spatial economics

Abstract:

This article provides a general overview of spatial economics, which covers location theory, spatial competition, and regional and urban economics. After a brief review of the main theoretical traditions, the fundamental role of non-convexities and imperfect competition is highlighted. The main challenges faced by theoretical and empirical research are also discussed, followed by a broader discussion of the relationship between this field of research and other sub-fields of economics and other disciplines.

What is spatial economics? In a nutshell, spatial economics is concerned with the allocation of (scarce) resources over space and the location of economic activity. Depending on how this definition is read, the realm of spatial economics may be either extremely broad or rather narrow. On the one hand, economic activity has to take place somewhere so that spatial economics may be concerned with anything that economics is concerned about. On the other hand, location analysis focuses mostly on one economic question, namely, location choice. This is only one decision among a large number of economic decisions.

Which boundaries for spatial economics?

In practice, we can distinguish three sets of questions for which the importance of the spatial dimension is very different. Consider first the core questions of spatial economics. For example, why are there cities? Why do some regions prosper while others do not? Why do we observe residential segregation? Why do firms from the same industry cluster? These are intrinsically ‘spatial’ questions, that is, questions in which the spatial dimension plays a dominant

role. For instance, it would be difficult to speak meaningfully about the existence and growth of cities without some explicit consideration of space. Then there is a second group of issues concerned with, *inter alia*, the analysis of technological spillovers, the determinants of trade flows, or even the functioning of social networks. These issues all have a spatial dimension but its importance remains to be determined. Put differently, these are 'contested' issues between spatial economists and other economists. Finally, there is an extremely broad range of economic questions for which the spatial dimension is likely to be less important. For example, what are the drivers of investment? How important are firing costs to explain unemployment? What are the returns to education? To answer these questions, the main role of space is to provide possibly a major source of variation for empirical research. If, for instance, different regions of a country have different education systems with, say, different age limits for compulsory schooling, this variation can be used to produce some meaningful estimates of the returns to education. Even if we take this third set of questions to lie outside spatial economics, it nonetheless remains the case that spatial economics is concerned with a broad and heterogeneous set of questions, which involve very different spatial scales (from the very small to the very big) with imprecise boundaries. It is quite possible that this breadth and heterogeneity has hindered the development of the field. This is also what makes it interesting.

The centrality of spatial frictions

To see what makes spatial economics specific, it is useful to reformulate the question about its definition in the following way. Is spatial economics only about adding a spatial dimension? In the *Theory of Value*, Debreu (1959) answers affirmatively. A commodity is defined by all its characteristics including its location: the same good traded in different locations must be treated as different commodities. This 'answer' runs into serious problems, as pointed most clearly by Starrett (1974). Consider the extreme case of homogenous space where firms face the same convex production set, and consumer preferences are the same (and locally not satiated). Transporting commodities between locations is costly. Then the *spatial impossibility theorem* states that, with a finite number of locations, consumers, and firms, no equilibrium involves transportation. The intuition behind this result is straightforward: since economic activities are perfectly divisible and agents have no objective reason to distinguish between

locations, each location operates in autarchy to save on transport costs. To avoid this very counterfactual result (no trade), one of the assumptions behind the spatial impossibility theorem needs to be relaxed. If one takes transport costs as an unavoidable fact of life, one must assume either some non-homogeneity of space or some non-convexity of production sets.

As shown by a first branch of trade theory, it is possible to develop a framework for spatial economics that builds only on local productivity differences. This approach was pioneered by Ricardo (1821), who developed a theory of land use based on relative fertility. This approach was later generalized to consider exogenous technological differences for all types of goods. A second branch of trade theory builds instead on differences in factor endowments over space. This is the so-called Hecksher–Ohlin theory of trade. The Ricardian and Hecksher–Ohlin approaches have led to sophisticated theories of location and trade that rely on the existence of (exogenous) 'comparative advantages' across locations. As shown by a large body of theoretical work in international trade, these approaches can be readily incorporated in the Arrow–Debreu framework. Although these approaches are central to the sister discipline of international trade, they played a much less important role in the development of spatial economic theory.

The pioneers

Instead, spatial economics has focused on the existence of non-convexities in the presence of transport costs. A key reason for this focus is that, although comparative advantage constitutes an appealing explanation for understanding trade flows at the world level, it provides at best a partial explanation for the location patterns of industries within countries, and it is at pains to explain major concentrations of population in large metropolitan areas. Instead, non-convexities in production or consumption seem to hold more promise for providing convincing answers to the core questions of spatial economics. The easiest way to model these non-convexities is to assume some indivisibility in a partial equilibrium framework. This type of work was pioneered by von Thünen (1826). In his model, a competitive farming sector bids for some homogenous land. The key non-convexity is that the output must be sold at a central market. With costly transport costs, farmers are willing to bid for land up to the point where the rent at a given distance from the market is equal to the gross revenue from the output minus the cost of non-land factors and minus transport costs. With a competitive land market, land goes to the highest bidder

and the equilibrium typically involves concentric rings of specific land use around the central market.

To understand land use patterns in cities, Alonso (1964) developed an approach that was based on similar principles. His model again assumes a homogenous space, but replaces von Thünen's market by a central business district to which residents must commute at a cost to find work. This very parsimonious microeconomic model manages to replicate key stylized facts about land use and land prices within cities using rigorous microeconomic modelling. This is a showcase for the power of microeconomic approaches. It has spawned a large literature, which first extended the basic model in a number of directions and then went on to model multi-centric cities (for further details, see Fujita, 1989).

Independently of the 'Thünen tradition' that relies on an exogenous focal point for trade or production, another tradition was developed following Weber's (1909) work. Weber deals with the location problem of an indivisible and competitive plant that faces transport costs in order to ship its inputs from their sources and its outputs to their markets. With the use of essentially linear-programming techniques, the optimal location (which minimizes total transport costs) can be derived. Like Alonso's monocentric model, Weber's approach has been extended in many directions to consider, among others, more flexible production functions and the optimal location of public facilities.

Hotelling (1929) also explored the location problem faced by producers but went in a very different direction. His fundamental insight is that, because of indivisibilities, there will not be infinitely many producers at each point so that Weber's price-taking assumption is not tenable. With a small number of producers the location decision will involve more than minimizing transport costs since location also affects the competitive process. To make his point, he assumes evenly distributed consumers over a finite segment, each consuming one unit of a homogenous good. The market is served by two firms that need to choose a location and each customer patronizes the firm that minimizes the sum of the 'mill' price and shipping costs. At a first stage, firms choose a location and then compete in price. This deceptively simple game has received a lot of attention. Firms face a fundamental trade-off between central locations, which allow them to capture a larger share of the market, and more peripheral locations, which allow them to mitigate the intensity of competition. The resolution of this trade-off depends on the fine details of the assumptions being made (and particularly how an increase in the distance affects the price

setting power of producers). Difficulties with existence of equilibrium has also turned out to be an important issue in this literature.

Modern approaches to spatial economic theory

Non-convexities in production lie at the heart of spatial economics. The literature discussed so far treats them as exogenous. It was not long before the literature started to worry about what these non-convexities were about. Nowhere was this worry stronger than in the 'new urban economics' literature, where Alonso's assumption of an exogenously given central business district quickly started to look very ad hoc. To understand central business districts or, more generally, why economic activity agglomerates, spatial economics had to provide microeconomic foundations for (local) increasing returns. Being able to generate increasing returns from plausible assumptions without leading to a degenerate market structure (for example, a monopoly firm for the entire economy) was a fundamental challenge for spatial economics. This was also true for many other fields such as industrial organization and international trade, where increasing returns were also needed to explain key stylized facts. Spatial economists were fortunate because they could draw on the insights provided by Adam Smith (1776) and Alfred Marshall (1890). Although Smith's argument about the division of labour being limited by the extent of the market pre-dates Marshall's *Principles* by more than a century, Marshall's 'magic trilogy' proved much more influential. Following Marshall, local increasing returns could arise because of knowledge spillovers, linkages between input suppliers and final producers, and thick local labour market interactions. What the modern literature on the microfoundations of increasing returns has achieved is a formalization of these insights (see Duranton and Puga, 2004, for an extensive review of this literature). Three main mechanisms can be used to generate local increasing returns: sharing, matching, and learning. Sharing mechanisms show how small non-convexities like small fixed costs paid by heterogeneous producers can be spread across larger quantities as market size increases and thus yield aggregate increasing returns. Matching mechanisms explore how larger markets might improve the probability and quality of matching. Finally, learning mechanisms explore the benefits of local size for the creation and diffusion of knowledge.

The second major problem faced by spatial economics is that many fundamental issues having to do with regional and urban development call for general equilibrium modelling. For

instance, some cities can afford to specialize because they can trade with other cities. Hence, looking at one isolated city in the tradition of Thünen and Alonso may not be enough for some purposes. Similarly, the agglomeration of economic activity in core regions may occur because firms find larger markets there and because consumers find cheaper and more diverse supplies. These two forces are mutually reinforcing. This is the famous circular and cumulative causation mechanisms first emphasized by Myrdal (1957).

To model spatial economies, two main approaches came to dominate the intellectual landscape. The first follows the work of Henderson (1974) and is know as the 'urban systems' approach. In this type of framework, cities arise endogenously as the result of a trade-off between agglomeration economies and urban crowding. Both types of forces are modelled with the use of various microeconomic foundations. Cities can also trade with one another and the workers decide where to work. This strand of literature has been successful in replicating many stylized facts about urban systems, from the tendency of many cities to specialize while others diversify to their role in the innovation process.

Following the work of Krugman (1991), the *new economic geography* is the second main general equilibrium approach in spatial economics. This approach puts trade costs (rather than commuting costs in urban systems) at the heart of the agglomeration–dispersion trade-off. Agglomeration in the larger market is beneficial for firms because it gives them better access to consumers. Following this, workers also want to be in the larger market in order to be able to buy goods without having to pay inter-regional trade costs. Krugman's model is based on Dixit and Stiglitz's (1977) model of product differentiation and offers a formalization of Myrdal's circular and cumulative causation. It goes beyond that because agglomeration is not always an equilibrium outcome. This is because, under agglomeration, most goods sold in the periphery need to be shipped from the core and thus prices there may be quite high. In turn, this can make it profitable for firms to locate in the periphery. When trade costs are high, the even dispersion of manufacturing is indeed the unique equilibrium in Krugman's model. On the other hand, when trade costs are low, serving the residual demand in the periphery can be achieved at a low cost and agglomeration occurs. This strand of literature has grown exponentially since 1990, culminating with Fujita, Krugman, and Venables's (1999) book.

The difficulties of spatial empirical work

What about the evidence, then? Ultimately, it is observation that should allow us to judge of the relevance of our theories. To discuss very briefly what the issues for empirical work are, it is useful to retain the partial versus general equilibrium distinction made above. A typical 'partial equilibrium' question that has attracted much attention over the years is that of location choices of new firms. To look at the determinants of location choice, one would like to somehow explain location choices in terms of a bunch of possible determinants. This is a difficult exercise, for several reasons. The first one has to do with the nature of the problem. Location choices are not continuous. Because they are discrete, firms decide to locate 'somewhere' rather than be spread continuously. Put differently, new firms choose between discrete alternatives so that one has to use discrete choice methods, which are more complex to implement than standard regression approaches. Then there is a whole range of possible determinants for location choices. This makes this type of exercise very data-intensive and particularly prone to missing variables biases. It is also likely that location decisions are made not only on the basis of the characteristics of the spatial units where firms locate, but are also influenced by what happens in neighbouring units. More generally, it is likely to be the case that different determinants of location matter at different spatial scales. To take these concerns into account, spatial econometrics has developed a set of tools. Spatial econometrics resembles standard time-series analysis in that it takes the values of the explanatory variables of the neighbouring spatial units (as well as their error term) into account. The fundamental complication is that spatial dependence can 'go both ways', unlike time dependence in time series. An alternative would be to ignore spatial units altogether and work directly on continuous space. Although there have been some developments in that direction (see Cressie, 1993, for a review), data limitations confine this type of approach to a small range of problems.

When one looks at more general equilibrium issues, many of the difficulties of partial equilibrium analysis are still there while other concerns also become prominent. Take the analysis of regional disparities as an example. A first issue is that such general equilibrium problems have several endogenous variables. In our example regional population and income are likely to be simultaneously determined. To analyse this type of question, two polar approaches (and everything in between) may be adopted. A more descriptive approach consists in focusing on one particular variable, say, local income, and trying to explain its spatial variation using a

range of potential factors as indicated by theoretical models. Many of these factors such as the local population are then likely to be endogenous. This requires finding appropriate instruments for such endogenous variables (since unfortunately natural experiments are even scarcer in this field than elsewhere in economics). In some cases, good instruments may not be available. In contrast to descriptive analysis, structural approaches require writing down a particular model and deriving a set of equations that can then be estimated. The main problem faced by this type of approach is that many possible models are likely to have some explanatory power. To return to our example, regional disparities in a country are likely to be caused by the factors highlighted by the urban systems approach (local external effects and so forth) and those highlighted by the new economic geography (trade costs and pecuniary externalities), as well as factor endowments, institutions, and so on. The list of plausible determinants for regional disparities is long and it is very problematic to impose strong priors regarding the validity of one specific model. For many issues in spatial economics, model selection is in fact a huge concern (see Sutton, 2000, for more). Despite these difficulties, it is fair to say that much has been learnt about cities and regions since the mid-1970s (see Rosenthal and Strange, 2004, and Head and Mayer, 2004, for recent reviews).

The road ahead (?)

What current challenges does spatial economics face? On the theoretical front, three main problems remain open. The first is to provide a unified general equilibrium approach to spatial economics and end the often uneasy coexistence between urban systems and the new economic geography. Despite some attempts, as of 2005 there is no such unified framework, and providing one will be difficult. The main obstacles are about modelling. General equilibrium models of spatial economics entail making detailed assumptions about the spatial structure, the production structure, and the mobility of people, goods and ideas, all this under increasing returns. In such cases, nonlinearities occur everywhere and analytical solutions are the exception rather than the rule. Despite this, a general but tractable model of cities and regions is probably worth fighting for. A second key challenge regards the microfoundations of trade costs. Trade costs play a fundamental role in many models but their microeconomic foundations have received only scant attention. This will probably involve looking beyond transport costs and open the black box of

the multiplicity of transactions costs associated with trade between different parties. A third major challenge regards the development of a 'theory of proximity' (for lack of a better name). Such theory would provide some answers as to why direct interactions between economic agents matter and how. Non-market interactions will no doubt loom large in any theory of proximity.

On the empirical front, a first key challenge is to develop new tools for spatial analysis. With very detailed data becoming available, new tools are needed. Ideally, all the data work should be done in continuous space to avoid border biases and arbitrary spatial units. We are still a long way from being able to do so. The second main challenge for empirical work is of a very different nature. Applied work has over the years managed to produce a reasonable set of estimates regarding a range of issues such as the intensity of local externalities or the determinants of urban growth, among others. No doubt, further progress is necessary and will occur but the main challenge is now to understand the mechanisms behind the elasticities or the decompositions that have been produced. For instance, the elasticity of local productivity to the density of economic activity is now well circumscribed between two and five per cent. We ignore nearly everything about the relative importance of the possible mechanisms behind such numbers. Finally, being able to distinguish between theories – for instance, between factor endowments, urban systems and new geography to explain regional patterns of economic activity – is also a fundamental task where research has barely begun to make progress.

To conclude, one may want to raise the issue of the position of this field within economics and its relationship with other areas of investigation. It is fair to say that, with the advent of Alonso's modern urban economics and that of strategic models of location, spatial economics traded its breadth of knowledge against some depth on much smaller subset of questions. Since the mid-1970s, spatial economics has managed to broaden again its focus by remaining open to outside influences. For instance, the new economic geography finds its roots in international trade theory, while much modern empirical work is heavily influenced by modern applied labour economics and industrial organization. For spatial economics, there is scope for further expansion. Over the years, housing and real estate economics have become fairly detached from the rest of spatial economics and the time may be ripe for new encounters and new cross-fertilizations. A similar statement also holds for local public economics. Finally, outside economics, the part of geography that deals with economic issues, 'economic geography', has a focus that considerably overlaps with spatial economics. The relationship between the two

disciplines has been fraught with difficulties. On the one hand many geographers react very negatively to the renewed interest by economists in spatial issues. On the other hand, economists tend to ignore the work done by economic geographers. Despite these difficulties, geographers may learn something from the economists' more rigorous approach while the greater breadth of geographers may offer a great source of inspiration for economists.

Gilles Duranton

See also central place theory; geography, economic; GIS data in economics; housing supply; international trade: theory; location theory; monocentric versus polycentric models in urban economics; neighbors and neighborhood; spatial econometrics; systems of cities; urban agglomeration; urban economics; urban growth; urban housing demand ; urban production externalities; urban transportation economics

Bibliography

Alonso, W. 1964. *Location and Land Use*. Cambridge, MA: Cambridge University Press.

Cressie, N. 1993. *Statistics for Spatial Data*. New York: John Wiley.

Debreu, G. 1959. *Theory of Value*. New Haven: Yale University Press.

Dixit, A. and Stiglitz, J. 1977. Monopolistic competition and optimum product diversity. *American Economic Review* 67, 297–308.

Duranton, G. and Puga, D. 2004. Micro-foundations of urban agglomeration economies. In *Handbook of Regional and Urban Economics*, vol. 4, ed. V. Henderson and J.-F. Thisse. Amsterdam: North-Holland.

Fujita, M. 1989. *Urban Economic Theory: Land Use and City Size*. Cambridge: Cambridge University Press.

Fujita, M., Krugman, P. and Venables, A. 1999. *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, MA: MIT Press.

Henderson, J. 1974. The sizes and types of cities. *American Economic Review* 64, 640–56.

Head, K. and Mayer, T. 2004. The empirics of agglomeration and trade. In *Handbook of Regional and Urban Economics*, vol. 4, ed. V. Henderson and J.-F. Thisse. Amsterdam: North-Holland.

Hotelling H. 1929. Stability in competition. *Economic Journal* 39, 41–57.

Krugman, P. 1991. Increasing returns and economic geography. *Journal of Political Economy* 99, 484–99.

Marshall, A. 1890. *Principles of Economics*. London: Macmillan.

Myrdal, G. 1957. *Economic Theory and Under-developed Regions*. London: Duckworth.

Ricardo, D. 1821. *The Principles of Political Economy*, 3rd edn. Homewood, IL: Irwin, 1963.

Rosenthal, S. and Strange, W. 2004. Evidence on the nature and sources of agglomeration economies. In *Handbook of Regional and Urban Economics*, vol. 4, ed. V. Henderson and J.-F. Thisse. Amsterdam: North-Holland.

Starrett, D. 1974. Principles of optimal location in a large homogeneous area. *Journal of Economic Theory* 9, 418–48.

Smith, A. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: Printed for W. Strahan and T. Cadell.

Sutton, J. 2000. *Marshall's Tendencies: What Can Economists Know?* Cambridge. MA: MIT Press.

von Thünen, J. 1826. *Isolated State: An English edition of Der isolierte Staat*. Trans. by C. Wartenberg and P. Hall, Oxford: Pergamon Press, 1966.

Weber, A. 1909. *Alfred Weber's Theory of the Location of Industries*. Trans. by C. Friedrich, Chicago: University of Chicago Press, 1929.