

Shifts in Economic Geography and Their Causes

Anthony J. Venables

Recent decades have seen momentous changes in the economic geography of the world. Political transitions and economic liberalization have brought formerly closed countries into the world economy. In Richard Freeman's phrase, this has amounted to a "doubling of the world labor force." The collapse of communism in the Soviet bloc brings 260 million workers into the world labor force; the opening up of China adds a further 760 million; and Indian liberalization adds 440 million. At the same time, technological change has continued to reduce the cost of interactions within and between countries. Any product that can be digitized can now be shipped at (almost) zero cost. Airfreight accounts for a third of U.S. imports by value, and 25 percent of African exports are now airfreighted. For the United States, the value of time saved by airfreight and containerization has been estimated as some 12 percent to 13 percent of the value of goods shipped (Hummels, 2000).

Some of these changes—falling trade and communications costs—have been going on for centuries, although interrupted by periods of increasing tariffs and closure of economies. Some have possibly come to an end. Technical change in shipping is no longer faster than technical change in goods shipped, so freight rates relative to shipment value are no longer falling. Transporting goods is oil- and security-intensive, two inputs with costs that are likely to increase in coming

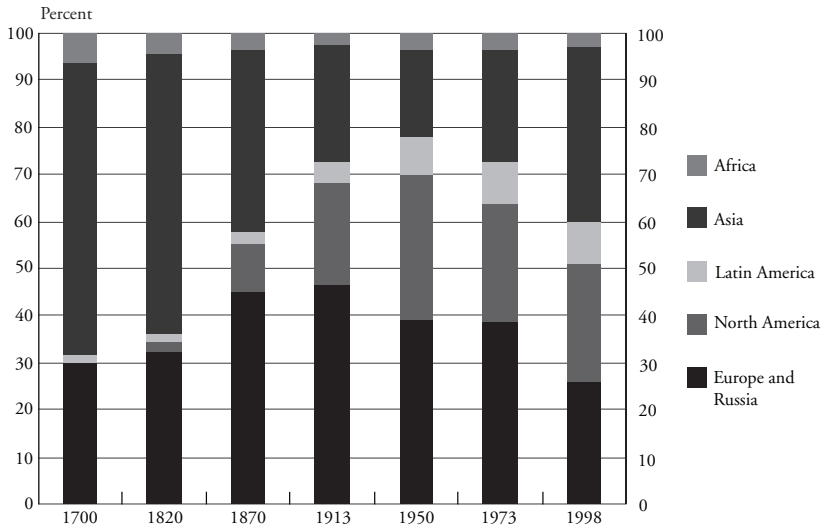
years. Digital transmission costs can fall no lower than zero, although no doubt the quality and range of electronic communication will continue to improve. Accompanying technological innovation has been business innovation. Multinational firms have expanded rapidly, with foreign direct investment (FDI) growing at approximately twice the rate of world trade, which itself has grown at twice the rate of world income. New forms of trade have emerged with the growth of outsourcing and production networks.

Accompanying these changes in spatial interactions have been changes in the location of activity. The historical record is illustrated in Chart 1, based on Maddison (2001), and giving shares of world gross domestic product (GDP) accruing to different regions. Although highly aggregate, the figure indicates four phases: the initial dominance of Asia, followed by the rapid growth of Europe during and after the Industrial Revolution, then the subsequent rise of North America, and now the resurgence of Asia, accelerating during recent decades. Part of the change is because of population, but the much larger part is because of changes in per capita income—the “great divergence,” which saw the ratio of per capita incomes of the richest to poorest nations increased from around 8:1 in 1870 to more than 50:1 in 2000.

Economic geography has changed at all spatial scales—not just the aggregate regions of Chart 1, but also within regions and countries. The most important of these subnational changes is urbanization. A majority of human beings are now urbanized, and China alone expects to see a doubling of its urban population to nearly 1 billion people by 2030. Population movements are not just rural to urban. They are often associated also with movement to booming coastal regions, and people leaving lagging regions in the interior or more remote areas of countries.

This sketch describes some of the forces driving change in the world economy, and some of the ensuing changes in economic geography. It also challenges our understanding of the location of economic activity and of the determinants of changes in the pattern of location.

Chart 1
Shares of Gross Domestic Product



Source: A. Maddison (2001)

The first question is, Why are economic activity and prosperity spread so unevenly? Is an American really 50 or 100 times more productive than an Ethiopian? Even within the United Kingdom, why are the earnings of a Londoner 70 percent higher than those of someone from Stoke-on-Trent (and 40 percent higher after controlling for both education and occupational mix)? Standard “neoclassical” economic theory suggests that while differences may arise as some countries or regions gain initial advantage, they should be rapidly arbitrated away. Capital will flow to where labor is cheap, and knowledge and new technologies will be transferred. Fundamentally, if there are diminishing returns to economic activities, then there will be a continuing process of convergence and a tendency for activity to be spread relatively uniformly across space. Yet these forces seem to operate in a manner that is, at best, selective. Even within reasonably well-functioning market economies, spatial disparities persist, and, in the wider world economy, they are astonishingly large.

The second question is, Does increasing trade—or spatial interaction, more generally—necessarily narrow these differences? The standard answer of international economics is that price and income differentials will be narrowed, although production structures may diverge. The most extreme form of the statement is the “factor price equalization” theorem of the Heckscher-Ohlin trade theory, giving conditions under which free trade in goods brings about complete equalization of factor prices between countries (even though these countries may be different in other important respects). This is a theorem that fails conspicuously in the cross-section of countries. And, in the time series, the great divergence of per capita incomes was taking place just as the costs of shipping were declining precipitously and world trade was going from 1 percent of GDP in 1790 to 8 percent in 1913 and 9 percent in 1929. Evidently, there have been episodes where falling transport costs and opening to trade have been associated with increasing rather than narrowing factor price differences between countries. We have to understand circumstances under which each of these possibilities may occur.

If we can develop satisfactory answers to these two questions, then we can address the third question: How should we think about future developments, for both developed and developing regions? We are currently in an era of globalization in which some economic activity is dispersing from existing centers, but what determines what sectors move and where the sectors go? Will prosperity be widely dispersed, or will some regions of the world continue to be left behind?

Answering these questions requires that we have a general theory of the location of economic activity, but this is an area that is not served well by mainstream economic theory. This theory is generally aspatial; the core theory of international trade treats countries as points with no spatial structure either between or within them. Growth theory is founded in closed economy models, and its application in the growth and trade literature has the same aspatial structure as trade theory. Yet it is clear that spatial relationships remain important. There is a spatial clustering of rich countries and of poor countries. Trade costs remain

high—looking at bilateral trade flows between countries, the median value of the CIF/FOB (cost, insurance, freight/free on board) ratio is 1.28; that is to say, trade costs nearly 30 percent of the value of goods shipped. Gravity models of trade tell us that the elasticity of trade with respect to distance is in the range -0.9 to -1.5 ; this means that an 8,000-kilometer distance chokes off more than 90 percent of the trade that would be observed over a 1,000-kilometer distance. Similar distance elasticities hold for other economic interactions, such as equity holdings, FDI, and technology transfer.

In addition to being aspatial, mainstream economic theory generally predicts convergence of performance. Differences between regions arise because of exogenous factors, and economic forces generally erode them, or at least do not amplify them. In trade theory, differences between countries—differences in endowments, institutions, or technology—are assumed, and analysis draws out implications for trade, the structure of production, and factor prices. In growth theory, technological leadership or institutional differences may create disparities between countries, but economic forces erode these disparities through a more-or-less steady process of convergence. Paul Krugman (1995) gave a powerful analogy with geology prior to the discovery of ocean spread and plate tectonics: The subject had a lot to say about the erosion of mountains, but nothing to say about the reason why they existed in the first place.

“New economic geography” (perhaps better called “geographical economics”) has worked toward addressing these deficiencies.¹ The objective is to offer an integrated theory of location, capable of explaining divergence as well as convergence in economic performance. The key building block is the recognition that proximity is good for productivity; dense configurations of economic activity work better than sparse or fragmented ones. Mobile factors—firms and possibly workers—will locate in order to take advantage of higher productivity, and this creates a positive feedback. Firms and workers go where productivity is high and, by so doing, tend to further raise productivity, creating an uneven distribution of activity and spatial income disparities.

There are several analytical challenges here. One is to have sound micro-based theories and evidence as to why proximity is good for productivity. Merely positing some form of spatial increasing returns to scale through some “black-box spillovers” is deeply unsatisfactory and essentially useless for policy. In this paper, we will devote some time to outlining what these microeconomic mechanisms are. The second challenge is to place the proximity-productivity relationship in a wider economic model and thereby identify the tradeoffs between forces for concentration or dispersion of economic activity. This will give a full theory of the location of activity. Once this is done, hypotheses can be formed about circumstances in which activities may concentrate or disperse and the associated shifts in economic geography.

While our focus in answering the questions outlined above is new economic geography, we do not want to claim primacy for geography over other factors determining prosperity. Growth takes hold in a country for a wide variety of (poorly understood) reasons. This paper is not going to resolve the relative importance of endowments, institutions, technology, or geography (particularly since most of them are, in any case, endogenous variables). Sound policy is essential for growth, and it is clear that good location is neither necessary (New Zealand) nor sufficient (Albania) for prosperity. The point is simply that a new economic geography perspective provides a number of additional insights of existing patterns of activity and the forces driving future changes.

The paper proceeds as follows. The next three sections discuss three key propositions. The first is that proximity to other economic agents—workers, consumers, and firms—is good for productivity (first section). The second is that large income disparities are a perfectly natural outcome of a world in which proximity matters (second section), and the third is that the effects of increased trade are potentially ambiguous—there are circumstances in which cheaper spatial interactions cause inequality, not convergence (third section). With these building blocks in place, the paper then turns to a discussion of forces shifting the present economic geography of the world and argues that this is taking the form of a “lumpy dispersion” of activity.

Proximity and productivity

The first proposition is that proximity to other people is good for productivity. What is the evidence, and what are the economic mechanisms that drive the effect? There is a long list of mechanisms which we organize under two headings: product markets and labor markets. The headings correspond loosely to different spatial scales; some of the product market effects might operate over long distances, while labor market effects are short-range, even coming down to the benefits of face-to-face contact.

Product markets

The most immediate effect of proximity is that it saves transport and other trade costs. Thus, if two producers have identical physical productivity, the one producing in the large market will have higher productivity in value terms because it does not have to bear the transport costs of shipping the good to remote consumers. This producer will also have lower-cost intermediate inputs, not having to absorb its share of shipping costs on these inputs. The combined effect of transport costs squeezing the producer's value added both from above and below (in other words, on final and intermediate products) can have quantitatively large effects on measured productivity.

Trade costs, of course, should be thought of in much more general terms than just freight charges. Time in transit is costly (as much as 0.5 percent of the value of goods shipped per day, according to work by Hummels, 2000). This high cost of time in transit comes partly from the costs of carrying stock, and also from the likelihood that long transit times reduce the reliability and predictability of deliveries. It also makes firms slower to respond to changing demand conditions or costs levels, and this by itself can be a force for the clustering of activities (Harrigan and Venables, 2006).

Transport and trade cost savings are a direct benefit of proximity, but its full economic impact comes from economies of scale associated with operating in an area of dense economic activity—close to consumers,

workers, and other firms. In a small or fragmented market, there is a tradeoff between having firms large enough to achieve economies of scale without becoming monopolists. A large or integrated market shifts this tradeoff, allowing benefits of both large-scale and more intense competition, and, as a consequence, firms will be larger, operating at lower average cost and setting lower prices. The more intense competition will weed out less efficient firms, concentrating production in efficient firms. A larger market will also support a greater variety of products. These price and variety effects benefit consumers and also, if the goods are intermediates, benefit firms in downstream sectors. For example, a larger market will support a greater variety of specialized input producers, tailoring their products to the needs of other firms. Notice that some of these effects are internal to firms, but most end up as pecuniary externalities—it is consumers who benefit from the lower costs and prices of firms' output.

Labor markets

In addition to efficiency gains deriving from the goods market, there are gains from operating in a large labor market. The larger the pool of workers that a firm can access, the more likely it is to be able to find the exact skills that suit its needs. It is sometimes argued that the fundamental “capabilities” of a firm are embodied in the workforce of the firm or of the local labor market (Sutton, 2002). If firms are subject to idiosyncratic shocks, then a larger labor market will expose workers to less risk, increasing the probability of reemployment if they are made redundant. More importantly, a large labor market will increase the incentives for workers to undertake training. This argument, like some of those we saw above, turns on increased intensity of competition. In a small market, workers who acquire specialist skills may be “held-up” by monopsonistic employers, in which case, there is no incentive for them to invest in skills. The presence of a large number of potential employers removes this threat of opportunistic behavior, and thereby increases training incentives (Matouschek and Robert-Nicoud, 2005).

A further set of arguments has to do with communication between workers. In many activities, face-to-face communication is extremely important. Face-to-face contact enables higher frequency interchange of ideas than is possible by e-mail, phone, or videoconference. Brainstorming is hard to do without the ability to interrupt and use parallel means of communication—oral, visual, and body language. Face-to-face is also important for building trust, once again by observing the body language and a wide range of other characteristics of one's interlocutor. By breaking down anonymity, face-to-face contact enables networks of the most productive workers to develop and promotes partnerships and joint projects between these workers. All these considerations are productivity-enhancing.

The final set of arguments is something of a catchall residual category. Knowledge spillovers are easier between proximate firms than remote ones. The mechanism may be labor mobility, face-to-face social contact between workers, or observation of the practices of other firms. Such effects are particularly important in innovation-intensive activities, and a large literature points to the spatial concentration of innovative activities (for example, Audretsch and Feldman, 2005). Location-specific knowledge spillovers also arise as firms learn about the characteristics of their location, and this knowledge spills over to other firms, as in the “self-discovery” story of Rodrik and Hausmann (2005). This may be learning about real economic characteristics of locations, or may just be a “herding” story, as firms simply choose to copy the location decisions of other (successful) firms. All of these knowledge spillover effects are probably best summarized in Alfred Marshall's statement that “the mysteries of the trade become no mystery; but are, as it were, in the air” (1890).

Short or long range?

The various proximity effects that we have outlined above operate over quite different spatial scales. The product market effects can be long-range—firms in New York may benefit from a large market in

California, and reductions in international shipping costs will increase market access for exporting firms. Much of European economic integration has been motivated by attempts to reap these gains. On the other hand, the labor market effects we have described operate within a much narrower area—indeed, some work suggests that 45 minutes driving time is the appropriate range for these effects (Rice, Venables, and Patachini, 2006). The literature on knowledge spillovers in innovation also suggests rapid spatial attenuation. For example, Jaffe, Henderson, and Trajtenberg (1993) establish that citations in patent applications are strongly skewed toward the same or neighboring metropolitan areas. We will argue that effects at all sorts of spatial scales are important in understanding shifts in economic geography.

Sectoral or aggregate?

Just as the effects we have described operate across different spatial scales, so they also operate across different sectoral scales. Some effects are driven by aggregate demand. Proximity to a large mass of consumers will cut trade costs and raise demand for all firms whose sales, direct or indirect, are concentrated in the area. All such firms will appear to have higher productivity near centers of high demand. Other effects are narrowly sector-specific; for example, a film actor benefits from proximity to a film producer, although does not care much about aggregate demand in Los Angeles.

The financial sector provides a good example of these varying sectoral and spatial scales. Within the financial sector, some backroom activities can be easily separated from the rest of the firm and operated from low-cost locations. Other parts of the business require proximity to final consumers; retail banking; and, perhaps—as suggested by recent controversy—call center operations. At the same time, the most skill-intensive parts of the sector are spectacularly prone to clustering, valuing face-to-face contacts, access to thick labor markets, and a dense network of firms offering complementary services. Similarly, within manufacturing, some stages of the production process can be outsourced and moved to low-cost locations. For other parts, this is not possible, partly because of disintegration costs within the firm (breaking the production flow,

transport costs, and so on) and partly because of the loss of proximity to complementary inputs, skilled labor markets, or the firm's consumers.

Evidence

There is a wide range of sources of evidence for these effects, and here we illustrate some examples. The most extensively researched source of evidence for the claim that proximity is good for productivity is from studies of the productivity of cities (or, more generally, areas of dense economic activity). A recent survey of the literature (Rosenthal and Strange, 2004) reports a consensus view that, over a wide range of city sizes, doubling city size is associated with a productivity increase of some 3 percent to 8 percent. This is a large effect—moving from a city of 50,000 inhabitants to one of 5 million is predicted to increase productivity by more than 50 percent. A good deal of work has used firm-level data to investigate the magnitude and determinants of spatial productivity variations. This work is surveyed by Rosenthal and Strange (2004) and typically finds significant effects, generally being larger in higher technology sectors of activity.

In the international context, proximity manifests itself in large trade flows—we have already noted the relationship between trade and distance. The gains from trade are widely (if somewhat controversially) documented, with perhaps the most widely accepted evidence being that of Frankel and Romer (1999), who find that a 1 percentage point increase in the share of exports in GDP raises per capita income by up to 1 percent. Because halving distance doubles trade flows, this suggests large proximity-productivity effects. Redding and Venables (2004) focus on measuring countries' access to markets and sources of supply and find that a 1 percent improvement in a country's market access—which has the effect of increasing its exports by 1 percent—raises per capita income by around 0.25 percent.

Equilibrium disparities

The second proposition is that large spatial disparities in income can be a persistent "equilibrium" outcome. To establish this, the arguments of the preceding section need to be combined with other forces to give

a theory of the location of economic activity and consequent wage and income differentials. The best way to do this is by thinking about the profitability of a firm that is choosing between various production sites. How do the potential profits of such a firm vary across alternative locations? They depend on three elements: one, productivity, the determinants of which we have discussed above, and defined broadly to include the benefits of transport cost saving; two, product market competition—in other words, the number of competitors that the firm will face in its chosen location; and three, input prices, including those of intermediate goods and primary factors. The equilibrium location of activity is the arrangement of firms that causes productivity levels, product market competition, and input prices to adjust until all firms are indifferent about their choice of location.

Now consider the following thought experiment: What happens to the profits of firms in a location when an additional firm establishes operations in the same location? If profits increase, then adding this firm increases the incentives for further firms to come, so there is an agglomeration process, with differences between locations becoming amplified. If profits fall, then activity will be dispersed and firms will tend to spread out. The productivity-proximity relationship is an amplification force because adding firms raises productivity and profits of existing firms. Product market competition and input prices are dispersion forces; adding another firm crowds the market, thus reducing revenue, and also bids up prices of immobile factors of production, raising costs.

Equilibrium location is therefore a balance between the productivity-proximity relationship, a force that amplifies initial differences, and product market competition and factor cost, forces which tend to dampen down effects. This is an important generalization of standard trade theory. If there is free trade, then standard theory determines the location of production from just factor costs and endowments, the third element in our list of the determinants of firms' profitability. If there are trade barriers, then supply and demand in national product markets also comes into effect, so location is determined also by product market competition. The new economic geography adds to these endogenous productivity differences. Let us now see how it operates in different contexts.

First nature geography and international wage differences

Consider first the implications of “exogenous” differences between countries. These may be institutional or policy differences or may be differences in natural geography. Geographers have a long-standing distinction between “first nature geography”—coasts, mountain ranges, natural endowments—and “second nature geography”—the geography of interactions between economic agents. The costs of bad first nature geography have been widely discussed by Sachs and his co-authors. Propensity to disease lowers productivity. Being landlocked raises transport costs—it has been estimated that the median landlocked country has transport costs 55 percent higher than the median coastal economy.

These are the direct disadvantages of bad geography, but what are the full equilibrium effects? Advantages and disadvantages of first nature geography become amplified as firms move into locations with good geography, and the proximity-productivity relationship causes further increases in productivity. Conversely, countries with bad first nature geography will have low levels of economic activity, reducing productivity further. Of course, the proximity-productivity relationship is not the only amplification mechanism. Jeff Sachs points to the endogeneity of institutions (Gallup and Sachs, 1999) and argues that the return to reform is lower in countries with bad geography, as such countries—even if they had good institutions—would remain handicapped.

Who bears the costs and benefits of these spatial variations in productivity? The answer is that they are borne entirely by immobile factors, meaning, in the international context, labor. Mobile factors have the same return (cost of living and risk adjusted) in different locations, so the impact falls entirely on that part of costs that is immobile. Since labor may be a small share of the costs of production, there can be a large multiplier effect here. If labor is 10 percent of gross costs, then a 50 percent difference in the productivity of all inputs will translate into a 500 percent wage difference. These large effects are confirmed by empirical work. Econometric work by Gallup and Sachs (1999) finds that 70 percent of cross-country variation in per capita income can be accounted for by just four measures of physical and economic geography: malaria, hydro-carbon endowment, coastal access, and transport costs.

Second nature geography and economic agglomeration

The relationship between proximity and productivity does not just amplify economic differences that arise because of other exogenous factors. If amplification effects are strong enough, then they can create disparities between locations that are identical in underlying characteristics. In the next section, we will draw out this argument in the international setting, in the context of changing trade costs. Here, we point to its importance as a driving force behind the existence of cities.

Cities are the most commonplace manifestation of the unevenness of economic activity. Yet a world with diminishing returns to activity would have no cities, as activity would be smeared across space. The apparatus outlined in this paper gives a very simple story for their existence.² The productivity-proximity relationship is a force for clustering all activity into a megacity. Pulling in the opposite direction are dispersion forces: product market competition, meaning that some firms remain dispersed in order to supply remote consumers, and high urban prices of immobile factors. If workers are free to migrate within a country, then the only immobile factor is land, the price of which is bid up, thus also raising urban wages as mobile workers are compensated for regional variations in the cost of living. Further dispersion forces may be provided by urban congestion and commuting costs. Notice that the dispersion forces are generally not sector-specific, although some of the agglomeration forces are because (as we saw in the first section) the proximity-productivity relationship can vary between sectors. This gives rise to sectorally specialized cities—London and Hollywood—the size of which depends on the importance of the sector in the world or regional economy.

Trade, location, and inequality

Our third proposition is that trade is not necessarily a force for convergence of incomes. We have seen from the historical record that 19th-century globalization was associated with substantial divergence of income between regions, and the impact of 21st-century globalization on international inequality remains hotly debated.

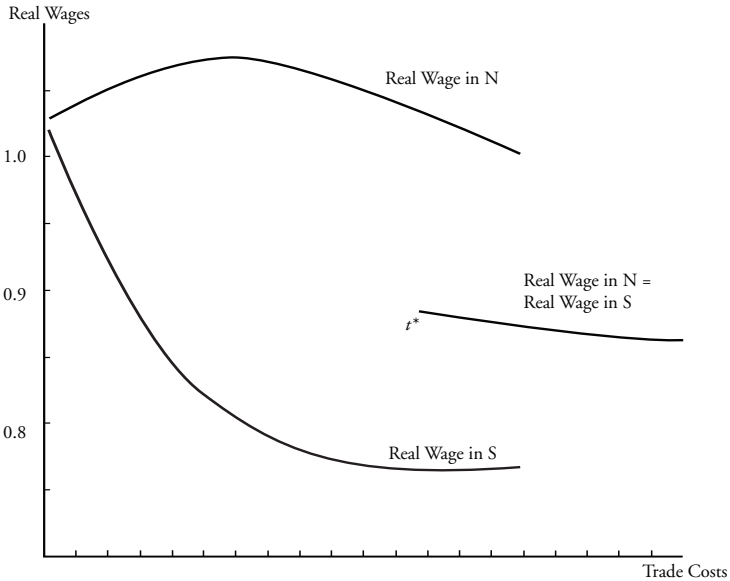
The interactions between trade and income divergence are quite complex, but the basic ideas can be developed in the context of a very stylized and aggregate model put forward by Krugman and Venables (1995). The model has just two countries, N and S, assumed to be *ex ante* identical, and with labor immobile between countries. Each country has an agricultural sector (shorthand for a sector that is tied to each country) and a monopolistically competitive manufacturing sector. Firms in the manufacturing sector produce final and intermediate goods and use labor and intermediates as inputs. The presence of intermediate goods creates a proximity-productivity relationship because the transport costs that firms incur on intermediate goods depend on their location relative to other firms.

Equilibrium outcomes are summarized in Chart 2, which has trade costs on the horizontal axis and real wages on the vertical axis. At very high trade costs, the economies have identical economic structures and identical real wages, as indicated. This is because when trade in goods is expensive, supply and demand in each country's product market (one of the dispersion forces of the second section) is the dominant force which determines the location of activity.

As trade costs fall (moving left on the chart), the possibility of supplying consumers through trade rather than local production develops, and the productivity-proximity relationship becomes relatively more important. Below some level of trade costs, t^* , these forces come to dominate, and one of the countries (N in the chart) gains most of the manufacturing and, consequently, has a high real wage. This clustering "deindustrializes" the other country (S), which experiences a fall in its real wage. The important point to note is that following this is a range of trade costs in which the world necessarily has a dichotomous structure. Wages are lower in S than in N, but it does not pay any firm to move to S, as to do so would be to forego the benefits of the productivity-proximity relationship arising from the large market and proximity to suppliers that are found in N.

In this model, the productivity-proximity relationship derives just from the costs of trading intermediate goods. This means that as trade costs fall further, the clustering force becomes weaker, and location comes to be determined by factor prices, a dispersion force. This is

Chart 2
Real Wages in a Two-Country Model



the era of globalization, in which manufacturing starts to move from N to S and the equilibrium wage gap narrows. Factor price equalization is attained when trade is perfectly free—the “death of distance.” Of course, the model goes all the way to factor price equalization simply because the only productivity-proximity mechanism captured in the model is trade in intermediates, a mechanism which is switched off once trade is free. If other mechanisms were included, then inequalities would persist.

Clearly, this model is highly stylized, but it illustrates the complex role of trade in determining the location of activity. Trade changes the balance between the dispersion forces of product and factor market competition and the clustering force of the productivity-proximity relationship. The model equips us with the apparatus to think through other shifts in economic geography, and it is to this that we now turn.

Lumpy dispersion

The preceding sections showed how we can think of world economic geography as a balance between concentration forces and dispersion forces. What light does this shed on thinking about future change in a globalizing world economy? We organize remarks under three headings. First, sectorally—what sort of activities are likely to remain concentrated, and which are likely to disperse? Second, by country—what might the cross-country pattern of location look like, and how will the international distribution of income change? And third, subnationally, concluding with a discussion of the growth of cities. The theme that will run through all three headings is that much activity will move out of existing centers, but relocation will be “lumpy,” benefiting some regions more than others and recoalescing into new patterns of agglomeration.

What sectors move?

What sectors are most likely to detach from existing centers of activity and relocate to lower-wage regions? One determinant is, of course, factor intensity. In line with standard trade and comparative advantage theory, unskilled-labor-intensive activities will tend to relocate to low-wage countries. It is helpful to extend this reasoning with a broader notion of comparative advantage. Comparative advantage theory states that specialization is determined by the interaction between country characteristics and commodity characteristics. In Heckscher-Ohlin, this is countries’ endowments of factors of production, interacting with commodities’ factor intensity. The principle generalizes, so to the list of country characteristics should be added various dimensions of institutional quality and business environment, these interacting with product characteristics. Thus, countries with good intellectual property protection will tend to attract sectors that value this protection, and so on.

Comparative advantage, even in this generalized form, is only part of the story. We also need to add some notion of the “linkage” intensity of the product. How easy is it to detach the activity from its

existing location, and how expensive is loss of proximity to related economic activities? This depends on all the elements of the productivity-proximity relationship that we outlined in the first section. If firms in a sector are highly dependent on a network of suppliers or on capabilities embodied in the local labor force, then it is unlikely that the sector will relocate. The strength of these interactions varies across sectors and, as we have seen, depends on the costs of transport and other spatial interactions. Thus, face-to-face contact may be crucial for some economic activities, but not for others. Skills may be embodied in the labor force and hard to transfer, or it may be very easy to train workers in a new location. Timely delivery is crucial for some goods, and there is evidence that production of fashion-sensitive garments has moved back to high-wage countries for this reason, despite the wage penalty incurred.

The profitability of relocation also depends on the extent to which the production process can be “fragmented” into different stages, with different factor endowments and different linkages to related activities. This is being studied in a rapidly expanding literature on fragmentation, production networks, outsourcing, and offshoring (for example, Grossman and Rossi-Hansberg, 2006). The argument is that globalization has created the possibility of a finer pattern of specialization, as it is now possible to do different parts of the production process in different countries. Component parts and semifinished goods can cross borders multiple times, and countries are able to engage in “vertical specialization,” producing just one very narrowly defined part of a product. Data on these activities is hard to obtain, as they do not correspond to the standard commodity classifications of trade. However, one of the fastest-growing elements of world trade has been trade in parts and components, now accounting for around 30 percent of world trade in manufactures (Yeats, 1998), and the share of imports in total inputs to U.S. goods-producing sectors has doubled to 18 percent over a 20-year period (Grossman and Rossi-Hansberg, 2006).

This is an area where a good deal more research is needed. From the point of view of high-income countries, is it possible to identify activities which are more or less likely to become detached? We need to develop a way of measuring whether a country has a deep or shallow

comparative advantage in a particular product or task. From the standpoint of developing countries, what are the sectors that are most footloose—apparel, electronic, assembly? And can we diagnose why a labor-intensive activity might not be willing to move? What exactly are the linkages that would be foregone in moving out of an established center of activity, and how easily can they be grown in a developing country?

A number of policy issues also exist. The productivity-proximity relationship creates a coordination failure, and so suggests a role for national industrial policy to act as a catalyst to overcome the failure. And internationally, it should influence the way we think about trade policy. For example, rules of origin are often thought of as minor technical details in trade agreements, but this may underestimate their importance. Both the United States and the European Union give duty-free access for imports of garments from low-income African countries, the United States under the African Growth and Opportunity Act (AGOA) and the European Union under Everything But Arms (EBA). How difficult is it to set up a garment industry to benefit from these rules? Probably not very, but what is hard to do is to set up a garment industry and simultaneously establish industries to supply textiles, yarn, fabric, and so forth. Relatively relaxed rules of origin under AGOA enabled garment firms to import these inputs from Asia, and led to substantial output and employment growth. More restrictive rules of origin under EBA mean that production of garments for the European Union market have stagnated at approximately zero.

Where will production go?

Turning to countries, what pattern of development is predicted? The fundamental point to come from the theory is that simultaneous development of similar countries is likely to be an unstable equilibrium. For example, suppose that activity is relocating from an established center into two similar emerging economies, and that productivity-proximity relationships operate in the sectors concerned. Then, whichever country gets slightly ahead will have higher productivity and become the more attractive location for

further investment, while the other country will fall behind. This observation has a number of implications.

First, we should expect growth and development to occur in sequence, not in parallel. Instead of all poor countries steadily converging to high-income status, as predicted in the world view of Lucas (2000), there is an inherent unevenness. Some countries will grow extremely fast as they transit from one “convergence club” to another, while other countries will be left out of the process. Which countries go first? Many factors count, including first nature geography and the institutional and policy environment. The models predict that economic development will spread out from existing centers, going to regions with low transport costs, such as the coastal regions of neighboring countries. This is a view of the world that fits well with recent growth patterns in Asia (as compared to Africa), and with the “flying geese” model of Asian development.

While this aggregate view is important, the phenomenon is seen even more sharply at the sectoral level. As sectors migrate from established centers of activity, their new location pattern turns out to exhibit clustering. A striking feature of growth has been the fact that many countries have done well in a few extremely narrow product segments, such as India’s success in software. Hausmann and Rodrik (2002) look in detail at the exports to the United States of Bangladesh, Pakistan, Honduras, Dominican Republic, Korea, and Taiwan, using data at the highly disaggregated six-digit level (for example, “hats and other headgear knitted or from textile material not in strips”). Even at this very fine level of disaggregation, there is a very high level of specialization. For each of these countries, the top four product lines account for more than 30 percent of exports to the United States, and there is little overlap in the top product lines of quite similar countries; only six product lines are in the top 25 for both of these countries. Bangladesh is successful in exporting shirts, trousers, and hats (but not bed linen or footballs), while Pakistan does well in bed linen and footballs. Hausmann and Rodrik conclude that “for all economies except possibly the most sophisticated, industrial success entails concentration in a relatively narrow range of high productivity activities.”

The story, then, is that sectors will relocate, but that this relocation will be “lumpy,” sectorally and in aggregate, with some countries being left out. A corollary of this is that small initial differences—the factors that first attract modern sector activity to a country—will generate large differences in outcomes. Once again, this points to the importance of policy. Bad policy environments can ensure that a country is left out. Creating a good business environment, institutionally and in terms of infrastructure provision, is essential. The role of proactive industrial policy remains intensely controversial (Rodrik, 2004).

Spatial concentration: Regions and cities

Finally, what shifts in economic geography are occurring within countries? High-income countries have an established city structure, but are nevertheless witnessing some changes. After decades of decline, cities are undergoing a renaissance as more knowledge-based activities seek to benefit from clustering.

The situation is more fluid in developing countries experiencing rapid structural change and migration. The finding that spatial inequality increases during development dates back at least to Williamson (1965) and has been confirmed in many studies since (for example, the studies of urban concentration by Shishido and Wheaton, 1982, and Henderson, 1999). This increase in spatial inequality often arises from spatial concentration in the development of manufacturing, and we see this most clearly in data for large countries. States in southern India have come to prominence in manufacturing, and, in Mexico, manufacturing has become highly concentrated in regions that border the United States, leading to large increases in spatial variation of per capita incomes. Increasing regional inequalities in China have been extensively documented.

While increasing spatial disparities are a problem for some developing countries, managing the process of urbanization is a problem for almost all of them. The number of cities in the world with a population of more than 1 million went from 115 in 1960 to 416 in 2000; for cities of more than 4 million, the increase was from 18 to 53; and

for more than 12 million, from 1 to 11 (Henderson and Wang, 2003). This indicates that, despite the massive diseconomies associated with developing-country megacities, there are even more powerful economies of scale making it worthwhile for firms to locate in these cities. This creates a major policy challenge. The view outlined in this paper is that the market system is riddled with imperfections—the clustering forces we have discussed are largely externalities, so the outcomes that they support are generally inefficient. Megacities may expand far beyond their efficient scale, but the clustering forces we have described make it difficult for new urban areas to compete and become established. There is a case for policy intervention to decentralize activity, but we remain woefully ignorant about what works and what doesn't.

Conclusions

There are many reasons for variation in the prosperity of countries and regions. Some factors are truly exogenous—first nature geography—and others are a function of political and institutional history. On top of these exogenous factors, we need to place a theory of the location of economic activity. International trade theory gets us part of the way, and the new economic geography approach broadens this out to capture (in a micro-founded and evidence-based way) endogenous variations in productivity. The approach offers an explanation of the emergence of disparities between countries and regions, and offers an explanation of their persistence. It suggests that even as globalization causes dispersion of activity, economic development will be in sequence, not in parallel; some countries will experience rapid growth while others will be left behind. At the microlevel, it points to the importance of overcoming coordination failures and threshold effects in growing new cities and in establishing new industries in developing economies.

Endnotes

¹For overviews, see Krugman (1991); Fujita, Krugman, and Venables (1998); or Henderson and Thisse (2004).

²Much of the urban economics literature simply assumes that economic activity takes place only in the “central business district” of cities. Endogenous modeling of productivity benefits of cities was developed by Henderson (1988), although initially with “black-box” spillovers between firms. See Duranton and Puga (2005) for fuller development of these ideas.

References

- Anderson, J., and E. van Wincoop. 2004. "Trade costs," *Journal of Economic Literature*, vol. 42, pp. 691-751.
- Duranton, G., and D. Puga. 2004. "Micro-foundations of Urban Agglomeration Economies," in V. Henderson and J. Thisse (eds.), *Handbook of Urban and Regional Economics*, vol. 4, North Holland, Amsterdam.
- Fujita, M., P.R. Krugman, and A.J. Venables. 1999. *The Spatial Economy: Cities, Regions and International Trade*. Cambridge, Mass.: MIT Press.
- Gallup, J.L., and J. Sachs. 1999. "Geography and Economic Development," in B. Pleskovic and J.E. Stiglitz (eds.), *Annual World Bank Conference on Development Economics, 1998*, World Bank, Washington, D.C.
- Grossman, G.M., and E. Rossi-Hansberg. 2006. "The Rise of Offshoring: It's Not Cloth for Wine Anymore," processed, Princeton.
- Harrigan, J., and A.J. Venables. 2006. "Timeliness and Agglomeration," *Journal of Urban Economics*, vol. 59, pp. 300-316.
- Hausmann, R., and D. Rodrik. 2003. "Economic Development as Self-Discovery," processed, Kennedy School, Harvard.
- Henderson, J.V. 1988. *Urban Development: Theory, Fact and Illusion*. Oxford University Press.
- Henderson, J.V., and J. Thisse. 2004. *Handbook of Urban and Regional Economics*, vol. 4, North Holland, Amsterdam.
- Henderson, J.V., and H.G. Wang. 2003. *Urbanization and City Growth*, processed, Brown University.
- Hummels, D. 2000. *Time as a Trade Barrier*, processed, Purdue
- Jaffe, A.B., R. Henderson, and M. Trajtenberg. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, vol. 108, pp. 577-598.
- Krugman, P.R. 1980. "Scale Economies, Product Differentiation and the Pattern of Trade," *American Economic Review*, vol. 70, pp. 950-959.
- _____. 1991. "Increasing Returns and Economic Geography," *Journal of Political Economy*, vol. 49, pp. 137-150.
- _____. 1995. *Development, Geography and Economic Theory*, Cambridge, Mass.: MIT Press.

- Krugman, P.R., and A.J. Venables. 1995. "Globalization and the Inequality of Nations," *Quarterly Journal of Economics*, vol. 110, pp. 857-880.
- Lucas, R.E. 2000. "Some Macroeconomics for the Twenty-First Century," *Journal of Economic Perspectives*, vol. 14, pp. 159-68.
- Maddison, A. 2001. "The World Economy: A Millennial Perspective," Organisation for Economic Co-operation and Development (OECD), Paris.
- Marshall, A. 1890. *Principles of Economics*. London: Macmillan (8th edition, 1920).
- Matouschek, N., and F. Robert-Nicoud. 2005. "The Role of Human Capital Investments in the Location Decisions of Firms," *Regional Science and Urban Economics*, vol. 35 (5), pp. 570-583
- Porter, M.E. 1990. *The Competitive Advantage of Nations*, New York, Macmillan.
- Redding, S., and A.J. Venables. 2004. "Economic Geography and International Inequality," *Journal of International Economics*, vol. 62, pp. 53-82.
- Rice, P.G., A.J. Venables, and E. Pattachini. 2006. "Spatial Determinants of Productivity: Analysis for the U.K. Regions," *Regional Science and Urban Economics*.
- Rodrik, D. 2004. "Industrial Policy for the Twenty-First Century," processed, Kennedy School, Harvard.
- Rosenthal, S.S., and W.C. Strange. 2004. "Evidence on the Nature and Sources of Agglomeration Economies," in V. Henderson and J. Thisse (eds.), *Handbook of Urban and Regional Economics*, vol. 4, North Holland, Amsterdam.
- Shishido, H., and W. Wheaton. 1981. "Urban Concentration, Agglomeration Economies, and the Level of Economic Development," *Economic Development and Cultural Change*, vol. 30, pp. 17-30.
- Sutton, J. 2002. "Rich Trades, Scarce Capabilities: Industrial Development Revisited," *Economic and Social Review*.
- Williamson, J. 1965. "Regional Inequality and the Process of National Development," *Economic Development and Cultural Change*, June, pp. 3-45.