

Scale Economies and Agglomeration



The most celebrated example in economics is perhaps the simplest. On the first page of *The Wealth of Nations*, published in 1776, Adam Smith wrote of the benefits of dividing labor to make pins. A single unskilled worker without the benefit of machines might make fewer than 20 pins in a day. But in a pin factory that Smith visited, 10 workers, who divided among themselves the 18 operations involved in making a pin, were producing 48,000 pins a day. Rather than struggling to produce just a few pins a day, each worker was turning out almost 5,000. Later in Smith's classic work are two important qualifiers: the gains from dividing labor are limited by market size, and not all activities exhibit increasing returns to scale.

The ability to transport products widens the market, so cities are located near the most natural and efficient of transport systems—waterways. Places blessed with this natural infrastructure often do well, while other places must bide their time. As Smith wrote,

There are in Africa none of those great inlets, such as the Baltic and Adriatic seas in Europe, the Mediterranean and Euxine seas in both Europe and Asia, and the gulphs of Arabia, Persia, India, Bengal, and Siam in Asia, to carry maritime commerce into the interior parts of that great continent: and the great rivers of Africa are at too great a distance from one another to give occasion to any considerable inland navigation.¹

Besides, not all activities exhibit scale economies, and some do not need large markets to thrive. Subsistence farming is

one such occupation, fruitfully carried out in villages. But such trades as manufacturing and commerce can be carried out only in bigger settlements, because they require access to both workers and customers.

Caveats notwithstanding, the benefits of producing large quantities in a single plant or place have increased as transport costs have fallen in the two centuries since Smith visited the pin factory. Those who doubt the awesome potential of scale economies and how access to world markets helps exploit them should visit Dongguan, a city half-way between Guangzhou and Shenzhen in Southeast China. Until the 1980s it was a collection of sleepy villages in China's Pearl River delta. Since then it has rushed headlong into the world of increasing returns (see box 4.1). Every year, millions of people in the developing world enter this new realm and the implications, for them and for policy makers, are nothing short of revolutionary.

This chapter summarizes the experience of entrepreneurs over the last two centuries in exploiting economies of scale in production. It focuses on "agglomeration economies," whose exploitation requires locating in areas densely populated by other producers. It next provides a brief synopsis of about two decades of work by economists seeking to understand these scale economies—work that has diminished the disconnect between research and the real world, and that yields valuable policy insights. It then assesses whether policy makers in the developing world have been learning from this experience and analysis.

BOX 4.1 *Scale economies in an almost unreal world: the story of Dongguan, China*

In 1978 what today is the city of Dongguan in China's Guangdong province was but a collection of villages and small towns spread over 2,500 square kilometers on the Pearl River, midway between Guangzhou to the north and Shenzhen and Hong Kong, China, to the south. The area's population of 400,000 relied on fishing and farming and—though not the poorest in China—was not especially prosperous.

Today Dongguan is home to about 7 million people. More than 5 million of its residents are migrants who work in the thousands of factories that dot the city, churning out a wide range of products in such huge volumes that recent media accounts have assigned Dongguan the label of "factory of the world." Dongguan's economy has grown at more than 20 percent annually since 1980, and in 2004 its gross domestic product (GDP) was about \$14 billion—greater than Iceland's. If one includes only registered urban residents (as in official statistics), Dongguan's GDP per capita of \$9,000 in 2004 made it the wealthiest city in China. Even if the city's floating population of migrant workers is included, its GDP per capita in 2004 was still more than \$2,000.

Dongguan's development since the 1970s, and particularly in the last decade, exemplifies (perhaps in exaggerated fashion) the economic forces shaping East Asia's middle-income economies (see the table below).

Location and favorable factor prices undoubtedly spurred Dongguan's early growth. For the first decade and a half after China's reforms began, small and medium enterprises from both Hong Kong, China, and Taiwan, China, were attracted to Dongguan by plentiful supply of land and low-cost labor, and by its proximity to both Guangzhou and Hong Kong, China. Despite these factors, Dongguan's rapid growth in the 1990s can best be understood through **economies of scale**, whether in the production of intermediate goods or differentiated products, and **agglomeration effects**, within and across industries. Combined with reductions in transport costs and improvements in logistics, technological progress demonstrates that such effects have emerged as important characteristics of global production.

The internal scale economies are obvious. In 2005 a single plant in Dongguan manufactured more than 30 percent of

the magnetic recording heads used in hard disk drives worldwide. Another produced 60 percent of the electronic learning devices sold in the U.S. market. A third produced nearly 30 million mobile phones, more than enough to provide a mobile phone for every man, woman, and child in Peru or República Bolivariana de Venezuela.

Agglomeration or external scale economies are equally visible. The knowledge spillovers and lower logistics costs from locating close to input providers and export traders have produced globally important industry clusters for knitted woollens, footwear, furniture, and toys. But the cluster that has dominated the industrial landscape of Dongguan since the mid-1990s is telecommunications, electronics, and computer components. Of the parts and components used in manufacturing and processing personal computers, 95 percent can be sourced in Dongguan, and for several products, Dongguan's factories account for more than 40 percent of global production.

Contributed by Shubham Chaudhuri.
Source: Gill and Kharas 2007.

Dongguan in numbers

Average annual GDP growth, 1980–2005 (%)	22.0	GDP (US\$, billions)	14.2
Population: registered residents (millions)	1.6	Population: total, estimated (millions)	7.0
GDP per registered resident (US\$)	8,999	GDP per capita (US\$)	2,070
Exports (US\$, billions)	35.2	Imports (billion US\$)	29.3
Government revenues (US\$, billions)	1.0	Government expenditures (US\$, billions)	1.2
Electricity consumption (kWh, billions)	35.2	Water consumption (ft ³ , billions)	1.5

Environmental impact indicators

Sulfur dioxide emissions (tons, thousands)	199.4	Industrial waste water (tons, millions)	225
Sulfur dioxide emissions meeting standards (%)	92.9	Industrial water discharge meeting standards (%)	90.1
Industrial solid wastes (tons, thousands)	28.6	Industrial solid wastes meeting standards (%)	86.5

Global market share in 2002 of computer and electronics components manufactured in Dongguan (%)

Magnetic heads and computer cases	40	Scanners and mini-motors	20
Copper-clad boards and disk drives	30	Keyboards	16
AC capacitors and fly-back transformers	25	Motherboards	15

Source: Dongguan Government 2005.

The main findings:

- **Developing economies are entering a new realm of agglomeration.** A century of experience indicates that as countries develop from agricultural to industrial to service-oriented production, entrepreneurs and workers leave behind not just their villages and their agrarian occupations, but also a world in which scale does not matter much. More and more of them enter not just larger and denser settlements, but also a world in which scale matters—where production and distribution enjoy scale economies, especially those associated with places. Proximity matters more, not just for access to markets for goods and services, but also for access to ideas.
- **A portfolio of places is needed for economic growth.** Research over the last generation indicates that different forms of human settlement facilitate agglomeration economies for different forms of production. A somewhat-oversimplified (but not altogether incorrect) generalization would be that market towns facilitate scale economies in marketing and distributing agricultural produce, medium-size cities provide localization economies for manufacturing industries, and the largest cities provide diverse facilities and foster innovation in business, government, and education services.
- **Policy makers have often misjudged the potency of market forces.** Many policy makers perceive cities as constructs of the state—to be managed and manipulated to serve some social objective. In reality, cities and towns, just like firms and farms, are creatures of the market. Just as firms and farms deliver final and intermediate goods and services, towns and cities deliver agglomeration economies to producers and workers. So city administrators are better advised to learn what their city does, and to help it do this well, rather than try to abruptly change the course of their city's destiny. Planners and policy makers should see their role as prudent managers of a portfolio of places, to get the most from agglomeration economies.

Table 4.1 A dozen economies of scale

Type of economy of scale			Example	
Internal	1. Pecuniary		Being able to purchase intermediate inputs at volume discounts	
	Technological	2. Static technological	Falling average costs because of fixed costs of operating a plant	
		3. Dynamic technological	Learning to operate a plant more efficiently over time	
External or agglomeration	Localization	Static	4. "Shopping"	Shoppers are attracted to places where there are many sellers
			5. "Adam Smith" specialization	Outsourcing allows both the upstream input suppliers and downstream firms to profit from productivity gains because of specialization
			6. "Marshall" labor pooling	Workers with industry-specific skills are attracted to a location where there is a greater concentration. ^a
		Dynamic	7. "Marshall-Arrow-Romer" learning by doing	Reductions in costs that arise from repeated and continuous production activity over time and which spill over between firms in the same place
	Urbanization	Static	8. "Jane Jacobs" innovation	The more that different things are done locally, the more opportunity there is for observing and adapting ideas from others
			9. "Marshall" labor pooling	Workers in an industry bring innovations to firms in other industries; similar to no. 6 above, but the benefit arises from the diversity of industries in one location.
			10. "Adam Smith" division of labor	Similar to no. 5 above, the main difference being that the division of labor is made possible by the existence of many different buying industries in the same place
		Dynamic	11. "Romer" endogenous growth	The larger the market, the higher the profit; the more attractive the location to firms, the more jobs there are; the more labor pools there, the larger the market—and so on
	12. "Pure" agglomeration		Spreading fixed costs of infrastructure over more taxpayers; diseconomies arise from congestion and pollution	

Source: Adapted from Kilkeny 2006.

a. For a formalization, see Krugman 1991a.

This chapter discusses, in general terms, the implications of experience and analysis for reshaping urbanization strategies in the developing world. Chapter 7 continues this task of reframing the debate over urban strategies.

A guide to scale economies

The benefits of increasing scale can be either internal or external to an individual firm or farm. External economies are synonymous with “agglomeration economies,” which include the benefits of localization (being near other producers of the same commodity or service) and urbanization (being close to producers of a wide range of commodities and services). Consumption externalities also are associated with agglomeration, but these are not yet well studied in the literature.² So, this chapter deals with production-related scale economies (see table 4.1).³

- **Internal economies** arise from the larger size of a plant to better exploit fixed costs (numbers 1 through 3 in table 4.1). A larger steel mill can get volume discounts from suppliers—implying fixed costs of transport and trade—and reap the benefits of dividing labor within the firm.
- **Localization economies** arise from a larger number of firms in the same industry and the same place (numbers 4 through 7 in table 4.1). Spatial proximity helps because immediate access to competitors in the same sector allow firms to stay abreast of market information in negotiating with customers and suppliers.⁴ Clustered firms can also share a larger and more dependable pool of specialized labor.
- **Urbanization economies** arise from a larger number of different industries in the same place (numbers 8 through 11 in table 4.1). A management consulting company can benefit from locating near business schools, financial service providers, and manufacturers.

Agglomeration economies depend not just on size (a big city or industry) but also on urban interactions. They are traditionally classified as localization economies arising from *within-industry* economic

BOX 4.2 *Sharing, matching, and learning*

Three reasons explain why firms in a particular industry often locate close to each other. Geographic concentration helps in—

- **Sharing.** Broadening the market for input suppliers, allowing them to exploit internal economies of scale in production (average costs decline as the scale of production rises). This sharing of inputs also permits suppliers to provide highly specialized goods and services tailored to the needs of their buyers. The result is higher profits for all, accompanied by easier access to a broader range of inputs.
- **Matching.** Expanding the availability of the range of skills required by employers to facilitate better matching to their distinctive needs. At the same time, workers find it less risky to be in locations with many possible employers.
- **Learning.** Accelerating spillovers of knowledge and allowing workers and entrepreneurs to learn from each other.

The ability to go beyond industry-specific sharing, matching, and learning (localization economies) to citywide processes (urbanization economies) requires additional mechanisms. These include the effects of cumulative causation and the interpenetration of production and trade across industries. They also include gains from the cross-fertilization of ideas. The concentration of workers and suppliers leads to a concentration of consumer demands.

If economies of scale are large and unexhausted, and if firms can compete not only on price but also through product differentiation, strong centripetal forces come into play. In addition, by formally introducing distance (the cost of shipping inputs and outputs), the framework used in this Report provides useful insights into the centrifugal forces that explain spatial dispersion in a country.

Sources: Gill and Khars 2007, based on Duranton and Puga 2004.

interactions, and as urbanization economies, arising from *between-industry* interactions.⁵ The reasons for producers to gain from proximity to others depend on the sharing of capital inputs, information, and labor. They also depend on improving the matches between production requirements and types of land, labor, and intermediate inputs—and learning about new techniques and products (see box 4.2).

Internal scale economies are higher in heavier industries

Internal increasing returns to scale are found in manufacturing and services, based on various sources of data. The internal scale economies range from negligible or low among light industries, to high among heavy and high-technology industries (see table 4.2). Based on engineering estimates, a summary of sector-specific studies that examines the minimum efficiency scale of production and cost-saving finds significant increasing returns in motor vehicles, other

Table 4.2 Internal scale economies are low in light industries and high in heavy industries

Findings	Data source
<i>Constant returns to scale:</i> apparel, leather, footwear, textiles, wood products	Based on trade data (Antweiler and Trefler 2002)
<i>High increasing returns to scale:</i> machinery, pharmaceuticals, instruments, iron and steel, petroleum and coal products	
<i>Constant returns or low increasing returns to scale:</i> leather goods, footwear and clothing, timber and wood, textiles	Based on engineering estimates to examine cost gradients and changes in minimum efficiency scale (Junius 1997, cited from Prateen 1988 and Emerson and others 1988)
<i>High increasing returns to scale:</i> Motor vehicles, other means of transportation, chemicals, engineering, printing and publishing	
<i>Low increasing returns to scale:</i> footwear, apparel, food products, leather	Based on markups in manufacturing industries for 14 OECD countries (Junius 1997, cited from Oliveira and others 1996)
<i>High increasing returns to scale:</i> tobacco, pharmaceuticals, office and computing machinery, railroad equipment	
<i>Low increasing returns to scale:</i> apparel, leather products, textiles	Based on markups of prices over marginal costs for two-digit sectors in the United States covering 1953–84 (24 sectors) (Roeger 1995)
<i>High increasing returns to scale:</i> electric, gas, and sanitary services, motor vehicles and equipment, chemicals, tobacco	
<i>Low increasing returns to scale:</i> textiles, milk products, lumber mills, fish oil and meal products	Based on production function estimates for 1963 Census of Manufacturing Establishments in Norway (27 industries) (Griliches and Ringstad 1971)
<i>High increasing returns to scale:</i> basic metal, transport equipment, cement products, fixtures, beverages	
<i>Low increasing returns to scale:</i> clothing, knitting, leather, textiles	Based on cost and profit data in (167 industries) four-digit SIC manufacturing industries for 1970 in Canada (Baldwin and Gorecki 1986) and labor productivity and output estimates for 90 four-digit industries in Canada between 1965 and 1970 (Gupta 1983)
<i>High increasing returns to scale:</i> petroleum, basic and fabricated metal, transport equipment	
<i>Low increasing returns to scale:</i> apparel, wood products	Based on estimates of firm-level production function estimates for 6,665 plants in Chile during 1979–86 (Levinsohn and Petrin 1999)
<i>High increasing returns to scale:</i> other chemicals, food products, printing and publishing	

Source: WDR 2009 team.

Note: OECD = Organisation for Economic Co-operation and Development; SIC = Standard Industrial Classification.

transport equipment, chemicals, machinery, engineering, and paper and printing. In the three-digit product category, the highest returns to scale are in books, bricks, dyes, and aircraft.⁶ By contrast, internal scale economies are negligible in rubber and plastics, leather and leather goods, footwear and clothing, and textiles.⁷

Based on cost and value added estimates, different sources point to similar findings. A sample of 5,000 manufacturing firms in Norway shows evidence of scale economies at the individual industry level.⁸ For Canadian industries at the four-digit level, returns to scale average 10 percent for 107

manufacturing sectors, with clothing, knitting, leather, and textiles at the lower end of the spectrum.⁹ Increasing output cuts costs in U.S. manufacturing, in the industries of middle-income countries (Chile), and in the European car, truck, and consumer durables industries.^{10,11}

Based on trade data, a third of all goods-producing industries have increasing returns to scale.^{12,13} Manufacturing industries with the highest plant-level economies and industry-level externalities are petroleum and coal products, petroleum refining, pharmaceuticals, machinery, and iron and steel. Industries with constant returns include footwear, leather, textiles, apparel, and furniture.

Markups are another source of information. Because increasing returns to scale confer market power on firms, markups of price over marginal cost can be a proxy for plant-level scale economies. Studies find a range of markups for U.S. manufacturing, from 15 percent in apparel to more than 200 percent in the electric, gas, and sanitary services. For 36 manufacturing sectors across 19 member countries of the Organisation for Economic Co-operation and Development (OECD), the highest markups are in tobacco, drugs and medicines, and office and computing machinery—and the lowest in footwear, apparel, and wood products.¹⁴

While manufacturing data dominate the literature, increasing returns in services also are evident. The best-studied sector is electric power generation, where the internal increasing returns to scale are considerable.¹⁵ The highest markups are in utilities and sanitary services.¹⁶ Scale economies also are found in banking and finance.¹⁷ A study of commercial banks in 75 countries shows that banks with larger loans and deposits have lower average costs—and that banks operating in larger financial systems require less proportionate increases in financial capital and have lower risk management costs.¹⁸

Localization economies arise from input-sharing and competition within the industry

Localization economies come from geographically concentrated groups of firms,

linked by the technology they use, the markets they serve, the products and services they provide, and the skills they require. Competitive pressures that force firms in the same sector to innovate or fail also lead to productivity growth. Conditions tend to be competitive when upstream and downstream firms and associated institutions in a particular industry (say, electronic machinery or petrochemicals)—including universities and trade associations—“cluster” together. Other channels for localization economies are the less easily measured “Marshall-Arrow-Romer externalities,”¹⁹ which come mainly from knowledge spillovers.

Proximity to similar firms influences the location decisions of firms. Consider the hosiery industry in the United States. Shortly after 1900 New York City became the U.S. center for garment production and distribution. But after World War II garment production moved south, to North Carolina.²⁰ Many knitting and weaving mills moved to be closer to the supply of yarn and to take advantage of cheaper power, labor, and land.

Today, the hosiery industry, localized in North Carolina, boasts many brands—among them, *Sheer Energy*, *Silken Mist*, *Just My Size*, and *No Nonsense*—all competing in a \$2 billion market. According to the U.S. Census Bureau, about 150 establishments producing women’s full-length and knee-length hosiery in the early 2000s, half the nation’s total, were located in North Carolina. They shipped \$973 million worth of hosiery, about 75 percent of the national total, and employed 13,497 people, including 11,567 production workers.²¹ Adding men’s socks and stockings, more than half of a \$6 billion industry is in North Carolina.

One reason textile producers went to North Carolina was to exploit productivity gains from proximity to upstream yarn producers. The yarn and pantyhose industries are tightly knit—in relationships delicately stitched together at each step of production—but fiercely competitive. Macfield, a textile giant and a leading producer of yarns for pantyhose, socks, outerwear, upholstery, and industrial products operates five plants in North Carolina and employs about a quarter of the yarn

industry’s labor force. About six of every 10 pairs of sheer hosiery sold in the United States were knitted with Macfield yarn.²² Together with other large North Carolina producers (Unifi, Regal, and Spanco), they make up more than three-quarters of the industry’s \$3.7 billion worth of textured yarn products.²³ The localization of the yarn and hosiery industries in North Carolina is a powerful manifestation of intra-industry external economies.

Urbanization economies come from industrial diversity that fosters innovation

As cities grow, urbanization economies become more important.²⁴ Urban diversity can foster the exchange of ideas and technology to produce greater innovation and growth.²⁵ Firms in different industries can share indivisible facilities or public goods, a wider variety of intermediate input suppliers, a larger pool of narrowly specialized workers, and risks. The evidence of greater importance of across-industry knowledge spillovers can be seen in established cities. In fairly mature cities, such as Los Angeles, and Philadelphia, competition and city diversity help employment growth, indicative of urbanization economies of between-sector innovation.²⁶ On New York’s Wall Street and in the city of London, financial firms, insurance companies, and banking syndicates benefit from being close to one another. And co-location stimulates the growth of other specialist services, such as legal, software, data processing, advertising, and management consulting firms. These clustered firms, by providing a thicker market for highly educated individuals, benefit from drawing on the same large pool of human capital. They also gain from the generation and diffusion of knowledge amongst one another.

Evidence of urban agglomeration economies comes primarily from developed countries.²⁷ But there is also evidence of external economies in developing countries, wherever data are available. A survey of 12,400 manufacturing firms in 120 cities in China points to the higher productivity of firms in more populous cities.²⁸ Agglomeration economies in Indonesian

BOX 4.3 *Agglomeration economies in Indonesia*

Much of the rigorous evidence of agglomeration economies comes from developed countries. An exception is Indonesia, where recent research helps to identify the determinants of industrial concentration. The analysis focuses on four broad groups—chemicals (including petroleum, rubber, and plastics); textiles (including garments, leather, and footwear); nonmetallic minerals (including glass, ceramics, and cement); and machinery (including electrical and nonelectrical machines, transportation equipment, and instruments). It sheds light on how the size and type of scale economies influence the extent and pattern of agglomeration in a developing country.

Localization economies—the benefits of locating near other firms in the same industry—have been more important than urbanization economies for manufacturing, and static agglomeration economies are more important than dynamic (or learning related) externalities. The sector-specific findings of tests for static externalities show that

- Localization economies are strong for textiles and chemicals.
- Urbanization economies are strong for nonmetallic minerals and machinery, though weak during some periods.

Activities subject to urbanization and dynamic economies are poor candidates for policies that seek to spread out economic mass within a country (see chapter 8 for a more detailed discussion). Such firms prefer to stay put, since this helps learning, and they thrive in fairly large and diverse cities. The agglomeration economies for textiles and chemicals (largely static and local) indicate that policies to deconcentrate production in these industries might succeed if accompanied by improvements in infrastructure and governance in the areas chosen for relocation. The agglomeration economies make the nonmetallic minerals and machinery (essentially static and urban) likely to resist relocation to smaller urban centers.

Source: Kuncoro, forthcoming.

manufacturing between 1980 and 2003 vary over time; however, in the broadest terms, these benefits are mainly static rather than dynamic and somewhat more likely to arise from localization than from urbanization (see box 4.3).

A different realm

Countries develop by shifting their economies from traditional subsistence-based agricultural activities to higher-value manufacturing and services. Along the way, firms rather than farms become the dominant production unit. The production of differentiated manufactured goods and services increases as a share of the economy's output. Between 1900 and 2000 the share of the global population in industrial or service-dominant urban localities rose from 15 percent to 47 percent. The global employment share in agriculture among working-age

people fell from just over 55 percent in 1960 to about 33 percent in 2004. Production technology shifts away from constant returns to increasing returns to scale. And over time, scale-augmenting technical change boosts scale economies. Imperfect and monopolistic competition become the dominant forms of market structure.

The world is more urban, and the concentration of economic mass in the densest urban centers is greater as well. In 1900 the number of people in the largest 100 cities added up to just 4.3 percent of the world's population. The same 100 cities now have 7.5 percent of the total, and the largest 100 cities, almost 10.5 percent. Despite ample open space, almost all recent development in the United States has been less than 1 kilometer from earlier developments.²⁹ Even today, only about 2 percent of the land area of the United States is built up or paved. Only agglomeration economies can explain this extreme clustering of firms and workers in cities.

As producers seek scale economies, agriculture disperses but manufacturing clusters

As economies develop, farms spread out to exploit scale economies in production. In the United States about 1,500 kilograms of agricultural products are produced annually to feed each American, whereas the Chinese make do with about 600 kilograms per person. In 2005 the average cropland in the United States was 20.4 hectares per farmer, in Australia it was 45 hectares, and in Canada it was 47 hectares. Average farm size in Brazil is about 19 hectares.³⁰ But scale economies in agriculture are generally difficult to obtain in low-income countries. The cropland per farmer was a fraction of that in developed countries: 0.16 hectares in China, 0.30 in Bangladesh and Indonesia, and 1.20 in Nigeria.³¹

As economies develop, manufacturing and services become more important, firms cram in closer together to harness agglomeration economies. In France, the United Kingdom, and the United States, 75–95 percent of industry is localized (clustered or concentrated relative to overall economic activity), while less than 15

percent is dispersed.³² In the United States more than a third of aerospace engines are produced in three cities: Hartford with about 18 percent of total employment, and Cincinnati and Phoenix with another 18 percent together.³³ Over time the spatial concentration of industries in U.S. states has increased.³⁴ Using continuous space without considering administrative boundaries and based on concentration of plants, more than half of the United Kingdom's 122 four-digit industries are localized, and only 24 percent are dispersed. The rest are randomly distributed.³⁵

Spatial clustering is more pronounced with high-skill and high-technology industries (electronic computing machinery, process control instruments, semiconductors, and pharmaceuticals) than light industries. This is consistent with the documented findings of higher-scale effects in heavier industries. High-skill and high-tech industries have more capital-intensive production technology. They are also likely to benefit more from the various mechanisms that generate external economies (discussed earlier).

In the Republic of Korea the ranking of industries by their localization economies follows the ranking of industries by their spatial concentration across cities. Heavy and transport industries (metals, chemicals, and transport equipment) tend to be concentrated in a few highly specialized cities to take advantage of local scale externalities, while traditional or light industries with low scale externalities (food and textiles) are more dispersed.³⁶ High-tech industries (computers, aircraft, medical instruments, and electronic components) tend to be more concentrated than durable-good, machinery-related industries (metal works, industrial, refrigeration, and machinery and equipment).³⁷ Cities in the Republic of Korea have also become more specialized.³⁸

Services become even more densely clustered than manufacturing

As countries move to a more mature phase of development, their economies become more knowledge based and service oriented. The spatial concentration of activity

also rises (see chapter 2).^{39,40} The important types of agglomeration economy change as development progresses. In particular, as an economy becomes more knowledge based, knowledge spillovers, which require proximity, become more important. Evidence suggests that knowledge industries are spatially concentrated.⁴¹

Services are even more spatially concentrated than manufacturing—for two reasons. First, they tend to use less land per employee. Banks, insurance companies, hospitals, and schools can operate comfortably in high-rise buildings that economize on land and allow for high density. Second, because of external economies, business services have even greater potential for agglomeration, as firms serve one another: every bank needs advertising, every advertising firm a bank account. The potential for codependence and agglomeration is thus intrinsic to services.⁴²

Services are prominent among the most agglomerated industries in the United States.⁴³ Larger cities have been amassing service jobs from areas less than 20 kilometers away.⁴⁴ Between 1972 and 1992, jobs in the United States became more spatially concentrated, driven primarily by the rising localization of service activities in larger cities,⁴⁵ as small and medium-size counties lost jobs to the more urban areas.⁴⁶ For instance, in Suffolk County, Massachusetts, which includes Boston, 35 percent of the workforce is in business services, nearly twice the national average of 18 percent.⁴⁷ In the United Kingdom nearly 60 percent of all venture capital offices are in London.⁴⁸ London-based venture capital offices favor investment in London-based small and medium enterprises to get better information: they can easily visit and monitor these enterprises. As communication costs fall, services become more tradable, allowing providers to take advantage of narrower specialization and agglomeration economies. For instance, financial services can be disaggregated into more refined categories of retail banking, consumer credit and financing, commercial and corporate banking, investment banking, and so on. And within investment banking, there is further specialization

in mergers and acquisitions, corporate finance, fixed income, debt management, and the like.

Cities facilitate scale economies of all types

A plant in an isolated location can benefit from internal scale economies, but unless it is situated in an area of density, it cannot enjoy the competitive benefits associated with localization or urbanization economies. Towns and cities bring together large pools of skilled labor and suppliers of specialized intermediate inputs and by doing so, enhance employer-employee and buyer-seller matches. Input-sharing is an important channel for agglomeration economies.⁴⁹ Density of activity allows more refined specialization and a wider variety of intermediate inputs. Averaging across industries, a firm's relocation from a less-dense location (of 499 or fewer neighboring employees in the same industry) to a denser location (of 10,000–24,999 neighboring employees) results in a 3 percent increase in purchased input intensity.⁵⁰ The composition of a city emerges from the scope for agglomeration economies and their interaction with other aspects of microeconomic behavior.

Large cities with more firms allow workers to hedge against sector-specific risks. Smaller specialized cities expose workers to greater industry-specific shocks but provide favorable match-specific advantages. In both cases the concentration of economic activity lowers the search costs between firms and workers, which results in fewer unfilled vacancies, lower risk of job loss, and shorter durations of unemployment. The large variety and quantity of inputs to share in cities also implies better quality-matching. For instance, because of the better matching possible, married couples with university education, are increasingly found in large cities, up from 32 percent in 1940 to 50 percent in 1990.⁵¹ Cities make it easier for producers to find inputs and for customers to experiment and discover new possibilities. Examples of easy diffusion of information and social learning range from the congregation of diners in certain restaurants, to the propagation of

rumors, to the word-of-mouth learning in neighborhoods.⁵²

Learning mechanisms also explain agglomeration in cities.⁵³ As Alfred Marshall implied, when knowledge spillovers exist, "The mysteries of the trade become no mysteries but are as it were in the air."⁵⁴ Knowledge spillovers are difficult to measure, because they can seldom be traced through transactions. With patent citations, however, it is possible to identify a paper trail for some knowledge spillovers. U.S. patent citations are spatially concentrated, with citations 5 to 10 times more likely to come from the same standard metropolitan statistical area as originator patents.⁵⁵ Another strand of research focuses on workers as the primary vehicles of knowledge, implying that economies with substantial labor mobility across industries will exhibit a greater spread of ideas and growth.⁵⁶

Agglomeration economies are amplified by density and attenuated by distance

Cities obviously reflect the demand for density. People choose to live close to one another, paying high rents and tolerating crime and congestion. This density helps reduce distances of all types. Cities are thus a natural market creator and a conduit for internal and external scale economies. Firms are drawn to dense areas concentrated with people and infrastructure by the possibility of serving a large local market from a large plant at low transport costs.⁵⁷ Increasing return-to-scale production technology leads to large factories with many workers. The sizable workforce forms a large local market. By reducing transport costs, cities with a large local demand attract firms in different industries. So a self-reinforcing process of agglomeration that begins with the expanding local market further raises industry productivity.

Plants in dense economic environments tend to be larger.⁵⁸ As local market scale increases, firms are more likely to outsource their service functions to local suppliers.⁵⁹ This outsourcing further encourages competition and diversity in the local business service market, which reinforces outsourcing. Firms are attracted to locations with

large concentrations of other firms in their industry and with large demand.⁶⁰ The large and growing academic literature suggests that doubling city size will increase productivity by 3–8 percent.⁶¹ In the Republic of Korea, a plant in a city with 1,000 workers could, without altering its input mix, increase output by 20–25 percent simply by relocating to a city that has 15,000 workers in the same industry.⁶² And the spatial concentration of people reduces the cost of producing knowledge because information transmission, competition, spying, imitation, learning, innovation, and the commercialization of new ideas are easier.⁶³ In the United States a staggering 96 percent of innovations occur in metropolitan areas.⁶⁴

Agglomeration economies are influenced by geographic scope, and the density of economic activity and the distance between economic agents influence the productivity gains from scale economies (see table 4.3). For example, doubling the density of economic activity in European Nomenclature of Territorial Units for Statistics (NUTS1) regions can increase total factor productivity growth by 0.42 percentage points a year.⁶⁵ Evidence from Brazil and the United States indicates that doubling the distance to dense metropolitan centers reduces productivity by 15 percent; doubling the distance from 280 to 550 kilometers reduces profits by 6 percent. The concept of distance can be generalized, in this context, from distance in physical space to distance in industrial space. For example, spillovers between industries are more likely if industries share related scientific facilities.⁶⁶ Furthermore, the extent to which distance attenuates agglomeration economies differs for different types of agglomeration. For example, knowledge spillovers that rely on face-to-face communication decay more quickly with distance than the home market effect.⁶⁷

A portfolio of places

Adam Smith introduced scale economies, factor mobility, and transport costs as central to understanding the nature and causes of the wealth of nations. But until the 1980s most economists were happier to anchor their inquiries on another concept

Table 4.3 Scale economies amplify with density and attenuate with distance

Finding	Data sources
Scale economies amplify with density . . .	
Doubling economic density increases productivity by 6 percent	1988 data on output per worker in U.S. states (Ciccone and Hall 1996)
Doubling employment density increases productivity by 4.5–5.0 percent	Data for the late 1980s on nonagricultural private value added per worker in European NUTS regions (Ciccone 2002)
A one-standard-deviation increase in the share of own-industry local employment in the first period will raise that industry's employment level by 16–31 percent in a later period	Data on five traditional manufacturing industries in 224 U.S. metropolitan areas between 1970 and 1987 (Henderson, Kuncoro, and Turner 1995)
A 10-percent increase in local own-industry employment results in 0.6–0.8 percent increase in plant output, for the same level of inputs	Republic of Korea city-industry data for 1983, 1989, 1991–93 (Henderson, Lee, and Lee 2001)
and attenuate with distance.	
Increasing distance from the city center by 1 percent leads to a 0.13 percent decline in productivity	1980 data for 356 new manufacturing firms in Brazil (Hansen 1990)
Doubling the distance to a regional market center lowers profits by 6 percent	Firm data in auto-component and agricultural machinery in Brazil and the United States (Henderson 1994)
Doubling travel time to a city center reduces productivity by 15 percent	Data for eight industries in Brazil (Sveikaukas and others 1985)
Own-county (lagged and contemporaneous) effect on plant productivity, but no effect from neighboring county	Plant-level data on productivity, 1972–92, in 742 U.S. counties (Henderson 2003b)
Effects of own-industry employment on new plant openings attenuate rapidly within the first five 1-mile concentric rings	12 million U.S. establishments from Dun & Bradstreet Marketplace database (Rosenthal and Strange 2003)

Source: WDR 2009 team.

introduced in *The Wealth of Nations*, that of the “invisible hand” of perfect competition.⁶⁸ But perfect competition is an artificial theoretical construct: it assumes a large number of infinitesimal firms with negligible influence over market prices, even in the immediate vicinity of the firm's location. Its assumption of constant returns to scale further implies the so-called problem of “backyard capitalism.”⁶⁹ That is, in the world of constant returns to scale, small-scale production is as efficient as large-scale production, so every household should be producing a fully diversified range of goods and services in its own backyard. Economics professors, when pressed by students to give a real-world example of such an industry, would offer subsistence agriculture—small farms producing wheat or rice, whose produce could not be distinguished from those of others. Never mind that most people no

longer worked on small farms in countries that had grown out of poverty. It led to convenient characterizations of the economy in which all firms and workers were identical, so one firm or worker could be considered representative of all. Scale economies were inconvenient—they required acknowledging that specialization differentiated people and products.

Occasionally, the contradiction between internal increasing returns and perfect competition would surface, but because of the technical difficulties it raised, it quickly would be buried again.⁷⁰ Then, during the 1970s, two economists at Princeton University proposed a technical solution to model increasing returns to scale, opening a door for researchers to the same realm that so many firms and workers had inhabited since the industrial revolution.⁷¹

By the late 1980s scale economies were standard features of the explanations for international trade. By the early 1990s, growth theorists had accepted the need to incorporate imperfect competition among firms into aggregate formulations of an economy. By the mid-1990s, theorists were beginning to show how these ideas could be used to understand the spatial distribution of economic activity, including the rise of towns and cities. With the new economic geography, researchers came to realize that the dichotomy between internal and external economies is often false. Why? Because, in modeling the microfoundations of agglomeration economies, the source of external economies have often been found in the interaction of internal scale economies with other influences, such as transport costs.

Table 4.4 Thirty years of theoretical advance recognize the importance of scale economies

Subject	Main insights	Key publications
Industrial organization, 1970s	Increasing returns to scale and imperfect competition can be incorporated into formal economic models	Spence 1976; Dixit and Stiglitz 1977
Urban economics, 1970s	External economies within cities and systems of cities; different levels of agglomerations are related to city functions	Mills 1972; Diamond and Mirrless 1973; and Henderson 1974
International trade, 1980s	Increasing returns and imperfect competition explain intraindustry trade between countries with similar endowments; initial endowments may, through trade and specialization, influence the long-run rate of growth; trade unleashes forces of both convergence and divergence	Krugman 1980, 1981; Ethier 1982; Helpman and Krugman 1985; Grossman and Helpman 1995
Economic geography, 1990s	Increasing returns-to-scale activities are characterized by agglomeration and imperfect competition, while constant returns-to-scale activities remain dispersed and competitive, helping to explain spatial distribution of economic activity and growth of cities	Krugman 1991; Fujita, Krugman, and Venables 1999; Henderson 1999
Endogenous growth, 1980s	Perfect competition and knowledge-related or human capital-related externalities imply aggregate increasing returns and explain why growth rates may not fall over time and why wealth levels across countries do not converge	Romer 1986; Lucas Jr. 1988
Endogenous growth, 1990s	Imperfect competition explains why the incentive to spend on R&D does not fall, and knowledge spillovers explain why R&D costs fall over time, resulting in more and better products that fuel growth	Romer 1990; Grossman and Helpman 1991; Aghion and Howitt 1992
Endogenous growth, 2000s	Imperfect competition and Schumpeterian entry and exit of firms, with entrants bringing new technologies, explain how a country's growth and optimal policies vary with distance to the technology frontier; knowledge accumulation in cities leads to growth	Aghion and Howitt 2005; Rossi-Hansberg and Wright 2007; Duranton 2007

Source: Adapted from Gill and Kharas 2007.

Recognizing scale economies: recent theoretical advances

The literature on the microeconomic foundations of agglomeration economies flourished in the last 20 years by combining models in the paradigms summarized in table 4.4 and insights about urban economics that emphasize the tension between benefits from the concentration of economic activity and costs arising from that spatial concentration.⁷² In general, researchers have progressively recognized that economic growth has different impacts on firms and workers depending on their sector and location. The underlying reason is the love for variety in consumption and the economies of scale in production; the proximate reasons are product differentiation, monopolistic power, specialization, and location externalities.

The formal recognition of scale economies, externalities, and imperfect competition makes economic theory conform more closely to the world in which policy makers live. The policy implications of this work arise from the way economic production relates to trade, ideas, and cities.

- *Intraindustry trade.* The main insight coming from a formal recognition of increasing returns to scale and product differentiation is that trade may take

place between economies that are similar in factor endowments: both interindustry and intraindustry trade may profitably take place. The main implication is that countries may, in theory, encourage some activities and ensure comparative advantage.

- *Idea-driven economies.* The insight is that the nonrival nature of ideas makes them different from other factors of production, such as capital, land, and labor, in that the market may underinvest in the creation of new ideas. The main implication is that governments should, theoretically, subsidize some strands of research and development (R&D), such as those that will ensure the continuance of the comparative advantage a country has acquired in certain areas.
- *City-based growth.* The main insight is that activities that display increasing returns generated by factors external to a firm tend to be concentrated in cities, while those displaying constant returns remain more dispersed. The main implication is that policies to keep cities business-friendly and livable become more important as economies develop.

Urban systems exhibit some stylized patterns. Larger cities tend to be more diversified and service oriented: they innovate, invent, breed new firms, and expel mature industries.⁷³ Smaller cities tend to be industrially specialized: they produce or manufacture and receive relocated industries from diversified cities.⁷⁴ The relative city-size distribution and industrial concentration in specific cities tend to be stable over time. An urban system tends to be made up of a few large diversified cities and many smaller, more specialized, cities.⁷⁵

The stylized observation in most countries is an urban hierarchy of a few large cities and many smaller cities with varied economic functions.⁷⁶ At the global level, “world cities” at the top of the hierarchy, such as New York, London, Paris, and Tokyo, are characterized by a diverse industrial structure, predominantly service based, and a labor force with a wide range of skills.⁷⁷

Smaller cities specialize, receiving industries as they mature and relocate

Even after controlling for natural comparative advantage, externalities are still important in explaining the patterns of specialization and diversity among cities (see table 4.5). The production of nontraditional items is more concentrated in diverse U.S. cities, while standardized traditional goods are concentrated in smaller specialized cities. Similarly, in Japan, smaller cities are specialized, while low-tech activity and standardized high-tech production processes are located offshore. Likewise, in the Republic of Korea, large cities are more service oriented and smaller cities, manufacturing oriented.⁷⁸

Mid-size cities tend to specialize in mature industries, not new ones, and larger cities specialize in services not manufacturing.⁷⁹ Improved infrastructure and falling transport costs have encouraged standardized manufacturing production to move out of high-rent centers to smaller cities.

Table 4.5 Agglomeration economies vary by city size and profile, and by the industry life cycle

Main finding	Data
Localization economies are more important for heavy industries; urbanization economies are more important for light industries	Data for two-digit manufacturing industries in Japan (Nakamura 1985)
Localization economies become less important, giving way to urbanization economies, as cities expand in size	Cross-sectional data for the United States and Brazil (Henderson 1986)
Scale economies from labor pooling are stronger in newer and expanding markets, while those from knowledge spillovers and specialized asset-sharing are more important in mature markets	Annual firm employment data for four U.S. metropolitan areas and three two-digit industries (Hammond and Von Hagen 1994)
For mature capital goods industries, there is evidence of localization economies but none of urbanization economies; for new high-tech industries, there is evidence of both localization and urbanization economies	Panel data of 742 urban counties for 1970–87 (Henderson, Kuncoro, and Turner 1995).
For all industries both localization and urbanization effects are important. For traditional industries most effects die out after four or five years, but for high-tech industries, the effects can persist longer. The biggest effects are typically from conditions of three to four years ago, in the county and metropolitan area	Data for five traditional and three new high-tech manufacturing industries in 224 metropolitan areas between 1970 and 1987 (Henderson 1997)
The historical industrial environment of cities matters. In fairly mature cities urbanization economies encourage industrial growth	Growth data for the largest industries (1956–87) in 170 U.S. cities (Glaeser and others 1992)
For high-tech industries a 1-standard-deviation increase in diversity of the local manufacturing base increases productivity by 60 percent, but diversity has no effect on standard industries (such as textiles, or food).	City-industry data for the Republic of Korea, 1983, 1989, 1991–93 (Henderson, Lee, and Lee 2001)

Source: WDR 2009 team.

Production in large cities focuses on services, nonstandardized manufacturing, and R&D.⁸⁰ The relocation of manufacturing to the suburbs has been documented in Colombia, Indonesia, the Republic of Korea, and Thailand.⁸¹ It is common to find that services do not deconcentrate from city centers to their surrounding suburbs.⁸²

Large cities diversify, incubate new ideas and firms, and push out mature industries

New firms often start in diverse cities, but they move to specialized ones after they mature. Of all new plants in France, for example, 84 percent were created in cities with above-median diversity.⁸³ Some 72 percent of firm relocations are from an area with above-median diversity to an area with above-median specialization. In the United States almost all product innovations are in metropolitan areas. Industrial diversity and city size are both good for innovative output.⁸⁴ Trial plants are based in large cities in Japan, but mass production plants are in small cities or rural areas. Young firms appear to need a period of experimentation to determine their ideal production process.⁸⁵ In the early learning phase, diversified cities act as “nurseries” for firms to try out a variety of processes. Once a firm identifies its ideal process, it can begin mass production in specialized cities, where all firms share similar processes or specializations (see box 4.4).

The different economic functions that cities serve can be seen in the clustering of headquarters from different sectors and concentrations of business services in a few large cities while production plants from each sector congregate in smaller specialized cities. In 1950 there was little difference across U.S. cities in their proportions of managers and production workers. Although the largest cities already housed more managers, there was no clear ranking by city size. By 1980, however, the differences across cities had increased substantially, and a clear ranking by size had emerged. Larger cities had become specialized in management and information-intensive activity, which benefit from face-to-face contacts, and smaller cities had become specialized

in production. This pattern became even more marked during the 1990s.⁸⁶

Many business and economic historians have argued that the extra costs of coordinating and monitoring multilocation firms relative to integrated firms have come down significantly following key developments in transport and communication technologies, as well as new management practices.⁸⁷ Technological progress in transport and telecommunication made it less costly for firms to separate their production facilities from their headquarters and management facilities. Firms can locate their production facilities in environments with same-sector specialization, and their headquarters in a metropolis with a concentration of business service employment. Furthermore, the reduced communication costs that make transportation of service industry outputs (through electronic transmittal) cheaper did *not* imply the “death of distance” and the fading of cities into obscurity, contrary to many predictions.⁸⁸ In this context, while distance has become less important for transmitting information, it has become more important for transmitting knowledge. Telecommunication can be a complement to, but it is certainly not a strong substitute for, face-to-face interactions, which involve several forms of communication simultaneously, notably body language and verbal conversation (see box 4.5).⁸⁹ The geographic distribution of commercial Internet domains suggests that the Internet is a complement to face-to-face interactions (primarily within-city) as well as a substitute for longer-distance communication, such as phone or postal mail.⁹⁰

Activities that cities specialize in are stable, and so are city-size distributions

Externalities imply that history matters. That is, modern-day location patterns for an industry are strongly influenced by the historical industrial environment of cities and thus by the localization economies. Such intangibles include the local stock of knowledge relevant for an industry or a labor force with specific acquired skills. Two otherwise-identical enterprises in the same city could benefit differently from the local agglomeration depending on how long

BOX 4.4 *When sowing and reaping happen in different places: rising interdependence of cities*

Urban specialists and economists have long debated whether specialized or diverse cities are more conducive to growth. Cities that are narrowly specialized create greater economies of agglomeration, so a firm's productivity increases with proximity to similar firms. Meanwhile, a diverse mix of activities makes cities more likely to grow, particularly in new sectors. The main conclusion: both diversity and specialization are important, but at different points in a firm's life cycle. A "balanced" urban system is not one in which all cities are similarly specialized or diversified, but one in which both diversified and specialized cities coexist.

For young firms, urban diversity is more important. A new businessman may not know all the details of the product to be made, what components to use, where to source them, which workers to hire, and how to finance the venture. Firms using similar technologies in different sectors are more likely to share information about new practices and technologies than firms in the same sector. For firms in more standardized or mature industries, urban specialization is more important. These firms typically benefit less from the flexibility from urban diversity, and by locating in a specialized environment; they can better reap the benefits of urban agglomeration economies. For example, auto firms in Detroit lower their costs by sharing parts suppliers, and garment manufacturers in cities like San Pedro Sula in Honduras benefit from thick labor markets that help workers move between fac-

tories as the market adjusts to the whims and fancies of fashion.

Clusters of similar firms are sometimes promoted as the best environment for innovation. But studies find instead that diverse metropolises do better in breeding new products and processes. For example, the adoption of computer-controlled machinery for cutting metals has been faster in situations in which many firms (ranging from furnace manufacturers to aircraft producers) have similar technical needs but are not direct competitors. Firms for which innovating is important (such as electronics producers) prefer diversity during the early innovative phases, and then they relocate to specialized cities for mass production. For manufacturing and services, unlike agriculture, "sowing" and "reaping" can take place in different locations.

Just as product development and mass production increasingly take place at different locations, so too does management and production. Half a century ago the difficulties associated with managing businesses from a distance made firms keep their headquarters and management offices close to their factories. Falling transport and communications costs have made it much easier to manage production from far away (see chapter 6).

As a result, many firms have separated management and production spatially, searching for the best possible conditions for each. For headquarters, this means locations with other headquarters where these firms can, for example, share legal

services or advertising agencies; for production facilities, this means places with other such plants. Headquarters are usually in bigger cities, because professional services tend to exhibit greater economies of agglomeration, are less land-intensive, and employ highly educated employees willing to pay for big-city amenities. If land markets work well, the ensuing increase in land prices prompts production establishments to relocate to smaller, more specialized towns and cities.

Cities in the United States provide a good illustration. In 1950 the ratio of managers to production workers was similar across cities of different sizes. By 1990, however, cities with between 75,000 and 250,000 people had 20 percent fewer managers per production worker than the national average; cities with 1.5 to 5 million people had 20 percent more managers per production worker; and those larger than 5 million people were 50 percent above the national average. A similar trend can be seen in other countries such as France and Germany.

Policy makers should be aware of these developments. Since this growing interdependence manifests itself in plant relocations away from large cities, governments may be tempted to take away resources from them. This would kill the goose that lays the golden eggs, since such relocations to smaller specialized cities are just a later part of a life cycle of firms that large, diverse cities helped give birth to.

Contributed by Diego Puga.

each has been in the city. Similarly, two otherwise identical cities would offer different types of external economies depending on their histories.⁹¹

The influences of history and specialization are consistent with the observed stability in the relative city-size distribution and the industrial concentration in specific cities over time. Within countries the relative sizes of cities tend to remain unchanged. Among urban specialists, this phenomenon is often represented as a recurring relationship between a city's size relative to the largest city in the country,

known as Zipf's law: a city's population size relative to the primate city is inversely proportional to its rank in the national hierarchy of cities.⁹² There is also persistence in the industrial concentration in specific cities.⁹³

Among mature industries the persistence in employment patterns across cities is high over time, and the convergence in individual industry employment across cities is slow. This persistence occurs despite high plant and employment turnover rates for individual manufacturing industries, and despite strong evidence that plants

BOX 4.5 *Cities continue to thrive as telecommunication costs fall*

As telecommunications improve, cities become more important as a platform for interactions and knowledge transfers. Recent studies in the United States and Japan document the complementary roles of telecommunications and face-to-face interactions: people closer to one another physically call each other more often.

One interpretation is that face-to-face interactions generate more demand for telephone interactions. Since the mid-1980s, when faxes and e-mail became prevalent, business travel has risen more than 50 percent. Another evidence of increased face-to-face interactions with falling telecommunication costs is the phenomenal growth of co-authored articles in economics—from 12 percent in the 1960s to 56 percent in the 1990s. Local, out-of-state, and international

co-authorships all rose. Better telecommunications increase long-range interactions, but not at the expense of local interactions.

As ideas become more complex and difficult to communicate, the value of intensive face-to-face interaction rises, and cities become even more important. And if cities are centers of telecommunication technology, improvements in information technology will increase their economic role. The rise of the New York multimedia industry may signal the comparative advantage of large cities in facilitating the difficult information flows in cutting-edge industries. In the developing world, the rise of Bangalore is a case in point.

Sources: Gaspar and Glaeser 1998; Huber 1995; Sassen 1991; and Gottman 1977.

relocate as local wages and demand conditions change.⁹⁴ Historically, some cities have undergone major sectoral overhauls, but they have tended to be the exceptions.⁹⁵

The persistence of an industry's employment concentration in specific cities, which implies the "lock-in" of industrial structure, can be explained by localization economies. These cities can better compete for and, over time, retain plants and employment in that industry. A larger scale of own-industry activity historically means that firms in that locality today will operate more productively with greater accumulated knowledge about technology, sources of supply of different quality inputs, and local culture and its effect on the legal, business, and institutional climate. These localization advantages are relevant for more traditional manufacturing industries.⁹⁶ They explain the longevity of many industrial clusters in certain locations—such as the world-class cutlery cluster in Solingen, Germany, since 1348.⁹⁷

There is also evidence of persisting concentration of particular services in specific cities. The American mutual fund industry

began in Boston in 1924, when the Massachusetts Investment Trust was founded. Today, Boston is still home to almost a third of U.S. employment in mutual fund and asset management services. The Hartford insurance industry began even earlier, in the late-eighteenth century. Local merchants insured each other's overseas trading expeditions by sharing their profits and losses. These informal arrangements eventually grew into large insurance companies, starting with the Hartford Fire Insurance Company in 1810. Other major Hartford insurers, including Aetna, Connecticut General, and Travelers, were founded in the early and middle 1800s. Hartford is still known today as the "insurance city," with a wide range of related services such as life insurance, medical insurance, fire/marine/casualty insurance, and pension funds.⁹⁸

Apprehension of market forces

Over the past century, producers and workers in the developing world have sought, and often found, their fortunes in towns and cities. In the past three decades, researchers have analyzed and increasingly understood the gains from urban agglomerations of all shapes and sizes. But it is not yet clear that policy makers appreciate the sheer strength of these market forces and the benefits that come from harnessing them.

More than half the developing world's governments surveyed in 2005 by the UN Population Division expressed a desire to make major changes to the spatial distribution of their populations. Almost three-quarters of developing country officials expressed a strong desire to implement policies to reduce migration into urban areas or to take actions to reverse rural-urban migration trends.⁹⁹ Many in developed countries are equally fearful of urbanization in developing countries. "The explosive growth of cities around the world—especially the rise of huge, nation-sized Third World metropolises—has U.S. scientists and officials worried. Chief among their concerns: "megacities increasingly will serve as incubators of diseases, economic disruptions, and endless political crises."¹⁰⁰ This worry was reflected in

the goal of the 2006 World Urban Forum, held to discuss “mega-cities with mega problems.”¹⁰¹ The prevalent view was that “cities in the developed world have historically been engines of economic growth. But many cities in the Third World are so dysfunctional that they have become drags on economic progress.”¹⁰²

Some of the favored solutions: slow the massive migration to cities, decongest the largest cities in the developing world by establishing new cities, and make the biggest cities centers for cleaner high-technology activities. These solutions all represent a potentially costly misreading of the market forces that drive the spatial transformations for economic development.

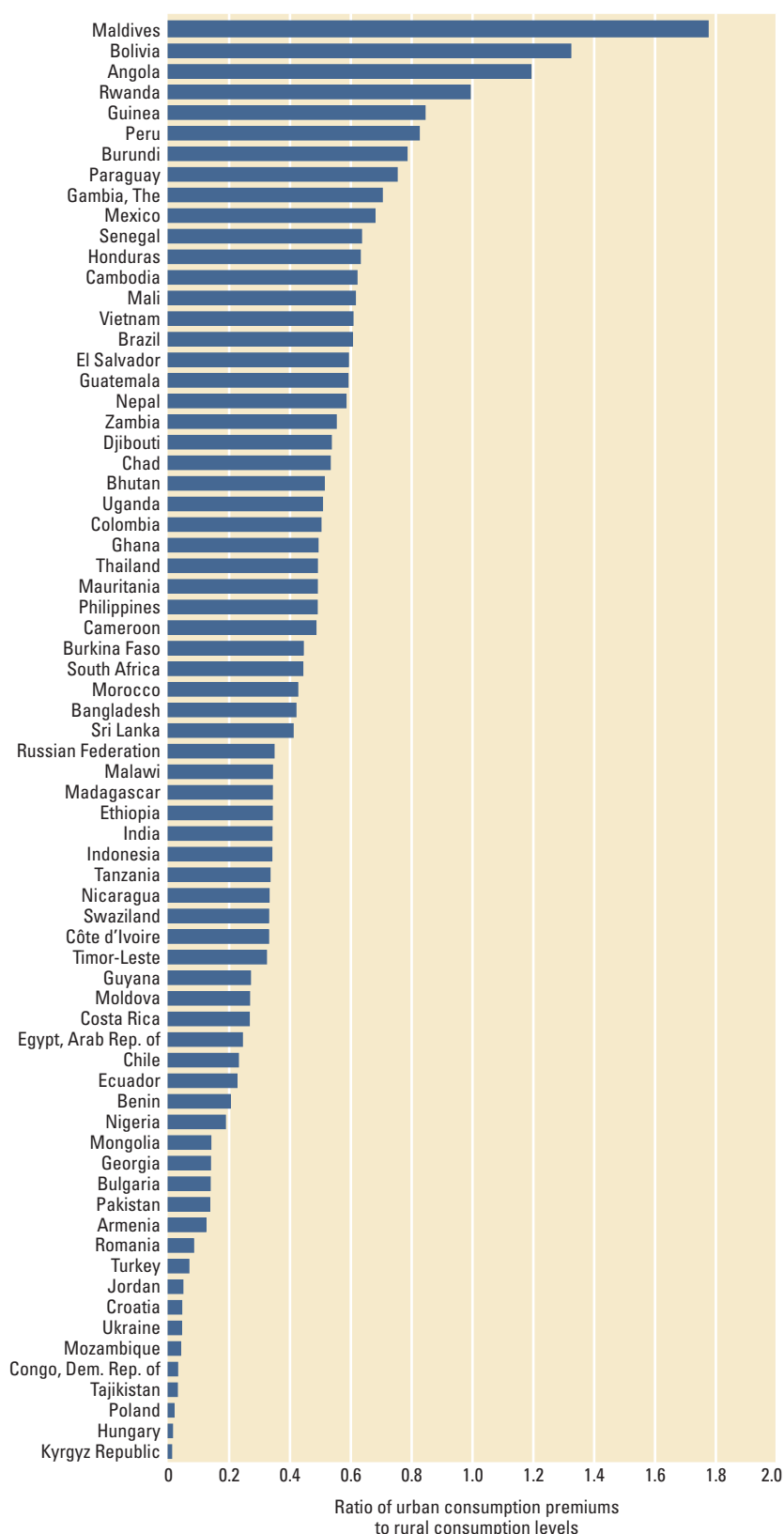
A misplaced fear of urbanization

Economic activities in urban areas account for as much as 80 percent of GDP in more urban and industrialized countries. The urban share of economic activity in less developed countries is about 50 percent. Just the 10 largest metropolitan areas in Mexico, which account for a third of the country’s population, generate 62 percent of its national value added.¹⁰³ In Vietnam, where the share of the urban population is 30 percent, the share of cities in national output is 70 percent. In China 120 cities account for three-quarters of the country’s GDP.¹⁰⁴ Clearly cities make a dominant contribution to economic production, even in poor and middle-income countries.

There is also ample evidence that urban areas in developing countries, including those in the poorest countries in Africa, deliver external economies. Consumption in urban and rural households in a broad cross-section of developing countries shows that people with similar observable characteristics enjoy higher consumption attributable purely to their urban location. The gains range from 2 percent in Hungary, the Krygyz Republic, and Poland, to 30 percent in Costa Rica, Ethiopia, India, Romania, and Tanzania, and to more than 80 percent in Angola, Bolivia, and Rwanda (see figure 4.1).

These magnitudes make it futile for policy makers to try to restrict the flow of people to urban areas. Even when restrictions have stemmed migration flows, the

Figure 4.1 The urban premium for household consumption can be considerable



Source: WDR 2009 team calculations, using 120 household surveys in 75 countries.

economic costs have been high. China's policies to restrict rural-urban migration until the late 1990s stunted urbanization, with between half and two-thirds of Chinese cities remaining too small. For the typical city in China, being too small is estimated to result in a loss of about 17 percent in net output per worker; for at least a quarter of the cities, these losses may range between 25 and 70 percent.¹⁰⁵

A misplaced preoccupation with size, not function, of cities

A city's prospects for prosperity and even survival are determined by how nimbly the same piece of land is adapted to changing market demands. Given that land is an immobile factor critical to the production of any activity, the real estate choices that cities provide influence the magnitude of external economies and the nature and specialization of city economies. To be

attractive to investors, a city must satisfy the demands of its dominant or growing industries for both real estate and facilities. For example, professional services and financial services require large amounts of office space, which can be more efficiently provided vertically in high-rise office buildings. Manufacturing requires large amounts of land for factories to produce goods, and for warehouses to store products and materials. And the recreation, tourism, and entertainment sectors require highly visible, pedestrian-friendly areas of cities and retail space.

The ability and ease of a city to adapt its land to different uses according to changing market needs will enable its sustainable growth. The last 800 years in Hong Kong, China, and the last 300 in New York show the importance of markets in signaling and implementing this urban renewal (see boxes 4.6 and 4.7). In New York the mercantile trade grew out of the early shipping industry. In turn, the mercantile trade industry would help give birth to the city's modern finance industry. Traders in New York City in the late-nineteenth century thrived by sharing access not only to physical transportation infrastructure (the harbor, canals, and railroads), but also to intermediate inputs of specialized services not available elsewhere (such as scheduled sailings, wholesalers, and ship brokers). Later, these inputs to trade became the foundations for shared inputs in finance, with maritime insurance underwriting the subsequent basis for other forms of investment.¹⁰⁶

Cities that provide fluid land and property markets and other supportive institutions—such as protecting property rights, enforcing contracts, and financing housing—will more likely flourish over time as the needs of markets change. Successful cities have relaxed zoning laws to allow higher-value users to bid for the valuable land—and have adopted flexible land use regulations to adapt to their changing roles overtime.

The benefits of agglomeration economies arise from the density of economic activity. These are the advantages for an information technology startup locating in Silicon Valley or a bookstall owner

BOX 4.6 *Hong Kong, China: market forces led the way, government followed*

Hong Kong, China, with a land area of about 1,000 square kilometers, less than a quarter the size of Rhode Island, started out as a fishing village. In the 1200s Hong Kong, a hilly and barren island, saw its first population boom as Chinese fled the mainland to escape war and famine. People made a living on salt production, pearl diving, and fishery trades. Between the 1650s and the 1800s, Hong Kong was also a military outpost and naval base, and its economy continued to rely on trade. By the end of World War II in 1945, the population in Hong Kong, China, had been reduced to less than half the prewar total of 1.6 million.

In the 1950s and 1960s, Hong Kong took up manufacturing buttons, artificial flowers, umbrellas, textiles, enamels, footwear, and plastics. Squatter camps provided homes for the masses. The camps led to disasters—like the Shek Kip Mei fire—until the governor responded by putting up multistory residential buildings. Conditions in public housing were basic, with communal cooking facili-

ties. For many decades, the private sector showed more commitment to and interest in urban redevelopment.

Between 1960 and 1980 the government experimented with urban renewal and comprehensive redevelopment to improve environmental conditions, traffic circulation, and community facilities. Over subsequent decades, the flexibility in land use planning and the participation of the private sector would prove crucial to satisfying the demands on land for housing, commerce, industry, transport, recreation, and community use. This combination enabled Hong Kong, China, to flourish into the regional center of business and financial services that it is today.

Consistent with the tradition of minimal government intervention in Hong Kong, China, the private sector has been the driving force behind urban transformation. The government contracted out urban redevelopment to a specialist organization dominated by private development interests.

Source: Adams and Hastings 2001.

BOX 4.7 *Reinvention and renewal: how New York became a great city*

New Amsterdam was founded as a Dutch colony in 1614. It passed into British hands and became New York in 1664. Manhattan, the Bronx, Brooklyn, Queens, and Staten Island were brought together in 1898 in the form we know today. Throughout its history, New York has continually rebuilt, reinvented, and renewed itself. Once a fur-trapping and shipping hub because of its natural harbor, New York City is today a global financial center and a regional powerhouse in mass media, arts, information and communication technology (ICT) innovation, and medical research. The New York metropolitan area is home to more than 18.7 million people with a GDP of \$1,133 billion, making it the second-largest urban agglomeration in the world, after Tokyo. New York had a gross metropolitan product of \$950 billion in 2005, making it the largest regional economy in the United States. If it were a country, New York City would be the world's seventeenth largest, ahead of Switzerland. At more than \$56,000, it has the second highest per capita production in the world.

A tour of four neighborhoods reveals the city's versatility and vibrancy.

SoHo. In the 1700s SoHo was farmland. By the early 1800s it was primarily residential, inhabited by the wealthy and, soon after, by the middle class. In due time rapid development attracted many businesses. Hotels, theaters, stores, mansions, minstrel halls, casinos, and brothels appeared along Broadway. Starting in the 1880s the textile industry settled in the area. By the 1950s artists flocked to the area because of low rents, a result of people, industry, and commerce shifting uptown. In October 1962 the City Club of

New York characterized SoHo as a commercial slum. But today the area, once called Hell's Hundred Acres, is a busy commercial and retail district and home to New York University.^a

Wall Street. The financial district is one of the city's best-known and oldest neighborhoods. Today's Wall Street neighborhood is part of Manhattan Community District 1, which extends south from Canal Street to the tip of Manhattan at Battery Park and includes Governor's Island. It is home to the New York Stock Exchange and the NASDAQ, the world's two largest stock exchanges.

The street's name was originally De Wall Straat in reference to the Walloons, Belgian farmers who were the majority of the residents living in New Netherland around Fort Amsterdam in 1630. The beaver belt was the single most important commodity in New Netherland. Trade encouraged new activity in the production of food, timber, tobacco, and eventually slaves. In the late-eighteenth century there was a buttonwood tree at the foot of Wall Street under which traders and speculators gathered to trade informally. In 1792 this arrangement was formalized with the Buttonwood Agreement, which laid the groundwork for the New York Stock Exchange.

Meatpacking district. In 1969, when Vincent Inconiglios moved to a loft on Gansevoort Street in the meatpacking district, it was a no man's land. The neighborhood was defined by a stench that overpowered the senses. Down the street from Mr. Inconiglios was a pickle factory, and an importer of Spanish melons occupied the shop downstairs. The area was teeming with barrels of bones, meat, and men in bloody white coats. Within

a generation, the transformation in the meatpacking district was as stark as the contrast between night and day.^b Today, more than 35 wholesale meat companies still operate there. But the area is now also home to world-class restaurants, art galleries, a fashionable retail corridor, and night clubs that take advantage of the enormous former factory spaces. Real estate prices have skyrocketed. Mr. Inconiglios paid \$50 a month when he moved to the meatpacking district. In 2007 the Carlyle Group and Sitt Asset Management acquired a pair of buildings on West 14th Street for \$70 million.^c

Williamsburg. This neighborhood reinvented itself from a booming trade port to a rich industrial town after the Civil War. With the construction of the Williamsburg Bridge in 1903, many Jewish families who lived in Manhattan's Lower East Side crossed the East River to a better life in Williamsburg. When industries left the area in the 1960s and 1970s, Williamsburg became an immigrant ghetto. But the cheap rent also made the neighborhood an artistic hub. The neighborhood evolved into a mix of Italian, Polish, Hispanic, and Hasidic residents. In 2005 New York City approved zoning changes that would allow for open spaces, parks, affordable housing, and light industry. Today, prices average \$700–\$900 per square foot, and prominent waterfront developments range in the millions.^d

Sources: Seeman and Siegfried 1978; Shaw 2007; Biedermann 2007; Lynch and Mulero 2007.

a. Seeman and Siegfried 1978.

b. Shaw 2007.

c. Biedermann 2007.

d. Lynch and Mulero 2007.

locating close to other bookstalls on Dada-bhai Naoroji Road in Mumbai, India. While the financial sector of London is largely concentrated in a few square miles of the City and Canary Wharf, financial firms also benefit from being located anywhere in Greater London. Firms benefit from locating close to other firms in either the same or different industries and unless all of them move together, they will become less profitable, even if the location

they are moving to has lower wages and cheaper land.

But bigger city size and economic density bring their own problems. For people and firms, city living comes at a price in both developing and developed countries. Traffic in central London moves at only 11 miles per hour¹⁰⁷—the same speed as horse drawn carriages a hundred years ago. Beijing is notorious for its pollution-induced smog. Land in Mumbai is among the

most expensive in the world. High levels of crime are an accepted feature of city living around the world. Millions of city dwellers live in overpopulated slum housing, with little or no access to basic amenities and services. These are the costs of density, the diseconomies of agglomeration.

The main source of diseconomies is the paucity of land in places where agglomeration economies take hold. Land is limited and as economic growth occurs, it has to be used with increasing intensity. Take Manhattan in New York City, which has an area of less than 35 square miles. In 1800 it had a population density of just under 3,000 people per square mile. By 1850 this had risen to about 23,500, peaking in 1910 with a population density of more than 100,000. Today, the population density is about 70,000. With land in fixed supply, its use eventually can offset any further benefits from agglomeration economies. The way to offset the fixed supply of a factor of production is to substitute other factors for it, and the rise of skyscrapers in many large urban areas is an illustration of this substitution of capital for land. The building of subway systems in many of the developed countries' larger cities is another example. But such substitution has its limits, and the increasing shortage of land in cities leads to higher rents and congestion costs for workers and firms.

Better transport can, by reducing the economic distance to density, in essence make land a less-binding resource. Indeed, with the long-term decline in transportation costs, cities have expanded. In 1680 London was only 4 square miles and, because of the difficulties of traveling, more than 450,000 people were crammed into this small area. By 1901 the city had expanded to 24 square miles, and the average population density had fallen to 79,000. In 2001 London's 627 square miles had a population density of 13,203 people per square mile. An expanding city meant that millions of commuters have to be transported from the suburbs, large volumes of retail goods have to be delivered to shops, and manufactured products have to be shipped out. All of this leads to congestion or diseconomies of scale

that reduce the gains from agglomeration economies.

But restricting the growth of cities is not the answer. There is no evidence that the agglomeration economies of megacities have been exhausted. Indeed, evidence suggests that the growth of vehicles in the developing countries is increasing with per capita income along a path similar to that followed by the richer countries.¹⁰⁸ The problem has more to do with the spatial structure of the city and investments in infrastructure. Vehicle ownership is rising 15 to 20 percent annually in much of the developing world.¹⁰⁹ But most countries have not matched this growth with a parallel expansion of transportation infrastructure, so traffic congestion is severe. Cities in developing countries only devote half as much land space to roads as in the United States. But it is not just a matter of increasing this capacity. In cities such as Bangkok and Manila, it is the management and the use of road space that is important. Part of the problem is that in many cities the responsibility for road infrastructure has devolved from central to local governments, which do not always have the necessary resources.

Combined with the differing propensities of industries to benefit from agglomeration economies, the resulting constraints explain why the spatial distribution of economic activity within a country is not restricted to a single center, but rather consists of multiple centers of differing sizes. For policy makers the challenge is to best relax the constraints generated by the congestion and overcrowding of land and resources so that the benefits of agglomeration can be maximized. In many cases these constraints have been tightened by misguided land use policies and planning failures, only adding to congestion (see chapter 7).

A misplaced fascination with "new" cities

The land Chicago was built on is not all that different from the more sparsely developed places around Lake Michigan. Yet the difference in economic production and household earnings between Chicago

and other settlements on the lakefront in Wisconsin and Indiana is stark. And along the 10-hour drive through Texas on Interstate Highway 75, wages and land rents spike in Fort Worth, Austin, and San Antonio and drop off sharply at points in between. It is hard to reconcile these huge differences in economic density with the minor differences in physical geography; it is as if the areas of Fort Worth and San Antonio cast a shadow over the points in between. A better understanding of economic geography, characterized by external economies, is required to harness economic forces. But it is not always obvious that developing country governments understand economic geography or appreciate these forces.

A survey of new city initiatives in the Arab Republic of Egypt, Brazil, Hungary, India, and República Bolivariana de Venezuela is sobering. Brazil transferred its capital city from the coast to the mid-western interior more than 900 kilometers away. República Bolivariana de Venezuela picked Ciudad Guayana in the 1950s, a city in the southern part of the country, to be the industrial “growth pole” of the central and southern region and to attract people and jobs from the already rapidly growing metropolitan region in the north. In many formerly planned economies, the more common practice was building industrial towns to accelerate industrialization. In Hungary, Dunaujvaró was designed as a “steel town,” Tiszaújváros as a “chemistry town,” and Kazincbarcika as a “mining and heavy industry town.” The Soviet Union built Magnitogorsk into a steel town in an area with huge reserves of iron ore to challenge its capitalist rivals.

Some new towns were built around metropolitan areas to alleviate the pressures that the large cities faced. Navi Mumbai was established in 1972 with the hope of developing a twin city for Mumbai, and to decongest Mumbai. Egypt started a comprehensive new town construction program around Cairo and away from Cairo to create a “new population map of Egypt” starting in the 1970s, and the construction is still ongoing. Many of these cities were created

for economic reasons, but some were created for political reasons. Have these new towns and cities, met their goals? Generally not.

- New cities do better when they are located near larger successful cities. But they often suffer from the same government-related failures that led the government to establish them, especially the failure to manage large cities well. That is, governments that do badly in managing large old cities also do badly in managing small new cities.
- New cities attract residents, sometimes even more than anticipated, but often not the people intended. That is, governments can set up (noncapital) cities, and they sometimes become viable, but not for the reasons the government envisaged.
- These cities attract people because of the circular causation that the new economic geography emphasizes: workers and entrepreneurs come to seek markets, and then more people come because this is where the markets are. But there may be huge opportunity costs, because the counterfactual could be more organic growth of settlements. That is, it makes sense for private agents to come to these cities since others are already there, but large efficiency losses may result from the country’s point of view. Once a “bad” location is picked, it may not fail entirely because of circular causation, but that means the economic costs of the mistake are greater, not smaller, since the country will pay these costs for a long time.
- New noncapital cities that seem to succeed are those where the purpose and location are chosen over time by markets and in cases in which the government hastens the pace of growth by coordinating investments in infrastructure, housing, and general governance.

For these reasons, cities and towns should be seen as market agents that, just like firms and farms, serve market needs.