NBER WORKING PAPER SERIES

SCALE ECONOMIES AND THE GEOGRAPHIC CONCENTRATION OF INDUSTRY

Gordon H. Hanson

Working Paper 8013
http://www.nber.org/papers/w8013

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2000

Scale Economies and the Geographic Concentration of Industry
Gordon H. Hanson
NBER Working Paper No. 8013
November 2000
JEL No. F2, J6, R1

ABSTRACT

In recent empirical literature on spatial agglomeration, many papers find evidence consistent with location-specific externalities of some sort. Our willingness to accept evidence of agglomeration economies depends on how well key estimation problems have been addressed. Three issues are particularly troublesome for identifying agglomeration effects: unobserved regional characteristics, simultaneity in regional data, and multiple sources of externalities. Two empirical results appear to be robust to problems created by the first two issues: (a) individual wages are increasing in the presence of more-educated workers in the local labor force, which is consistent with localized human-capital externalities, and (b) long-run industry growth is higher in locations with a wider range of industrial activities, which suggests that firms benefit from being in more diverse urban environments. Other evidence is supportive of agglomeration effects related to regional demand linkages and short-run, industry-specific externalities.

Gordon H. Hanson
Department of Economics
University of Michigan
Ann Arbor, MI 48109
and NBER
gohanson@umich.edu

I. Introduction

During the last decade or so, economists have rediscovered geography. While urban and regional economics have been and continue to be vibrant areas of research, other fields within economics have increased their use of regional economies as laboratories in which to examine different questions. Central to these lines of inquiry has been an attempt to understand why economic activity tends to concentrate geographically. As it turns out, if we can explain geographic concentration, then we can go a long way towards explaining important aspects of international trade and economic growth.

A common theme in recent theory is that increasing returns to scale cause industry to spatially agglomerate. Scale economies may be external to firms, such as those related to knowledge spillovers, or internal to firms, which in combination with transport costs create location-specific externalities. In either case, distance weakens the impact of external effects, giving agents an incentive to locate near each other. Scale economies also feature prominently in recent empirical work on economic geography. The literature is unified by a quest to estimate the impact of agglomeration effects on regional outcomes. Most papers find evidence which is presented as being consistent with agglomeration economies of some sort.

In this paper, I survey recent empirical work on the geographic concentration of economic activity. That theory proposes scale economies as an explanation for spatial agglomeration and that empirical work claims to find evidence in support of theory would seem to be a major success for the field. The point of my review is to see whether this superficial impression is confirmed upon closer examination of the literature.

To begin with a note of caution, empirically identifying the impact of agglomeration economies on economic activity is a task wrought with difficulties. Our willingness to accept

evidence of agglomeration effects depends in large part on how well various empirical problems have been addressed. Assessing the literature in light of the key estimation issues will be an important function of this review. It is useful to begin by briefly discussing three issues that complicate identification of agglomeration effects: unobserved regional characteristics, simultaneity in regional data, and multiple sources of externalities.

It goes without saying that agglomeration economies are difficult to observe empirically. Except in unusual cases, spillovers are not measurable, which leaves us to infer their existence by indirect means. Indirect inference raises the risk of misidentification. While it is plausible to say that the movie industry is concentrated in Los Angeles because of location-specific externalities (actors are inspired by each others' presence), it is no less plausible to attribute the industry's location to natural amenities that Los Angeles offers (actors are inspired by desert canyons and sunsets on the beach). We would be justifiably skeptical of any test of agglomeration economies that did not control for geographic concentration due to the supply of exogenous site-specific resources (Ellison and Glaeser, 1997).

A related problem is that where industry agglomerates may be indeterminate. In principle, we have little reason to doubt that but for an accident of history (Hewlett and Packard's friendship and initial location choice) Silicon Valley would be located in Massachusetts or New Jersey today rather than Northern California. Such indeterminacy accords with recent theory, in which models are subject to multiple equilibria in terms of the location (or even number) of cities or industry clusters. What this means for empirical work is that there may be little exogenous variation in the characteristics of regions which we can use to identify the causal effects of agglomeration economies on regional behavior. In this line of research the potential for simultaneity problems to confound statistical inference is acute.

Finally, finding evidence of agglomeration economies is not the same as identifying the source of these economies. Theory gives us an embarrassment of riches in terms of explaining agglomeration. While there is a fundamental difference in how and why agglomeration occurs in models based on internal scale economies versus models based on purely external scale economies, the two approaches share many predictions for the spatial distribution of economic activity. Also, identifying the contribution of scale economies to agglomeration requires that we control for the impact of congestion effects, such as limited supplies of non-traded goods (e.g., housing) or amenities (e.g., clean air), which may work against agglomeration. Without a well-specified model as the basis for empirical work we may have little guidance for how to interpret empirical results. If we cannot identify the source of externalities that cause agglomeration, or the attenuating effects of congestion costs, then we also cannot evaluate the welfare consequences of agglomeration or conduct policy analysis.

The remainder of the paper contains four sections. Section II gives a brief overview of recent theories of agglomeration. The following three sections discuss empirical work. I categorize papers by the empirical phenomena they are trying to explain: the spatial variation in wages and labor productivity (section III), the spatial variation in industry employment, production, and innovation (section IV), and the evolution in the size and structure of cities (section V). Section VI suggests directions for future research by way of conclusion.

Several caveats are in order. I make no attempt at an exhaustive survey, but instead focus on work which has been influential or is useful for illustrating methodological issues. Case studies and research by non-economists largely escape mention. This will hopefully sharpen the analytical focus, but at the expense of excluding some worthwhile papers. Additionally, most papers I discuss use U.S. data. This reflects the fact that studies on the U.S. economy account

for a disproportionate share of the literature. Growing empirical interest in regional economic issues in Europe and elsewhere will hopefully soon rectify this state of affairs.

II. Theory

Recent theory attributes the spatial agglomeration of industry to increasing returns to scale in production. Urban and regional economists have long emphasized a relationship between scale economies and geographic concentration. To provide a basis for evaluating empirical research on agglomeration effects, I briefly discuss several prominent formal models. This body of research builds on many ideas from older well-established literature, which I do not cite.[1] My intent is solely to highlight the different formal mechanisms through which scale economies contribute to agglomeration.

In a relatively early and influential line of work, Henderson (1974, 1988) suggests that agglomeration economies are the result of positive spillovers between firms that share the same location. While individual firms are perfectly competitive and perceive there to be constant returns to scale, the agglomeration of economic activity generates externalities that raise the productivity of all firms in a particular industry that share a given geographic location. These externalities are assumed and so their source is not specified. The standard story, following Marshal (1920), is that the geographic clustering of firms promotes learning and the exchange of ideas between agents. The existence of localized externalities implies that firms prefer to be near large agglomerations of other firms in their own industry or related industries. An urban hierarchy arises in which cities specialize in different industries and the size of cities is determined by the size of their respective export activities. Costs associated with congestion,

[1] For recent surveys of theoretical research on agglomeration in spatial economic models, see Fujita and Thisse (1996), Krugman (1997), Quigley (1998), and Fujita, Krugman, and Venables (1999).

which I describe in more detail below, act as a centrifugal force which prevents economic activity from becoming agglomerated in a single location or region.

The externalities that Henderson (1974) describes are constant over time and so the resulting spatial equilibrium is static. Lucas (1988) offers a related, dynamic model for a one-region closed economy, in which the accumulation of human capital generates positive spillovers. If one worker acquires a new skill, then all workers that share her location become more productive. Though the mechanism through which this occurs is not made explicit, Acemoglu (1996) shows that it is possible for spillovers in human capital accumulation to occur endogenously if it is costly to match skilled workers to firms.[2]

To relate human capital spillovers to spatial agglomeration, Black and Henderson (1999a) combine elements of Lucas (1988), Henderson (1974), and Eaton and Eckstein (1997). Theirs is a dynamic model of city formation, in which there are location-specific industry external economies and location-specific human-capital externalities. Industry agglomeration makes all firms within the same local industry more productive and the agglomeration of labor makes all local workers, regardless of industry, more productive. In this environment, the urban hierarchy that forms, again in which each city specializes in a single export industry, is stable over time. In steady state, each city and each industry grows at the same rate.

In an alternative view of geographic concentration, Krugman (1991) suggests that industry agglomeration is the result of demand linkages between firms, which are created by the interaction of transport costs and fixed costs in production.[3] In this case, scale economies are internal, rather than external, to the firm, and there are transport costs in shipping goods between regions. The basic model is one that is familiar from international trade theory (Krugman, 1980),

[2] Duranton and Puga (1999a) and Keely (1999) give microfoundations for technology spillovers in a spatial setting.

extended to a regional setting. Individuals prefer to consume the widest possible variety of products, but fixed costs in production limit the number of goods that can be produced. In response to consumers' love of variety, firms differentiate their products, such that each good is produced by a single, monopolistically-competitive firm. Given fixed production costs firms prefer to concentrate production in a single location, and given transport costs firms prefer to locate their plants near large markets. Firms are thus drawn to densely concentrated regions by the possibility of serving a large local market from a single plant at low transport costs.

One can obtain similar results if there are scale economies in producing non-traded intermediate inputs (Fujita, 1988; Rivera-Batiz, 1988) or if industries have vertical stages of production in which firms produce both consumer and industrial goods (Venables, 1996).[4] In the Venables model, vertical linkages between industries help make the location decisions of buyers and suppliers interdependent. As the upstream industry in a location expands the range of intermediate goods that it produces, the downstream industry benefits both because it values the specialized inputs and because it obtains these inputs at low transport cost. Such cost and demand linkages between industries further raise the incentive for agglomeration.

Whether scale economies are internal or external to firms, location-specific externalities contribute to spatial agglomeration.[5] Congestion costs, associated with limited local supplies of housing or other non-traded goods or factors, work against agglomeration. As cities form, the price of housing is bid up in urban areas relative to outlying areas. To attract workers to cities, firms must compensate them for the relatively high cost of urban living. The higher productivity of labor in agglomerated regions justifies these higher wages. If agglomeration economies are

[3] For contrasting views of recent economic geography models, see Henderson (1996) and Krugman (1996). See Tabuchi (1999) for a synthesis of geography models based on internal and external scale economies.
[4] For related work, see Puga (1999), Ottaviano and Thisse (1998), and Fujita, Krugman, and Venables (1999).
[5] Quah (1999) models spatial agglomeration without location-specific externalities or transport costs.

sufficiently strong, most production occurs in industry clusters, where wages and housing prices are relatively high, and large tracts of space are relatively idle.

At first pass, Henderson-style models of external economies and Krugman-style models of regional demand linkages appear to have similar implications for the spatial distribution of economic activity. Both predict that the spatial distribution will be lumpy and that across regions nominal wages and housing prices will be positively correlated with the agglomeration of economic activity.[6] The models differ, however, in terms of *how* agglomeration occurs. In the original Henderson (1974) model, agglomeration occurs because firms benefit from being near other firms in their own industry; in Black and Henderson (1999a), an additional motivation for agglomeration is that workers benefit from being near other workers; and in Krugman (1991) and related work, agglomeration occurs because firms benefit from being near large consumer and industrial markets. To distinguish between the competing theories, we must allow for each of these mechanisms to contribute to the geographic concentration of industry.

One final issue is that theory pays relatively little attention to how the exogenous characteristics of regions influence where industry clusters arise. Theory predicts *how many* cities or clusters will form but not necessarily *where* they will form.[7] In many models, all locations are ex ante identical such that it is impossible to determine which regions will be home to industry clusters and which will not. This type of multiple equilibria may complicate empirical estimation.[8] Also, in reality the natural features of regions are likely to impact industry location. In taking theory to the data, we want to control for how the supply of exogenous site-specific resources might affect the location of economic activity.

[6] The Krugman model, in its original form, does not consider housing. Helpman (1998) modifies Krugman (1991) by replacing an agricultural sector with a housing sector, which then generates predictions for spatial variation in wages and housing prices. Other features of the model are unaffected by this extension. See also Puga (1999).
[7] Exceptions include Fujita and Mori (1996, 1997).

III. Agglomeration, Wages, and Labor Productivity

A common prediction of spatial models based on increasing returns is that equilibrium wages and land prices will be higher in more densely concentrated regions. These theories are not alone in implying such outcomes. Assessing whether localized externalities contribute to the spatial fluctuations in factor prices is an important task for empirical work.

There is a large literature on the variation in nominal wages across regions within the United States. It is well-known that wages are higher in urban areas than in rural areas (Glaeser and Mare, 1994), that wages vary substantially across metropolitan areas (Montgomery, 1992), and that wage differences across U.S. regions have diminished over time (Treyz, 1991; Blanchard and Katz, 1992). Recent literature devotes much attention to two explanations for long-run variation in regional wages: (i) regional wages vary because of lumpiness in the regional supply of amenities, and (ii) regional wages (and labor productivity) vary because of location-specific externalities.[9] I discuss each in turn.

III.A Exogenous Amenities and Regional Wages

Rosen (1979) and Roback (1982) suggest that regional differences in wages (and land rents) reflect regional variation in exogenous amenities, such as climate, air quality, or physical geography, which influence consumer utility and/or labor productivity. These amenities are typically assumed to be non-exclusive in consumption, such that all agents in a location benefit from them equally. If consumers prefer warm, sunny weather to cold, cloudy weather, they will

[8] A second type of multiple equilibria, in which agglomeration and regional autarky are possible outcomes (see Krugman, 1991; Puga, 1999), is unlikely to be relevant empirically.

[9] While business-cycle phenomena may contribute to transitory differences in regional wages (Blanchard and Katz, 1992; Topel, 1986; and Abraham, 1996), I focus on long-run sources of regional wage divergence.

require a wage premium to live and work in Detroit over San Diego. By the same logic, consumers will be willing to pay a premium for housing in San Diego over housing in Detroit. If the only effect of weather on the location of economic activity is through its impact on consumer utility, we expect pleasant climes to have low wages and high land rents relative to unpleasant climes, such that real consumption wages are equalized across locales.

The story gets more complicated if weather, or other exogenous regional characteristics, also affect labor productivity. If workers are more productive in good weather, firms in San Diego will be willing to pay workers higher wages than firms in Detroit. Migration of labor to high-wage regions will capitalize these productivity differences into land rents. In contrast to the previous case, in which weather affected utility but not productivity, sunny climes will have relatively high land rents *and* wages. Since it is impossible to say in principle whether amenities influence behavior through consumer preferences or producer technology, their net effect on wages and land rents can only be resolved empirically.

The standard technique to estimate the impact of exogenous regional characteristics on wages and land prices is Rosen's (1974) framework for hedonic price estimation. In this approach, the dependent variable is either the wage of an individual worker or the imputed rent for a particular housing unit. The independent variables include the characteristics of an individual worker -- education, age, gender, marital status -- or housing unit -- size, age, number of rooms, heating/cooling system -- and the exogenous regional characteristics that influence firm or consumer behavior. Some studies also include local tax rates, local crime rates, or quality of local schools among regional characteristics, but it is probably more appropriate to treat these variables as jointly determined with local wages and land rents.

To cite results that are representative of this literature, Roback (1982) finds that in a cross section of U.S. metropolitan areas labor earnings are positively correlated with local measures of the crime rate, particulate air matter, heating degree days, total snowfall, and the number of cloudy days, but that metropolitan land rents are uncorrelated with these variables. One interpretation of these results is that workers dislike poor weather, crime, and pollution and so require a wage premium to live in regions with higher levels of such disamenities.

Subsequent research extends the hedonic approach. Beeson and Eberts (1989) find that regional productivity effects explain a larger fraction of variation in wages across U.S. cities than do regional consumption amenities. Gyourko and Tracy (1991) find that local fiscal conditions, such as property taxes, income tax rates, and corporate tax rates, also influence regional wages and land prices. And though wages vary across regions in part because the return to additional years of schooling is higher in some U.S. cities than others (Farber and Newman, 1987; Peri, 1998), exogenous regional characteristics can account for only a small fraction of the variation in returns to schooling across U.S. urban areas (Beeson, 1991).

There is abundant evidence that exogenous regional characteristics contribute to spatial variation in factor prices. However, there also appears to be abundant evidence that within regions for which exogenous characteristics are relatively constant there remains wide dispersion in these prices. Along the roughly ten-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 this pattern with the relatively minor changes in climate and physical setting encountered along the route. A complete treatment of spatial variation in factor prices would require accounting for the endogenous features of regions.

This point may seem obvious, but other implications of this line of research are more subtle. Changes in consumer preferences or production technology will in principle change the implicit prices attached to exogenous amenities. Consumers may be willing to pay more for good weather, in terms or lower wages or higher housing prices, as technological change (faster bicycles, more comfortable running shoes) makes good weather more enjoyable. There is no reason, then, to presume that the impact of exogenous regional characteristics on the spatial distribution of economic activity will be constant over time.

III.B Agglomeration, Education, and Labor Earnings

Spatial theories based on localized human-capital spillovers, including Eaton and Eckstein (1997) and Black and Henderson (1999a), predict that wages and land rents will be higher in regions with larger stocks of public knowledge. Recent empirical work tests this prediction by using average regional education to measure local knowledge stocks.

In one widely-cited study, Rauch (1993) estimates hedonic-price equations for wages and housing rents using data on a cross-section of individual workers and individual housing units in U.S. metropolitan areas. He finds that increasing the city-wide average education level by one year raises an individual worker's wage by 3% and an individual housing unit's rent by 13%.[10] Raising city-wide average labor experience has smaller but still positive effects. Both sets of estimation exercises control for the observable characteristics of either workers or housing units and for other observable regional characteristics.

[10] Peri (1998) finds that the effect of average education on wages that Rauch estimates is due mainly to the fact that the returns to education for individual workers are increasing in the average local education level. In related work, Black (1998) suggests that increased segregation of workers across cities by education level could partially account for rising wage inequality in the United States.

Rauch's results are consistent with localized human-capital spillovers, as workers appear to be more productive where their co-workers are more educated or more experienced. The variation in growth rates across cities is supportive of this reasoning. Glaeser, Scheinkman, and Shleifer (1995) and Black and Henderson (1999a) find that population growth is faster in U.S. metropolitan areas that begin with a more educated population.[11] Migration is one mechanism which could sustain these growth patterns. Migrants with higher education levels appear to be attracted to regions with higher returns to education (Borjas, Bronars, and Trejo, 1992).

Another interpretation of these results is that industries subject to agglomeration effects are relatively intensive in the use of more-educated labor. In this case, firms in agglomerated industries would pay workers high wages to compensate them for high housing costs. The positive correlation between wages and the skill composition of the local labor force would then reflect a deeper correlation between wages and local industry composition. A similar outcome would obtain if agglomerated industries were relatively intensive in the use of high-quality labor, where quality was observed by firms but not by the econometrician.

Are there other plausible alternative explanations for why wages are higher in cities with higher average education levels? One source of concern is that there may be unobserved causal factors which account for both higher city-wide education levels and higher returns to unobserved individual characteristics. In this case, the estimated social returns to education could be the byproduct of omitted-variable or simultaneity bias. Suppose, for instance, that some cities invest more in schools and so provide a higher quality education to local students. Students in these cities might stay in school longer, since their returns to education would be

[11] These results are vague about whether the externalities at work are static or dynamic. While the attraction of workers to locations where average education levels are higher may affect the spatial distribution of employment, there is no evidence on whether it effects accumulation of human capital (and so rates of economic growth). That

relatively high, and as workers might earn more than individuals who have the same years of schooling but attended low-quality schools in other cities. If individuals tend to work in the city in which they went to school, then the correlation between individual wages and average local schooling could be simply an artifact of variation in urban school quality.

To control for the possible endogeneity of average regional education levels, several recent papers instrument for the variable in hedonic wage regressions. Acemoglu and Angrist (1999) treat both individual years of schooling and average local years of schooling as endogenous. Defining the state to be the geographic unit over which human-capital spillovers extend, they use quarter of birth (Angrist and Krueger, 1991) to instrument for individual schooling levels and state compulsory schooling laws, which mandate school attendance and were enacted earlier in some states than in others, to instrument for average state education. Compulsory schooling laws appear to affect educational attainment in middle and high schools but not at higher levels. In contrast to Rauch's results, Acemoglu and Angrist estimate the excess social return to education to be small and statistically indistinguishable from zero.

Moretti (1999) regresses individual wages on the share of college graduates in the labor force of the city in which an individual works. He instruments for the college-graduate share using the presence of a land-grant college in the local area, the cost of tuition at state colleges and universities, and the city demographic structure. Similar to Rauch, Moretti finds that the social returns to education exceed private returns. Increasing the share of college graduates raises wages for all workers, with less-educated workers enjoying the largest effects.

These apparently contradictory sets of results can be reconciled if college educated workers are the primary source of human-capital externalities. In this case, the approach of

---

cities with more educated populations grow faster may reflect no more than the relocation of workers to take advantage of static human capital externalities in an economy where the population is growing.

Acemoglu and Angrist, who in effect identify excess social returns to education off of cross-state variation in high-school and middle-school education, would yield no evidence of human-capital externalities. Moretti's approach would yield such evidence, and, to the extent that the variation in average schooling levels is correlated with the presence of college graduates, Rauch's approach would also. One shortcoming of the most recent studies is that they do not examine the correlation between average local human capital levels and housing rents. This would provide a key source of corroborating evidence for spillovers associated with human capital.

III.C Spatial Linkages, Labor Productivity, and Wages

The literature on human capital externalities goes to considerable lengths to address two of the three identification issues mentioned in the introduction -- exogenous regional characteristics and simultaneity in regional data. The third issue, multiple sources of externalities, receives less attention. The implicit assumption is that average education levels are a sufficient statistic for agglomeration effects. Externalities may of course be transmitted through other mechanisms, such as firm production levels, R&D spending, etc. It would be simple to extend hedonic estimation techniques to allow for such possibilities.

A more serious problem would arise if externalities are transmitted across regions, as would be occur if spillovers die out slowly over space, or if regional linkages are themselves a source of externalities, as in Krugman (1991). In either of these cases, it would be difficult to identify the impact of spillovers on regional economies without an explicit model of how they are created and transmitted. Several recent papers attempt to explain regional variation in wages or labor productivity using such explicit models.

Ciccone and Hall (1996) assess whether labor productivity is higher in U.S. states where economic activity is more densely concentrated. In their framework density is a reduced form for several possible types of agglomeration economies. They first posit that the local density of employment influences productivity and then derive and estimate an equation in which state labor productivity is a function of the education level of workers in the state and employment densities for counties within the state. To control for the endogeneity of employment densities, they instrument for the variable using historical data on population and other state characteristics. They find that doubling the employment density in a region raises local labor productivity by 6% and interpret this result to mean that the close interaction of workers raises productivity.

An additional source of variation in wages across space are the cost and demand linkages proposed in Krugman (1991) and related work. A key prediction of this model is that firms are willing to pay workers higher wages in regions that are close to large consumer or industrial markets, since firms in these regions benefit from relatively low transport costs. This idea has its roots in an older literature on market potential (e.g., Harris 1954). Hanson (1997, 1999) tests for the presence of regional demand linkages by examining the correlation between wages and proximity to markets across states in Mexico and counties in the United States. In both cases, he finds that, controlling for regional education levels and other regional characteristics, wages are higher in locations that are closer to large consumer markets. These results are consistent with the hypothesis that that demand linkages between firms create location-specific externalities which contribute to spatial agglomeration. In related work, Dekle and Eaton (1999) use data on Japanese prefectures to estimate the rate at which agglomeration effects decay over space.

In a less structured setting, Quah (1996) examines the covaration in per capita incomes across regions within Europe. A region's per capita income is more strongly correlated with per

capita incomes in neighboring regions, regardless of nationality, than with other regions within the same country. Also, convergence in incomes over time is much stronger across neighboring regions than across regions that share the same nationality. Overman and Puga (1999) obtain similar results for regional variation in European unemployment rates. In both of these papers, the spatial distribution of economic activity is itself the unit of analysis. The point of the exercise is to estimate the distribution of incomes or unemployment conditional on outcomes in other regions. Most other empirical work we have mentioned, in contrast, examines spatial outcomes parametrically from the perspective of the representative region.

IV. Agglomeration and Industry Behavior

The empirical work we have seen so far appears to be consistent with the idea that localized externalities help explain the positive covariance between geographic concentration and wages and housing prices. Except for recent applications of the Krugman model, these strands of literature impose little structure on the environment and so do not uncover the source of externalities that might contribute to agglomeration. Theory proposes explicit reasons why firms would prefer to locate near other firms in their same industry or in related industries. Among factors that make industry clusters attractive are industry-specific spillovers that raise factor productivity and forward and backward linkages that permit firms to obtain inputs or deliver goods to market at relatively low cost. In this section, I examine recent literature on spatial agglomeration and industry behavior to see whether there is empirical evidence of such specific advantages to industry clustering. I again start with how exogenous regional characteristics might influence spatial economic behavior.

IV.A Regional Resource Supplies and Industry Location

Perhaps the simplest and best-known theory of industry location is the Heckscher-Ohlin (HO) model (see, e.g., Dixit and Norman, 1980), which attributes patterns of industrial specialization to relative resource availability. Though the HO model is usually applied to countries, rather than regions inside countries (Davis et al., 1997), it offers a parsimonious explanation for why industries locate where they do. In a two-good, two-factor, two-region setting, the HO theorem says that a region will specialize in the industry that uses relatively intensively its relatively abundant factor. Where factors are perfectly mobile across space, this statement is tautological, but as long as at least some factors are immobile the HO model helps predict which regions will produce which goods (Courant and Deardorff, 1992).

To gauge the importance of regional resource availability for regional patterns of specialization, Kim (1999) uses data on two-digit U.S. manufacturing industries to examine the correlation between state output levels and state supplies of capital, labor, mineral resources, and agricultural land between 1880 and 1987. Factor supplies can account for a large share of the cross-sectional variation in output levels across states, with relatively little change in explanatory power over the 100-year period. Though it is reasonable to treat regional supplies of land and natural resources as exogenous, by the second half of the twentieth century at least, it is probably best to think of long-run regional supplies of capital and labor as being determined by national factor markets. This makes it difficult to determine whether Michigan produces a lot of automobiles because it has large supplies of physical capital or the reverse. Consistent with regional factor supplies being endogenous, migration has contributed to convergence in regional

factor prices (Blanchard and Katz, 1992; Topel, 1986) and regional patterns of specialization have become more similar over time (Kim, 1995).[12]

IV.B Agglomeration and Industry Location

One feature of industry location which is difficult to attribute to exogenous resource supplies is that the location of industry clusters appears to change over time. Dumais, Ellison, and Glaeser (1997) find that, although the geographic concentration of industry declined only slightly between 1972 and 1992, the birth and death of manufacturing plants produced large shifts in industrial activity across regions. The death of manufacturing plants contributes to the decline of an industry cluster in a particular location, but the birth of new plants helps recreate clusters of firms in the same industry in entirely new locations. These dynamic features of firm location patterns suggest that the acquired characteristics of regions, rather than their endowed characteristics, are an important part of what attracts firms to particular locations.

Carlton (1983) examines the attractiveness of industry clusters using data on the location decisions of new plants in industries that tend to ship their goods long distances. Looking at new firms presumably nets out any effect of sunk investments on business location decisions. Controlling for local labor costs, energy costs, and tax rates, Carlton finds that new plants are more likely to choose a city the larger is own-industry employment in the city. [13] In related work, Rosenthal and Strange (1999) find evidence of highly localized own-industry agglomeration effects on births of new firms at the level of U.S. zip codes, and Wheeler and

[12] The extent of regional specialization and industry agglomeration rose during the late nineteenth century, peaked around 1914, stabilized during the inter-war period, and then declined after 1947 (Kim, 1995).

[13] There are many papers which assess how local tax rates, unionization, and other factors affect business location decisions. While early studies tend to find little impact of local taxes on local business activity (Bartik, 1985), few of these explicitly address the endogeneity of local government policies. Several recent papers account for sources of variation in local government policies and find that higher local tax rates do appear to discourage local business activity (Feldstein, Hines, and Hubbard, 1995; Hines, 1996; Holmes, 1998).

Mody (1992) find that, controlling for differences in labor cost, infrastructure, market size, economic openness, and political risk across countries, foreign investment by U.S. multinationals is higher in countries with larger initial concentrations of foreign firms.

In an approach that relates productivity to industry agglomeration more explicitly, Nakamura (1985) and Henderson (1986) estimate production functions for Japanese, and U.S. and Brazilian manufacturing industries, respectively, and examine whether factor productivity is higher in a regional industry when regional own-industry scale is higher or the regional population is higher.[14] Nakamura finds that for Japanese cities value added per worker is increasing in the scale of local industry output for most industries that produce capital goods and for a few industries that produce consumer or intermediate goods. Henderson finds that gross output per worker is increasing in local industry employment for capital goods industries in Brazilian cities and for capital and consumer good industries in U.S. cities. In terms of urbanization effects, Nakamura finds some evidence that industry labor productivity is higher in larger cities, while Henderson finds little evidence of such effects. Both authors interpret their results as evidence of industry-specific agglomeration economies. A large literature uses this general approach to estimate the strength of agglomeration effects.

One way in which agglomeration may raise industry productivity is if it expands the local demand for goods, either through market-size effects or input-output linkages between industries (Fujita, 1988; Rivera-Batiz, 1988; Krugman, 1991; and Venables, 1996). Local demand linkages matter when transport costs cause most output to be traded locally. Wolf (1997) finds that for the United States goods shipped domestically travel an average of 255 miles, with consumer

[14] Early contributions in this literature include Sveikauskas (1975), Carlino (1979), and Moomaw (1981). Much of the literature distinguishes between localization economies, in which industry productivity is correlated with industry scale, and urbanization economies, in which industry productivity is correlated with city size. See Quigley (1998) for a discussion of these and related papers.

products travelling longer distances and intermediate inputs travelling shorter distances. These patterns could be the result of downstream and upstream industries locating near each other. In support of this idea, Holmes (1999) finds that plants in more geographically concentrated industries have high input purchases relative to total sales, which may mean that they are less vertically integrated than plants outside industry clusters.

Davis and Weinstein (1999a) examine whether regional demand linkages contribute to spatial agglomeration. They find that regional industrial production across Japanese prefectures increases more than one for one with regional absorption of an industry's output.[15] This suggests that there is an excess concentration of production in regions where the demand for a good is relatively high. Such an effect could work through consumer markets – with firms concentrating production near relatively large sources of final demand, as in Krugman (1980) – or through markets for intermediate inputs – with firms concentrating near their buyers or suppliers, as in Venables (1996). In regressing regional industry output on regional industry absorption, Davis and Weinstein in effect assume that regional variation in the absorption of goods is due to exogenous regional variation in preferences. This assumption is probably more appropriate for consumer goods, such as beer or apparel, than for industrial goods, such as machine tools or chemical solvents. In related work, Justman (1994) finds strong positive co-movements between local supply and local demand for manufacturing industries in U.S. cities.

Firms appear to be both attracted to locations with large concentrations of firms in their industry and more productive in these locations. While cross-section correlations between industry agglomeration, new firm location, and labor productivity are consistent with localized externalities, they are also subject to one or more of the identification problems mentioned in section I. First, few studies control for regional characteristics beyond labor costs, tax rates, and

aggregate industry characteristics. This raises the possibility that industry agglomeration is proxying for omitted variables which influence where firms prefer to do business, such as the supply of factors besides labor and capital (Kim, 1999), the supply of exogenous amenities (Beeson and Eberts, 1989), or other characteristics (e.g., quality of local labor force, quality of local infrastructure, etc.).[16] Second, according to theory the level of industry activity in a location is jointly determined with industry demand and productivity. Since where industry agglomerates may be indeterminate ex ante, there may be few valid instruments for the scale of local industry. Without such instruments it is difficult to estimate consistently the magnitude of agglomeration effects. Third, many studies examine either a single type of agglomeration effect or unspecified agglomeration effects, and so do not resolve the issue of which types of externalities contribute to geographic concentration.

In an attempt to control for the first two identification problems, Head, Ries, and Swenson (1995) examine the location decision of new Japanese manufacturing plants in the United States. Controlling for the local concentration of existing U.S. plants in the same industry (and other region and industry characteristics), Japanese plants are more likely to choose a location the higher is the existing local concentration of existing Japanese manufacturing plants in the same industry. While this finding reproduces Carlton's (1983) result -- that new plants are drawn to clusters of plants in their industry -- their approach, by holding constant local U.S. industry activity, controls for a much wider range of unobserved site-specific characteristics.[17] These results remain silent, however, on the source of agglomeration effects.

IV.C Agglomeration and Industry Growth

[15] Davis and Weinstein (1999b) extend this approach to an international setting.
[16] See Carlino and Voith (1992) for estimation of state industry productivity as a function of state characteristics.

An alternative strategy to identify the effects of agglomeration economies on industry location is to examine variation in industry growth or innovation across regions. One common approach is to estimate regional industry growth as a function of initial regional conditions. This has two advantages. First, looking at growth allows one to net out the effects of time-invariant regional characteristics which affect the distribution of economic activity in the cross section but not over time. Second, initial conditions are pre-determined and so may be reasonably assumed to be uncorrelated with future shocks to regional industries.[18] One disadvantage is that this approach is silent about whether the estimation results reflect a steady-state growth process or the transition to a steady state. This complicates interpreting parameter estimates.[19]

In an influential paper, Glaeser et al. (1992) examine changes in industry employment from 1956 to 1987 for U.S. metropolitan areas. Controlling initial relative establishment size, initial relative industry size, initial wage levels, and overall city employment growth, they find that employment growth in a city industry is positively correlated with the initial diversity of industry employment in the city but not with initial own-industry employment in the city.[20] This result is interpreted to mean that firms benefit from agglomeration due to interaction with firms in a wide range of industries, but not firms in their own industry.[21] In complementary work, Henderson, Kuncoro, and Turner (1995) examine employment growth over the period 1970 to 1987 for a subset of manufacturing industries across U.S. cities. Similar to Glaeser et al., they find that for new industries employment growth in a city is positively correlated with initial

[17] For related work see Smith and Florida (1994) and Aitken, Hanson, and Harrison (1997).
[18] Initial conditions may, of course, be the result of endogenous outcomes. But, to the extent they are pre-determined, they are uncorrelated with current regional shocks may be treated as exogenous in a statistical sense.
[19] A further disadvantage is that this approach negates the possibility of estimating the effects of static externalities which may contribute to spatial agglomeration.
[20] This paper and others in this strand of literature focus on localization economies, agglomeration economies specific to a particular industry. To net out the effects of urbanization economies, agglomeration economies common to many industries, the typical approach is control the impact of initial city size on industry growth.
[21] See Duranton and Puga (1999b) for a discussion of theoretical issues related to urban diversity.

industry diversity in a city, but, in contrast to the earlier paper, they find that for mature industries employment growth is positively correlated with initial own-industry employment in the city and uncorrelated with initial industry diversity.[22] They interpret these results to mean that benefits to agglomeration vary over the industry life cycle. Firms in newer industries benefit from exposure to ideas drawn from many sources, while firms in industries with established production techniques benefit from proximity to firms in similar lines of activity.

Henderson (1999) uses a production-function approach to assess the relationship between industry agglomeration and growth.[23] He examines output growth in U.S. manufacturing plants from machinery and high-tech industries at five-year intervals over the 1963-1992 period. Controlling for plant inputs, time fixed effects, and plant fixed effects, he finds that plant output (or, more precisely, deviations from mean plant output over time) are (a) positively correlated with the contemporaneous number of own-industry plants in the same county, with this effect being stronger for high-tech than for machinery; (b) positively correlated with the lagged number of own-industry plants in the same county, with this effect holding for high-tech only; and (c) uncorrelated with city industrial diversity or total city manufacturing employment. There appear, however, to be limits to the robustness of these results. Measuring agglomeration using employment instead of number of plants lowers estimated agglomeration effects, as does controlling for city-industry-year fixed effects. Also, instrumenting for plant inputs and industry agglomeration produces implausible results or imprecisely estimated agglomeration effects.

It appears that for mature industries growth is faster where industry diversity is greater, while for young industries growth is faster where own-industry agglomeration is greater. The concern about results for short time horizons is that endogeneity problems are likely to be much

[22] Hanson (1998) also finds no evidence of industry diversity effects on industry growth in the medium run.

more severe in this setting. In an attempt to identify specific types of agglomeration effects, Dumais, Ellison, and Glaeser (1998) examine the correlates of industry employment growth over five-year intervals for U.S. cities over the period 1972-1992. Industry employment growth is higher in cities whose industries (a) use workers in similar occupations to the industry in question, and (b) are relatively specialized in producing inputs demanded or supplied by the industry in question. It appears that firms are attracted to regions that have relatively abundant supplies of workers they are likely to hire, as would be consistent with Marshall (1920), and of firms in upstream and downstream activities, as would be consistent with Venables (1996) and Fujita, Krugman, and Venables (1999). At the level of national industries, Bartelsman, Caballero, and Lyons (1994) and Paul and Siegel (1999) also find evidence that external scale effects are transmitted through industry input-output linkages.

Another way to identify the source of agglomeration benefits is to focus on a particular set of externalities. Jaffe, Trajtenberg, and Henderson (1993) examine the geographic localization of U.S. patent citations. New patents are required to cite previous patents which relate to the innovation the new patent embodies. Jaffe et al. find that new patents and cited patents are much more likely to originate in the same city than are new patents and a control group of patents, where the control group is selected randomly from patents issued at the same date and in the same industry as the new patent. They interpret this result to mean that there are location-specific spillovers associated with innovation, which contribute to the incentive for industry localization. Consistent with this idea, Audretsch and Feldman (1996) find that new

[23] A related literature estimates production functions using data on national industries over time (Caballero and Lyons 1990, Bartelsman, Caballero, and Lyons 1994, Paul and Siegel 1999).

innovations are concentrated in locations with relatively high spending on R&D, spending on university research, and employment of skilled labor.[24]

V. Spatial Interactions and Urban Growth

In the formal theoretical models discussed in section II, agglomeration in one region affects outcomes in other regions. As the economy of Los Angeles expands, for instance, firms in neighboring cities, such as Anaheim or San Bernadino, may benefit. Growth in Los Angeles may expand the regional market for their goods or increase the local public stock of knowledge, raising the productivity of the factors they hire.[25] More distant regions, however, are less likely to benefit. All else equal, growth in Los Angeles necessitates the in-migration of labor, which means that regions losing workers contract.

While regional interactions of this kind are a fundamental aspect of any general equilibrium spatial model, they are missing from most empirical literature we have discussed so far. Only a few of the cited papers (Quah 1996, Hanson 1999, and Overman and Puga 1999) bring spatial interactions into the analysis. The standard approach in most other literature is to see whether in the representative region the local agglomeration of economic activity affects local wages, productivity, growth, etc. This approach ignores how either regional demand linkages, as in Fujita, Krugman, and Venables (1999), or cross-city patterns of specialization, as in Henderson (1988), influence outcomes in different locations. An emerging body of literature is beginning to address the impact of spatial economic linkages on urban economies.

Black and Henderson (1999b) examine the correlation between population growth in U.S. metropolitan areas and initial conditions over the period 1950-1990. Population growth is faster

[24] See Keller (1999) for evidence on international technology spillovers.
[25] There may also be negative effects (e.g., if growth in Los Angeles makes surrounding areas more congested).

in cities with warmer climates, less precipitation, and closer proximity to a coast. In terms of agglomeration effects, population growth is also faster in cities which are closer to cities with larger initial populations, with this effect weakening as neighboring population masses become larger. This is consistent with the idea that market potential (Harris, 1954; Fujita, Krugman, and Venables, 1999), or the potential demand for goods and services created by proximity to population masses, contributes to spatial agglomeration.

In related work, Dobkins and Ioannides (1999) examine the relationship between city population growth and neighboring regions in more detail. Using data for U.S. metropolitan areas over the period 1900 to 1990, they find that population growth is decreasing in a city's distance to the nearest neighboring larger city. This result depends on whether a city is spatially isolated or adjacent to another city (e.g., Albuquerque is isolated, Fort Worth is not). For cities with adjacent neighbors, there is no effect of distance to larger cities on population growth. These results are also consistent with the market-potential hypothesis, but may, alternatively, capture less well-understood features of how systems of cities evolve over time.[26]

The previous two studies exploit the correlation between city growth and lagged city characteristics to examine how city sizes changes over time. While this approach controls for simultaneity in regional outcomes, there is no guarantee that initial conditions are a reliable predictor of city growth (Dobkins and Ionnides, 1999). An alternative way to address simultaneity problems is through calibration. Fan (2000) calibrates models based on Henderson (1988) and Krugman (1991) to U.S. data. The theoretical models treat the United States as a rectangle composed of many symmetric sub-regions with randomly varying levels of exogenous amenities. Trade between regions occurs over land and/or through a fixed number of coastal

[26] Ioannides and Overman (1999) examine issues related to city growth using a non-parametric approach similar to that in Quah (1996) and Overman and Puga (1999).

ports. Fan examines whether parsimoniously specified Henderson or Krugman-style models can replicate the spatial distribution of U.S. employment or labor earnings over the period 1970-1990. While he does not conduct formal hypothesis tests, the data appear to be more supportive of the Krugman model than of the Henderson model.

Recent literature uses a variety of different approaches to examine spatial interactions between regions. As most of these papers are unpublished it is probably too early to pass judgement on this strand of literature, but it does appear that linkages between regions matter for how cities grow over time and for the spatial distribution at any point in time.

VI. Conclusion

Economics is experiencing one of its periodic waves of interest in spatial economic issues.[27] The timing of this development is a bit ironic. Just as many academic economists begin to recognize, yet again, that the spatial distribution of economic activity holds clues for the deep structure of the economy, popular literature begins to suggest that advances in information technology will soon divorce geography from business location decisions (Cairncross, 1997; Coyle, 1997). What is amusing about this contrast in perspectives is that technology is central to both lines of argument. Academic economists claim that space is important because of its impact on how technology is created and diffused, while popular writers suggest that it is technology that is making space irrelevant. Recent academic literature has shouldered the burden of trying to demonstrate that space matters for how the economy is organized.

In this paper, I have reviewed recent empirical research on agglomeration economies and spatial economic behavior. One goal of the review has been to assess whether the literature has

[27] The current is the fourth such wave this century according to Quigley (1998).

addressed estimation problems which confound the identification of external effects. There appear to be two empirical results that are reasonably robust to estimation problems created by exogenous regional characteristics and simultaneity in regional data. First, individual wages are increasing in the presence of highly-educated workers in the local labor force, which is consistent with the idea that localized human-capital externalities make the social returns to education higher than the private returns. Second, for at least some industries long-run growth is higher in regions with a wider range of industrial activities, which suggests that firms benefit from being in more diverse environments.

Other patterns of correlation are also present, but subject to greater concerns about identification. One such pattern is that short run, but not long run, industry growth and productivity are higher where own-industry agglomeration is higher. That own-industry agglomeration effects are temporary (Dumais, Ellison, and Glaeser, 1997) suggests they are not solely a product of exogenous regional characteristics. Empirical modeling of the dynamics of industry agglomeration is clearly an area where more work is needed. A second evident pattern of correlation is that regional demand linkages, as measured by market potential, are associated with higher wages and faster urban population growth. This result is broadly consistent with the new economic geography models in the vein of Fujita, Krugman, and Venables (1999). To this point, however, there has been little consistency across papers in how the market-potential concept is applied. This strand of literature would benefit from exploiting the well-specified structural relationships identified by theory as a basis for empirical work.

Taken together, the body of empirical results suggest that location-specific externalities exist and influence the spatial distribution of economic activity. The empirical literature is noticeably short, however, on attempts to estimate the impact of different types of externalities or

how different types of congestion costs constrain agglomeration. The evidence which supports specific models, such as own-industry agglomeration effects which are consistent with Henderson (1974, 1988) and regional demand linkages which are consistent with Fujita, Krugman, and Venables (1999), is also some of the evidence subject to relatively severe estimation problems. The only strong evidence in support of a particular model is that on localized human capital externalities, as first proposed by Lucas (1988).

In its current state, the empirical literature is then a poor guide for which theoretical model would be most appropriate to use for evaluating the welfare consequences of agglomeration or to evaluate the impact of specific policy changes on the spatial distribution of economic activity. An important direction for future research is to develop models which integrates multiple sources of externalities and which could be used to estimate the relative importance of different types of agglomeration economies for spatial outcomes.

In closing, it is worth identifying some phenomena the literature has missed entirely, or nearly so. In section V, I discussed the absence of spatial linkages in empirical work on economic geography. Current research efforts (e.g., Black and Henderson 1999b, Dobkins and Ionnides 1999, Overman and Puga 1999) will hopefully make this point moot.

Another issue which is underrepresented in empirical research is the role of services in how cities are organized. In reading recent empirical work on spatial agglomeration, one could easily get the impression that manufacturing is what drives dynamic urban economies. This may have been true once for the United States, and may still be true for some U.S. cities and many cities in developing countries, but the largest and most dynamic cities in terms of income and income growth -- Chicago, Los Angeles, New York (not to mention Hong Kong, London, and Tokyo) -- are all centers for business services. Kolko (1999) shows that from 1910 to 1990 the

share of U.S. urban employment declined for manufacturing and rose for business services in all cities, with the biggest shift occurring in the largest cities. For large metropolitan areas business services have surpassed manufacturing in terms of employment. Business services, which include accounting, computer, engineering, finance, insurance, and legal services, span many export activities and so in principle may be subject to the agglomerative forces described in section II. We know embarrassingly little about why these activities appear to concentrate in large cities and whether they are subject to agglomeration economies in a manner similar to manufacturing. There remains, at least, no shortage of work to be done.

References

Abraham, F. 1996. "Regional Adjustment and Wage Flexibility in the European Union." *Regional Science and Urban Economics* 26: 51-75.

Acemoglu, D. 1996. "A Microfoundation for Social Increasing Returns in Human Capital." *Quarterly Journal of Economics* 111: 779-804.

Acemoglu, D. and J. Angrist. 1999. "How Large Are The Social Returns to Education? Evidence from Compulsory Schooling Laws." NBER Working Paper No. 7444.

Angrist, J. and A. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics* 106: 979-1014.

Audretsch, D. and M. Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review* 86: 630-640.

Aitken, B., G.H. Hanson, and A.E. Harrison. 1997. "Spillovers, Foreign Investment, and Export Behavior," *Journal of International Economics*, 43: 103-132.

Bartelsman, E.J., R.J. Caballero, and R.K. Lyons. 1994. "Customer- and Supplier-Driven Externalities." *American Economic Review* 84: 1075-1084.

Bartik, T.J. 1985. "Business Location Decisions in the United States." *Journal of Business and Economic Statistics* 14-22.

Beeson, P.E. 1991. "Amenities and Regional Differences in Worker Characteristics." *Journal of Urban Economics* 30: 224-241.

Beeson, P.E. and R. Eberts. 1989. "Identifying Productivity and Amenity Effects in Interurban Wage Differentials." *Review of Economics and Statistics* 71: 443-452.

Black, D. 1998. "Local Human Capital Externalities: Educational Segregation and Inequality." Mimeo, London School of Economics.

Black, D. and V. Henderson. 1999a. "A Theory of Urban Growth." *Journal of Political Economy* 107: 252-284.

Black, D. and V. Henderson. 1999b. "Spatial Evolution of Population and Industry in the United States." *American Economic Review Papers and Proceedings* 89: 321-327.

Blanchard, O. and L. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity* 1: 1-75.

Borjas, G.J., S.G. Bronars, and S.J. Trejo. 1992. "Self-Selection and Internal Migration in the United States." *Journal of Urban Economics* 32: 159-185.

Caballero, R.J. and R.K. Lyons. 1990. "Internal versus External Economies in European Industry." *European Economic Review* 34: 805-830.

Cairncross, F. 1997. *The Death of Distance*. Cambridge, MA: Harvard Business School Press.

Carlino, G. 1979. "Increasing Returns to Scale in Metropolitan Manufacturing." *Journal of Regional Science* 19: 363-373.

Carlino, G. and R. Voith. 1992. "Accounting for Differences in Aggregate State Productivity." *Regional Science and Urban Economics* 22: 597-617.

Carlton, D.W. 1983. "The Location and Employment Choices of New Firms." *Review of Economics and Statistics* 65: 440-449.

Ciccone, A. and R. Hall. 1996. "Productivity and the Density of Economic Activity." *American Economic Review* 86: 54-70.

Courant, P.N. and A.V. Deardorff. 1992. "International Trade with Lumpy Countries." *Journal of Political Economy* 100: 198-210.

Coyle, D. 1997. *The Weightless World*. Cambridge, MA: MIT Press.

Davis, Donald R., David E. Weinstein, Scott C. Bradford, and Kazushige Shimpo. 1997. "Using International and Japanese Regional Data to Determine When the Factor Abundance Theory of Trade Works." *American Economic Review*: 421-446.

Davis, D.R. and D.E. Weinstein. 1999a. "Economic Geography and Regional Production Structure: An Empirical Investigation." *European Economic Review* 43: 379-407.

Davis, D.R. and D.E. Weinstein. 1999b. "Market Access, Economic Geography, and Comparative Advantage: An Empirical Assessment." Mimeo, Columbia University.

Dekle, R. and J. Eaton. 1999. "Agglomeration and Land Rents: Evidence from the Prefectures." *Journal of Urban Economics* 46: 200-214.

Dixit, A.K and V. Norman. 1980. *Theory of International Trade*. Cambridge University Press.

Dobkins, L.H. and Y. Ioannides. 1996. "Spatial Interactions among U.S. Cities: 1900-1990." Mimeo, Emory University and Tufts University.

Dumais, G., G. Ellison, and E. Glaeser. 1997. "Geographic Concentration as a Dynamic Process." NBER Working Paper No. 6270.

Duranton, G. and D. Puga. 1999a. "Nursery Cities: Urban Diversity, Process Innovation, and the Life-Cycle of Products." CEPR Discussion Paper No. 2376.

Duranton, G. and D. Puga. 1999b. "Diversity and Specialization in Cities: Why, Where and When Does It Matter?" CEPR Discussion Paper No. 2256.

Eaton, J. and Z. Eckstein. 1997. "Cities and Growth: Theory and Evidence from France and Japan." *Regional Science and Urban Economics* 27: 443-474.

Ellison, G. and E.L. Glaeser. 1997. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy* 105: 889-927.

Fan, W. 2000. "Internal versus External Economies of Scale in Economic Geography." Mimeo, University of Michigan.

Farber, S. and R. Newman. 1987. "Accounting for South/North Real Wage Differentials and for Changes in those Differentials over Time." *Review of Economics and Statistics* 69: 215-223.

Feldstein, M., J.R. Hines, and R.G. Hubbard, eds. 1995. *The Effects of Taxation on Multinational Corporations.* Chicago: University of Chicago Press.

Fujita, M. 1988. "Monopolistic Competition and Urban Systems." *European Economic Review* 37: 308-15.

Fujita, M., P. Krugman, and A. Venables. 1999. *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, MA: MIT Press.

Fujita, M. and T. Mori. 1996. "The Role of Ports in the Making of Major Cities: Self-Agglomeration and Hub-Effect." *Journal of Development Economics* 49: 93-120.

Fujita, M. and T. Mori. 1997. "Structural Stability and Evolution of Urban Systems." *Regional Science and Urban Economics* 27: 399-442.

Fujita, M. and J.F. Thisse. 1996. "Economics of Agglomeration." *Journal of the Japanese and International Economies* 10: 339-378.

Glaeser, E.L., H. Kallal, J. Sheinkman, and A. Shleifer. 1992. "Growth in Cities." *Journal of Political Economy* 100: 1126-1152.

Glaeser, E.L. and D.C. Mare. 1994. "Cities and Skills." NBER Working Paper No. 4728.

Glaeser, E.L., J.A. Scheinkman, and A. Shleifer. 1995. "Economic Growth in a Cross-Section of Cities." *Journal of Monetary Economics* 36: 117-143.

Gyourko, J. and J. Tracy. 1991. "The Structure of Local Public Finance and the Quality of Life." *Journal of Political Economy* 99: 774-806.

Hanson, G. 1996. "Localization Economies, Vertical Organization, and Trade," *American Economic Review* 86: 1266-1278.

Hanson, G. 1997. "Increasing Returns, Trade, and the Regional Structure of Wages." *Economic Journal* 107: 113-133.

Hanson, G. 1998. "Regional Adjustment to Trade Liberalization." *Regional Science and Urban Economics* 28: 419-444.

Hanson, G. 1999. "Market Potential, Increasing Returns, and Geographic Concentration." Mimeo, University of Michigan.

Harris, C.D. 1954. "The Market as a Factor in the Localization of Industry in the United States." *Annals of the Association of American Geographers* 44: 315-348.

Head, K., J. Ries, and D. Swenson. 1995. "Agglomeration Benefits and Location Choice." *Journal of International Economics* 38: 223-248.

Helpman, E. 1998. "The Size of Regions." In D. Pines, E. Sadka, and I. Zilcha, eds., *Topics in Public Economics*. Cambridge: Cambridge University Press, pp. 33-54.

Henderson, J.V. 1974. "The Sizes and Types of Cities." *American Economic Review* 64: 640-56.

Henderson, J.V. 1986. "Efficiency of resource usage and city size." *Journal of Urban Economics* 19: 47-70.

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

Henderson, J.V. 1996. "Ways to Think about Urban Concentration: Neoclassical Urban Systems versus the New Economic Geography." *International Regional Science Review* 19: 31-36.

Henderson, V. 1999. "Marshall's Scale Economies." NBER Working Paper No. 7358.

Henderson, J.V., Kuncoro, A., and M. Turner. 1995. "Industrial Development and Cities." *Journal of Political Economy* 103: 1067-1081.

Hines, J.R. 1996. "Altered States: Taxes and the Location of Foreign Direct Investment in America." *American Economic Review* 86: 1076-1094.

Holmes, T.J. 1998. "The Effects of State Policies on the Location of Manufacturing Industry: Evidence from State Borders." *Journal of Political Economy* 106: 667-705.

Holmes, T.J. 1999. "Localization of Industry and Vertical Disintegration." *Review of Economics and Statistics* 81: 314-325.

Ioannides, Y. and H. Overman. 1999. "Spatial Evolution of the Urban System." Mimeo, Tufts University and London School of Economics.

Jaffe, A., M. Trajtenberg, and R. Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics* 108: 577-598.

Justman, M. 1994. "The Effect of Local Demand on Industry Location." *Review of Economics and Statistics* 76: 742-753.

Keely. L.C. 1999. "Exchanging Good Ideas." Mimeo, Oxford University.

Keller, W. 1999. "Geographic Localization of International Technology Diffusion." NBER Working Paper No. 7509.

Kim, S. 1995. "Expansion of Markets and the Geographic Distribution of Economic Activities." *Quarterly Journal of Economics* 110: 881-908.

Kim, S. 1999. "Regions, Resources, and Economic Geography: Sources of U.S. Regional Comparative Advantage, 1880-1987." *Regional Science and Urban Economics* 29: 1-32.

Kolko, J. 1999. "Can I Get Some Service Here?" Mimeo, Harvard University.

Krugman, P. 1980. "Scale Economies, Product Differentiation, and the Pattern of Trade." *American Economic Review* 70: 950-959.

Krugman, P. 1991. "Increasing Returns and Economic Geography." *Journal of Political Economy* 99: 483-499.

Krugman, P. 1996. "Urban Concentration: The Role of Increasing Returns and Transport Costs." *International Regional Science Review* 19: 5-30.

Krugman, P. 1997. *Development, Geography, and Economic Theory*. Cambridge: MIT Press.

Lucas, R.E. 1988. "The Mechanics of Economic Development." *Journal of Monetary Economics* 22: 3-42.

Marshall, A. 1920. *Principles of Economics.* New York: Macmillan.

Montgomery, E. 1992. "Evidence on Metropolitan Wage Differences across Industries over Time." *Journal of Urban Economics* 31: 69-83.

Moomaw, R.L. 1981. "Productivity and City Size: A Critique of the Evidence." *Quarterly Journal of Economics* 96: 675-688.

Moretti, E. 1999. "Estimating the Social Return to Education: Evidence from Repeated Cross-Sectional and Longitudinal Data." Mimeo, UC Berkeley.

Nakamura, R. 1985. "Agglomeration Economies in Urban Manufacturing Industries: A Case of Japanese Cities." *Journal of Urban Economics* 17: 108-124.

Ottaviano, G. and J. Thisse. 1998. "Agglomeration and Trade Revisited." CEPR Discussion Paper.

Overman, H. and D. Puga. 1999. "Unemployment Clusters across European Regions and Countries." CEPR Discussion Paper No. 2225.

Paul, C.J.M. and D.S. Siegel. 1999. "Scale Economies and Industry Agglomeration Externalities: A Dynamic Cost Function Approach." *American Economic Review* 89: 272-290.

Peri, G. 1998. "Human Capital Externalities and U.S. Cities." Mimeo, University of California.

Puga, D. 1999. "The Rise and Fall of Regional Inequalities." *European Economic Review* 43: 303-334.

Quah, D. 1996. "Regional Convergence Clusters across Europe." *European Economic Review* 40: 951-58.

Quah, D. 1999. "Internet Cluster Emergence." CEPR Discussion Paper No. 2293.

Quigley, J.M. 1998. "Urban Diversity and Economic Growth." *Journal of Economic Perspectives* 12: 127-138.

Rauch, J.E. 1993. "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities." *Journal of Urban Economics* 34: 380-400.

Rivera-Batiz, F.L. 1988. "Increasing Returns, Monopolistic Competition, and Agglomeration Economies in Consumption and Production." *Regional Science and Urban Economics* 18: 125-53.

Roback, J. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90: 1257-1278.

Rosen, S. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82: 34-55.

Rosen, S. 1979. "Wage-Based Indexes of the Urban Quality of Life." In P. Mieszkowski and M. Straszheim. Baltimore: Johns Hopkins University Press.

Rosenthal, S.S. and W.C. Strange. 1999. "Geography, Industrial Organization, and Agglomeration." Mimeo, University of British Columbia.

Smith, D.F. and R. Florida. 1994. “Agglomeration and Industrial Location: An Econometric Analysis of Japanese-Affiliated Manufacturing Establishments in Automotive-Related Industries.” *Journal of Urban Economics* 36: 23-41.

Sveikauskas, L. 1975. “The Productivity of Cities.” *Quarterly Journal of Economics* 89:393-413.

Topel, R.H. 1986. "Local Labor Markets." *Journal of Political Economy* 94: S111-S143.

Treyz, G.I. 1991. “Causes of Changes in Wage Variation among States.” *Journal of Urban Economics* 29: 50-62.

Tabuchi, T. 1998. “Urban Agglomeration and Dispersion: A Synthesis of Alonso and Krugman.” *Regional Science and Urban Economics* 44: 333-351.

Venables, A.J. 1996. "Equilibrium Locations of Vertically Linked Industries." *International Economic Review* 37: 341-360.

Wheeler, D. and A. Mody. 1992. "International Investment Location Decisions." *Journal of International Economics* 33: 57-76.

Wolf, H.C. 1997. “Patterns of Intra- and Inter-State Trade.” NBER Working Paper, No. 5939.