Geography and Development.

J. Vernon Henderson *(Brown University)*

Zmarak Shalizi *(World Bank)*

Anthony J. Venables *(London School of Economics and CEPR)*

Abstract

Economic development and underdevelopment is one aspect of the uneven spatial distribution of economic activity. This paper reviews existing literature on geography and development, and argues that rigorous theoretical and empirical analysis is needed to increase understanding of the role of geography in development and to better design development policy. The analytical issues are: why does economic activity cluster in centers of activity? How do new centers develop? And what are the consequences of remoteness from existing centers? Empirical evidence comes both from the international context and from studies of internal economic geography and urbanization.

Introduction.

The most striking fact about the economic geography of the world is the uneven distribution of activity. High-income regions are almost entirely concentrated in a few temperate zones, 50% of world GDP is produced by 15% of the world's population, and 54% by countries occupying just 10% of the world's land area. The poorest half of the world population produces 14% of world GDP, and 17 of the poorest 20 nations are in tropical Africa. The unevenness is also manifest within countries, with metropolitan concentrations of activity. The share of the population of developing countries in urban areas has increased dramatically in recent years; Latin American countries (as European and North American ones) are 75% urbanized, and while Asian countries are under 30% urbanized their urban populations are growing at around 4% pa.

Why do these spatial inequalities exist, and why are spatial differences in land rents and wages not bid away by firms and individuals in search of low cost or high income locations? The answers to this question have to do partly with spatial variations in institutions and endowments (natural endowments and accumulated human and physical capital), and partly also to do with geography – the spatial relationship between economic units.[i] There are two key (and related) geographical questions. Why do so many economic decision takers choose to locate close to each other? And, for those that cannot locate in an economic center, what are the consequences of being outside, and possibly remote from, existing centers?

This paper argues that understanding these two issues is central for understanding many aspects of economic development and underdevelopment. At the international level it is important for understanding the forces shaping the location decisions of firms, and the consequent demand for labor and pattern of cross-country wage differentials. At the national and sub-national it is important for understanding the processes driving urbanization and the evolving internal economic geography of countries and cities during their development. And at the policy level, it is important for understanding the contributions of international trading arrangements, of regional policies, and of public infrastructure investments in shaping the location of economic activities and thereby promoting development.

This paper reviews some of the recent theoretical and empirical work that illuminates these issues. In the next section we overview analytical issues, and then turn to a review of empirical work. This looks first (section 3) at the consequences of being remote from established centers, concentrating on the international setting. Section 4 turns to empirical work on the forces that drive agglomeration, where evidence is from intra-country, particularly urban, studies. Section 5 delves further into urban issues, looking both at the evolution of cities in developing countries, and at the issue of over-urbanization. Section 6 turns to policy. Here we are much more speculative, but argue that a number of key policy issues need to be analyzed from a rigorous geographical perspective, and that further research needs to be done on these issues.

2: Analytical issues:

2.1 Agglomeration; sources and consequences

Why is economic activity so concentrated? The presence of transport costs suggests that industry might spread out to minimize the costs of reaching consumers in different locations, and if production takes place under conditions of constant or diminishing returns to scale, then this is exactly what economics predicts. The 'folk theorem' of spatial economics says that under these conditions there will be very many small plants supplying local markets. It is only the presence of increasing returns to scale which forces firms to concentrate production in relatively few locations, and thus confronts them with the choice of where to operate [ii],[iii].

Agglomeration forces and dispersion forces:

The increasing returns that are necessary for agglomeration may be either external to the firm or internal. External mechanisms include knowledge spillovers and externalities arising in the labor market. For example, information spillovers can arise with neighboring firms; by observing them and learning about what they are doing, firms learn about technological developments, whom to buy from and sell to, whom to hire, what product lines are selling, and the like (see Eberts and McMillen 1999 for a review). In the labor market, there may be gains from locating in a thick labor market, and in a

location where other firms have already trained a supply of skilled workers (Marshall, 1890, Krugman 1991b)

External economies create incentives for firms to locate close to each other, and so too can internal economies of scale. Firms' location decisions are based both on input price considerations and on ease of access to markets.[iv] Consider first market access. Firms want to locate close to demand (or, more generally, in locations from which transport networks make it relatively cheap to reach markets) and models generally yield the result that increasing returns activities are pulled *disproportionately* towards locations with good market access. For example, if there are 9 locations, 8 of which have 10% of final expenditure and 1 of which has 20% then, other things being equal, more than 20% of manufacturing supply will be met from this larger location as firms locate to exploit the benefit of proximity to the large market. This immediately creates a force for agglomeration of activity. As a disproportionate share of manufacturing is attracted to a location so either the wage rate in the location is bid up or labor is attracted to immigrate – either of which will tend to increase this location's share of total expenditure still further. The market access effect is sometimes called the 'home market effect', and this combined with labor mobility is the basis of Krugman's seminal 1991a paper.[v]

A second force comes from combining market access with intermediate goods production. Demand for manufacturing comes not just from final consumers but also from intermediate demand, so a location with a lot of firms will have a high demand for intermediates, making it an attractive location for intermediate producers. This in turn makes it an attractive location for firms that use these intermediate goods, as they can economize on transport costs on inputs. There is thus a positive feedback between location decisions of upstream and downstream firms, tending to draw both types of firms together in the same location, so leading to agglomeration.[vi] These forces are just the backwards (demand) and forward (cost) linkages that figured so prominently in an earlier generation of development economics (in particular the writings of Hirschman 1958 and Myrdal 1957). However, as we have already remarked, these effects can only really matter in an environment of increasing returns to scale, without which upstream and downstream firms could be broken into many small plants.

Agglomeration forces can operate across more or less broad ranges of activity. For example, the key externalities and linkages might occur between firms in a particular industry or between firms that engage in a narrow field of R&D. Alternatively they might operate at a much broader level – through aggregate demand as a whole, the development of general labor skills, or the provision of basic business infrastructure and inputs used by wide sectors of the economy. It is also argued that they may stem not from specialization but from diversity in the activities of a location (Jacobs 1969).

Pulling in the opposite direction are forces for dispersion. These are of essentially three types. One is negative externalities from congestion. Another is the supply of immobile factors, the prices of which will be bid up in centers of activity, encouraging firms to move to lower factor cost locations. And the third is the extent of the market, limited by the presence of geographically dispersed demand for output. Thus, if labor is dispersed it encourages a dispersed location of firms for both supply and demand reasons.

The importance of these dispersion forces depends critically on what factors are immobile, and what mobile. In a regional context labor might be mobile, and land the only immobile factor. Agglomeration then causes labor movement (eg to cities), until choked off by congestion costs or land prices. In an international context most sorts of labor are immobile, so agglomeration will bid up the price of labor as well as land. This discourages agglomeration, but means that when it occurs it will be associated with international income inequalities.

Outcomes:

Outcomes are determined by the balance between agglomeration and dispersion forces. Theoretical modeling establishes the dependence of this balance on model parameters, and shows how small parameter changes can lead to discontinuous changes in the configuration of equilibria. Thus, for some values of parameters a model may predict that economic activity will be dispersed between locations. This configuration is robust until a bifurcation point is reached, at which point the dispersed equilibrium becomes unstable. If one location gains further activity then positive feedback (which in the earlier spatial literature was referred to as the process of cumulative causation) causes further activity to be drawn in, forming an agglomeration.

For example, if transport and communication costs are very high then activity must be dispersed; (under autarky every location must have its own industry to meet final demand). And if transport costs are extremely low, then firms will not care whether they are close to markets and suppliers; (if transport and communications are costless we encounter the end of geography). So it is at intermediate levels of transport costs that the likelihood of agglomeration is greatest. This is often presented in the models as an application of 'symmetry breaking'. At high enough transport costs all locations are identical, but as costs fall below a critical value the model goes through a bifurcation, at which point the economic geography of the world self-organizes into a structure of centers of activity, with intermediate hinterland areas. Further declines in transport costs may lead to erosion of this structure.

Typically many locations are candidates for hosting the agglomeration, and small initial differences (or historical chance, or self-fulfilling expectations) determine which gains it. However, once a site has become a center of activity, then a 'lock-in' effect operates. Even if exogenous circumstances change (perhaps reducing the attractiveness of the site) economic agents will not want to move away and forego the benefits of the agglomeration. This tendency will be accentuated by the durability of sunk cost investments, such as plant and infrastructure. There is therefore a path dependency in the structure of the equilibrium, with history being as important as current circumstances.

In our discussion of agglomeration and dispersion forces we distinguished according to the breadth of activities drawn together by agglomeration forces, and according to the mobility of factors of production (often corresponding to an intra- vs inter-national distinction). Different combinations of these cases apply in different situations, generate different outcomes, and correspond to different strands in the literature. Table 1 illustrates some of the possibilities.

If agglomeration forces operate primarily within particular industries and most factors are mobile, then the likely outcome is agglomeration of industries in specialized locations (top left cell of table 1). Inter-locational factor price differences are small, both because each of these centers only contains a small fraction of possible activities, and because many factors are mobile. The classic model of this type is that of Henderson (1974), who constructs a general equilibrium model of a system of specialized cities.

Two extremes are analyzed -- a world of developers who set up competitive cities potentially achieving efficient outcomes, and a world without "large agents " (developers) where cities (of generally excessive size) form through "self-organization."

At the other extreme, if linkages operate at a much broader level and factors (especially labor) are immobile then agglomeration, if it occurs, will be associated with inequalities in factor prices and real incomes (bottom right cell of table 1). Thus, in the international model of Krugman and Venables (1995) industrial activity concentrates in 'north', even though wages may be many times higher there than in 'south'; firms are deterred from moving south because agglomeration benefits foregone might outweigh labor cost savings. This view of the world is radically different from that of conventional international economics, predicting that the world divides into rich and poor regions, even if there are no international differences in factor endowments, skill levels, institutional quality or other underlying economic characteristics. Development and under-development are simply manifestations of agglomeration of economic activity.

Table 1: Agglomeration: forces and outcomes

		Dispersion forces	
		Weak (eg factors mobile)	Strong (eg factors immobile)
Agglomeration forces	Narrow, (eg intra-industry)	City specialization (Henderson 1974)	Industrial clusters vs. comparative advantage (Fujita, Krugman, Venables 1999, chapter 16)
	Broad, (eg aggregate demand)	City formation (Fujita 1988)	World income inequalities (Krugman and Venables 1995)

2.2: The formation of new centers.

Agglomeration mechanisms are one way to explain the observed unevenness in the spatial distribution of activity and income. Development must then take the form either of mitigating the disadvantages of being outside existing centers, or of the creation of new centers of activity. What does this approach have to say about the birth of new centers? This question has typically been addressed by supposing that there is some

exogenous growth process – population or technical change – and showing how this will create new centers of activity.

In the urban context, Fujita, Krugman and Venables (1999) show how rising population will lead to the birth of new cities. Population growth expands the agricultural hinterland surrounding existing cities, and at some point it becomes profitable for a new frontier city to develop. Continuing growth will lead to the development of multiple cities, spaced out from each other and locked into their locations.

In the international context, Puga and Venables (1999) model the spread of an agglomeration from country to country.[vii] They suppose that world demand for manufactures is increasing (due perhaps to exogenous technical progress), and tending to widen the wage gap between countries with industry and those without. There comes a point at which the wage gap is too large to be sustainable, and manufacturing begins to move out of established centers to low wage regions. However, just one (or a few) new industrial centers will become established at a time. The logic is as we have already seen. An equilibrium with dispersed industry is unstable; any location that gets just slightly ahead of the others gains from forward and backward linkages, these positive feedbacks causing the location to develop faster and the others to fall back. Development therefore takes the form of enlargement of the set of countries that are in the 'center', while most countries remain outside, largely unaffected. As the growth process continues, so enlargement of the set of 'central' countries proceeds sequentially, adding countries in turn. The approach predicts that development is not a process of steady convergence of poor countries to rich ones, but instead the rapid transition of selected countries from the poor club to the rich club.

Which countries are most likely to make this transition? It may be determined by very small initial cross-country differences (indeed, if all countries were identical, it would simply be a matter of chance). The pertinent dimensions of difference are those which determine the profitability of the first firms to relocate, so include labor market factors, internal infrastructure (Martin and Rogers 1995), as well as institutional characteristics of the country. Since the first entrants will be highly dependent on

imported intermediate and capital goods and on export markets for final sales, they will tend to go to locations close (or with good transport links) to established centers.

A further issue concerns the industrial structure of these newly industrializing economies. What sectors do they attract first, and how does their industrial structure change during development? The first sectors to become detached from an existing agglomeration will typically be those that are intensive in immobile primary factors (the prices of which are high in the center), and that are not too heavily dependent on linkages with other firms. These may be firms with low usage of intermediate goods, low levels of sales to other industrial sectors, or that do not need to cluster with related activities to gain new technology. As these sectors relocate, so they may begin to create linkages and attract other sectors. The sequence in which industries enter then depends on their factor intensities, their tradability, and the way in which they benefit from linkages to other activities, and create their own linkage effects.

The message then, is that new centers of activity can develop, but the process is not one of steady convergence of all locations. Instead, it is rapid development of a few locations, leaving others essentially unaffected. This fits well with the historical record. Recent decades have seen a small group of countries make a rapid transition from being amongst the low income group to join the middle- or high-income countries, while divergence has continued between high-income and the great majority of low income countries (Quah 1997). Furthermore, growth performance is much more variable across countries than is accumulation of either physical or human capital (Easterly and Levine 1999).

2.3: Regional structure and the costs of distance:

Although new centers can form, most locations remain outside. What determines the structure of activity outside established centers, and the magnitude of the income penalty to being outside?

The costs of distance from an established center arise essentially because of the costs of trading goods with, and receiving information and technology from, the center. These costs will impact entirely on immobile factors, and if these account for a small share of production costs, then even quite low transport costs can have a large effect on

their prices. The classic analysis of this is von Thunen (1826). A city is located in the center of a 'featureless plain' and labor is mobile between the city and agricultural employment in the surrounding area. Regions specialize – forming concentric circles of activity – according to the transport intensity of the products, and transport costs determine the rent gradient. Rents diminish steadily with distance since land – the only immobile factor – bears all the costs.

If labor is immobile – or has frictional costs in moving – then it too will bear some of the costs of distance. An international application of the von Thunen model is developed in Venables and Limao (1999) in which there are several immobile factors in countries at increasing distances from an economic center. The countries specialize according to the interaction between two pairs of forces. One is products' transport intensity interacting with distance, as in von Thunen; the other is products' factor intensity interacting with countries' factor endowments, as in Heckscher-Ohlin trade theory. Real incomes decline with distance, although the prices of individual (immobile) factors need not, as changing patterns of specialization influence factor demands.

3. Geographical remoteness and underdevelopment

We turn now to empirical evidence, looking first at the international evidence on the implications of being outside established centers. We then move on in following sections to present evidence on agglomeration (based largely on intra-country studies) and also to draw out the way in which the internal economic geography and urban structure of economies change during development.

3.1: Transport costs, trade and income.

Transport costs incurred on traded goods are only one of the direct costs associated with distance, although they are perhaps the one that is most readily observable. They can be measured by the c.i.f./f.o.b. ratio giving the 'carriage, insurance and freight' costs of countries' imports, which typically range from a few percent of the value of trade, up to 30-40% for the most remote and landlocked (and typically African) economies. More direct measures provide a clearer measure of the cross-country variation in shipping

costs. For example, estimates of the cost of shipping a standard container from Baltimore to selected West African destinations range from $3,000 to Cote d'Ivoire, to $7,000 to Burkina Faso, up to $13,000 for the Central African Republic. Limao and Venables (1999) find that being landlocked raises transport costs by more than 50% (comparing the median landlocked country with the median coastal economy), and an extra overland kilometer costs as much as 7 additional sea kilometers. Infrastructure (and, for landlocked economies, transit countries' infrastructure) also matters, so shipping to Austria or Switzerland costs around $4,000, compared to $3,000-$3,500 for Germany and Belgium.

What are the consequences of transport costs of these magnitudes? First, they are a real cost, using up scarce resources. Second, they choke off trade. Gravity estimates of bilateral trade flows use distance as a proxy for transport costs (and possibly also control for countries sharing a common border and for language and cultural links), and typically find elasticities of trade volumes with respect to distance of between –1 and –1.3. This is a large effect, indicating that doubling distance cuts trade volumes by between ½ and 2/3. Combining gravity results with estimates of the elasticity of transport costs with respect to distance, indicates that the elasticity of trade volumes with respect to transport costs is around –2.5 (Limao and Venables 1999). Thus, doubling transport costs reduces trade volume by around 80% and the median landlocked country has less than 40% of the trade volume of the median coastal economy. The trade reducing effect is strongest for transport intensive activities – i.e., activities that are dependent upon exports for sales and/or imported intermediate goods for production. Radelet and Sachs (1999) find that increasing a country's c.i.f./f.o.b. ratio from 12% to 17% reduces the long term growth of the share of non-primary manufactured exports in GDP by around 0.2% per annum.

Can levels of transport costs of these magnitudes go far towards explaining real income differences of the magnitude we observe in the world, varying by up to 50 to 1? Some simple arithmetic helps. Suppose that a product sells for $100 in the 'center' and uses intermediate inputs that cost $40, giving value added of $60, perhaps $15 of which goes to capital and $45 to labor. In a location which imports the intermediates and exports the final product to the center, both at 30% ad valorem transport costs, the cost of intermediates rises to $52 and receipts from output fall to $77, giving value added of $25. Capital costs $15 (at least, supposing perfect access to the center's capital market),

cutting the maximum possible return to labor to $10. Raising ad valorem transport costs to 40% reduces value added to $15, less than the cost of the capital input. These numbers are not that extreme – firms in East Asian export processing zones typically have imported inputs accounting for 60% of the value of their output, ranging up to nearly 80% in electronics.[viii] The point is that transport costs which seem quite modest relative to the value of gross output can be very large relative to value added attributable to immobile factors of production. Distance can then have a major impact on wages and per capita incomes.

3.2: Investment and technology.

The spatial relationship between countries affects not only goods trade, but also other forms of interaction. Foreign direct investment flows follow a gravity relationship, similar to trade flows. Thus US FDI tends to be located relatively close to the US, and estimates of the gravity coefficient on this investment are typically around –1, slightly less than for trade. Brainard (1997) uses the ratio of sales by affiliates to direct exports to measure US FDI activity, and finds that the share of affiliate sales is lower the lower are trade costs, and the lower is per worker income in the host country. Estimates based on Swedish outflows of FDI suggest that the distance coefficient is more negative for FDI than for trade, implying even greater sensitivity to distance (Ekholm 1998).

Just as trade and investment fall off sharply with distance from established economic centers, so too does the transmission of technology. Coe and Helpman (1995) show how total factor productivity depends on both domestic and foreign (trade weighted) stocks of R&D, and how the beneficial effects of foreign R&D on domestic productivity are larger the more open the domestic economy. In their approach geography enters via the trade weighting of foreign R&D stocks, although this weighting is challenged by Keller (1998). In recent work Keller (2000) relates total factor productivity directly to distance from R&D producing countries. He finds that, on average, being 10% further away from a major R&D producing economy (such as the US) reduces total factor productivity by around 0.15%. While these studies give us some valuable insights they are, unfortunately, restricted to OECD countries, and less is known about the transmission of technology to developing countries. Often FDI is such a

vehicle, both directly and via spillovers to local firms (see Blomstrom and Kokko 1997 for a survey), and we have already seen the geographical concentration of FDI.

3.3: Geography and per capita income:

How much of the cross-country inequality of per-capita income levels can be attributed to a set of geographical variables, including some distance measures? A statistical answer to this question is provided by Gallup and Sachs (1999), who regress national per capita income on four variables; a measure of the endowment of hydro-carbons per capita; a dummy variable for incidence of malaria; the proportion of population who live within 100km of the coast; and international transport costs, as measured by the c.i.f./f.o.b. ratio on imports. They find that these four variables alone account for an astonishing 69% of the per capita income variation across their sample of 83 countries. Looking at the relationship between countries' per-capita income and distance from one of the three core regions (taken to be New York, Rotterdam and Tokyo), they find that doubling distance reduces income by around 25%.

Redding and Venables (2000) use a trade and geography framework to construct a single summary measure of the maximum wage that firms in a country can afford to pay, given their access to markets and to intermediate goods. They use a gravity model to estimate the parameters of this 'wage equation', which becomes (like a traditional market potential measure) essentially a function of the weighted average of country incomes, weights being inversely related to distance and other geographical characteristics such as landlockedness. The relationship between this constructed wage measure and actual per capita income is very tight, with the measure explaining over 66% of the cross-country variation in per capita incomes.

Pulling together the evidence on the costs of being outside existing centers, it seems clear that distance matters, impacting on trade, investment and income. New technologies, as well as trade liberalizing policies, are undoubtedly mitigating some of these costs, but it is worth pointing out that many aspects of trade costs are not falling. Hummels (1999) charts the path of ocean shipping costs and shows how these costs – relative to other goods prices in the economy – have fluctuated, with no trend decrease over the last few decades.[ix]

4. Urbanization and internal structure

The previous section looked at the costs of being geographically distant from existing economic centers. We now turn to the flip side of the coin, and review the evidence on the productivity benefits that can be derived from being in a concentration of activity. The literature is derived almost entirely from intra-country studies, usually focussing on cities.

4.1 Bases of agglomeration and urbanization

There is a large and increasingly sophisticated empirical literature on the magnitude of the productivity advantage gained by being located in a center of activity. Evidence suggests the externalities in specific manufacturing industries derive mostly from firms doing similar activities -- localization economies. Since this paper is concerned primarily with development we cite results from a recent study of Korea. However, the findings are similar for industrial countries, such as the USA (e.g., Sveikauskas 1978, Henderson 1988, and Ciccone and Hall, 1995), and Japan (Nakamura, 1985), as well as other developing countries, such as Brazil (Henderson, 1988) and Indonesia (Henderson and Kuncoro, 1996).

Table 2 shows the magnitude of localization economies for different industries in Korea. There, a 1% increase in local own industry employment results in a .06-.08% increase in plant output, for a typical industry. So a plant in a city with 1,000 workers in other firms in the same industry would, without changing its own inputs, increase its output by 20-25% by moving to another city with 15,000 workers in the same industry, creating a big incentive to agglomerate. In Korea, the rankings of industries by the magnitude of their externalities follows exactly the ranking of industries by the extent to which they are spatially concentrated across cities. So heavy and transport industries tend to be concentrated in a few highly specialized cities to take advantage of these local scale externalities, while traditional industries with low scale externalities are more dispersed. Similarly, Henderson (1988) analyzes the close connection between the magnitude of

scale externalities for different industries in the USA and Brazil and the sizes of cities that specialize in different industrial activities.

Recent work on the USA looks at details of how externalities affect different types of firms. For high tech industries, using plant level productivity data, Henderson (1999a) finds that single plant firms benefit more than corporate plants from localization economies, since corporate plants have their own internal information networks and are less reliant on the external environment. He also finds evidence of dynamic externalities, especially for single plant firms, in addition to static ones. With dynamic externalities the past local industrial environment of a city affects productivity today, by contributing to local knowledge accumulation-- a stock of local trade secrets. Locations with no history of an industry are disadvantaged because there is no accumulated body of knowledge for new plants to draw upon. Finally, he finds that for high tech production, local diversity and scale of either the overall or the business service environment have no effect on productivity, in contrast to results for less developed countries such as Korea and Indonesia. Similarly for new plants that are *foreign owned* in Portugal, apart from localization economies, Guimaraes, Figueiredo, and Woodward (2000) find such firms benefit from locations with diversity and scale in business and financial services, but not, most surprisingly in other foreign owned activity. For foreign owned firms, being able to access domestic business services in a country they are unfamiliar with may be critical.

Table 2: Evidence on Agglomeration Economies
Magnitude of External Economies of Scale for Korea

Industry	**Localization economies (elasticity)**	**Diversity economies (% increase in productivity for one standard deviation increase in diversity)**
Traditional (food, textile, apparel, wood and paper products and furniture)	.021**	NA
Heavy (basic and fabricated metals, chemicals and plastics)	.082**	NA
Transport	.096**	NA
Machinery and Electrical Machinery	.053**	NA
High Tech (computer, communication, TV, radio and scientific instruments)	.056**	.599**

**Coefficient significant at 5% level.

Table based on Henderson, Lee, and Lee (1999) [HLL] who estimate scale economies using city-industry data for 1983, 1989, and 1991-93. They examine the determinants of value added per production worker across cities by industry, controlling for capital per worker and accounting for time, city and sub-industry fixed effects, with results on scale externalities reported in the table. In the column on localization economies, a coefficient of .06-.08 means that a 1% increase in local own industry employment results in a .06-.08% increase in plant output. HLL found that measures of metro scale had no effect on productivity in any industry. However for high tech industries a diversity measure did is important, a 1 standard deviation increase in diversity of the local manufacturing base increasing productivity by 60%. This accords with the intuition that "the bright lights" of diverse large metro areas, with cross-industry fertilization of ideas, are important for new evolving high tech industries, but not standard ones.

In related work, indirect evidence on patterns of patent citations in Jaffe et al. (1993) and on patterns of plant births in Rosenthal and Strange (1999) suggest that information spillovers attenuate rapidly over space, requiring close spatial clustering--dense industrial neighborhoods-- to be realized. Moreover technological developments in communications are not reducing this need for "face-to-face" proximity and communication. In fact, Gasper and Glaeser (1999) suggest that the technological developments in communications complement and accentuate the need for more face-to-face interactions!

In addition to these relatively narrow within industry links, there is also work suggesting that the general education level of workers improves productivity in a plant, accentuating the scale benefits of agglomeration by enriching the quality of local information spillovers (Black and Henderson 1999b). In recent empirical work Moretti (1999) explores this for the USA, solving some of the difficult data and econometric

issues. He finds that a 1% increase in the supply of college graduates in a city raises the productivity of high school drop-outs by 2.3%, of high school graduates by 1.4% and of college graduates by 1.2%. Productivity in cities specialized in production intensive in high skill workers benefits from localization economies, knowledge spillovers, and their interaction.

These studies leave two main gaps in our knowledge. The first is to do with service activities. Econometric investigations of the productivity effects of scale externalities have focused on manufacturing. We have no direct evidence on the nature or magnitude of scale economies in service activities, although some reasonable inferences on these could be attempted by examining patterns of service agglomerations.

The second gap is that, as we saw in section 2, agglomeration can arise not only because of direct technological spillovers, but also because of benefits from access to a thick labor market, access to suppliers, and access to large markets. Empirical work quantifying these benefits still needs to be undertaken.[x]

4.2 Limits to City Sizes

Dispersion forces operate against agglomeration to constrain city sizes. As we have seen, they include the congestion costs that can arise in cities, the prices of immobile factors, and the extent of the market.

A number of studies quantify the costs to workers and firms of living and operating in cities. Based on UNCHS data, for a cross-section of 80-90 cities worldwide (both OECD and LDCs), Henderson (1999b) calculates elasticities of 0.25 for both average housing prices and commuting times with respect to metro area populations. Thus housing prices and commuting times are each more than 100% higher in a metro area of 5 million compared to one of 100,000 population. These numbers are similar to those in Rousseau (1995) which indicate that costs-of-living are about 90% higher in Paris at 9 million people than in the typical French city. Rousseau also notes that wages for the typical worker are also about 90% higher (so real incomes across cities of different sizes are similar). These differentials for both wages and costs of living have also been found in other studies for the USA and Brazil (Henderson, 1988) and some other countries in Latin America (Thomas, 1980).

While urban theory focuses on land and labor costs as they increase with city size, there are infrastructure and other capital costs associated with locating activity in dense cities-- costs of road networks and utilities, as well as environmental costs. Using data on Pakistan, Egypt, Indonesia, and Bangladesh, Richardson (1987) calculates that the investment costs of housing a family in a large metro area are at least three times those in a village. Henderson (1999b) shows that the ratio of public investment to GDP in countries rises sharply at the stage when rates of urban population growth in countries accelerate. Quality of life measures (e.g., child mortality, pupil/teacher ratios in primary school) deteriorate as urbanization accelerates and national public resources are strained. A variety of studies link environmental degradation to metro area size; for example, dense output of emissions of autos and plants leads to excessive concentrations of ground level ozone and carbon oxides in large metro areas. Once rates of urban population growth drop off in countries and country incomes are correspondingly higher, then these conditions tend to improve.

5. Divergence and Convergence: Cities in Developing Countries.

Given the forces outlined above, what evidence is there on outcomes? How has the balance between agglomeration and dispersion shaped city growth in developing countries? In a 1965 paper, Williamson hypothesized that, in growing from low-income levels, countries go first through a period of regional divergence and concentration of development and industrialization in just a restricted portion/region of the country. The restriction of development to a limited region conserves on scare economic infrastructure (Hansen, 1990), such as roads, technical skill workers and managerial resources. As development proceeds, the original area of development becomes congested and subject to diminishing returns to further investment and the country can afford to develop other regions, with the requisite investments in infrastructure, institutions, and human resources. That leads to industrial deconcentration, growth of hinterland regions, and a move towards regional convergence. This general hypothesis has been investigated both for regions of a country and for patterns of convergence among the cities of a country.

5.1 Regional Convergence.

Regional convergence in more developed contexts of the USA, Europe, and Japan has been the subject of wide analysis.[xi] For the USA and Japan, Barro and Sala-i-Martin's papers (1991, 1992) suggest that steady states are similar across regions, but rates of convergence are slow. In Japan, they find that low-income regions are catching up to original higher income regions, through internal growth (not out-migration) of more remote parts of Japan. However Fujita and Tabuchi (1997) find evidence of more recent core-periphery divergence. Barro and Sala-i-Martin (1991) find that European states seem to be converging to somewhat different steady states. More recent econometric advances in analyzing growth patterns find faster rates of conditional convergence, but perhaps greater steady-state differences (Caselli, Esquire, and Lefort, 1996). Unfortunately there is a lack of comprehensive studies of regional convergence in different countries at different stages of development, which would indicate for what regions under what policy regimes and at what points in time there will be convergence.

5.2 Urban Convergence.

Evidence supports the view that the degree of national urban concentration rises then falls with development. Shishido and Wheaton (1982), Henderson (1988), Ades and Glaeser (1995) and Henderson (1999c) all examine the pattern of national urban concentration. Shishido and Wheaton (1982) and Henderson (1999c) find national urban concentration rises with growth from low income levels, peaks at low-middle income levels (around 1987 PPP income per capita of $2500), and then declines. So it seems that Williamson's hypothesis holds for cities-- there is increasing concentration of resources in one city (or a few cities in a large country) from low income levels, which at some point in the growth process peaks and then declines. Below in Section 5.4, we explore some of the details of the process of deconcentration, discussing suburbanization and ex-urban versus hinterland development and population versus industrial deconcentration. In this section we focus on key national issues concerning the level of urban concentration.

It is also useful to review what other factors affect urban concentration. Geography matters. Ades and Glaeser (1995) and Henderson (1999c) find primacy is increased significantly by increases in national land area and population, or by having the

primate city be a national capital or a port. In Henderson (1999c), openness of the economy has a very small and ambiguous effect on national concentration. The effect of an increase in political decentralization has a surprisingly small effect in reducing concentration (under 10% of a standard deviation of primacy), although previous studies Henderson (1988) and Ades and Glaeser (1995) find much bigger effects.

The policy instrument that seems critical in influencing national urban concentration is investment in inter-regional transport infrastructure. As Gallup, Sachs, and Mellinger (1999) suggest (historical) investments in national navigable waterways induce inland habitation, significantly reducing urban concentration (Henderson 1999c). Rosen and Resnick (1978) also find rail investment reduces national urban concentration. In a more modern version, Henderson (1999c) finds that investment in national roads and highway systems significantly reduces national urban primacy, with the effect rising with income.

There is a presumption in the academic literature and in the view of international agencies (Renaud 1981) that there is a tendency to over-concentration in many countries, especially rapidly urbanizing low-income countries. The UN (1993) asks how bad "the negative factors associated with very large cities" need to get before [it is in the] self-interest of those in control to encourage development of alternative centers." The academic literature on the subject has three strands. The theoretical literature tends to argue that in theoretical models stable city sizes can only be either efficient or too big, so market failure may lead to too big cities. There is a cost benefit literature on the USA (Tolley, Gardner and Graves 1979) and on developing countries (eg Richardson (1987) which argues that the social marginal costs (congestion and commuting, environmental costs, etc) of increasing city size exceed the social marginal benefits of scale at equilibrium city sizes, again suggesting that cities are too large. Finally, and most critically, there is a political-economy perspective on the subject.

Renaud (1981), Henderson (1988) and Ades and Glaeser (1995) argue that often the political institutions in countries encourage over-concentration. The idea is that, in many countries, there is a lack of a level playing field in the allocation of resources across cities. The national government can choose to favor one (or two) cities over others. Typically such cities are national capitals (Bangkok, Mexico City, Jakarta, or

Seoul, not to mention Paris); but may also be a Sao Paulo, the seat of national elites. Such favoritism can involve the allocation of local public services in favor of national capitals, where decision-makers live. That problem can be exacerbated if hinterland cities do not have the power (i.e. both the authority and financial resources) to determine their own public service levels, either because of a unitary national constitution or because local autonomy has been suspended (as in Korea from 1961 to the 1990's).

Favoritism can take the form of the national government choosing not to invest sufficiently in interregional transport and telecommunications, so that hinterland cities are less competitive locations for private producers. That favors producers and investors (who may include national politicians) in the national capital. Favoritism, as in Indonesia (Henderson and Kuncoro, 1996 and Kaiser, 1999) or in China or Brazil in the recent past, can also take the form of restrictions in capital markets, export/import markets, and licensing of production rights. These restrictions all favor firms which locate in the national capital (or other favored mega-cities), allowing central bureaucrats and politicians to extract rents in the allocation of loans and licenses without competition from lower ranked bureaucrats in other locations. There may be also an "innocent" bias towards locating production in mega-cities, as in Brazil or China, based on notions of hierarchies in the allocation of technologies across cities and the inherent scale benefits of larger cities. As stressed above, the issue is not whether larger cities offer greater scale economies for all types of production, but, rather, what types of production benefit sufficiently from being in large cities to compensate for the high costs of such cities.

All analyses tell us favored cities are oversized. Migrants and firms flow to a favored city, until it becomes so congested and costly to live in that these costs offset the advantages of the favoritism. Moreover, the excessive resources devoted to one or two favored cities detract from the quality of life in the rest of the urban system. Based on the UNCHS data set for 80-100 cities in 1996 worldwide, Henderson (1999b) shows that high urban concentration in a primate city increases child mortality, pupil-teacher ratios, use of non-potable water and other poor quality of life dimensions in *typical medium size metro areas*, after accounting for size, income, and growth differences among cities. For example, a one-standard deviation increase in the national urban concentration measure raises child mortality in typical cities by 1/3 of a standard deviation of the child mortality

rate across cities in the sample. So the costs of excessive urban concentration in a primate city are felt throughout the whole urban system, not just in very large cities.

5.3 Urban concentration and economic growth.

Putting all these analyses together, they suggest that there is an optimal degree of urban concentration in a country, where too little means scale economies are inadequately exploited and too much means that cities are too crowded and congested.

The effects of urban primacy on economic growth are tested directly in Henderson (1999c), using panel data on primacy (interacted with per capita income and national scale) and growth. Henderson finds that the growth maximizing degree of primacy depends on income and national scale, with tight statistical fit. Table 3 gives the growth maximizing primacy levels (% population in largest city), and illustrates that it increases up to an income per capita of about $5000 (1987 PPP) and then declines. Best primacy levels in a medium size country are quite high at the peak, amounting to 1/4 of the national urban population being in the largest city. The list of countries with highly excessive primacy includes the usual suspects such as Argentina, Chile, Peru, Bangladesh, Thailand, Japan, Congo, France, and Greece. Remaining columns in Table 3 demonstrate the effect of excessive urban primacy on economic growth and the role that transport policy can play in the process. The economic growth rate loss from excessive primacy (primacy one standard deviation above the best level) rises with income and then flattens out at a very high level. The growth rate losses of 1.6% correspond to the losses from reductions in investment or education rates by one standard deviation in growth model econometrics. Finally the indirect effect on economic growth rates of increasing road investment to reduce primacy in excessive primacy countries rises with income to a high level. Transport investment can play a key role in reducing excessive concentration and that has very positive effects on growth rates. The form of national urbanization-- the degree of urban concentration-- matters very much in a country and inter-regional infrastructure investment plays a key role in the process.

Table 3: The Effect on Annual Economic Growth Rates of Urban Concentration and National Transport Infrastructure in Medium Size Country.

	The optimal degree of urban primacy (% population of medium size country in largest city).	**Loss in growth rate from excessive primacy (one standard deviation)**	**Growth effect of a one standard deviation increase in road density in a country with excessive primacy**
Low income ($1100)	15	.71	.23
Medium income ($4900)	25	1.6	.68
High income ($13400)	23	1.6	.68

The table looks at a medium size country --national urban population of 22 million. Numbers for countries with urban populations of up to 50-60 million are similar. The first column calculates the degree of urban primacy that maximizes growth rates and steady state income levels. Error bands about this for medium or higher income countries are quite tight (standard error of .018). The growth losses of excessive primacy are high, although more so, as income rises. The role of transport investment (length of the national road system divided by national land area) is quite significant, particularly as countries enter middle income phases when deconcentration becomes critical.

5.4 The anatomy of deconcentration of industry and population

In the initial stages of industrialization countries tend to concentrate industry in one or two major cities, following the Williamson hypothesis, and then at some point industry starts to deconcentrate. The details of this process are of interest. Deconcentration of industry is more dramatic than that of population, and it transforms the industrial bases of the largest metro areas. Initial industrial deconcentration is generally into the suburban and ex-urban areas of the major city and then into nearby satellite cities, but all within, say, 60 kilometers of the original city. Such deconcentration usually raises issues of suburban sprawl and environmental degradation. Industry and workers move en masse, often in just a 2-3 year year period, into nearby rural areas that do not have the administrative capacity and skills to plan for appropriate infrastructure investments and land use development, nor for regulation of industrial polluters. The second wave is into hinterland areas, both cities and rural areas, away from the domain of the major metro area(s). The process has been studied for Brazil (Townroe, 1981 and Hansen, 1968) and Korea (Chun and Lee, 1985 and Henderson, Lee, and Lee, 1999).

Recent evidence is available for Korea (again). Part 1 of Table 4 shows Seoul's peak in urban primacy occurred around 1970. Part 2 shows that, while Kyonggi province (in which Seoul is located) has not changed its share of population greatly, its share of

manufacturing has declined enormously since 1970. That decline involved massive industrial movement in the late 1970's and early 1980's from the main Seoul metro area to nearby satellite cities, involving most industries, but especially chemicals and primary and fabricated metals.

Table 4: Urban Deconcentration in Korea

Part 1. Metro Seoul's Primacy

	1960	**1970**	**1980**	**1990**
Share of national urban population	34	41	38	33

Part 2. Metro Seoul's Share of Population and Manufacturing in Kyonggi Province*

	1970	**1980**	**1983**	**1993**
Population	62	63	67	61
Manufacturing	76	61	45	30

*excludes Inchon metro area

Part 3. Share of National Manufacturing Employment

	1983	**1993**
Seoul	21	14
Pusan and Taegu	23	14
"Satellite" metro areas	30	30
Other cities, other rural areas	12	17
Rural areas of satellite city provinces	14	25

The key reason to move out of Seoul was lower wage and lands rents; but the impediment was the need for access to the bureaucracy in Seoul, in a context where the economy was still highly regulated. Kwon (1985) suggests plants would only relocate to places within a 3/4 hour drive from Seoul. Part 3 of table 4 shows the deconcentration of industry from the three major metro areas of Seoul, Pusan, and Taegu and their satellite cities into the rest of the country. This deconcentration occurred after the massive investments in roads and communications blanketing Korea in the late 1970's and early 1980's and after the economic liberalization of the early 1980's that helped diminish industrial ties to the government. In just 10 years, the shares of the three metro areas in national manufacturing employment fell by 36% and the shares of satellite cities remained the same. The shares of other cities and rural areas grew dramatically,

increasing from a 26% share in 1983 to 42% in 1993, this in a time period when their populations fell by 9%! This points out a key feature of the deconcentration process. While actual and desired (see table 3) population deconcentration is fairly modest after the peak, manufacturing deconcentration can be dramatic. Hinterlands, if accessible by transport and modern communications, offer lower wage and rent costs, proximity to natural resources including traditional, albeit lower skill populations, and improved congestion and environmental conditions.

While Korea is a not a large country geographically, this deconcentration of industry into hinterlands of larger developing countries has been studied for both Brazil (Townroe, 1981 and Hansen 1983) and Mexico (Hanson, 1996), quite apart from the analysis of historical USA (Cronon, 1992). For Brazil the deconcentration of industry from Grande Sao Paulo to lower wage populated hinterland cities followed the transport corridors first through Sao Paulo state and then into Minas Gerais, the interior state with the main iron ore and other mineral reserves. Deconcentration may involve provinces (each achieving wide diversity in production) so that many are producing a similar range of products, while trade increases across provinces (in intra-industry parts and components) and increases within provinces across cities that become more industrialized *and* specialized.

5.5 Long run city dynamics: lessons from developed countries

Based on work covering the last century in Japan and France (Eaton and Eckstein 1997) and in the USA (Black and Henderson 1999a and 1999b), we know that the populations of cities tend to grow continuously over time. In the USA, over the last century, the average, median, and maximum size city have all increased in size by 4-5 fold. Technological developments affecting commuting and other service provision lower the costs of having larger city populations. In addition, on-going local knowledge accumulations that interact with scale externalities enhance the benefits of larger cities, as noted earlier and as modeled in Black and Henderson (1999b). While growth in city sizes in the USA has slowed in recent years, it is reasonable to expect typical sizes to continue rapid growth in developing countries.

Growth in city sizes tends to be "parallel", with the relative size distribution of cities being remarkably stable over time. Eaton and Eckstein show that the larger cities and towns in France and Japan tend to grow at the same rate, maintaining the same relative size distribution. Henderson (1988) notes the stability of city size distributions for other countries such as Brazil and India. Black and Henderson (1999a) show that, for the USA, the 1900 and 1990 relative size distributions almost perfectly overlap, although there is a modest increase in concentration near the top end in recent decades. In addition to city sizes increasing, the numbers of cities has grown as well. Depending on definitions of a metro area (in 1900 and in 1990) the number of metro areas in the USA has grown by 50-200% over the period.

Beeson et al (1999) and Black and Henderson (1999b) point out a key feature of city growth and the size distribution of cities. While there is remarkable stability of the overall size distribution of cities over time, there is also mobility of cities through that size distribution. Over the last century, some new and existing small cities have grown to be very large. Some medium size cities have stagnated and shrunk in relative size, although remarkably few metro areas ever have absolute population losses. However, consistent with Eaton and Eckstein's observations on France and Japan, in the USA there is almost no downward relative size mobility at the top end. Cities in the top ten percentiles of the size distribution in 1900 are all there in 1990, although since the number of cities has expanded, there have been additions to the list of cities in the top ten percentiles.

The big stay big for two sets of reasons, even though they typically radically change production compositions over the decades. First they have "history" on their side (built-up physical infrastructure and housing that would have to be abandoned in order for cities to shrink absolutely) and second, they have a base of accumulated local knowledge, institutions and traditions that gives them a strong competitive advantage. And not inconsequentially, they have the political influence to help shape national policy to their advantage. For the USA, Black and Henderson note that the older major cities in the North-East and mid-West have held their relative sizes because they are in more densely populated regions with high market demand, or demand for each other's

products. New major metro areas in the South and West have arisen to take advantage of the natural coastal and climatic amenities in parts of those regions.

6. Geography, development, and policy.

Pulling all these strands together we conclude that geography matters for development, but that economic growth is not governed by a geographical determinism. Costs of remoteness can be reduced, and new economic centers can develop. Ireland's GDP grew by 70% between 1987-97 as it benefited from integration in the EU, and major new economic hubs can form, as in the coastal regions of China. Policy is instrumental in shaping these changes, although the design of policy is not well understood. What sorts of policy are effective in altering the economic geography of countries or cities? Given the presence of multiple market failures, what are the welfare implications of such policies, both for directly affected locations and for the world as a whole?

Policy issues arise in a number of contexts, the first of which is international. Here it seems clear that the costs of remoteness can be reduced by policy to facilitate trade and investment flows and to bring countries into the world trading system. However, it is also clear that the package of policy measures required goes well beyond simple reductions in border tariffs and quotas. Tariff liberalization may perhaps be necessary for successful participation in the world economy, but it might not be sufficient, as real costs of distance remain. These can be mitigated by infrastructure improvements, measures to improve port facilities and policies to reform customs and other border procedures, although the cost-effectiveness of such measures remains to be studied.

Many countries are attracted by regional integration schemes as ways of opening their economies to trade. Thinking geographically sheds new light on these regional trading arrangements. South-South arrangements between developing countries have the effect of enlarging markets and possibly thereby promoting industrialization. But if industry concentrates due to agglomeration forces, then gains are likely to go to one member country at the expense of others. Both analytical and empirical study of this issue suggests that the likelihood of integration causing divergence is greatest in 'south-

south' agreements, and developing countries are likely to be better served by entering 'north-south' regional agreements (World Bank 2000).

Perhaps the most important international issue relates to attracting footloose industry and developing a viable cluster of activities. Here, it is much easier to point to necessary conditions than to sufficient ones. Geographical analysis alerts us to the importance of linkages and density of activity, and the difficulty of attracting activity to remote locations. A large array of policy instruments have been used in the past to try to influence the location decisions of footloose industries. These instruments include fiscal subsidies (including tax holidays) with their potential for race to the bottom type tax competition, the provision of infrastructure in industrial estates and free trade zones, the improvement of trunk infrastructure and key transshipment nodes, the provision of a trained labor force, etc. None of these instruments have been systematically successful, and many have been very costly. More empirical research, with better specified counter-factuals, is necessary to evaluate the efficacy of these policies and programs and the contexts in which they are likely to succeed or fail. One of the areas in which more research is needed before policy instruments are activated is the international relocation of industry, identifying which sectors move from clusters most readily, and which sectors create the greatest linkages in the host economy.

The second policy context is the management of urban structures. Here the literature gives two important messages. The first is the importance of concentration. We saw the efficiency gains associated with density, and policy should not inhibit achievement of these gains (particularly in an environment where to do so would risk international relocation of industry). Second, deconcentration does occur at a certain stage in a country's development. Policies that promote this – positively, rather than by damaging the existing centers – can be effective. Here again from a policy perspective it is important to determine when the level of concentration is insufficient or excessive. Many countries have had active policies and programs to discourage migration of labor or investment to the largest cities (from outright prohibition, to limiting the provision of housing and other supporting infrastructure). They also had active programs to encourage deconcentration of economic activity to targeted areas outside the major agglomerations by improving the acquired characteristics of the targeted locations. Some

of these programs have been successful and others not. But rarely have the effectiveness and the costs and benefits of the programs been evaluated critically (ex-ante or ex-post). Moreover, many ostensibly non-spatial policies, such as policies and programs to benefit specific sectors or household groups, have spatial consequences because the targeted sectors and households are not distributed uniformly across space. Sometimes these non-spatial policies can dominate explicit spatial policies thereby offsetting or negating the intended consequences of the latter.

One of the most difficult questions is; what to do with lagging regions, often remote and perhaps sparsely populated? Here we see most starkly the fundamental spatial policy issue of whether it is more cost-effective to move people or to move jobs. Can we identify thresholds of remoteness (distance) and critical mass of economic activity below which a region is unlikely to ever stand a chance of viable development (i.e. exploiting increasing returns and the potential for clustering for some subset of economic activities)? Despite many studies on regional convergence, we lack comprehensive studies of convergence in different countries at different stages of development, which would indicate for what types of region, under what policy regimes, and at what points in time convergence is likely. The question of whether to move people or jobs is most evident within countries, but also has an international dimension. What is the role of international migration – and of brain-drains – in changing international inequalities, and what should be the policy stance to such flows?

Developing answers to these and other policy issues requires a good deal of further work. Theory is needed, to see through the full equilibrium implications of policy measures, and to be able to assess the likely effects of other changes, such as new technologies that change the costs of distance. New applied analytical tools are needed, such as a cost-benefit analysis that can take into account the linkages – technological and pecuniary -- that lie at the basis of many location decisions, and that can address the discontinuous and non-marginal changes that policy might induce. And more empirical work is needed to further quantify the strength of many of the forces that we have discussed. As stated above, we know little about the strength of linkages due to pecuniary externalities, little about the location of service activities, and little about the complementarities between services and the rest of the economy. Agglomeration can

arise from direct technological spillovers, thicker labor markets, better access to suppliers and larger product markets, but many of these benefits have not yet been quantified through empirical work. Even for topics on which we have cited empirical studies such as how trade, investment, and the transmission of technology fall off sharply with distance, little is known about the strength of these relationships in developing countries.

Finally, one useful exercise that has yet to be undertaken is to relate the theoretical implications of the new economic geography literature in a more systematic way to the observations and insights of traditional urban and regional economics. Which of the latter results have a clearer theoretical underpinning now, which are inconsistent with the new theory, and what earlier observations and findings remain still without adequate theoretical foundations?

Acknowledgements:

Thanks to R. Arnott, K.H. Midelfart-Knarvik, and referees of this journal for helpful comments.

References

Abdel-Rahman, H. and M. Fujita (1990), 'Product varieties, Marshallian Externalities, and City Sizes,' *Journal of Regional Science*, 2, 165-183.

Ades, A.F. and E.L. Glaeser (1995), 'Trade and Circuses: Explaining Urban Giants,' *Quarterly Journal of Economics*, 110, 195-227.

Arthur, B. (1990), 'Silicon Valley Location Clusters,' *Mathematical Social Sciences*, 235-251.

Barro, R. and X. Sala-i-Martin (1991), 'Convergence Across State and Regions,' *Brookings Papers on Economic Activity*, 1, 107-82

Barro, R. and X. Sala-i-Martin (1992), 'Regional Growth and Migration: A Japan-USA Comparison,' NBER Working Paper No. 4038.

Bartelsman, E.J., R.J. Caballero and R.K. Lyons, (1994), 'Customer and supplier driven externalities' *American Economic Review*, 84, 1075-84.

Beardsell M., and J. V. Henderson (1999), 'Spatial Evolution of the Computer Industry in the USA,' *European Economic Review*, 43, 431-456.

Beeson, P. E., D. N. DeJong, and W. Troesken (1999), 'Population Growth in U.S. Counties, 1840-1990,' mimeo, University of Pittsburgh.

Bergsman, J., P. Greenston and R. Healy (1972), 'The Agglomeration Process in Urban Growth,' *Urban Studies*, 9, 263-88.

Black, D. and J. V. Henderson (1999a), 'Urban Evolution in the U.S.A.,' mimeo (September), Brown University.

Black, D. and J. V. Henderson (1999b), 'A Theory of Urban Growth', *Journal of Political Economy*, 107, 252-284.

Blomstrom, M. and A. Kokko (1997), 'How Foreign Investment Affects Host Countries,' Policy Research Working Paper 1745, Washington DC, World Bank..

Brainard, S.L. (1997), 'An empirical assessment of the proximity concentration trade-off between multinational sales and trade', *American Economic Review*, 87, 520-44.

Caselli, F., E. Esquire, and F. Lefort (1996), 'Re-opening the Convergence Debate: A New Look at Cross-Country Growth Empirics,' *Journal of Economic Growth*, 1, 363-389.

Chun, D.H. and K.S. Lee (1985), 'Changing Location Patterns in Population and Employment in the Seoul Region,' World Bank Report No. UDD-65.

Ciccone, P. and R. Hall (1995), 'Productivity and Density of Economic Activity,' *American Economic Review*, 86, 54-70.

Coe, D.T. and E. Helpman, (1995), 'International R&D spillovers', *European Economic Review*, 39, 859-87.

Cronon, W. (1992), *Nature's Metropolis: Chicago and the Great West*, W. W. Norton & Co., New York.

Easterly, W. and R. Levine, (1999) 'Its not factor accumulation: stylized facts and growth models', processed, World Bank, Washington DC.

Eaton, J. and Z. Eckstein (1997), 'Cities and Growth, Theory and Evidence from France and Japan,' *Regional Science and Labor Economics*, 27, 443-474.

Eberts, R.W. and D. McMillen (1999), 'Agglomeration Economies and Public Infrastructure' in P. Cheshire and E.S. Mills (eds.), *Handbook of Urban and Regional Economics*, Vol. 3, North Holland, N.Y.

Ekholm, K. (1998), 'Proximity advantages, scale economies and the location of production', in P. Braunerhjelm and K. Ekholm (eds), *The geography of multinational firms*, Kluwer, Boston.

Ellison, G. and E. Glaeser (1997), 'Geographic Concentration in US Manufacturing Industries: A Dartboard Approach,' *Journal of Political Economy*, 105(5), 889-927.

Fujita, M. (1988), 'A monopolistic competition model of spatial agglomeration; differentiated product approach', *Regional Science and urban economics*, 18, 87-124.

Fujita, M., P. Krugman and A.J. Venables (1999) *The spatial economy; cities, regions and international trade*, MIT Press, Cambridge MA.

Fujita, M. and T. Tabuchi (1997), 'Regional Growth in Post War Japan,' *Regional Science and Urban Economics*, 27, 643-670.

Gallup, J., J.Sachs, and A. Mellinger (1999), 'Climate, Navigability, and Economic Development,' mimeo, Harvard University.

Gallup, J.L and J. Sachs (1999), 'Geography and economic development', in B. Pleskovic and J.E. Stiglitz (eds) *Annual Bank Conference on Development Economics*, World Bank, Washington.

Gaspar, J. and E.L. Glaeser, (1996), 'Information Technology and the Future of Cities,' National Bureau of Economic Research Working Paper: 5562.

Glaeser, E.L., H.D. Kallal, J.A. Scheinkman, and A. Shleifer (1992), 'Growth in Cities,' *Journal of Political Economy*, 100, 1126-1152.

Guimanaes, P., O. Figueiredo and D. Woodward (2000), 'Agglomeration and the Location of Foreign Direct Investment in Portugal,' *Journal of Urban Economics*, 47, 115-135.

Hansen, E. R. (1983), 'Why Do Firms Locate Where They Do?' Water supply and urban development department discussion paper, UDD 25, World Bank.

Hansen, N. (1990), 'Impacts of Small and Intermediate-Sized Cities on Population Distribution: Issues and Responses,' *Regional Development Dialogue*, 11, 60-76.

Hanson, G., (1996), 'Agglomeration, Dispersion, and The Pioneer Firm', *Journal of Urban Economics*, 39, 255

Henderson, J.V. (1974), 'The Sizes and Types of Cities,' *American Economic Review*, 64, 640-656.

Henderson, J. V. (1988), *Urban Development: Theory, Fact and Illusion*, Oxford University Press.

Henderson, J. V. (1997), 'Medium Size Cities,' *Regional Science and Urban Economics*, 27, 583-612.

Henderson, J. V. (1999) 'How urban concentration affects economic growth' Policy research working paper 2326, World Bank.

Henderson, J. V. (1999a), 'Marshall's Scale Economies,' NBER Working Paper No. 7358.

Henderson, J. V. (1999b), 'Notes on the Costs of Urban Primacy' mimeo (10-24-99), Brown University.

Henderson J.V. (1999c) 'The Effects of Urban Concentration on Economic Growth' NBER Working Paper No. 7503.

Henderson, J. V. and A. Kuncoro (1996), 'Industrial Centralization in Indonesia', *World Bank Economic Review*, 10, 513-540.

Henderson, J. V., A. Kuncoro and M. Turner (1995), 'Industrial Development in Cities,' *Journal of Political Economy*, 103, 1067-1090.

Henderson, J. V., A. Kuncoro, and P. Nasution (1996), 'Dynamic Development in Jabotabek,' *Indonesian Bulletin of Economic Studies*, 32, 71-96.

Henderson, J. V., T. Lee and J-Y Lee (1999), 'Externalities and Industrial deconcentration under Rapid Growth,' mimeo, Brown University.

Hirschman, A. (1958), *The strategy of economic development*, New Haven, CT: Yale University Press

Hochman, O. (1977), 'A Two Factor Three Sector Model of an Economy with Cities,' mimeo.

Holmes, T. (1997), 'How Industries Migrate When Agglomeration Economies Are Important,' *Journal of Urban Economics*, 45, 240-63.

Holmes, J. (1999), 'Location of Industry and Vertical Disintegration,' *Review of Economics and Statistics*, 81, 314-25.

Hummels, D. (1999), 'Have international trade costs declined', processed University of Chicago Business School.

Jacobs, J. (1969), *The economy of cities*, New York, Random House.

Jaffe, A., M. Tratjenberg and V. Henderson (1993), 'Geographic Location of Knowledge Spillover as Evidenced by Patent Citations,' *Quarterly Journal of Economics*, 108, 577-598.

Kaiser, K. (1999), 'Pre- and Post-Liberalization Manufacturing Location in Indonesia (1975-96),' mimeo june 99, London School of Economics.

Keller, W. (1998) 'Are international R&D spillovers trade related? Analyzing spillovers among randomly matched trade partners', *European Economic Review*, 42, 1469-81.

Keller, W. (2000) 'Geographic localization of international technology diffusion', NBER working paper no 7509, Cambridge, MA.

Krugman, P.R. (1991a), 'Increasing returns and economic geography', *Journal of Political Economy*, 99, 483-99.

Krugman, P.R. (1991b), *Geography and trade*, MIT press, Cambridge MA

Krugman, P.R. (1995), *Development, geography, and economic theory*, Bradford Books.

Krugman, P.R. and A.J. Venables (1995), 'Globalization and the inequality of nations', *Quarterly Journal of Economics*, 110, 857-880.

Kwon, W.G. (1985), 'Issues and Problems in Planning and Implementing Industrial Location Policies in Korea,' discussion paper, World Bank.

Limao, N. and A.J. Venables, (1999), 'Infrastructure, geographical disadvantage and transport costs', Policy Research Working Paper no 2257, World Bank.

Marshall, A. (1890), *Principles of Economics*, MacMillan, London.

Martin, P. and C.A. Rogers (1995), 'Industrial location and public infrastructure', *European Economic Review*, 39, 335-351.

Moomaw, R. (1983), 'Spatial productivity variations in manufacturing. a critical survey of cross-sectional analysis', *International Regional Science Review*, 8, 1-22.

Moretti, E. (1999), 'Estimating the Social Return to Education: Evidence from Cross-Sectional and Longitudinal Data,' Working Paper No. 22, UC Berkeley.

Murphy, R., A. Shleifer and R. Vishny (1989) 'Industrialization and the big push', *Journal of Political Economy*, 97, 1003-1026.

Myrdal, G. (1957), *Economic theory and underdeveloped regions*, London, Duckworth.

Nakamura, R. (1985), 'Agglomeration economies in urban manufacturing industries. a case of Japanese cities,' *Journal of Urban Economics*, 17,108-124.

Puga, D. and A.J. Venables, (1999), 'Agglomeration and economic development; import substitution versus trade liberalisation', *Economic Journal*, 109, 292-311.

Quah, D. (1997) 'Empirics for growth and distribution; stratification, polarization and convergence clubs', Centre for Economic Performance discussion paper no 324, LSE, London.

Radelet, S. and J. Sachs (1998), 'Shipping costs, manufactured exports and economic growth', paper presented at American Economic Association meetings.

Redding, S.J. and A.J. Venables (2000), 'Economic geography and international inequality', discussion paper, Centre for Economic Policy Research, London.

Renaud, B. (1981), *National Urbanization Policy in Developing Countries*, Oxford University Press.

Richardson, H. (1987), 'The Costs of Urbanization: A Four-Country Comparison,' *Economic Development and Cultural Change*, 33, 561-580.

Rosen, K. and M. Resnick (1978), 'The Size Distribution of Cities, the Pareto Law, and the Primate City,' mimeo, Princeton University.

Rosenthal, S. and W. Strange (1999), 'Geography, Industrial Organization and Agglomeration,' mimeo 1999, University of British Columbia.

Rousseau, M.P. (1995), 'Les Parisiens sont surproductifs,' *Études Foncières*, 68, 13-18.

Shishido, H. and W. Wheaton (1981), 'Urban Concentration, Agglomeration Economies, and the Level of Economic Development,' *Economic Development and Cultural Change*, 30, 17-30.

Sveikauskas, L. (1978), 'The Productivity of Cities', *Quarterly Journal of Economics*, 80, 393-413.

Thomas, V. (1980), 'Spatial Differences in The Cost of Living', *Journal of Urban Economics*, 8, 108-122.

Tolley, G., J. Gardner and P. Graves (1979*), Urban Growth Policy in a Market Economy*, New York: Academic Press.

Townroe, P. (1981), 'Location Factors in the decentralization of industry, a survey of Metropolitan Sao Paulo,' Staff working paper 517, World Bank.

UN (1993), *World Urbanization Prospects: the 1992 Revision*, New York: United Nations.

UNIDESA (1998), *World Urbanization Prospects*, New York.

Venables, A.J. (1996a) 'Equilibrium locations of vertically linked industries', *International Economic Review*, 49, 341-359.

Venables, A.J. (1996b) 'Trade policy, cumulative causation and industrial development', *Journal of Development Economics*, 37, 179-197.

Venables, A.J. (2000) 'Winners and losers from regional integration agreements' discussion paper, Centre for Economic Policy Research, London.

Venables, A.J. and N. Limao (1999), 'Geographical disadvantage: a Heckscher-Ohlin-von Thunen model of international specialisation', with N. Limao, Policy Research Working Paper no 2256, World Bank.

Von Thunen, J.H. (1826) *Der isolierte Staat in Beziehung auf landtschaft under nationalokonomie*, Hamburg, (English translation by C.M. Wartenberg, *Von Thunen's Isolated State*, Oxford, Pergamon press, 1966).

Williamson, J.G. (1965), 'Regional Inequality and the Process of National Development,' *Economic Development and Cultural Change*, 13, 3-45.

World Bank (1999), *Entering the 21st Century, World Development Report 1999/2000*, Oxford University Press.

World Bank (2000), *Trade blocs*, Policy Research Report, Washington.

Endnotes

[i] Throughout the paper we use the term geography to mean the spatial relationship between economic units, not countries' endowments of land, resources, mountains, etc.
[ii] Krugman (1995) argues that the difficulty in modelling increasing returns to scale was largely responsible for the marginalisation of geography by mainstream economists.
[iii] There could still be spatial concentration in the sense of Ellison-Glaeser (1997), where, for example, rural populations might concentrate near mineral fields to save on transport costs of the raw materials for their family steel production.
[iv] This and the following section draw heavily on Fujita, Krugman and Venables (1999).
[v] The importance of aggregate demand and increasing returns for development was analysed by Murphy, Shleifer and Vishny (1989), although not in a spatial context.
[vi] This argument is developed in Venables (1996a). Having more suppliers may also make the market structure more competitive and thereby reduce price cost margins. For a development application of this argument see Venables (1996b).
[vii] See also Fujita, Krugman and Venables (1999).
[viii] See Radelet and Sachs (1998)
[ix] Shipping costs fell with the introduction of containerization, and vary with the oil price. In recent decades technical change in this sector has been no faster than in other sectors.
[x] Bartelsman, Caballero and Lyons (1994) relate industry productivity growth to activity levels in customer industries, although this work has no spatial dimension.
[xi] What is meant by convergence is not always well understood. There is general mean reversion where, on average high income [low] regions grow slower [faster], but the variance in income across regions doesn't narrow over time. There is conditional convergence where regions are converging to different steady states and there is absolute convergence where regions are converging to roughly the same steady state.