GRAVITY MODELS

Abstract: Basic gravity models state that economic interactions between two geographically defined entities are proportional to the size of these entities and inversely related to the distance between them. They have great empirical explanatory power. The impact of distance is strong and not diminishing over time. Extended gravity models incorporate borders and contiguity effects and more sophisticated interaction cost measures. They can be theory grounded, which makes each country's location vis-à-vis the rest of the world play a role in the bilateral relationship. Various empirical approaches have been proposed to tackle the econometric issues at stake in these more sophisticated frameworks.

The simplest gravity model states that economic or social interactions between two geographically defined economic entities are proportional to the size of these entities and inversely related to the distance between them. The system of interactions that results from these bilateral relationships shapes the spatial organization of the global economy. Initially borrowed from the universal law of gravitation that Newton established in 1687 for heavenly bodies, this model has undergone many refinements in economics, in particular in order to better match some underlying theoretical models and data. Some of the first applications were proposed by social physicists such as Ravenstein (1885) for migration flows, and Reilly (1931) for consumers' shopping behaviour. Stewart (1947), an astronomer, suggested that the gravity law could be applied to a very wide class of social interactions. Tinbergen (1962) initiated what continued to be the main application of gravity models, namely, the study of the determinants of trade.

The basic framework

Typically, country *i*'s exports to country *j*, F_{ij}, are modelled as a function of the distance between these countries, d_{ij}, and of their economic mass, $\left(M_i, M_j\right)$, which is most often proxied by their GDPs. Thus a basic gravity trade model estimates parameters α, β, and δ (expected to be positive) such that

$$F_{ij} = G\frac{M_i^{\alpha} M_j^{\beta}}{d_{ij}^{\delta}}\varepsilon_{ij}, \qquad (1)$$

where G is a constant and ε_{ij} an error term capturing what is left unexplained by the model. M_i can be interpreted as the supply of the good traded and M_j the demand, while d_{ij}^{δ} captures trade costs, which encompass all costs incurred in transferring goods. These costs add to the price of goods when they are not sold locally and are assumed to increase with spatial distance. Alternatively, if F_{ij} is the number of migrants from i to j, regional populations are often more relevant as measures of M_i and M_j, while d_{ij}^{δ} reflects a moving cost. Similar interpretations can be proposed for other kinds of flows.

According to gravity models, proximity is the main engine of trade, of migration or of any precisely defined social interaction between spatially distinct economic entities. This could appear as an obsolete view of the world if one believes in the 'death of distance', as touted by popular accounts. However, Disdier and Head's (2006) meta-analysis over 1,467 estimations of δ on trade flows indicates an average value around 0.9: halving distance increases trade by 45 per cent. These authors report even larger δ's for recent periods, which means that the distance decay effect has actually increased in recent years.

In addition to the reliable estimates of the impact of distance they lead to, the success of gravity models is due to their great explanatory power for flows, and this holds true whatever the geographical scale (countries, large or small regions), the period of study or the goods considered. This makes gravity one of the most stable relationships in economics and a useful predictive tool. It can be also used for obtaining predictors of variables used in a second stage to explain income, productivity, or growth dispersion across space (for example, Redding and Venables, 2004).

A wide range of applications

Gravity models are applicable to many other endogenous quantities in addition to trade flows. We have mentioned migration. Urban planners use them for traffic forecasting. There is also evidence of gravity effects in explaining foreign direct investment. More novel are the estimations on equity flows. In that case, the explanatory power of the model is as great as it is for trade, and halving distance increases flows by 25 per cent (Portes and Rey, 2005). In a supposedly financially integrated world, this is high. The impact of distance is still significant, although around six times lower, for flows of ideas, identified for instance as the citations of patents in a country that have been taken out in another country (Peri, 2005).

McCallum (1995) shows that borders also matter a lot, as well as distance. Trade between two Canadian regions was found to be 20 times larger than trade between a Canadian region and a US state of the same size and at the same distance, even if this drops to around six times once some statistical problems are removed. Discrete gaps in trade flows are also systematically observed between areas that are contiguous, relative to those that are not. Such effects suggest that the impact of distance on trade is not log-linear, or even smooth, and that a wider class of spatial proximity measures is necessary to fully encompass the effects of space.

Proximity matters for spatial interactions because it proxies for many of their determinants. Transport costs are the most obvious one. Clearly, the energy consumption or the time spent in transport, which results in opportunity costs, increase with distance. For migration, moving costs (both monetary and psychological) increase with distance and jump upward once borders are crossed. International trade flows are clearly reduced by trade policies, but trade agreements are typically first established between nearby countries. More original is the idea that preferences and tastes may be biased towards local goods, which may result from better information about them. Information costs are also critical for firms that want to access distant markets. They need to find local retailers, and then to work with them (which also clearly matters for foreign direct investment). Prior to this, they must assess market size, find out about local tastes and possibly adapt their products and marketing. All of this may explain not only the impact of distance but also the role of additional proximity measures. Moreover, on top of a possible composition effect (relatively more goods that are less easy to trade are traded), this suggests another explanation for the recently increasing impact of distance. While transport costs and trade barriers have clearly strongly diminished, preference biases or information costs may have risen, possibly due the increasing number and complexity of the goods available.

Simple gravity models have been extended to control for economic factors that would directly capture trade costs. Large data-sets on trade agreements or on transport costs are now available. The fraction of the population sharing the same language is used to capture closeness of tastes or reduced information costs. More generally, the positive role of business and social networks (among migrants from the same country or among firms belonging to the same business groups) on trade is currently being studied. These measures reduce the impact of distance, borders, and contiguity, without completely eliminating them. Naturally, not all the reasons why space matters for interactions have yet been identified. Extensive references on trade costs proxies and the various applications of gravity models can be found in Anderson and van Wincoop (2004) and in Combes, Mayer and Thisse (2006, ch. 5).

Extended frameworks and the role of the rest of the world

The bilateral nature of the gravity model is somewhat surprising. One would expect the actual location of the respective economies vis-à-vis the rest of the world to matter also. Trade between Australia and New Zealand is certainly much larger than it would be if these countries were not so isolated. But nothing in basic gravity models takes this into account. Similarly, if the relationship results from equilibrium between supply and demand, the absence of the goods prices in the model is also striking. Moreover, economists have long been challenged by the strong empirical support of gravity models and, simultaneously, their lack of theoretical foundations. Therefore, from Anderson (1979) on, a number of approaches have been proposed to derive the gravity relationship from fully specified theoretical models. Anderson and van Wincoop (2004) show that for trade flows the class of models that can be used is fairly large. They are based on either comparative advantage or on imperfect competition; the most popular uses monopolistic competition. They all lead to specifications more complex than (1) but such microfoundations significantly improve the understanding of gravity models and of their underlying mechanisms. Their main feature is that the interactions between two economic entities, such as regions, do depend on their location relative to other areas because of interesting price effects. Indeed, in an economy with more than two regions and costly trade, the supply, demand and price of goods traded between two given economies depend on the relative costs of all firms and consumers to access all the markets. For instance, if Australia and New Zealand reduced their trade costs to all other developed countries, the price of the goods they exchange bilaterally would increase relative to the price of all other goods, which would reduce their bilateral trade. On the other hand, the overall saving in trade costs and the increase in competition it induces would also imply a greater purchasing power, which would increase trade with all partners, including bilateral trade. Interestingly, in these models local incomes can be shown to be function of the area's market potential. This potential takes a form similar to the economic version proposed by Harris (1954), which is reminiscent of gravitational or electric potentials, which must be enriched by price effects again. This is the sum of all regions' income discounted by trade costs and weighted by complex price effects.

Econometric issues

Unfortunately, such formulations modify (1) in ways that make estimation more cumbersome. Typically the equation becomes nonlinear in some unknown parameters (such as price elasticity) that must be estimated simultaneously with the impact of trade costs. Various approaches, detailed in Feenstra (2004, ch. 5), have been proposed to deal with that difficulty.

The first one consists of using nonlinear estimation procedures. Another solution takes as the left-hand-side variable the ratio of the bilateral flow to the flow to a reference destination, which makes the right-hand-side variables depend again on the origin and destination only. People sometimes use real price data, but these rarely match their exact definitions in the theoretical model. The least data-demanding strategy, which remains compatible with a large number of theory-grounded approaches, consists in estimating fixed effects for each origin and destination. However, the impact of the determinants of the fixed effects, the mass of the economies and their locations for instance are no longer identified.

Other econometric problems remain. First, the fact that each country trades with many destinations induces correlations between error terms, and therefore possible heteroskedasticity biases. More problematic are the zero trade flows towards a large number of destinations that often characterize many countries, which may induce selection biases. Santos Silva and Tenreyro (2006) propose a pseudo-maximum likelihood estimator to deal with both issues. The literature based on heterogeneous firms and the presence of fixed export costs might help provide more theory-grounded approaches to these problems (Helpman, Melitz and Rubinstein, 2006). The fact that economic masses and prices are simultaneously determined with trade flows might also bias the estimations. Appropriate instrumental procedures should help, however. Last, the trade cost proxies might themselves be endogenous. For instance, new infrastructure can be built in response to an increase of trade flows, trade agreements may be signed preferably between privileged trade partners and networks may emerge once trade is large. This is more difficult to handle and these possible reverse-causality biases have yet to be really investigated.

Hence, the long history of gravity models does not prevent them from stimulating a lot of current research, both theoretical and empirical. Microfounded frameworks move the model further and further away from Newton's law, adding a lot to the understanding of the mechanisms shaping spatial interactions. Additional challenges lie ahead.

Pierre-Philippe Combes

See also GIS data in economics; international trade and heterogenous firms; new economic geography; regional development; spatial economics; systems of cities; trade costs; tradeable and non-tradeable goods

Bibliography

Anderson, J. 1979. A theoretical foundation for the gravity equation. *American Economic Review* 69, 106–16.

Anderson, J. and van Wincoop, E. 2004. Trade costs. *Journal of Economic Literature* 42, 691–751.

Combes, P.-P., Mayer, T. and Thisse, J.-T. 2006. *Economic Geography*. Princeton: Princeton University Press.

Disdier, A.C. and Head, K. 2006. The puzzling persistence of the distance effect on bilateral trade. Mimeo, University of British Columbia.

Feenstra, R.C. 2004. *Advanced International Trade*. Princeton: Princeton University Press.

Harris, C. 1954. The market as a factor in the localization of industry in the United States. *Annals of the Association of American Geographers* 64, 315–48.

Helpman E., Melitz, M. and Rubinstein, Y. 2006. Trading partners and trading volumes. Mimeo, Harvard University.

McCallum, J. 1995. National borders matter: Canada-US regional trade patterns. *American Economic Review* 85, 615–23.

Peri, G. 2005. Determinants of knowledge flows and their effects on innovation. *Review of Economics and Statistics* 87, 308–22.

Portes, R. and Rey, H. 2005. The determinants of cross-border equity flows. *Journal of International Economics* 65, 269–96.

Ravenstein, E.G. 1885. The laws of migration. *Journal of the Royal Statistical Society* 48, 167–227.

Redding, S. and Venables, A. 2004. Economic geography and international inequality. *Journal of International Economics* 62, 53–82.

Reilly, W.J. 1931. *The Law of Retail Gravitation*. New York: Pilsbury.

Santos Silva, J. and Tenreyro, S. 2006. The log of gravity. *Review of Economics and Statistics* (forthcoming).

Stewart, J. Q. 1947. Suggested principles of ‘social physics’. *Science* 106, 179–80.

Tinbergen, J. 1962. *Shaping the World Economy: Suggestions for an International Economic Policy*. New York: Twentieth Century Fund.

Index terms

borders
business networks
comparative advantage
distance
equity flows
fixed-effects estimation
foreign direct investment
gravity models
heteroskedasticity bias
imperfect competition
information costs
market access
market potential
maximum likelihood
microfoundations
migration
monopolistic competition
nonlinear estimation
patents
preference bias
proximity
regional and preferential trade agreements
reverse causality
selection bias
social interaction
social networks
spatial interaction
trade
trade costs
transport costs