Geographical Information Systems (GIS) and Economics[1]

Henry G. Overman (London School of Economics)

5th January 2006

Abstract: Geographical Information Systems (GIS) are used for inputting, storing, managing, analysing and mapping spatial data. This article consider the role each of these functions can play in economics. GIS can map economic data with a spatial component; generate additional spatial data as inputs to statistical analysis; calculate distances between features of interest and define neighbourhoods around objects. GIS also introduce economics to new data. For example, remote sensing provides large amounts of data on the earth's surface. This data is of inherent interest, but can also provide an exogenous source of variation and allowing the construction of innovative instrumental variables.

[1] Thanks to Steve Gibbons, Yannis Ioannides, and Diego Puga for comments. This article will appear in S. N. Durlauf and L. E. Blume, The New Palgrave Dictionary of Economics, forthcoming, Palgrave Macmillan, and is reproduced with permission of Palgrave Macmillan. This article is taken from the author's original manuscript and has not been reviewed or edited. The definitive published version of this extract may be found in the complete New Palgrave Dictionary of Economics in print and online, forthcoming.

Geographical Information Systems and Economics

Geographical Information Systems are used for inputting, storing, managing, analysing and mapping spatial data. In this article, we consider each of these functions to help assess the role that GIS can play in economic analysis. Of course, a wide range of software can provide similar functions for quantitative data so it is the geographical, or spatial, element that separates GIS. That spatial dimension will be the focus of this piece. One important aspect of GIS that will not be covered, is the choice of software. Standard texts, such as chapter 7 of Longley, Goodchild, Maguire and Rhind (2005) consider the question of appropriate software in some depth.

At the outset, note that while GIS are widely used in business, government and a range of academic disciplines their application in economics has, to date, been more limited. The most frequent application in economics is the use of GIS to visualise or map economic data with a spatial component. Most entry level courses in econometrics begin with a plea to “plot the data” at an early stage of analysis to help identify trends, outliers etc. Much the same could be said of the role of mapping spatial data and GIS provide a simple and efficient way to do this.

Less common, but arguably more interesting, is the use of GIS storage and management functions to generate additional data as inputs in to further statistical analysis. In the simplest case this will involve using GIS to manage spatial data from a variety of sources. Many of these sources, for example sampling and census data, will be familiar to economists. Some, such as aerial photography and remote sensing data, less so. The spatial nature, or format, of the data will depend on the geographical data model used. The two most common models are raster format (assigning a code to each cell on a regular grid) and vector format (assigning a code to, and providing coordinates for, irregular polygons). GIS provide tools for moving between these different geographical data models. While the methods used to do this are rather intuitive, the devil is in the detail. As with the implementation of pre-packaged econometric routines one should understand the underlying basis for these transformations before proceeding. These issues are covered at depth in most, if not all, of the standard references and we do not consider them further here.

More generally, it is the ability of GIS to reconcile spatial data from different sources that allows the creation of new data sets. In the simplest case, this may involve combining socio-economic data from different spatial units. For example, population data from US census tracts with employment data for US zip codes. Many economists will be used to using ready made concordances (i.e. mappings from one set of spatial units to another) for undertaking such data merges. GIS bring the flexibility of allowing the user to define their own concordances between different geographical units of observation when faced with data from different sources.

The construction of more ambitious data sets is possible if one is willing to draw on a range of analytical functions available in the more advanced GIS. GIS can be used to identify whether observations occur at particular locations and if so to identify the characteristics of observations at those locations. For example, one of the most frequent application of GIS in economics to date has been to identify and characterise properties for use in hedonic analysis (see Bateman, Jones, Lovett, Lake and Day (2002) for a review). At its most basic, this will simply involve the merging of different data sets as described above. However, much more complex analysis is possible. Given that GIS data are spatial, a natural use is to measure the distance between observations or between observations and other features of interest. These distances could be physical distances, network distances (e.g. along a transport network) or involve some more general concept of social distance.

Observation to observation distance calculations have seen widespread application in the fields of biology and biomedical sciences through the statistical analysis of spatial point patterns (see Diggle, 2003). Knowing the distance between observations is useful if we think that there may be interactions between them and that the strength of these interactions is mitigated by distance. For example, in industrial organisation models of spatial competition, the intensity of competition may depend on the distance between firms. Observation to observation distances are also useful when the underlying entities are free to choose their location and we want to assess whether there are systematic patterns to those location decisions. For example, the study of localisation asks whether firms in a particular industry tend to be spatially concentrated relative to overall economic activity. If they do, one would expect the observation to observation distances to be less for firms in that industry than for a randomly chosen firm from the economy at large. The increasing availability of geo-referenced economic data suggests that the application of appropriately adapted procedures will become more common in economics (see Duranton and Overman, 2005). Hedonic analysis again provides the most frequent application of GIS to calculate distances of observations from other features. For example, Gibbons and Machin (2005) use GIS to measure the proximity of properties to rivers, coasts, woodlands, roads, railway lines and airports in their study valuing rail access.

In addition to the calculation of distances, GIS can be used to construct measures of area or to define neighbourhoods (or "buffers") around objects. For example Burchfield, Overman, Puga and Turner (2005), in their study of urban sprawl use GIS to calculate the percentage of the urban fringe - defined as a 20 kilometer buffer around existing development - that lies above water yielding aquifers.

These examples cover the main types of spatial analyses that are undertaken to construct spatial data in economic applications, but others are possible and should be covered in any of the standard texts. It should be noted that in advanced GIS these operations can be done both interactively and automatically using batch files (i.e. where the user writes a sequence of commands in a file that the computer implements one by one). Both approaches, but particularly the latter, involve fairly large fixed costs in terms of both purchasing software and learning how to implement the relevant procedures. There are other methods for conducting many of these analyses that do not imply the use of GIS. For example, great circle distances can easily be calculated on the basis of latitude and longitude (see Overman and Ioannides 2004). Whether the fixed cost investment is worthwhile will depend on individual circumstances. The benefits can be substantial. In many circumstances, GIS calculations should be more accurate than short cuts implemented using non-spatial software and some analysis such as the calculation of areas and buffers is much easier to implement in GIS.

GIS also introduce economics to new sources of data. In particular, remote sensing from either satellite or aerial photography, or digitised geological maps can provide a huge amount of data on the earth's surface. Early applications using these kinds of data tended to focus on issues arising from natural resource management such as valuing timber yields from forested areas. However, data on land cover and land use (i.e., the physical features that cover the land and what those features are used for), soil type, geological and landscape features, elevation and climate are opening up new avenues of research. These data sources allow the description of different features of the economic landscape that one might seek to explain. For example, Burchfield et al (2005) use remote sensing data to track the evolution of land use on a grid of 8.7 billion 30×30 metre cells covering the conterminous United States and then seek to explain differences in land development patterns across cities. Another example is Rappoport’s (2004) study of the role that weather plays in explaining changes in population for U.S. counties. The meteorological data that he uses comes from 6,000 meteorological stations and covers 20 winter, summer and precipitation variables. GIS analysis by the Spatial Climate Analysis Surface at Oregon State University applied to this meteorological data allows the construction of weather variables for a 2 kilometre grid covering the continental U.S.

GIS data also have the potential to contribute to a range of established areas of study, particularly because data on the earth’s surface can provide an exogenous source of variation and thus allow researchers to construct instrumental variables using GIS. Some examples should help to make this idea concrete. Hoxby (2000) is interested in whether competition among public schools improves school outcomes. That is, do cities with more school districts have better public schools and less private schooling? The problem that the analysis needs to confront is that for a city of a given size, better public schools and less private schools in a city should imply more school districts. That is, the number of districts is endogenous to public school quality. What is needed is an instrument that should determine the supply of school districts but that is independent of the local public school quality. The paper argues that the number of streams in a metropolitan area provides such an instrument. Cities with a large number of streams end up with more school districts for reasons that are surely nothing to do with public school quality. This paper provides a well known example of the strategy, although not of the use of GIS as her work is based on the study of detailed paper maps.

Rosenthal and Strange (2005) provide an example of the use of GIS to implement such an instrumental variables strategy. They are interested in whether density of employment helps determine wages. The problem is that higher wages should attract more workers and lead to higher employment densities. That is, density may be caused by wages and not vice-versa. The paper argues that the density of employment will be partly determined by the height of buildings in a location. They point out that the height of buildings is, in turn, partly dependent on the underlying geology of the site. Given that geology should not determine wages directly (they are studying cities, not agricultural production) the underlying geology can be used as an instrument. Locations with a suitable underlying geology can have higher buildings, higher employment density and should thus have higher wages. They use GIS data on the type of underlying bedrock, seismic and landslip hazard as instruments for the density of employment in their regressions of wages on employment density. Such examples suggest a potentially important role in future work for GIS data as a component in novel instrumentation strategies.

This piece has only skimmed the surface of the potential applications of GIS in economics. As spatially referenced socio-economic data becomes more widely available it is to be expected that the scope for applications can only increase.

References

Bateman, I. J. and A. A. Lovett (1998), 'Using Geographical Information Systems (GIS) and Large Area Databases to Predict Yield Class: A Study of Sitka Spruce in Wales', Forestry 71 (2), 147–168.

Bateman, I., A. Jones, A. Lovett, I. Lake and B. Day (2002) Applying Geographical Information Systems (GIS) to Environmental and Resource Economics. Environmental and Resource Economics 22: 219–269.

Burchfield, M., H. G. Overman, D. Puga and M. Turner (2005) The Determinants of Sprawl: A Portrait From Space. Forthcoming Quarterly Journal of Economics.

Diggle, P. (2003) Statistical Analysis of Spatial Point Patterns, Arnold, London.

Duranton, G. and H. G. Overman (2005) Testing for Localisation Using Micro-Geographic Data, Review of Economic Studies, 2005, 72(4), 1077-1106.

Gibbons, S. and S. Machin (2005) Valuing Rail Access Using Transport Innovations, Journal of Urban Economics, 57: 148-169.

Hoxby, C. (2000), "Does Competition Among Public Schools Benefit Students and Taxpayers?", American Economic Review, 90(5), 1209-38.

Longley, P., M. Goodchild, D. Maguire and D. Rhind (2005) Geographical Information Systems and Science, 2nd edition, John Wiley and Sons.

Overman, H. G. and Y. Ioannides (2004) The Spatial Evolution of the US Urban System, Journal of Economic Geography 4 (2), 131-156.

Rappaport, J. (2004), Moving to Nice Weather, Federal Reserve Bank of Kansas City Working Paper #03-07.

Rosenthal, S. and W. Strange (2005), The Attenuation of Agglomeration Economies: A Manhattan Skyline Approach, Processed University of Toronto.