THE “NEW” ECONOMIC GEOGRAPHY: WHERE ARE WE?

Paul Krugman

Nov. 26, 2004

Importance is an eigenvector. Really. The search engine Google uses links among websites to rank sites by importance, through a seemingly circular process in which a site’s importance is determined by the number of links it receives from other sites, weighted by their importance. The process bears a strong formal resemblance to the process that generates geographical concentrations of economic activity in some of the models that Masa Fujita, Tony Venables, and I have worked on; in both cases what emerges is the eigenvector with the largest eigenvalue. And anyone who uses Google routinely knows that the eigenvector embodies the truth: Google is almost always right about what is important.

I mention this because Google has a new service, which lets one search scholarly work the way one searches the web; citations play the same role that links play in regular Google. When I learned about the service, I did what any academic would do, and Googled myself. And it turns out that the work I did on economic geography is, in the judgment of the eigenvector, the most important of my scholarly career. That’s good to know – especially because of the fields I have worked in, geography is my personal favorite, and the collaboration with Masa and Tony among the most fruitful.

But where does the field stand, almost 15 years after I began working in it? Where is the “new” economic geography today? For this conference, I want to provide an overview of the state of empirical play.
Four propositions

The new economic geography began with simple, stylized models designed for tractability rather than realism; in Fujita, Krugman, and Venables we described our approach as depending on “Dixit-Stiglitz” – a highly unrealistic but tractable model of increasing returns and imperfect competition – “icebergs” – a highly unrealistic but tractable way to incorporate transportation costs – “evolution” – an ad hoc but tractable way to think about dynamics – and “the computer”, because even with all this sacrifice of realism to tractability the models weren’t all that tractable, and numerical examples were an essential guide to theorizing. But the specifics of the models weren’t the point; what we and the many other theorists who have entered this field were really trying to do was find a way to clarify a world-view about how economic interactions over space work. That world view rested on four propositions:

1. Transportation costs, or more broadly transaction costs across distance, play a crucial role in shaping international and interregional trade. In contrast to traditional trade theory, and even traditional urban economics, we argued that distance matters.

2. The interaction of market size with increasing returns plays an important role in determining the location of production. That is, we argued that some kind of home-market effect, as opposed to localized resources or more amorphous externalities, was at least one major explanation both of differences in population density and localized specialization.
3. A cumulative process in which large markets attract production, which increases the size of markets, leads to agglomeration – and possibly to multiple equilibria. Much of the excitement surrounding the new economic geography came from its suggestion that historical accident might play a major role in shaping the location of production, and also that cities and regions might be subject to discontinuous change.

4. The same processes that shape economic geography within countries also shape international trade. We hoped to finally justify Ohlin’s claim that international trade theory is simply international location theory – or my version from 1990, that I was like Moliere’s character who was startled to learn that he had been speaking prose all his life; I thought I had learned that I had been doing economic geography all my years as a trade theorist.

It was an exciting and inspiring world view, but was it right? We now have quite a lot of empirical evidence. So let’s look at the propositions.

*Distance matters*

One of the major areas of empirical research in international economics over the past 15 years has been the estimation of “gravity models” of trade flows. Such models are mainly useful as a way of identifying anomalies: if you want to know whether a common currency, or a common language, or a trade agreement really increase trade and if so how much, you estimate a gravity model and look for trade in excess of what that model
predicts; if you believe that restrictions deter trade, you look for trade falling short of what the model predicts.

One side result of these models, however, has been the confirmation that distance matters. Despite all the advances in transportation and communication, trade between any two countries drops off dramatically as the distance between these countries increases.

Figures 1, 2, and 3 offer an illustration of just how powerful distance remains as a deterrent to trade. Figure 1 shows the income aspect of the gravity relationship at work: it shows how US trade with EU members depends on the size each country’s GDP. GDP is measured as a share of the EU total; trade, the sum of US exports to and imports from each country, is measured as a share of total US trade with the EU. As you can see, a simple 45-degree relationship, as implied by a simple gravity model, fits pretty well. And the anomalies are, as usual, illuminating: Ireland’s special role as host to US-based multinationals shows up as a big trade share given its GDP, while the roles of Rotterdam and Antwerp as Europe’s first and second ports are reflected in the trade shares of the Netherlands and Belgium.

But now take the same data, and add America’s neighbors, and you get Figure 2. Canada and Mexico do far more trade with the US than more distant countries of equal economic size. In fact, Canada, with a Spain-sized economy only 8 percent the size of the EU, does as much trade with the US as the whole of the EU. I can’t show it here, but this outsized trade relationship predates NAFTA and the various US-Canada trade agreements; mainly, we’re talking about the role of distance.

Finally, Figure 3 shows an illuminating set of data on the trade of the Canadian province of British Columbia. It shows BC’s trade with a set of other Canadian provinces,
and also with a set of US states, in each case as a share of the province or state’s GDP.

Within each sub-sample, there’s a clear negative relationship. There’s also a striking
difference in trade within Canada and trade across the border; I’ll come back to that later.

Empirical estimates of the effect of distance on trade volumes typically find an
elasticity of trade with respect to distance of between -0.7 and -1. (That’s consistent with
the data in Figure 3.) It’s hard to justify such a large effect in terms of literal
transportation costs – so much for icebergs – leading many trade economists to talk
loosely about personal contact across space. Anyway, it’s clear that the role of distance in
deterring trade, central to the new economic geography but ignored in most standard
trade theory, is strongly confirmed by the data.

One further note: the importance of distance, and the success of gravity models,
explains one of the noticeable features of recent trends in world trade: the rapid growth of
trade among East Asian economies, and particularly Asian trade with China. Rapid
growth in any economy is, gravity tells us, the source of rapid growth in its trade; if two
economies grow fast, their mutual trade will grow very fast. And if they are relatively
close geographically, their mutual trade will quickly become a major part of world trade,
whether or not there are any special affinities or links.

*Increasing returns and home-market effects*

When production is subject to increasing returns, there is an incentive to concentrate
that production in only a small number of locations. When there are transport costs, there
is an incentive to choose locations that are close to large markets, and service smaller
markets at long distance. Hence the home-market effect: countries and/or regions should, other things equal, tend to export goods subject to increasing returns for which they have large domestic/local demand.

It’s a compelling story, but does it work in practice? Although I first wrote up that story in 1980, and it plays a crucial role in the geography work that began around 1990, until recently there wasn’t much evidence that the home market effect makes any difference in world trade. It’s gratifying to note that this has changed. A growing empirical literature finds evidence that home market effects do indeed play a role in shaping the location of production. Papers include Davis and Weinstein (1999, 2002), Feenstra, Markusen, and Rose (1998), Hanson and Xiang (2002), Trionfetti (2001). I won’t try to summarize this literature here, except to say that it relies mainly on somewhat indirect indicators. Nonetheless, it looks as if the home market concept has been vindicated.

*Cumulative processes and multiple equilibria*

The possibility of multiple equilibria, path dependence, a crucial role for history, etc.. is part of what makes economic geography an exciting and appealing field. At the level of individual industrial localizations, the evidence for path dependence is overwhelming in the sense that locations are evidently arbitrary to some degree, and the historical accidents that gave rise to particular concentrations can easily be traced. But how important is the phenomenon at the level of regional or urban agglomerations?

The answer so far is that the evidence is both scarce and contradictory. I was very impressed by two recent studies that exploited the drastic events of World War II to
search for evidence one way or the other. Unfortunately, they came up with opposite results.

Davis and Weinstein (2001) looked at the effects of the U.S. bombing campaign against Japanese cities – a horrifying story, but also a natural experiment. What they found was that the extent of damage, which varied widely among cities, had no effect on a city’s population once the postwar recovery was fully achieved – prewar population predicted postwar population, regardless of how severe the damage in between. As they say, this means that there is no evidence of the persistent effects of temporary shocks we should have expected if multiple equilibria were widespread.

On the other hand, Rhode (2003) looks at the temporary Pacific Coast boom induced by the war in the United States, and finds evidence that it kicked California and the west into a higher-level equilibrium that persisted after the war.

With only two studies along these lines, it’s premature to draw large conclusions. The Davis-Weinstein results are consistent with one feature of new economic geography models that becomes apparent when one works with numerical examples: the range of multiple equilibria narrows drastically given even small natural advantages to particular locations. To the extent that Japanese cities are or were mainly in locations determined by natural harbors (or rail lines reconstructed after the war), their stability should not be too surprising.

On the other hand, it’s not hard in models of regional development to get situations in which a growing region spends a substantial time with the potential for much larger population, but with that potential unrealized; that’s pretty much what Rhode is
suggesting. Maybe the rise of California was inevitable, but WWII accelerated its advent by a decade or two.

This needs more work; we should look for more natural experiments, though we should also hope that no more ones like the experiment exploited in these papers arise.

*Interregional versus international trade*

When I first began writing about geography, I made a point of emphasizing how artificial national boundaries are: any major city in the EU is closer to any other major city than NY is to LA, and Canada, as I liked to point out, is closer to the US than it is to itself. At times I came close to asserting that borders are irrelevant, and that we should think about international trade in a borderless framework. But I never did write that explicitly, and that’s a good thing. For it turns out that borders still do matter, a lot.

The best evidence comes from Canada, which collects data on interprovincial trade – the basis of Figure 3. As you can see from the figure, Canadian provinces do much more trade with each other than they do with comparably situated US states. A rough scan of that figure suggests that the border deters trade about as much as the distance between Montana and Ohio; a number of papers have suggested that the border is, in effect, a 1500-2500 mile wide moat.

Why that should be the case, especially when the US-Canada border is as innocuous as a border can be (although the madness now sweeping my country may change that) is an unsettled question. But it’s clear that for now, anyway, countries remain relevant economic units.
The end of distance?

We hear a great deal about globalization, and about new technologies that allow long-distance communication and trade in ways that were previously impossible. Everyone knows that the technical assistance number you call from the US or the UK may put you in touch with someone in Bangalore. But is a shrinking world changing the rules of economic geography?

Let me start by setting up a straw man. A simplistic view would be that the progress of transportation and communications technology is steadily abolishing the tyranny of distance. Once upon a time, the story would go, most economic interaction was local, with neighboring regions or nations. Now, because air travel brings all the world within a few hours, and the Internet links everywhere instantaneously, regions 10,000 miles apart are as likely to trade as regions a few hundred miles apart.

One does hear versions of this straw man story quite often. But the facts, if anything, go the other way. Despite the evident decline in transportation and communication costs, international trade has become more localized rather than more globalized.

The most striking evidence comes from the UK; Figure 4 shows the direction of UK exports in 1910 and in 1996. Long-distance trade, as measured by trade not with other European nations, has diminished rather than increased in relative importance. To some extent this reflects a sharp increase in trade with Europe as a share of GDP, but non-European trade declined even as a share of GDP.
How is this possible? In a simple differentiated-products model of world trade, the type of model often used to rationalize a gravity equation, a decline in transportation costs should lead to a rise in long-distance trade and a decline in the most local trade. Obviously that model isn’t good enough to fit what we actually see.

If we introduce some realistic complications into the model, however, it may be possible to make sense of the data. These complications also help us understand the difference, emphasized by Baldwin and Martin (1999), between the first wave of globalization – based on steam power and the telegraph – and the second wave, which took place after World War II and gained pace in the last 25 years.

Here’s the hypothesis: in the first wave of globalization, transport costs declined enough to make large-scale trade possible – but only where there remained fairly strong comparative advantage, giving rise to substantial production cost differentials. As it happened, big comparative advantage differences tended to be associated with long distances, because regions with very different climates and land abundance from those of Western Europe tended to be a long way away from Western Europe. So long-distance trade between resource-based economies and advanced countries with a strong comparative advantage in manufacturing flourished: English machinery was traded for Indian tea, Argentine beef, and Australian wool. But the costs of even local trade remained high enough to discourage intra-industry trade between countries with similar factor endowments and technological bases, even if they were close to each other.

In the second wave of globalization, transport costs have fallen low enough that small differences in products and tastes fuel trade between similar countries and regions. This
meant that from the 1950s on, advanced countries began taking in each others’ washing on a large scale – hence the rise both of intra-industry trade and intra-regional trade.

But it has also meant that new forms of long-distance trade have emerged. So far, the most important of these is the rise of developing-country manufacturing exports. Figure 5 shows the shares of manufactures and agricultural goods in developing-country exports over the past 40 years; there has been an almost complete role-reversal. Long-distance trade continues to be based on strong comparative advantage, but the source of the comparative advantage of developing countries with large exports has shifted: where once it was based largely on tropical climate, now it’s based largely on abundant labor.

Why has this happened since the 1970s, when it didn’t happen before? A best guess is that it has a lot to do with the intangible costs of making economic transactions at a distance. In 1960, wages in developing Asia were already very low compared with wages in North America or Europe, but the difficulty of coordinating production at such long distances made exporting manufactures from that region unprofitable despite that cost differential. Long-distance trade was restricted mainly to standardized commodities where the control issues mattered much less. With the falling cost of telecommunications, the fax machine, and so on, all this changed. That, at least, is the hypothesis.

The current version of this story is, of course, the rise of long-range trade in information services. It has long been possible to employ well-educated, English-speaking Indians for much less than the wages of equivalent workers in Britain or the US. Only recently, however, has it been possible to put those Indians on the other end of a 1-800 phone call.
The shift of developing country exports toward manufactures (and now services) may be having an influence on the internal economic geography of advanced nations – one reason why agglomeration economies may be weakening.

*Are centripetal forces weakening?*

There are some reasons to believe that the centripetal forces emphasized by the new economic geography – forward and backward linkages driven by the interaction of increasing returns and transport costs - actually had their peak influence some time ago, and are weakening in the 21st-century economy. Kim and Margo (2003), in their survey of economic geography in the U.S., generally seem to find that measures of regional specialization as well as disparities in income peak in the early 20th century, if not before. This is consistent with models that suggest that declining transport costs should have a U-shaped impact on agglomeration: as costs fall, they first make agglomeration possible, then make it unnecessary.

There may now be a globalization component to this weakening of intranational centripetal forces. Livas-Elizonda and I (1996) suggested, inspired by the case of Mexico after trade liberalization, that increased access to foreign markets might weaken core-periphery patterns within developing countries. Recently Tomiura (2003) has provided evidence that increasing import penetration is weakening industrial concentration within Japan.

*Where geography is taking us*
A dozen years ago, it seemed to some of us that we were facing a stark choice of world visions. One vision was the traditional vision of international trade theory, in which countries are discrete economic points, whose location in space is irrelevant. Another was the pure geography vision, in which location in space is all and borders are irrelevant. Finally, there was the vision of a spaceless, borderless world in which distance had been abolished – not a world that yet exists, but possibly one just over the horizon.

What seems to have emerged from the empirical work of the past dozen years is a compromise vision. Distance matters a lot, though possibly less than it did before modern telecommunications. Borders also matter a lot, though possibly less than they did before free trade agreements. The spaceless, borderless world is still a Platonic ideal, a long way from coming into existence.

The compromise vision isn’t as radical as some would like. But it’s a significant change from the way most of us viewed the world economy not too long ago.


