yes	Yes	Yes	Yes
ASCII	Yes	Yes	Yes	Yes
ARC/INFO	Yes	Yes	Yes	No
ODBC	No	No	Yes (allows import of Lotus, Excel, . . .)	No
Polygon and vector				
TIGER/Line	No	No	Yes (import only)	No
MapInfo	Yes	Yes	Yes	No
ARC/INFO	Yes	Yes	Yes	No
ARCView	Yes	Yes	Yes	No
Atlas GIS	Yes (AGF to Shapefile converter available free from ESRI www site)	No	Yes	Yes
Relational database management				
Link files	Yes	Yes	Yes	No
Link attributes to object	Yes	Yes	Yes	No
SQL	Yes	Yes	Yes	No
Conditional select	Yes	Yes	Yes	No

Layering and aggregation				
Polygon to polygon	Yes	Yes (accessed via thematic mapping feature)	Yes	No
Polygon to point	Yes	No	Yes	No
Polygon operations				
Larger polygon to smaller	Yes	Yes (accessed via thematic mapping feature)	Yes	No
Smaller polygon to larger	Yes	No	Yes	No
Point and graphics				
Creation				
Includes a layout form	Yes	Yes	Yes	No
Ease of producing good thematic maps	Excellent	Excellent	Excellent	Excellent
Exports screen image	Yes	Yes	Yes	Yes
Includes true type fonts for symbolization	Yes	Yes	Yes	Yes
Type				
Graduated color	Yes	Yes	Yes	No
Graduated symbol	Yes	Yes	Yes	No
Unique value	Yes	Yes	Yes	No
Dot density	Yes	Yes	Yes	No
Contour lines and 2D surface	Yes (with the 3D analyst extension)	No	No (available with GIS+)	Yes
3D surface	Yes (with the 3D analyst extension)	No	No	No
Scaled symbol	Yes	No	Yes	No
Pie and bar for each map object	Yes	No	Yes	No
Automatic data normalization	Yes	Yes	Yes	No
Built-in statistical functionality				
Moments about the mean	Yes	Yes	No	No
Bar and pie charts	Yes	No	Yes	No
Regression analysis	No	No	No	No
Morans I	No	No	No (included with GIS+)	No
Univariate statistics for selected records	Yes	Yes	Yes	No
Object creation: On screen				
Object reshape or move	Yes	Yes	Yes	No
Polygon/line join	Yes	Yes	Yes (part of topological editing of map areas)	No

continued

TABLE I—*Continued*

GIS functions	ArcView	MapInfo professional	Maptitude	Surfer
Polygon/line split	Yes	Yes	Yes (part of topological editing of map areas)	No
Adding and deleting objects				
Polygons	Yes	Yes	Yes	No
Points	Yes	Yes	Yes	Yes
Lines	Yes	Yes	Yes	No
External digitizer support	Yes	Yes	Yes	Yes
Feature snapping support	Yes	Yes	Yes	No
Auto completion of polygons	Yes	No	Yes	No
Multi-part feature support	Yes	No	No	No
Point georeferencing				
Street level address matching	Yes (requires purchase of data)	Yes (requires additional purchase of geographic data)	Yes (includes as standard the required geographical data and software)	No
Five digit ZIP code matching	Yes (requires purchase of data)	Yes (requires additional purchase of geographic data)	No (user must supply the required geographic data from third party vendor)	No
ZIP+4 matching	No (available as an add-on)	No (but available as an add-on from many vendors)	No (user must supply the required geographic data from third party vendor)	No
Conversion between different projections	Yes	Yes	Yes	No
Use of externally derived points	Yes	Yes	Yes	Yes
Query				
Polygon, points, vectors	Yes	Yes	Yes	No
Spatial relationships query	Yes	Yes	Yes	No
Buffering				
Around polygon, points, vectors	Yes	Yes	Yes (can do multiple buffers at same time)	No
Raster base map				
Display	Yes	Yes	Yes	No
Registration	No	Yes	Yes	No
Rubber sheeting	No	Yes	No	No
Automated capability				
Macro creation	Yes	No (available with purchase of MapBasic from MapInfo Corp.)	No (available with purchase of GISDK from Caliper Corp.)	No

Customizable interface	Yes	No (available with purchase of MapBasic from MapInfo Corp.)	No (available with purchase of GISDK from Caliper Corp.)	No
Third party applications	Yes	Yes	Yes	No
Labeling				
Automatic labeling	Yes	Yes	Yes	Yes
Automatic labeling—multiple layers	Yes	Yes	Yes	No
Labeling conflict resolution	Yes	Yes	Yes	Yes
Balloon text boxes	Yes (requires use of free software downloaded from ESRI's www site)	No	No	No
Network (routing) analysis				
Point to point routing	Yes (requires purchase of ESRI's Network Analyst)	No (available through purchase of third party application)	Yes	No
Best route between *N* number of points	Yes (requires purchase of ESRI's Network Analyst)	No (available through purchase of third party application)	Yes	No
Drive time/distance calculation	Yes (requires purchase of ESRI's Network Analyst)	No (available through purchase of third party application)	Yes (each link can have its own average speed)	No
Closest facility routing	Yes (requires purchase of ESRI's Network Analyst)	No (available through purchase of third party application)	No (available with Transcad from Caliper Corp.)	No
Data				
Geographic data				
Street	No (available with StreetMap Extension or from third party vendors)	No (available for purchase from MapInfo or third party vendor)	Yes (national street data on one CD)	No
Major roads	Yes	No (available for purchase from MapInfo or third party vendor)	Yes (included with street data CD)	No
Highways	Yes	No (available for purchase from MapInfo or third party vendor)	Yes (included with street data CD)	No
State, city, and county boundary	Yes	No (available for purchase from MapInfo for third party vendor)	Yes	No
Landmark points	Yes	No (available for purchase from MapInfo or third party vendor)	Yes	No
Census tract	Yes	No (available for purchase from MapInfo or third party vendor)	Yes	No (requires Atlas GIS files)
Census block group	No	No (available for purchase from MapInfo or third party vendor)	No (available at additional cost)	No (requires Atlas GIS files)
Census block	No	No (available for purchase from MapInfo or third party vendor)	No (available at additional cost)	No (requires Atlas GIS file)

TABLE I—*Continued*

GIS functions	ArcView	MapInfo professional	Maptitude	Surfer
Attribute data				
U.S. Census of population	Yes (a subset)	No (available for purchase from MapInfo or third party vendor)	Yes	No
U.S. Census of housing	Yes (a subset)	No (available for purchase from MapInfo or third party vendors)	Yes	No
1996 population estimates	Yes	No (available for purchase from MapInfo or third party vendors)	No (available from third party vendors)	No
World Data				
International and sub-national boundaries	Yes	No (available for purchase from MapInfo or third party vendor)	No (available for additional charge)	No
International landmark	Yes	No (available for purchase from MapInfo or third party vendor)	No (available from third party vendors)	No

Note. See the Appendix for a glossary with definitions of technical terms.

addresses within a ZIP+4 code. The postal service does not publish a ZIP+4 polygon map as it does a ZIP code map.

Geocoding may be done at any level of spatial aggregation that is useful to the researcher; for example, some data may be summarized at the town or county levels; geocoding would assign them to the geographical centers of the appropriate area.

As an example of geocoding, the authors worked with nearly 80,000 transactions of single-family residential property in Fairfax County, Virginia. For each property, they wanted to find the straight-line distance to significant points such as Washington, DC, and Dulles Airport. The GIS software enabled them to use street address to assign a latitude and longitude coordinate to about 65,000 of the transactions in one batch; the remainder could not be geocoded accurately.[9] This sort of spatial analysis would have been exceptionally time consuming using the old system of paper maps.

GIS also allows spatial data to be assigned to districts such as census tracts or labor market areas. Data within these districts can be assigned to the geographical centers of the districts. Furthermore, the analyst can draw district boundaries of her own; for example, census tracts can be joined to form districts, or neighborhood boundaries of the analyst's choice can be drawn on the map; the attributes of the tracts can be added or averaged for the new, larger polygon. Each of the 65,000 geocoded transactions in the Fairfax County data set was assigned to one of 166 census tracts, a process that would not have been feasible without GIS.

GIS provides a flexible means for assigning latitude and longitude coordinates to important points such as the CBD or the airport. Also, spatial dummy variables can be developed rather easily. For example, the Fairfax County observations were assigned spatial dummies depending on whether they were inside the beltway or inside the Dulles access corridor. This was done by a simple point and click technique to identify coordinates bounding these areas; most GIS software allows the user to draw polygons and select the subset of points within polygon boundaries.

GIS software has the potential for calculating great arc distances (i.e., distances that allow for the curvature of the earth) based on latitude and longitude coordinates. Most GIS software makes this rather easy to do for any pair of points by using a mouse. However, researchers typically need to calculate great arc distances for a large number of pairs of points; spatial autoregressive analysis requires flexible computation of distances less than

[9] The primary reason why data at the property level might not be easily geocoded is that there may be incomplete or inaccurate information in the real estate data or the underlying map files. If ungeocoded data are randomly distributed throughout the area being studied, then efforts to further geocode may not be warranted. However, if there is a systematic bias in the ungeocoded data (such as ignoring new developments due to missing new roads in the underlying maps) efforts should be made to remove these biases.

a given number. Some GIS products have built-in functionality to quickly compute a matrix of distances among points in a database while others require the distance from each data pair to be computed one at a time (see Table I). The ability to calculate road distances and travel time from points in a database also differs across GIS products.

GIS allows plotting spatial data or analytical results on a map. Points of interest can be emphasized, only major roads can be included or excluded as desired, and graphical presentations (e.g., dot density or thematic maps) can be constructed rather easily. In addition to facilitating the presentation of results, this provides a type of visual analysis of spatial relationships. Table I indicates that the midrange of GIS software has good to excellent mapping capabilities, except that most are not able to produce contour or three-dimensional maps.

The Availability of Maps and Data Suitable for GIS. Maps are increasingly available to GIS systems. Street maps are available from the Bureau of Census; these are the so-called TIGER files, now updated to 1994. Some software packages (e.g., Maptitude) make these available on a single CD. Likewise, state boundaries and interstate roads are often distributed free of charge with GIS software. The boundaries of census tracts and zip code areas are available relatively inexpensively.[10]

The 1990 Census of population and housing is now readily available in a format suitable for GIS systems. Several companies use statistical techniques to update and forecast the census information. Furthermore, they tap other data sources to estimate sales for different kinds of products in each Census tract. These data can be obtained for a small geographical area at a price affordable to many researchers (e.g., a few hundred dollars).

Limitations of GIS Software and Data

Calculation of straight-line and road distances is an important part of spatial analysis. The mid-range of GIS software summarized in Table I allows straight line distance to be calculated with point and click techniques. But, this is not really adequate for the purposes of academic research.

Real estate and urban economic analysis that uses explicit distances needs to calculate distances from each observation in the data set to each point of interest. For example, the distance from each of thousands of transactions to the CBD needs to be calculated. The GIS software summarized in Table I can do this after some programming. However, weighting of road links is not readily available to most GIS software. For instance, researchers often need to weight interstate roads half as much as secondary roads;

[10] Zip code boundaries can now be purchased for a few hundred dollars for the entire U.S.

additional weights are needed for congested areas, construction zones, and bridges.

Spatial autoregressive (SAR) methods are even more demanding of distance calculations. Here, distances from each transaction to every other transaction need to be repeatedly recalculated. For SAR, researchers often need to export geocoded observations from the GIS software to other software which is specifically designed to calculate great arc distances. Software designed by Luc Anselin is able to calculate these distances along with relevant spatial statistics such as Moran's I statistic.[11] Thus, the researcher is involved in a cumbersome process, where the GIS is used to geocode observations and to display results; most GIS software does not have the functionality required to actually perform spatial statistical analysis.

Road distances can be calculated by some GIS software, but it is not feasible to calculate these distances for large numbers of transactions. Similarly, it is generally not possible to weight different types of linkages (e.g., interstate roads versus secondary roads) in a way suitable for real estate and urban economic analysis. Thus, the use of this aspect of GIS must await further software development.

Three-dimensional graphics are generally not supported by the mid-range of GIS software presented in Table I. This is a limitation from the point of view of researchers; it is desirable to display a price or rent surface on top of a map showing roads and places of interest.[12]

Some GIS software supports easy connectivity with other applications. For example, Mapinfo supports relatively easy connectivity with Microsoft Office. Object-embedded linkages (e.g., a map embedded in a word processing document) is also supported by Mapinfo, whereas SAS–GIS does not support this to the same degree.

In general, GIS software is not user-friendly. It can often be very difficult to do simple tasks such as calculating straight-line or road distances. Standard database management functions such as queries often can be very cumbersome and time consuming.

Limitations of the Data Available to GIS. Large amounts of data are becoming available in a format suitable for use by GIS. However, some of these data, such as employment, income, and retail sales may be available

[11] Anselin is with the Regional Research Institute, West Virginia University, Morgantown, West Virginia. IDRISI software from the Department of Geography, Clark University, Worcester, Massachusetts, also calculates a range of spatial statistics.

[12] However, for every shortcoming that the mainstream GIS software has, the larger geography technology market has provided a solution. Contour maps can be made from attribute data geocoded using GIS software; we reference one such product, SURFER, for comparison purposes.

only after a substantial lag. The researcher will generally need to find her own sources of transactions data at the individual person or household level; it will undoubtedly be costly and time consuming to keep these data current.

Researchers frequently want to work with time series data; but, most GIS data are not available on a time series basis. Even where transactions data (e.g., house sales) are available in time series form, it may be very difficult to geocode these data correctly. Road networks change over time but GIS vendors rarely maintain the "old" versions of these maps. Similarly, about 3–5% of zip codes change every year but historical information is not readily available.

Electronic boundary files for Census geography (e.g., blocks, block groups, or tracts) are not typically available for 1970 and 1980 Censuses. These boundary files are needed to match historical information (e.g., on transactions) to census data available for one of the census geographical units. Since these boundary files are missing, one must engage in a difficult process of tracking splits and consolidations from one decennial census to another.[13]

The format of 1970 and 1980 Census data further complicates the process of developing a time series for these data; 1970 and 1980 Census data must be obtained from magnetic tape and translated into a form suitable for GIS analysis. Alternately, for small geographical areas it is feasible to collect 1970 and 1980 Census data from the printed reports and key punch them in a form suitable for GIS analysis. In the case of the Fairfax project, we translated 1980 Census data from the tapes and copied a very limited amount of 1970 data. We then went through the torturous process of matching the boundaries of 1970 Census tracts to 1980 and 1980 to 1990, adjusting data for splits and consolidations.

4. CONCLUSIONS

Geographical information systems (GIS) are database management systems with a spatial reference. GIS is a powerful tool for real estate and urban economic studies using explicit distance (e.g., from each house to the CBD). Straight-line distances can be readily calculated by a mid-range of GIS software (see Table I). Road distances with flexible weights (e.g., interstate roads weighted half of secondary roads) awaits further software development.

[13] When tracts from two Censuses are not spatially congruent, population and employment data must be allocated to maintain consistency. For income data, weighted averages are needed.

Spatial autoregressive models need to repeatedly calculate large numbers of distances from each observation to every other observation within a given radius. In this context, GIS software is primarily useful for assigning a latitude and longitude location to each observation based on the best of street, ZIP, or ZIP+4. These geocoded observations may then be exported to other software for batch processing of SAR statistics. Future development of GIS software should eliminate the need for multiple software packages.

The mid-range of GIS software reviewed here is capable of displaying data and results on a map; typically, this takes the form of dot density, shaded, or color-coded thematic maps. However, researchers may need to map three-dimensional surfaces (e.g., rent or price surfaces). Further development of GIS software is required.

The availability of data in a form suitable for GIS is expanding rapidly. For example, 1990 Census data and the data from the City and County Data Book are now available. Private vendors provide a large amount of additional data such as updates of the census data and estimates of sales for different types of products. However, time series data are generally lacking. For instance, 1970 and 1980 Census data are difficult to use with GIS software.

APPENDIX: GLOSSARY[14]

Address matching: see Point Georeferencing.

Contour elevation or contour surface is comprised of geographic coordinate pairs represented as (x, y), and an attribute represented as the z value. Hence, a georeferenced point with an attribute is represented as (x, y, z).

Choropleth is a form of thematic map where polygons (e.g., states or towns) are shaded according to a numeric attribute (e.g., averaging income).

Digitizer is a computer instrument that allows the input of geographic coordinates (e.g., from a paper map) into a database.

Lines require a minimum of two sets of geographic coordinate pairs and may have a set of attributes assigned to the resulting line object. Streets are normally represented as lines. Topologically a line has a left and right side, as well as above and below positions.

Manhattan distances. If straight-line distance, or "as-the-crow-flies" distance, is the hypotenuse of a right triangle, then the sum of the other two sides of the right triangle are known as the Manhattan distance. This is often a good approximation for street travel distance in urban settings that conform to a rectangular grid system.

[14] Also see Marks, Stanley, and Thrall, (1994) and Thrall (1998).

Morans I is a statistic used to measure the presence or absence of spatial autocorrelation. If geographic proximity between two observations is the cause for similar attribute values, such as house price, then spatial autocorrelation is likely present.

Point has one geographic coordinate pair (x, y) that defines its position on a map. The point object does not have area, width, or length; however, a point may have a set of attributes that provides associated information. The location of a house may be represented as a point object where the point object has attributes such as size and price.

Point georeferencing is a procedure whereby points are positioned onto a map. "Address matching" is a point georeferencing procedure that uses the street address of a property matched to information within the TIGER/Line database for the assignment of geographic coordinates.

Polygons (e.g., counties or towns) have at least three geographic coordinate pairs that enclose an area and are used to define their position on a map. Polygons have area, width, and breadth and can have other attributes.

Projections are a means of translating geographic information from a sphere to a flat plane. Distortion in the form of distance, direction, and shape (area) occurs whenever a sphere is projected onto a flat plane. Any projection system is a compromise of distortion.

Raster imagery from satellites or air photography can be used as a base map over which other forms of vector (point, line, polygon) information can be layered.

Registration is a process of positioning a raster image onto a vector map in its correct geographic location.

Rubber sheeting allows a raster image to be bent, stretched, shrunk, and repositioned so that as a base map the vector objects may be more geographically accurately positioned relative to information on the raster image.

Scaled symbol is a map object whose size is made to be proportional to some attribute value.

Snapping support is useful when placing objects such as text onto a map. A rectangular grid with some designated intervals is associated with the map and if snapping support exists the mouse cursor will jump to the closest grid line or closest map object.

SQL, or structured query language, is a standardized and commonly used programming language used to search, order, and otherwise manipulate large databases.

Tiger Line Files are often referenced as TIGER/Line from the acronym for Topologically Integrated Geographically Encoded Reference line database published by the U.S. Bureau of the Census. The database was intended to be accurate as of January 1990, though there was a significant update in 1992. Information contained in the database includes coordinate boundaries of states, counties, Census tracts, Census block groups, Census blocks, municipalities, political jurisdiction, streets, waterways, and many other important features. The street database also includes the name of the street, the range of addresses on the street, and the 1990 five-digit ZIP code.

Vector is one form of geographic database and is most commonly used for the analysis of urban phenomena. Vector-based GIS uses geographically referenced points, lines, and polygons as map objects that may have attribute information tagged to those objects.

REFERENCES

Arbia, G. (1989). *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems.* London: Kluwer Academic.

Can, A., and Megbolugbe, I. (1996). "The Geography of Undeserved Mortgage Markets." Draft paper prepared for the mid-year meeting of the American Real Estate and Urban Economics Association.

Can, A., and Megbolugbe, I. (1995). "Spatial Dependence and House Price Construction." Draft paper presented at the AREUEA mid-year meetings.

Case, K. E., and Mayer, C. J. (1996). "Housing Price Dynamics Within a Metropolitan Area," *Regional Sci. Urban Econ.* **26,** 387–407.

Do, Q., Wilbur, R. W., and Short, J. L. (1994). "An Empirical Examination of the Externalities of Neighborhood Churches on Housing Values," *J. Real Estate Finance Econ.* **9,** 127–136.

Dubin, R. A., and Chein Hsing Sung (1990). "Specification of Hedonic Regressions: Non-nested Tests on Measures of Neighborhood Quality," *J. Urban Econ.* **27**(1), 97–110.

Dubin, R. A. (1988). "Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms," *Rev. Econ. Statistics* Aug., 466–474.

Goodman, A. C., and Thibodeau, T. G. (1995). "Age-Related Heteroscedasticity in Hedonic House Price Equations," *J. Housing Res.* **6**(1), 25–42.

Huxhold, W. E. (1991). *An Introduction to Urban Geographic Information Systems.* Oxford and New York: Oxford Univ. Press.

Marks, A. P., Stanley, C., and Thrall, G. (1994). "Criteria and Definitions for the Evaluation of Geographic Information System Software for Real Estate Analysis." *J. Real Estate Lit.* **2**(2), 227–241.

Pace, R. K., and Gilley, O. W. (1997). "Using the Spatial Configuration of the Data to Improve Estimation," *J. Real Estate Finance Econ.* **14**(3), 333–340.

Quigley, J. M., VanOrder, R., and Deng, Y. (1994). "The Competing Risks for Mortgage Termination by Default and Prepayment: A Minimum Distance Estimator." Draft paper, NBER Summer Institute, Cambridge, MA.

Ripley, B. D. (1981). *Spatial Statistics.* New York: Wiley.

Rodriguez, M., Sirmans, C. F., and Marks, A. P. (1995). "Using Geographic Information Systems to Improve Real Estate Analysis," *J. Real Estate Res.* **10**(2).

Rosenthal, S. S., and Helsley, R. W. (1994). "Redevelopment and the Urban Land Price Gradient," *J. Urban Econ.* **35**(2), 182–200.

Sivitanidou, R. (1996). "Are Service Agglomerations Weakening? Evidence from Metropolitan Los Angeles." Draft paper, School of Urban and Regional Planning, University of Southern California.

Star, J., and Estes, J. (1990). *Geographic Information Systems: An Introduction.* Englewood Cliffs, NJ: Prentice Hall.

Thorson, J. A. (1994). "Zoning Policy Changes and the Urban Fringe Land Market," *J. Amer. Real Estate Urban Econ. Assoc.* **22,** 527–538.

Thrall, G. I. (in press). *Business Geography: GIS for Economics, Real Estate, and Market Analysis.* Oxford: Oxford Univ. Press.

Thrall, G., and Marks, A. P. (1993). "Functional Requirements of a Geographic Information System for Performing Real Estate Research and Analysis," *J. Real Estate Lit.* **1,** 49–61.

Turnbull, G. K. (1995). *Urban Consume Theory.* AREUEA Monograph Series, No. 2, The Urban Institute Press, Washington, DC.

Waddell, P., Berry, B. J., and Hoch, I. (1993). "Residential Property Values in a Multinodal Urban Area: New Evidence on the Implicit Price of Location," *J. Real Estate Finan. Econ.* **7,** 117–141.

Wilson, A. G. (1970). *Entropy in Urban and Regional Modeling.* London, Pion Limited.