NBER WORKING PAPER SERIES

CITIES, REGIONS AND THE DECLINE OF TRANSPORT COSTS

Edward L. Glaeser
Janet E. Kohlhase

Cities, Regions and the Decline of Transport Costs
Edward L. Glaeser and Janet E. Kohlhase
NBER Working Paper No. 9886
July 2003
JEL No. R12, R14, R23, R41, J30

ABSTRACT

The theoretical framework of urban and regional economics is built on transportation costs for manufactured goods. But over the twentieth century, the costs of moving these goods have declined by over 90% in real terms, and there is little reason to doubt that this decline will continue. Moreover, technological change has eliminated the importance of fixed infrastructure transport (rail and water) that played a critical role in creating natural urban centres. In this article, we document this decline and explore several simple implications of a world where it is essentially free to move goods, but expensive to move people. We find empirical support for these implications.

Edward L. Glaeser
Department of Economics
Littauer Center 315A
Harvard University
Cambridge, M 02138
and NBER
eglaeser@harvard.edu

Janet E. Kohlhase
Department of Economics
University of Houston
Houston, TX 77204-5019
jkohlhase@uh.edu

1 Introduction

The new regional economics has been built on transport costs. The models of Krugman (1991a,b), Fujita, Krugman and Venables (1999), Fujita and Thisse (2002) follow the classic firm location literature of Weber (1909, 1929), the central place theories of Christaller (1933, 1966), Losch (1944, 1954), and the spatial economy approach of Isard (1956), and develop models where economic decisions are both created – and limited – by the costs of moving goods over space.[1] What is the benefit of being close to one another? Being able to buy goods that are produced locally. What is the cost of agglomerating? Having to ship basic commodities farther over space.

The foundational models of urban economics are also, if less clearly, founded on transportation technologies. The classic monocentric urban models developed by Alonso (1960, 1964), Muth (1969) and Mills (1967, 1972), hereafter referred to as the Alonso-Mills-Muth model, are consistent with a world where people walk and take public transportation.[2] The central business district (CBD) can be viewed as the hub for transportation technologies. Public transportation brings people to the hub and people walk from that point to their work places and use their feet to interact during the workday. The workers' physical output then gets shipped from the hub to consumers using rail and water transport. As such, a monocentric model, where firms are extremely close to one another in the central business district (CBD), is natural when thinking about cities built around feet and trains.

Historically, these models capture the essence of urban economies. In the United States transport costs before 1900 were enormously high. People moved by foot and goods were

---

[1] Several texts develop the classic regional economics models and location theory models. See Beckmann (1968), Greenhut (1970), Nourse (1968) or Hoover and Giarratani (1985). The last is available online at the web page of regional science http://www.rri.wvu.edu/regscweb.htm.

[2] The classic monocentric model is discussed in detail in the text by Fujita (1989). Several summary articles, including Anas, Arnott and Small (1998) and Brueckner (1987), discuss the basic monocentric model and comparative statics. The agricultural land use model developed by von Thünen (1826, 1966) can be considered one of the antecedents to the monocentric urban land use models.

carried by water. Both the structure and location of cities reflect high transport costs. Because roads and rail were rare and costly, every large city in 1900 was located on a waterway. Of the 20 largest cities in America in 1900, seven were ocean ports where rivers meet the sea (Boston, Providence, New York, Jersey City, Newark, Baltimore, and San Francisco); five were ports where rivers meet the Great Lakes (Milwaukee, Chicago, Detroit, Cleveland, and Buffalo); three were on the Mississippi river (Minneapolis, St. Louis and New Orleans); three were on the Ohio river (Louisville, Cincinnati and Pittsburgh); and the remaining two were on east coast rivers, close to the Atlantic (Philadelphia and Washington). America's largest city, New York, was clearly its best port.

Moreover, high transport costs created a series of cities scattered throughout the country, each oriented towards exploiting the rich American hinterland. Cronon (1991) details how Chicago was built as a transport hub enabling the movement of commodities such as lumber, wheat and cattle from America's vast hinterland to the east coast and to Europe.[3] Moreover, transport costs dictated the existence of a hierarchy of cities (Christaller 1933, 1966; Losch 1944, 1954; Henderson 1974). Small cities dotted the interior of the country and were specialised in providing basic services for men making their living from the earth. Larger cities served as depots for goods coming from and headed to the east coast and Europe. They also increasingly became centres of manufactured goods, which were moved naturally, exploiting their comparative advantage in transportation.

We know this urban world because it still surrounds us. The durability of physical infrastructure ensures us that cities built around Great Lakes still remain even though the lakes' importance as a transport mechanism has declined. Baltimore, Buffalo, Cleveland, Detroit, Pittsburgh and St. Louis (in the Midwest) and Baltimore and Philadelphia (on the coast) excelled in moving the products of the American country by water or rail. These are seven of the 15 largest cities in the country in 1950, and every one of them has lost population in every decade since 1950. They continue to exist, but their gradual

---

[3] Other classics which emphasise the link between transport technologies and urban form include the multiple nuclei model of Harris and Ullman (1945) and Blumenfeld (1955).

weakening should remind us that the world has changed, and the factors that made these places centres of productivity in 1900 seem unimportant 100 years later.

The great force that has reshaped the city in the twentieth century is the engine; it has become both more powerful and noticeably lighter over the last 100 years.[4] As a result, people have increasingly been able to propel themselves and their goods over long distances with better and better machines. The average cost of moving a ton a mile in 1890 was 18.5 cents (in 2001 dollars). Today, this cost is 2.3 cents. At their height, the transportation industries represented 9% of GDP. Today, if we exclude air travel, they represent 2% of national product. Two factors have acted to decrease the importance of transportation costs for goods. First, the technologies designed for moving goods have improved. Second, the value of goods lies increasingly in quality, rather than quantity, so that we are shipping far fewer tons of goods relative to GDP than we have in the past.

These reduced costs, and the declining importance of the good-producing sector of the economy, means that in our view, it is better to assume that moving goods is essentially costless than to assume that moving goods is an important component of the production process. The implication of this claim is that the prevailing urban models are poorly suited for the twenty-first century-city, although they do help us to understand the 19th century cities that still surround us. As we contemplate the future of urban and regional economics, it becomes crucial to build a new theoretical paradigm built on forces other than the costs of moving physical products across space.

In this article we do three things. First, we review the evidence on the decline of transportation costs. The evidence indicates a large change that appears to be continuing. Certainly it is an exaggeration to claim that moving goods is free, but it is becoming an increasingly apt assumption. We also note that moving people is not free and that the twentieth century has seen a switch from infrastructure-heavy transport (railroads, ports) to infrastructure-light transport (cars, trucks). This also has implications for urban form.

---

[4] Mumford (1934, 1963) emphasised the powerful role of technology in transforming social organisation.

Second, we discuss the implications of this change to the analysis of urban and regional economics. The new regional economics is built around fixed cost technologies with substantial transport costs. Population is anchored by a desire to be in close proximity to natural resources. In Krugman (1991) it is the fixed agricultural sector that ties workers to both regions of the economy. These assumptions fit nineteenth century America perfectly, but are inappropriate for the twenty-first century. Instead, models should continue to emphasise agglomeration effects, which come ultimately from the benefits of easy access to other people, not from saving transport costs for goods. Natural resources in production are becoming increasingly irrelevant, and the items that provide features to the landscape are consumption-related natural amenities (e.g., warmth) and state or city specific government policies.

Third, we empirically test, in Sect's. 4 and 5, a number of implications stemming from our approach. People should leave areas that were once centres for natural resource extraction and instead should increasingly live in a small handful of metropolitan areas. The main factors that stop convergence to a single area are heterogeneity of tastes for weather, government policies and historically developed infrastructure. Manufacturing should be located in lower density areas; services should be in areas with greater density. Service firms should locate near their suppliers and customers; manufacturing firms should not. Extreme congestion on urban roads may have negative productivity impacts. In the last section of the article, we synthesize our discussion and offer suggestions for future modelling efforts.

2   Transportation costs over time

A standard theoretical element in the new regional economics-new economic geography literature follows Samuelson (1954) and assumes iceberg transportation costs.[5] The

---

[5] See Fujita, Krugman and Venables (1999) (FKV) for a discussion of iceberg transport costs and its use in regional science, urban economics, economic geography, and international trade models. FKV point out (p. 59) that the agricultural land-use model of von Thünen (1826, 1966) contains the predecessor to Samuelson's iceberg transport costs; in that model one of the assumptions was that the oxen pulling the loaded carts ate some of the grain being shipped to the market.

iceberg specification maintains that the cost of moving goods involves the loss of some proportion of the product during shipment. Obviously, this assumption is a simplification and misses much that is interesting in the economics of moving goods, but nonetheless it provides us with a convenient starting point for thinking about the magnitude of transportation costs. If these costs were truly iceberg, then we could estimate the size of the cost parameter just by examining the share of GDP spent on moving goods. As such, one starting point for thinking about the magnitude of transportation costs is the share of GDP involved in the transportation industry. At a more detailed level, a measure of transport costs is the price of moving a commodity as a fraction of the total value of the commodity.

Figure 1 shows the overall share of transport in GDP from the late nineteenth century until today. This sector of the economy includes rail, water and pipeline transportation, trucking and warehousing, air transport, transportation services, and local and interurban passenger transit. Prior to 1929, we use the Martin Series (Historical Statistics of the United States, F-250-261), which includes both transportation and public utilities. Since this series includes components unrelated to transportation, we multiplied the series by .67 – the ratio of transportation spending to transportation plus public utilities in 1929. If this ratio was not constant, but rather decreasing between 1870 and 1930 (which seems probable), then transportation spending would be declining more sharply before 1929.

As late as 1929 (the first year we have Department of Commerce data available), transportation represented 8% of gross domestic product. By 1990, only 3% of GDP is being spent on transportation. This figure understates the true decline of transportation because air travel, which is overwhelmingly involved in transporting people, not goods, is a major component of transportation expenditures during the later time period. The triangles in the figure represent the transport cost series without air transportation. This figure is, unsurprisingly, almost the same as total transportation expenditures in 1949, but by the 1990s, more than one-quarter of total spending in this category was on air transport. Without that category, transportation represents only 2.3% of GDP in the 1990s.

Of course this figure does not truly represent an estimate of iceberg costs, even in the best of circumstances, because a significant fraction of GDP is not shipped. Services tend to involve little freight shipment. Other more physical goods only involve small amounts of shipping (e.g., construction). Moreover, many physical goods are actually consumed at home and not shipped. Since only a fraction of GDP (perhaps one-half) is in physical goods that are traded, the share of GDP spent on transportation is something of an underestimate of the hypothetical iceberg costs, perhaps by as much as one-half.

Another reason that these numbers may tend to underestimate the overall importance of shipping costs in the economy is that they exclude shipping that is done in-house. When a manufacturing firm hires an external shipper, that payment is included in the share of GDP in the transportation industries. When a firm uses its own trucks, the salaries of the trucks will not be attributed to the transportation industry. Furthermore, to the extent that the government subsidises the trucking industry through the construction and maintenance of roads, those costs will not be included in these figures. The Eno Foundation has attempted to calculate a national transportation bill to include all of the different components of transportation expenditures. This bill is not exactly comparable to GDP, because if this procedure were done for all of the different functions of the economy, the sum of the results would add to more than GDP.

Figure 2 uses the Eno Foundation data to show the shift in the ratio of the nation's freight bill to GDP between 1960 and today. This graph shows a decline from 0.09 in 1960 to 0.06 today. Most of the decrease took place between 1960 and 1990, and little change in the share has occurred in the 1990s. While this is certainly a larger number than the 2.3% of GDP cited earlier, it nevertheless displays a significant downward trend over time.

These aggregate movements combine together several different changes: (1) movements in the real cost of moving goods within modes of transport (2) overall changes in the ratio of goods shipped to GDP and (3) changes in modes of transport. Indeed, the overall decline in transportation costs has been quite muted because people have moved from a

7

cheaper technology (rail) to a more expensive technology (trucking). We now turn to these different components of the movement in transport costs.

Figure 3 shows our longest series on transport costs over time: costs per ton-mile on railroads between 1890 and today. Figures are in 2001 dollars and show a decline from more than 18 cents per ton-mile in 1890 to 2.3 cents today. The decline has been essentially continuous, except for an increase during the 1920s and 1930s. Of course, this graph itself does not control for the average length of the haul and other factors, which will tend to influence the cost of rail shipping. Nonetheless, the data does suggest a remarkable reduction in the real cost of shipping goods over the twentieth century.

Figure 4 shows the trends in costs for other industries. We have included data since 1947 for trucks and pipeline (water is the missing major mode). These figures illustrate nicely the huge gap in shipping costs between trucks and the other modes of transportation. It also illustrates that trucking costs remained essentially constant over much of the time period. Rising fuel prices and a regulated industry kept trucking prices at essentially their 1947 levels through 1985. Since 1985, deregulation has enabled technological change and trucking costs have fallen from 38 cents a ton-mile (in 2001 dollars) to 28 cents a ton-mile in 1999. Since the Motor Carrier Act of 1980, which effectively decontrolled the industry, trucking costs have been falling by 2% per year, which is similar in magnitude to the 2.5% per year decline that rail experienced over the entire time period.

Although the low costs of pipe transport make the graph difficult to understand, between 1978 and 1999, the real costs of pipeline transport fell 25% from 2 cents per ton-mile to 1.5 cents per ton-mile. Both before 1975 and after 1978, real pipeline costs fell by about 2% per year. Only during the mid-1970s, when pipeline costs shot up by one-third, did this trend reverse. Overall, across all modes there have been declining costs, and in the absence of outside factors (the oil crisis, government regulation) costs per ton-mile, within each mode, appear to be declining by about 2% per year.

Figure 5 combines all of the modes and shows a steady downward trend, with the exception of the remarkable year of 1978. Between 1960 and 1992, costs per ton-mile fell from 16 cents to 11 cents, or an average of 0.15 cents per year, or 1.1% per year. This average is declining by somewhat less than the within-mode numbers – in part because of the increasing importance of trucking in the overall share of transportation.

The rise of trucking has been a major factor in the postwar transportation industry. As late as 1947, more than 50% of total transportation spending was on rail. Today trucking represents 77.4% of the nation's freight bill (Bureau of Transportation Statistics 1994). Of course, as Fig. 6 shows, rail is still the dominant technology measured in terms of ton-miles and ton-miles by rail are still rising. However, since trucking is more than ten times more expensive on average than rail, it accounts for the lion's share of overall spending on transportation.

These numbers tell us the costs of moving a ton of goods one mile (on average), but to understand how big a cost this actually represents, we need to connect this with average length of hauls and with the value of goods transported. Using the 1997 Commodity Flow Survey (Table 1-52, National Transportation Statistics 2002), we have been able to calculate for selected industries the relationship between average transport costs and average value. The Commodity Flow Survey tells us both the average length of haul, by industry group, and the average value per ton in this industry grouping. In Table 1, we then multiply that average haul by 2.4 cents (for rail transport) and 26 cents (for truck transport) to give two different estimates of the costs of transporting the goods.

The first column of Table 1 describes the industry; fuller descriptions are available in the commodity flow survey. The second column gives the total value of shipments of these industries in 1997. The third column shows the total ton-miles travelled by this industry and the fourth column gives the value per ton. This is calculated by dividing total value by total tons. Column five shows the average length of haul. In Columns six and seven, we multiplied column five by 2.4 cents and 26 cents, respectively, and then divided by

the average value per ton. This calculation is meant to give us the transport cost, relative to value, if the good is shipped by rail and truck, respectively.

Naturally, the length of haul is itself endogenous. Commodities with lots of bulk tend not to be shipped far. Indeed, the relationship between value per ton and average length of haul is comfortingly tight (shown in Fig. 7). The regression line is:

$$\text{Log (Miles per Average Haul)} = \underset{(0.318)}{3.22} + \underset{(0.045)}{0.32} \times \text{Log (Average Dollars per Ton)} \quad (1)$$

where $R^2 = 0.56$, the standard errors in parentheses, and the number of observations is 42. Dollars per ton is the inverse of tons per dollar or the average weight of a fixed value of goods. If the costs of shipment are roughly proportional to weight, then this suggests that as transport costs rise by 10%, the average length of distance between supplier and consumer falls by –3.2%.

Despite the endogeneity, these numbers can inform us about the importance of transport costs across a number of industries. Transport costs for some industries still appear to be quite important. For example, if wood products were shipped their average haul of 287 miles by truck, this would cost approximate one-fifth of the value of the shipment. If base metal was shipped its average haul of 276 miles by truck, transport would eat up 8.4% of the value of the commodity. Other commodities, such as basic chemicals or plastics and rubber, also feature significant transport costs, at least if shipped by truck.

However, many bigger industries all face trivial transportation costs. For machinery, electrical equipment and transportation equipment costs are always less than 1.2% of total product if shipped by truck and one-tenth of 1% of total product if shipped by rail. These three industries together account for one-quarter of the value of all shipments within the US, and 36% of all shipments (measured by value) fall in this very low cost category. Indeed, these calculations suggest that only 18% of all shipments occur in industries where transport costs are more than 6% of total value – even if all transport was by truck. If we assume that all industries with an average haul above 400 miles uses

rail, then 80% of all shipments (again by value) occur in industries where transport costs are less than 4% of total value.

The previous discussion clarifies that effective transportation costs depend as much on value per ton as on the real costs of transport. In general, just as there is a secular trend towards lower real transport costs, there is a secular trend in the value of dollars per shipment. While Fig. 5 showed that shipments were rising, this figure was not normed relative to GDP. In Fig. 7, we show total ton-miles in the economy divided by GDP. As the figure shows, a downward trend reflects increasing value of commodities shipped (and an increasing reliance on the service sector). Not only is transport becoming cheaper, it is also becoming less important relative to GDP. Another aspect of the cost of moving goods is the time cost; and other authors have documented the rapid decline of time costs over a 300-year period, 1658-1966, in moving goods between cities. (For an example of the time costs in shipping between Edinburgh and London, see Janelle 1968). These two effects reinforce each other so that the full-cost of transporting goods – money costs plus time costs – is also rapidly declining over time.

Transport costs of course still remain expensive in one area – the movement of human beings. While the car has simplified the movement of people within cities, and the airplane has greatly facilitated movement between cities, these technologies still represent large amounts of our national resources. Indeed, according to the 2001 Consumer Expenditure Survey, 18% of total expenditures for the average household is spent on vehicle purchases, gasoline and other vehicular expenses (e.g., insurance). This cash cost fails to include the far more important time costs of moving people, and these time costs are not withering away with technological progress. Instead, as wages continue to rise, these time costs should rise roughly in proportion to wages.

Table 2 reports trends in the movement of people within cities[6] over the three-decade period 1980-2000. Data are shown for the 10 largest urbanised areas in the United States

---

[6] In this table we use the concept of "urbanised area" as defined by the Census where the geographic extent of the city is defined by minimum population density.

and for size-class averages for 75 small, medium, large and very large cities. Various measures of congestion are taken from the Texas Transportation Institutes's (TTI) *2002 Urban Mobility Report* (Schrank and Lomax 2002) and include the travel time index, the percentage of daily travel in congestion, and the average annual hours of delay. The travel time index shows the amount of additional time (in percentage terms) that would be required to make a trip because of congested conditions of the roadways. For example, the largest travel time index is 1.90 for Los Angeles in 2000, which implies that it would take a traveller about 90% longer to make a given trip during peak periods than if the person could move at freeflow speeds. Also reported are average travel times for work trips.

An overwhelming pattern emerges from examining the table: over the past three decades, congestion and delay have been increasing in all size classes of cities, not just the very large urban areas. For all 75 urbanised areas taken together, on average the annual delay increased by over 280%, with the most dramatic increase occurring for large cities – those between 1 and 3 million – about 450% percent. For small cities, ranging from 100,000–500,000, the delay has increased by over 300%. Real incomes per capita have, at the same time, been rising, about 25% for the combined 75 metropolitan areas. Both trends reinforce our contention that people-moving costs are not declining within US cities. The pattern with respect to commute times is similar, but not as dramatic. Between 1980 and 2000, commute times rose by about 13% for the combined 75 areas. On average commute times are greater in larger cities for each of the decades, but the time pattern differs. Average commute times rose in all 10 of the very large urbanised areas and in all size classes between 1990 and 2000, but a few (New York, Chicago, and Houston) experienced shorter average commutes between 1980 and 1990.

We end this section by concluding that transport costs for goods are not negligible, but they are low and getting lower. Across all modes, transport costs are declining and the economy is moving away from producing high bulk products and toward more expensive products where transport costs are less relevant. As such, transportation costs, at least for goods, should play an increasingly irrelevant role in the urban economy. Conversely, as

the costs of moving people are certainly not disappearing, these should continue to be a dominating presence in the structure of urban form. An open question remains whether improvements in the internet and other forms of information technology will reduce the demand for face-to-face contact altogether.[7]

3   Implications for cities and regions of the decline of transport costs for goods

We documented in the previous section that transport costs in 1900 were high and have been falling throughout the past century. Other facts also support the relevance of the new economic geography as a tool for understanding nineteenth century American geography. Natural resources were a large share of the economy. In 1900, 40% of the labour force worked on farms (Historical Statistics of the United States, D1-10). Another 600,000 worked in mines. In the non-agricultural sector, manufacturing dominated services. There were 5.4 million manufacturing workers and only 1.7 million workers in services in 1900. Manufacturing, unlike services, displayed then (as now) the large fixed costs and returns to scale that are a critical element in the Krugman model.

Furthermore, there are two other important aspects of transportation technologies not highlighted in the new economic geography, but which are also important for the structure of cities. Water-borne transport remained important. As late as 1924, water borne domestic tonnage was about one-fifth that of rail tonnage. All relevant forms of transportation technologies in 1900 also displayed significant scale economies that created centralisation. Having dozens of railroad stations spread throughout a city did not pay, and major stations were centred in a particularly large area. Rail is optimally organised in a hub and spokes network that naturally led to large rail centres (such as Chicago). The transportation technologies without these fixed costs in 1900 were walking and, to a much less significant degree, riding. Individuals primarily had to reach their homes on foot.

---

[7] Gaspar and Glaeser (1998) argue that the implications of information technology for the city depend on whether face-to-face contact is a substitute or complement to electronic interactions (see also Kolko, 2000a).

13

Together these facts help us better understand American economic geography at the end of the last century. People are situated across the hinterland where the proximity to natural resources and saving on transport costs by crowding manufacturing activities together, and are close to ports or rail depots. Moreover, the rail depots and ports create the natural city centres, which then produce the monocentric cities modelled by Alonso, Muth and Mills. Fixed infrastructure, which facilitates the shipping of manufactured goods and the commuting of workers, created a natural centre for the city. Workers then spread out around that centre in patterns suggested by the classics of urban economic theory.

Much of this has changed over the twentieth century. We have already discussed the declining importance of transport costs and the switch from fixed cost intensive means of shipping (rail) to more flexible shipping (truck). Likewise for individual consumers there has been a decreased importance of public transportation, which also created natural agglomerations – and a rise in the car, which had no need for hub or spokes. Natural resources have become far less important; in the 2000 census 1.9% of workers laboured in agriculture, fishing, forestry, and mining. Manufacturing has declined as well and now represents only 14.1% of the economy. More than 50% of the economy now works in various service industries.

These facts are well known, but their implications for economic geography are surely understudied. To these facts we would add several other points. While natural resources are an increasingly irrelevant part of production, they are an important part of consumption. Space differs significantly in its natural endowment of warm weather and rainfall. These factors are the exogenous forces that make the US something other than a featureless plane. Differences in government policy will also differentiate space, although they are certainly less exogenous. Holmes (1998) uses discontinuities at borders to show how state policy impacts local employment. In the housing market as well, differences across areas in the degree of regulation are extremely important. Since the 1970s, some states, such as California and Oregon, have sharply limited new housing production;

Texas has not. In a service economy where transport costs are small and natural productive resources nearly irrelevant, weather and government stand as the features, which should increasingly determine the location of people.

At the top of these layers of "innate" factors are production and transportation technologies. Agglomeration economies appear to be as important as always. People in dense areas are more productive and earn more (Ciccone and Hall 1996). In a competitive labour context income should reflect the marginal product of labour. Figure 9 shows the relationship between and density across counties and suggests that a 10% increase in county density corresponds to a 206 dollar increase in median household income. Some of this relationship is due to higher skilled people living in dense areas, but if we control for schooling we find:

$$\text{Log (Income)} = \underset{(0.02)}{8.76} + \underset{(0.002)}{0.06} \times \text{Log(Density)} + \underset{(0.0005)}{0.016} \times \text{Share w. BA}$$

$$+ \underset{(0.0004)}{0.015} \times \text{Share w. HS} \tag{2}$$

where $R^2 = 0.64$, standard errors are in parentheses, and there are 3,109 counties in the regression. Income refers to median household income in 1989, density is people per square mile, share w. BA refers to the percentage of the population over age 25 with college degrees, share w. HS, refers to the percentage of the population over age 25 with high school degrees only.[8] There are many reasons to be sceptical about evidence of this kind. Controlling for these obvious schooling measures does not guarantee that we have controlled adequately for all forms of omitted ability (Glaeser and Mare 2001, for an extensive discussion of these issues in the urban context).[9] Nonetheless, it is impossible not to start at least with the view that there remain massive economies from agglomeration.

---

[8] The connection between average income and density is also strong but the estimated elasticity is smaller.

[9] The interpretation of the schooling coefficient is difficult also because it will include any direct effect of schooling and human capital spillovers (as in Rauch 1993).

Where do these agglomeration economies come from? If there really were no transport costs, it is certainly true that agglomeration economies could not exist. There would be little possible reason for there to exist any differences across space, because without transport costs, anyone could costlessly access any other person or firm on the planet. As such, transport costs are still important, but the relevant transport costs are likely to be for moving people, not goods. The advantages from proximity to other people appear to come from saving the costs of providing and acquiring services and from improving the flow of knowledge. There has been much useful work on this in the past, and hopefully further work will help us better understand why these agglomeration effects are so critical.[10] Regardless of the cause, it seems certain that proximity will remain important and likely continue to be true well into the future.

However, while proximity matters, the form of proximity certainly has changed. The dominant nineteenth century mode of individual transportation – walking – has allowed people to travel one to two miles in a 30-minute walk. The automobile allows people to move between 15 and 30 miles during the same time. While the reduction of transport costs for goods has freed people from living close to natural resources and has facilitated concentration within a region, the reduction of transport costs for people has meant that individuals can live at much lower densities and still enjoy the advantages of proximity.

The second significant transport technology change is the reduced importance of public transportation, ports, rail hubs, and other massive infrastructure centres.[11] These centres provided cities with natural cores, and their absence means that cities can sprawl without limit. As such, we have moved from monocentric cities to polycentric regions, which are polycentric or rather uncentred entirely.[12] This change is a major reason why increasingly relevant economic units are regions, rather than urban cores.

---

[10] Many citations to the vast literature on agglomeration economies can be found in Fujita and Thisse (2002), several chapters Henderson and Thisse (forthcoming), and in the review article by Anas, Arnott, and Small (1998).

[11] Some ports, particularly ocean-oriented ports, may continue to flourish. See chapters 8 and 13 of Fujita, Krugman and Venables (1999) or Fujita and Mori (1996).

[12] Fujita and Ogawa (1982) develop the first theoretical model with simultaneous location of firms and households in a linear city. Lucas and Rossi-Hansberg (2002) generalize the Fujita-Ogawa paper to a

Together these changes suggest a natural structure for urban models of the twenty-first century. The only innate advantages that matter are consumer amenities, and perhaps government policies (which presumably should be treated as endogenous). Individual productivity is a function of the number of people within reasonable driving distances. There are no natural centres or cores of urban areas. While it remains for theorists to develop these models more thoroughly, the implications of any model fitting the above description seems clear: we should expect to see much less population in the natural resource related sectors of the country and much more in pleasant inhabitable areas. We should see continuing agglomerations, but these agglomerations should be built around the automobile and should have few natural centres. There are no natural limits to the sprawl of these agglomerations and there is no reason why the population cannot be centred in a few big areas.[13]

Indeed, the only thing to limit growth of the areas in the long run is continuing demand by some consumers for various forms of natural amenities or government policies. Without these factors we could possibly all end up in southern California. However, because some people like winters and because California regulations limit both building and certain forms of economic activity, there are likely to be a number of large, sprawling urban areas that dominate the American landscape in future. Moreover, because housing and other infrastructure are durable, these changes are happening only slowly, and we still see the remnants of cities built around different transport technologies.

4   Testing the implications of declining transport costs for goods

---

symmetric circular city. Both papers develop various equilibriums including monocentric, polycentric or completely mixed cities.

[13] We will define sprawl simply as decentralised population and employment. Empirically, high levels of sprawl would be captured by a large share of a metropolitan area's population and employment that is more than five miles from the central business district.

We now turn to our evidence on the current state of America's regions and on the twentieth century transformation of economic geography. We review some natural implications of the decline in transport costs and assess the evidence that relates to them.

Implication #1: People are no longer tied to natural resources

This is perhaps the most natural implication of a decline in the importance of transport costs. If in 1900, it was advantageous to be nearby natural resources, in 2000 this is no longer relevant. Areas once populated because of their natural resources should have lost large numbers of inhabitants over the twentieth century.

We have no perfect measure of these resources; instead we use two proxies. First, we use the share of employment in a county that works in agriculture, fishing, forestry, or mining. Admittedly, we would prefer to have this measure at the beginning of our time period, instead of the end, but we think that the inter-temporal correlation between these measures is high enough to represent a reasonable measure of the importance of innate natural resources in the area. Figure 10 shows the relationship between this employment share and the logarithm of population in 2000 in the county divided by population of the county in 1920.

The estimated regression is:

$$
\text{Log}\left(\frac{\text{Population in 2000}}{\text{Population in 1920}}\right) = \underset{(0.02)}{0.95} - \underset{(0.15)}{4.52} \times \frac{\text{Natural Resource Employment}}{\text{Total Employment}} \qquad (3)
$$

where $R^2 = 0.22$, standard errors are in parentheses, and the number of observations is 3,056. The coefficient implies that as the share of employment in natural resources rises by 10%, the growth of the county between 1920 and 2000 should be expected to fall by 45.2%. This coefficient is strongly robust to other controls.

A second method of showing this change is to examine the relationship between population growth and longitude. In 1990, and we believe in 1900, the centre of the US specialised in the production of natural resource based commodities. Indeed, the peopling of America was based largely on the demand for agricultural land and the desire to exploit America's rich natural wealth. However, as transport costs fell, we should expect to see America hollow out. People should ostensibly leave the middle states, which have always had harsh environments, and move to the coasts, which are more temperate and provide easier access to Europe and Asia.

To test this implication Fig. 11 indicates the relationship between population growth and longitude. We have estimated a spline with a break at −100 degrees longitude. This number was chosen fairly arbitrarily – it is the longitude of central Nebraska. The graph shows that the population increased on both coasts and declined in the centre. The estimated regression is:

$$\text{Log}\left(\frac{\text{Population in 2000}}{\text{Population in 1920}}\right) = \underset{(0.35)}{-7.3} - \underset{(0.003)}{0.07} \times \underset{(\text{Less than -100 degrees})}{\text{Longitude}} + \underset{(0.002)}{0.03} \times \underset{(\text{More than -100 degrees})}{\text{Longitude}} \quad (4)$$

where $R^2 = 0.14$, standard errors are in parentheses, and the number of counties is 3056. Again, this estimated relationship is robust to many other factors. For example, latitude also has a significant effect on growth over this period, but including this does not materially impact the coefficients on longitude. We are witnessing the rise of the US as a coastal nation, which is emphasised by Rappaport and Sachs (2000). While both of these regressions and graphs represent rough proxies, they suggest that natural advantages are becoming increasingly irrelevant to the location of people and economic activity.

Of course not every county in the hinterland is declining in relative importance. Some communities, especially those with remarkable natural beauty or other consumer amenities, are actually gaining in population. We explore this effect in the next section.

Implication #2: Consumer-related natural advantages are becoming more important

19

Implication #2 is the natural counterpart to implication #1. If innate productive advantages are becoming increasingly irrelevant, then innate consumption advantages should become more important. This helps us again to understand the hollowing of America. Living in the hinterland has become less valuable, but people would not have moved if the coasts did not have other innate attractions. Here we show the importance of weather variables in predicting the success of different areas.

Because our weather variables are at the city, not county level, we look at the relationship between metropolitan area growth and mean January temperatures. Data availability limits our focus to the 1980 to 2000 period. Figure 12 shows the basic connection. The estimated regression is:

$$\text{Log}\left(\frac{\text{Population in 2000}}{\text{Population in 1980}}\right) = \underset{(0.02)}{-0.08} + \underset{(0.0005)}{0.0054} \times \text{Jan. Temp.} \tag{5}$$

where $R^2 = 0.30$, standard errors are in parentheses, and there are 275 observations. As January temperatures rise by 10 degrees, expected growth over this time period is expected to increase by 5.4%. Again, the result is robust to the use of alternative controls, and the results are robust to exclusion of cities in California or any other individual state.

Other weather variables, such as average precipitation, are also potent predictors of metropolitan growth over this time period. Using county level population data and the average January temperature of the largest city in the state, we also see a large effect of warm weather on growth over the entire time period. For example, a ten-degree increase in state January temperature increases county level population growth between 1920 and 1950 by 8%. This is not merely a post-war phenomenon.

This is not prediction that everyone will move to California. Of course there is no innate problem with all of America living there. California's total land area is approximately 100 million acres, which could comfortably house every American family on a one-half acre lot. Two factors tend to break the growth of that area. First, some consumers may actually prefer the environmental bundle on the east coast or in the south. Second,

California itself appears to have decided to use growth controls to limit the expansion of the housing stock in the state. Growth controls have significantly slowed the development of that state over the past twenty years.

Implication #3: Population is increasingly centralised in a few metropolitan regions

We have argued that the spread of population throughout the hinterland of the United States at the beginning of the twentieth century was motivated by a desire to be near natural resources. As these resources become less important, there is no longer any reason for an urban hierarchy spread across the country. Instead, people need only congregate in a few large metropolitan areas where they can reap the benefits of agglomerated service economies. We would expect there to be an increasing agglomeration of population in a few large areas.

Table 3 shows the pattern of agglomeration across time. We rank counties by their density levels in each decade and ask what share of population lived in the 50% of counties with the lowest density levels, what share of population lived in the 10% of counties with the highest density levels, and what share of population lived in the 1% of counties (approximately 30 counties) with the highest density levels. The first two figures inform us about the spread of lower density areas. The last figure is of more importance to the concentration within particular urban areas, and we consider this last column in the next implication.

The table shows a continuing decline in the share of US population living in the least dense counties and a continuing increase in the share of US population living in the densest 10% of counties. In 1920 19% of the population lived in the least dense half of counties. Eighty years later, that fraction has dropped to 9%. Most of this decrease occurred between 1940 and 1960 when the share of the population living in low-density counties fell from 17 to 11%.

This fall has been offset by an increase in the medium to high-density counties. The second column shows that the share of population living in the top decile of counties (ranked by density) but not in the 1% of most dense counties has risen from 30% in 1920 to 49% today. Some of this rise is also surely driven by the decline in population in the very densest counties, but there remains an impressive increase in the proportion of the population living at middle densities.

Table 3 reflects the overall rise in the great metropolitan regions which have come to dominate American geography. The top six CMSAs (New York, Los Angeles, San Francisco, Chicago, Philadelphia, and Washington) alone comprise approximately one-quarter of the US population. More generally, it is probably correct to think that eight great population centres exist in the US: the Boston-Washington corridor, the Los Angeles area (perhaps even including Las Vegas), the San Francisco Bay area, the Portland-Seattle corridor, Chicago, Atlanta, Miami, and southeastern Texas. Each of these areas is relatively compact and it seems likely that altogether these areas will continue to attract an increasing share of the US population.

We think that there are three reasons why there will continue to be a set of major metropolitan regions, rather than one.[14] Historical infrastructure is so vast that abandoning regions altogether would be extremely costly. This surely helps us understand why, despite its bad weather and increasingly inappropriate location. Detroit continues to have five million residents. Heterogeneity of preferences for weather will continue to ensure some demand for the East Coast and some for the West. Finally, differences in state policies will continue to ensure that not everyone moves to Los Angeles.

Implication #4: People are increasingly decentralised within those regions

---

[14] Empirically, the rank-size rule has been found to hold for many countries, at least for the largest cities in the city-size distribution. See Fujita, Krugman, and Venables (1999, ch.12) and Gabaix and Ionnides (forthcoming).

While metropolitan regions have been expanding and people have been leaving places in the hinterland, the story within cities is one of decentralisation rather than centralisation. As we argued earlier, this can be seen as a natural outcome of changes in transportation technology. Rail and water required centralisation around hubs. Cars and trucks do not. As a result, there is no reason to centralise around a given centre. Moreover, the automobile requires much more space than rail or the feet and can facilitate living at much lower densities.

This helps us understand column three of Table 2, which shows a decline in the proportion of the US population living in the very densest counties (approximately the top 30). In 1920, 20% of the population lived in those counties. Today, 11% of the population lives in these densest areas. To put this fact another way, in both 1920 and today there are eight counties with more than 10,000 persons per square mile. In 1920, those eight counties represented 9% of America's population. In 2000, these eight dense counties represent only 4% of the nation.

Of course, the strongest evidence on decentralisation can be seen within counties and there is a vast array of papers and books, documenting both suburbanisation and the decentralisation of employment (some of the literature includes Anas, Arnott, and Small 1998; Scott and Soja (1998); Small 1996; Song and Small 1994; Garreau 1991; McDonald 1989). McMillen and Smith (2003) explain how polycentric cities tend to form as the population grows. Craig and Kohlhase (2003) document the links between employment subcentres and Glaeser and Kahn (2003) document that, in the average metropolitan area, less than 20% of employment is in the city centre. As argued above, sprawl is ultimately the result of changing transportation technologies. The automobile enables people to enjoy living on larger lots and requires lower densities for effective operation. While declining transport costs have brought people together in a few central metropolitan areas, these lower costs have also permitted them to spread out.

Implication #5: High-density housing and public transportation are becoming increasingly irrelevant

As late as 1960 only 64% of the US population drove to work; 12.1% used public transportation and 9.9% walked. Certainly there was much more driving than at earlier times during the century, but alternative modes of transportation remained important and viable. These modes, walking and public transportation, were meant for high-density living and indeed they were used primarily in high-density locales.

As of 2000, according to the 2001 American Community Survey, 88% of Americans drove to work (either alone or in carpools). Five percent of Americans used public transportation and 2.5% walked to work. Across metropolitan areas, there are exactly four metropolitan areas in 1990, where more than 10% of the population used public transportation to get to work. There were only 13 MSAs in which more than 5% of the population used public transportation in their community. The change in transportation technology has finally become complete. Moreover, as Glaeser and Kahn (2003) document, the car has now become ubiquitous for almost all trips, not just long commutes to work. This technology has restructured urban areas and facilitated lower densities, and as a result, it is the only technology that can be used for transportation in those new areas.

Just as the past century has seen a shift from high-density modes of transportation to the single low-density mode of transportation, the revolution in transport technology has also meant practically an end to apartment living. Because housing is durable, we will have apartments with us for centuries to come, but the data on new starts suggests a very strong pattern. Indeed, during much of the late 1960s through the early 1970s, multi-unit construction was about as important as single-family construction. As late as 1972 there were more units in multi-unit buildings being constructed than single-family units.

But in 2002 there were 1.36 million single-family houses that were started. Of these 1.2 million were detached, and the median floor area (of all single-family starts) was 2,117 feet. Meanwhile, there were 347,000 apartment building units started, and 182,000 in apartment buildings with 20 or more units. These units had a median square footage of 1,093 feet. The rise of sprawl has facilitated a switch in the housing industry and a

relative decline in apartment buildings. This switch has, of course, been accompanied by a significant rise in the size of homes occupied by Americans, so that today the median house contains 1,692 square feet.

Implication #6: Services are in dense areas; manufacturing is not

A natural implication of the claim that transportation costs for people are still high, but transportation costs for goods are not, is that services will be located in dense areas, but manufacturing will be located in places of medium or low density. Since manufacturing still requires workers, it seems unlikely that it will be located in the lowest density areas, The most likely locations are where land is relatively cheap and firms do not have to pay for proximity to consumers. Conversely, services will locate in the densest counties, especially those with the most value added.

Figure 13 shows the relationship between the share of adult employment in finance, insurance and real estate, and the logarithm of population over land area at the county level. Both variables are at county level. The relationship shown in the graph is:

$$\frac{\text{Employment in FIRE in 1990}}{\text{Total Employment}} = \underset{(0.0007)}{0.023} + \underset{(0.00016)}{0.0057} \times \text{Log}\left(\frac{\text{Population in 1990}}{\text{County Land Area}}\right) \quad (6)$$

where $R^2 = 0.27$, standard errors are in parentheses and there are 3,109 observations. The coefficient means that as density doubles, the share working in this industry increases by 57%. This is a small sector of the economy, but it is particularly likely to be located in high-density areas.

The relationship for the larger service sector is:

$$\frac{\text{Employment in Services in 1990}}{\text{Total Employment}} = \underset{(0.002)}{0.19} + \underset{(0.0005)}{0.0058} \times \text{Log}\left(\frac{\text{Population in 1990}}{\text{County Land Area}}\right) \quad (7)$$

where $R^2 = 0.04$, standard errors are in parentheses, and there are 3,109 observations. Services are spread much more evenly than finance, insurance and real estate, but there is still a strongly significant tendency for services to be disproportionate in high-density areas. The magnitude of this effect is that services represent 20% of employment in the lowest density counties and rise to include 27% of employment in the densest areas.

As shown in Fig. 14, the relationship between manufacturing and density is non-monotonic and appears to be highest in middle-density regions. As discussed earlier, only 10% of the population lives in those counties with the lowest density levels, and manufacturing does not locate there either. Indeed, these low density places are heavily based in the agricultural, fishing, forestry and mining sector of the economy. On average, 16% of the employment in counties with density levels below the median are in this sector. By contrast in the counties with density levels in the top ten-tenth of U.S. counties, only 1.6 % of employment is in this sector. Once we exclude these unpopulated areas, the relationship between manufacturing and density is strongly negative. Across the densest one-half of counties, we estimate:

$$\frac{\text{Employment in Manufacturing}}{\text{Total Employment}} = \underset{(0.01)}{0.31} - \underset{(0.002)}{0.02} \times \text{Log}\left(\frac{\text{Population in 1990}}{\text{County Land Area}}\right) \qquad (8)$$

where $R^2 = 0.06$, standard errors are in parentheses, and the number of observations is 1,554. The relationship is not overwhelming, but it is generally true that manufacturing is not located in the highest density tracts, just as we would expect if manufactured goods are inexpensive to ship.

Implication #7: The location of manufacturing firms is not driven by proximity to customers or suppliers, the location of service firms is determined by proximity.

Dumais, Ellison and Glaeser (1997) examine the determinants of co-location of manufacturing industries. The article identifies which industries locate near one another

using the Longitudinal Research Database of the Census Bureau. In particular, the authors use input-output tables to determine which industries buy from and sell to each other. Then using these relationships, they ask whether industries that interact in this way also have a tendency to co-locate. They find that a one-standard deviation increase in the degree of inter-industry purchases increases the amount of co-location by one-twentieth of a standard deviation. The coefficient is quite significant but small in magnitude.

This finding can be compared to the tendency of industries that use the same labour to locate near one another. The article finds that a one-standard deviation increase in the degree to which industries hire the same type of workers is correlated with a 0.4 standard deviation increase in the degree of co-location. Thus, using the same workers is approximately eight times more likely to increase the degree of co-location than are trade relationships. To us, this suggests the importance of the costs of moving people, but that moving goods is sufficiently cheap so that it does not largely affect the location decision.

Kolko (2000b) repeated this experiment with service industries. As in the case of manufacturing industries, he finds that service industries using the same type of labour locate near one another. However, he also finds that service industries buying from or selling to each are also likely to co-locate, and the magnitude of these trade variables is essentially comparable to the magnitude of the labour-related variables. Because services involve face-to-face contact, we interpret these results as a confirmation of the continuing importance of transport costs for people.

Implication #8: Density and education go together

If cities are about facilitating the contact of individuals, not moving goods, then we should expect people with more human capital to live in highly dense areas. This implication occurs because these higher human capital people have higher wages and thus a higher cost of transporting themselves. Thus we should expect them to live in closer proximity. There is also the likelihood that the gains from interaction will be higher among the more skilled who also tend to concentrate in more social occupations.

To test this implication, we analysed the relationship between the share of a county's population with college degrees and the density of the county. Figure 15 shows the strong positive relationship. This estimated regression line is:

$$\frac{\text{Population with B.A. Degrees}}{\text{Total Population}} = \underset{(0.0026)}{0.079} + \underset{(0.00066)}{0.015} \times \text{Log}\left(\frac{\text{Population in 1990}}{\text{County Land Area}}\right) \qquad (9)$$

where $R^2 = 0.14$, standard errors are in parentheses, and there are 3,109 observations. This is a strong and robust result. People with more human capital live in denser counties. Although there are certainly other explanations for this phenomenon beyond those sketched above, this certainly stands as a significant feature of density in today's urban world. Future models and empirical work will help us better understand this phenomenon.

5   Testing the implications of the increase in time costs for moving people

While most of our implications centre around the consequences of falling transport costs for goods, we would like to end with a conjecture about the potential impacts on productivity as people-moving costs increase. As one of the negative aspects of high density, congestion may work to counteract the benefits of proximity. For small values of congestion, productivity effects are unlikely to be found, but as congestion and delays increase, there may eventually be an effect.

Implication # 9 Productivity will decline as congestion exceeds some threshold level

We conjecture that after some point, congestion increases are likely to be associated with a measured decline in worker productivity. How to measure productivity and congestion are topics requiring research, but we suggest a first look at the data for the year 2000 by

using median earnings as the measure of productivity and measuring congestion by the variable "annual delay per person." We use the cities in the TTI mobility study (Shrank and Lomax 2002).

The estimated regression is:

$$\begin{aligned}
\text{Log(median earnings)} = \quad & 8.7565 + 0.0506(\text{log population density}) + 0.0124(\%\text{BA+}) \\
& (0.3491)\ (0.0383) \qquad\qquad\qquad\qquad\quad (0.0019) \\[6pt]
& + 0.0178(\%\text{HS}) + 0.0124(\text{delay}) - 0.00013(\text{delay})^2 \qquad (10) \\
& \ \ (0.0035) \qquad\quad (0.0030) \qquad\quad (-0.00006)
\end{aligned}$$

where $R^2 = 0.63$, standard errors are in parentheses, and the number of urbanised areas is 71. The variable "%BA+" is the percent of the population over age 25 whose highest degree attained is a bachelors or higher; "%HS" is the percent of the population over age 25 with a high school diploma as the highest level of education attained. Congestion effects are measured by linear and quadratic terms of "delay," which is the annual hours of delay per person as calculated by the TTI (Schrank and Lomax 2002). The education and congestion variables are strongly significant while population density is insignificant (*p*-value = 0.19). Based on the coefficients on the delay-quadratic form, we calculate that negative impacts of congestion may become apparent after about 47 hours of delay per person hours [= 0.0124 / (2×0.00013)]. In our dataset the variable *delay* ranges from 2–62 hours, but only one city, Los Angeles, averages more than 47 hours of delay per person each year. Our results are preliminary, but suggest that if the trend of increasing congestion continues, more and more cities may become subject to productivity declines.

6 Conclusion

In this article we have documented that the twentieth century witnessed a remarkable decline in transportation costs, at least for the shipment of goods. In the last three decades, there appears to have been a rising cost of moving people within cities mainly due to increases in road delay. Cities are best regarded as the absence of physical space between people and firms. Since cities are ultimately only proximity, the demand for cities must come from the desire to eliminate transportation costs for goods, people and ideas. A revolution in transportation technologies unfolded during the twentieth century; it would have been surprising had a revolution in urban form and location not occurred as well.

There are two major ways in which cities are changing. First, the reduction in transportation costs for goods has allowed cities to change their locations and their primary functions. Today, cities serve mainly to facilitate contact between people. As such, there is little reason for cities to be near natural resources or natural transport hubs. Instead, cities should locate where it is pleasant to live or where governments are friendly. We think that the movement away from the hinterland should best be understood as a flight from natural resources towards consumer preferences. The location of manufacturing firms in medium density, not high-density counties, is also a natural implication of low transport costs for goods.

The second way in which cities are changing is internal to the city. The automobile has allowed cities to sprawl and eliminated any tendency towards a single city centre. These are both features of the twenty-first century. Both of the classics of urban economics, the Alonso-Muth-Mills model and the Krugman-Fujita-Thisse new economic geography, need updating. Both frameworks are analytically beautiful and remarkably apt characterisations of the city of the past. But a new regional model, without centres and without transport costs for goods, will better capture the future of the city.

Our view is that such a model would have the following basic elements. First, productivity would be a function of agglomeration because there are gains from people being able to interact. Second, the key transport mode – the automobile – travels much faster on highways than on city streets and is subject to congestion effects. Third, physical output is generally relatively costless to ship. Fourth, even though output is almost costless to ship, most people produce services that require face–to–face interaction. Fifth, land is heterogeneous and some places are nicer than others. We hope that the regional models of the future are built around these elements.

References

Alonso W (1960) A theory of the urban land market. *Papers and Proceedings Regional Science Association* 6: 149 – 157

Alonso W (1964) *Location and land use*. Cambridge, MA: Harvard University Press

Anas A, Arnott R, Small K (1998) Urban spatial structure. *Journal of Economic Literature* 36: 1426 – 1464

Beckmann M (1968) *Location theory*. New York: Random House

Blumenfeld H (1955) The economic base of the metropolis. *Journal of the American Institute of Planners* 21: 114 – 132

Brueckner J (1987) The structure of urban equilibria: A unified treatment of the Muth-Mills model. In: *Handbook of regional and urban economics: Urban economics Vol. 2*. Mills, ES (ed) Amsterdam: North Holland Press

Ciccone A, Hall RE (1996) Productivity and the density of economic activity. *American Economic Review* 86 (1): 54 – 70

Christaller W (1933) Die zentralen Orte in Suddeutschland, Jena: Gustav Fischer. English translation: (1966) Baskin CW, Central Places in Southern Germany, Engelwood Cliffs, NJ

Craig SG, Kohlhase JE (2003) Employment subcenters and the distribution of economic activity. Working paper, Department of Economics, University of Houston

Cronon W (1991) *Nature's metropolis: Chicago and the great west*. New York: WW Norton and Company

Dumais G, Ellison G, Glaeser EL (1997) The determinants of geographic concentration. *Quarterly Journal of Economics* (in revision)

Fujita M (1989) *Urban economic theory: Land use and city size*. Cambridge: Cambridge University Press

Fujita M, Krugman P, Venables AJ (1999) *The spatial economy: Cities, regions and international trade*. Cambridge, MIT Press

Fujita M, Mori T (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, Ogawa H (1982) Mutiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics* 12: 161 – 196.

Fujita M, Thisse J-F (2002) *Economics of agglomeration: Cities, industrial location and regional growth*. Cambridge: Cambridge University Press

Gabaix X, Ionnides YM (2003) The evolution of city size distrbutions. In: *The handbook of urban and regional economics Vol. 4*. Henderson JV, Thisse J-F (eds) Oxford University Press (forthcoming)

Gaspar J, Glaeser EL (1998) Information technology and the future of cities. *Journal of Urban Economics* 43: 136 – 156

Garreau J (1991) *Edge city: Life on the new frontier*. New York: Doubleday

Glaeser EL, Kahn ME (2003) Sprawl and urban growth. In: *The handbook of urban and regional economics Vol. 4*. Henderson JV, Thisse J-F (eds) Oxford University Press (forthcoming)

Glaeser EL, Kolko J, Saiz A (2001) Consumer city. *Journal of Economic Geography* 1: 27 – 50

Glaeser EL, Mare D (2001) Cities and skills. *Journal of Labor Economics* 19 (2): 316 – 342

Greenhut ML (1970) *A theory of the firm in economic space*. New York: Appleton Century. Reprinted in 1992 by Gregg Revivals: Hampshire, England

Harris C, Ullman E (1945) The nature of cities. *Annals of the American Academy of Political and Social Science* 242: 7 – 17

Henderson JV (1974) Size and nature of cities. *American Economic Review* 64: 640 – 656

Henderson JV, Thisse J-F (eds) (2003) *Handbook of regional and urban economics Vol. 4: Urban economics*. Amsterdam: North Holland Press (forthcoming)

Holmes TJ (1998) The effects of state policies on the location of industry: Evidence from state borders. *Journal of Political Economy* 106 (4): 667 – 705

Hoover EM, Giarratani F (1985) *An introduction to regional economics*, 3rd ed. New York: Alfred A. Knopf. Full text available on-line at the WebBook of Regional Science site, http://www.rri.wvu.edu/regscweb.htm

Isard W (1956) *Location and space economy: A general theory relating to industrial location, market areas, land use, trade, and urban structure*. Cambridge, MA

Janelle DG (1968) Central place development in a time-space framework. *The Professional Geographer* 20 (1): 5 – 10

Kolko J (2000a) The death of cities? The death of distance? Evidence from the geography of commercial internet usage, Ch. 4. Vogelsang I, Compaine BM (eds) *The internet upheaval*. Cambridge, MA: MIT Press

Kolko J (2000b) Three essays on urban economics. Harvard University, Ph.D. dissertation

Krugman P (1991a) Increasing returns and economic geography. *Journal of Political Economy* 99: 483 – 499

Krugman P (1991b) *Geography and trade*. Cambridge, MA: MIT Press

Losch A (1944) *Die raumliche Ordung der Wirtschaft*, 2nd ed. Jena: Gustav Fischer English translation 1954 Woglom WH: *The economics of location*. New Haven: Yale University Press

Lucas RE Jr., Rossi-Hansberg E (2002) On the internal structure of cities. *Econometrica* 70 (4): 1445 - 1476

McDonald JF (1989) Econometric studies of urban population density: A survey. *Journal of Urban Economics* 26: 361 - 385

McMillen D, Smith SC (2003) The number of subcenters in large urban areas. *Journal of Urban Economics* 53 (3): 321 – 338

Meyer J, Kain J, Wohl M (1965) *The urban transportation problem*. Cambridge: Harvard University Press

Mills ES (1967) An aggregative model of resource allocation in a metropolitan area. *American Economic Review* 57: 197 – 210

Mills ES (1972) *Studies in the structure of the urban economy*. Baltimore: The Johns Hopkins Press

Mumford L (1963) *Technics and civilization*. New York: Harcourt, Brace and World

Muth R (1968) *Cities and housing*. Chicago: University of Chicago Press

Nourse HO (1968) *Regional economics*. New York: McGraw-Hill

Rappaport J, Sachs JD (2003) The United States as a coastal nation. *Journal of Economic Growth* 8: 5 – 46

Rauch JE (1993) Productivity gains from geographic concentration of human capital: Evidence from the cities. *Journal of Urban Economics* 34 (3): 380 – 400

Samuelson PA (1954) The transfer problem and transport costs, II: Analysis of effects of trade impediments. *Economic Journal* 64: 264-289

Schrank D, Lomax T (2002) *The 2002 urban mobility report*. Texas Transportation Institute. Report and data available online at http://mobility.tamu.edu/

Scott AJ, Soja EW (1998) *The city: Los Angeles and urban theory at the end of the twentieth century*. University of California Press

Small K (1992) *Urban transportation economics: Fundamentals of pure and applied economics, Vol. 51*. Harwood Academic Publishers

Small K (1996) Urban transportation economics (revised). In: *Urban and regional economics*. Arnott R (ed) Vol. 1 of *Encyclopedia of economics*, pp. 251 – 439. London: Harwood Academic Publishers

Small K, Song S (1994) Population and employment densities: Structure and change. *Journal of Urban Economics* 36: 292 - 313

US Department of Transportation, Bureau of Transportation Statistics (2002) *National transportation statistics 2002*, BTS02-08. Available online at http://www.bts.gov/publications/national_transportation_statistics/2002/index.html

Von Thünen, JH (1826) *Der isolierte Staat in Beziehung auf Nationalalökonomie und Landwirtschaft*. Reprinted 1966 by Stuttgart: Gustav Fischer. English translation by Wartenberg CM: *von Thünen's isolated state*. Oxford: Pergamon Press

Weber A (1909) *Über den Standort der Industrie*. Tubingen: Mohr JCB. English translation by Friedrich C: *The theory of the location of industries*. Chicago: Chicago University Press, 1929, reprinted 1957

Zipf GK (1949) *Human behavior and the principle of least effort: An introduction to human ecology*. Reading, MA: Addison-Wesley

Table 1. Transportation costs and commodity value, selected industries

Commodity Description	Value ($ billion)	Ton-miles (billion)	Value per ton ($)	Average miles per shipment	Shipping costs/value (Rail)	Shipping costs/value (Truck)
Meat, fish, seafood, and their preparations	183.8	36.4	2,312	137	0.001	0.015
Milled grain products, preparations, and bakery products	109.9	48.5	1,069	122	0.003	0.029
Alcoholic beverages	87.9	27.8	1,085	58	0.001	0.013
Tobacco products	56.4	1.0	13,661	296	0.0005	0.006
Gasoline and aviation turbine fuel	217.1	136.6	225	45	0.005	0.052
Basic chemicals	159.6	136.8	539	332	0.014	0.160
Pharmaceutical products	224.4	5.6	22,678	692	0.0007	0.008
Chemical products and preparations (NEC)	209.5	45.0	2,276	333	0.004	0.038
Plastics and rubber	278.8	69.1	2,138	451	0.005	0.054
Wood products	126.4	96.9	384	287	0.018	0.194
Printed products	260.3	22.8	3,335	431	0.003	0.033
Textiles, leather, and articles of textiles or leather	379.2	24.7	8,266	912	0.003	0.028
Base metal in primary or semi finished forms and in finished basic shapes	285.7	117.5	851	276	0.008	0.084
Articles of base metal	227.2	48.7	2,133	403	0.005	0.049
Machinery	417.1	27.0	8,356	356	0.001	0.010
Electronic and electrical equipment, components and office equipment	869.7	27.1	21,955	640	0.0007	0.008
Motorised and other vehicles (including parts)	571.0	45.9	5,822	278	0.001	0.012

Source: National Transportation Statistics 2002 and authors' calculations assuming that the cost per ton-mile is 26 cents by truck and 2.4 cents by rail.

Table 2. Trends in commuting and congestion for US metropolitan areas[a]

Metropolitan Area[a]	Year[b]	Population (× 1,000)	Travel time index	Daily travel in congestion (%)	Delay per person (hrs/yr)	Mean travel time to work (min)	Real income per capita ($1982–1984)
Very Large (> 3 million in 2000)							
New York NY-Northeastern NJ	1980	15,500	1.13	14	6	32.3	11,088
	1990	15,925	1.31	27	18	31.3	14,887
	2000	17,090	1.41	35	23	35.1	15,616
Los Angeles CA	1980	9,900	1.34	31	22	24.1	11,738
	1990	11,420	1.91	42	63	26.2	13,587
	2000	12,680	1.90	45	62	28.7	12,615
Chicago IL-Northwestern IN	1980	7,080	1.19	23	7	31.6	11,716
	1990	7,510	1.37	35	18	28.5	13,298
	2000	8,090	1.47	40	27	31.2	14,812
Philadelphia PA-NJ	1980	4,090	1.11	16	5	26.0	10,168
	1990	4,370	1.18	23	9	24.9	12,959
	2000	4,590	1.28	30	15	28.2	14,286
San Francisco-Oakland CA	1980	3,290	1.21	27	12	25.2	13,219
	1990	3,675	1.50	41	37	26.5	16,293
	2000	4,030	1.59	41	41	29.9	18,151
Detroit MI	1980	3,810	1.12	17	7	22.9	11,767
	1990	4,000	1.28	30	20	23.1	12,723
	2000	4,025	1.34	35	25	26.0	14,702
Dallas-Fort Worth TX	1980	2,450	1.07	9	6	22.4	11,660
	1990	3,150	1.18	19	18	23.5	13,135
	2000	3,800	1.33	29	37	26.8	14,262
Washington DC-MD-VA	1980	2,700	1.18	25	10	28.1	14,355
	1990	3,100	1.34	37	22	29.0	17,578
	2000	3,560	1.46	40	35	32.2	18,948
Houston TX	1980	2,400	1.28	26	19	26.2	12,540
	1990	2,880	1.31	28	18	25.9	12,404
	2000	3,375	1.38	33	36	28.4	13,115
Boston MA	1980	2,850	1.14	16	9	23.1	11,113
	1990	2,955	1.27	30	18	24.1	15,456
	2000	3,025	1.45	38	28	28.3	16,760

Table 2. Continued

Metropolitan Area[a]	Year[b]	Population (× 1,000)	Travel time index	Daily travel in congestion (%)	Delay per person (hrs/yr)	Mean travel time to work (min)	Real income per capita ($1982–1984)
Average by Size Class in 2000							
Very Large (> 3 million)	1980	5,407	1.20	22	10	27.3	11,937
	1990	5,899	1.47	23	28	27.5	14,232
	2000	6,427	1.53	38	35	30.6	15,327
Large (1 – 3 million)	1980	1,214	1.08	13	4	21.8	10,952
	1990	1,398	1.18	23	12	22.4	12,393
	2000	1,664	1.30	32	22	25.5	13,629
Medium (0.5 – 1 million)	1980	542	1.05	8	2	19.8	10,148
	1990	615	1.09	14	6	20.3	11,644
	2000	710	1.18	23	14	22.8	13,231
Small (0.1 – 0.5 million)	1980	196	1.03	6	2	18.1	9,732
	1990	224	1.06	9	4	18.8	10,374
	2000	274	1.11	15	7	21.3	11,513
All sizes	1980	1,395	1.14	17	7	24.3	10,626
(75 metro areas)	1990	1,560	1.31	26	19	24.6	12,052
	2000	1,772	1.39	33	27	27.5	13,350

Notes: [a] We use the concept "urbanised areas" as defined by the Census.
[b] For rows labelled "1980" data in columns 2–5 are for 1982, the earliest year reported by TTI.
Sources: Columns 2–5 from *The 2002 Urban Mobility Study*, Texas Transportation Institute (TTI), Texas A&M University System, June 2002, http://mobility.tamu.edu/ums/. Columns 6–7 from various US Censuses: Table 118,124 from Vol. 1, Ch. C *1980 Census of Population*; 1990 STF3C and 2000 SF3C available from http://factfinder.census.gov/servlet/BasicFactsServlet

Table 3. Distribution of US population by county density level, in percents

Year	Share of population in the least dense counties (bottom 50%)	Share of population in the dense counties (90–99th percentiles)	Share of population in the most dense counties (top 1%)
1920	19	30	20
1930	17	33	21
1940	17	34	20
1950	14	38	19
1960	11	43	17
1970	10	45	16
1980	10	45	13
1990	9	46	12
2000	9	49	11

Source: US Population Census, various years.

Share of GDP
in transport
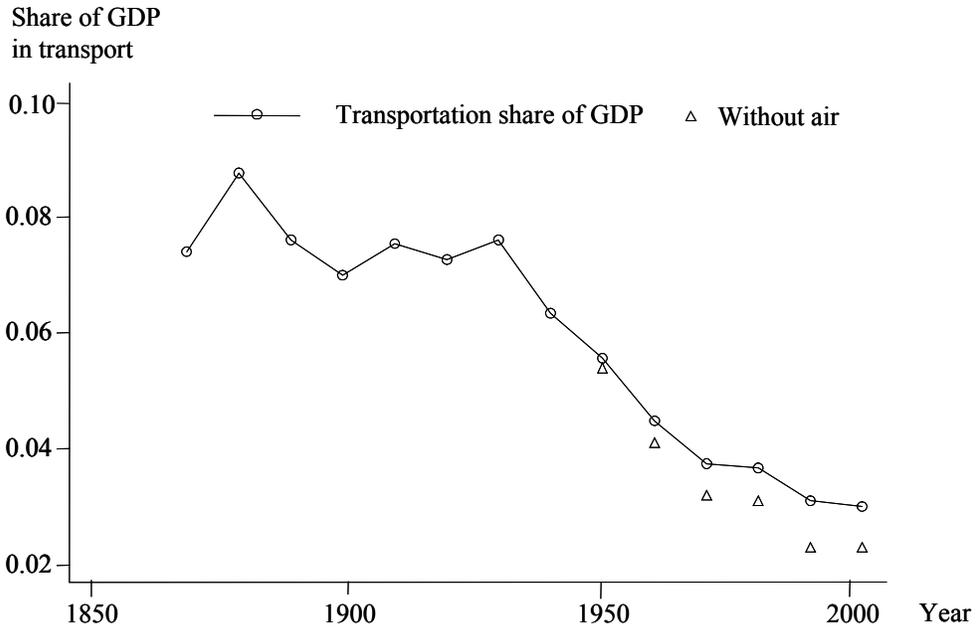
Fig. 1. The share of GDP in transportation industries
Source: Department of Commerce (since 1929), and Historical Statistics of the U.S. (Martin Series) before then.
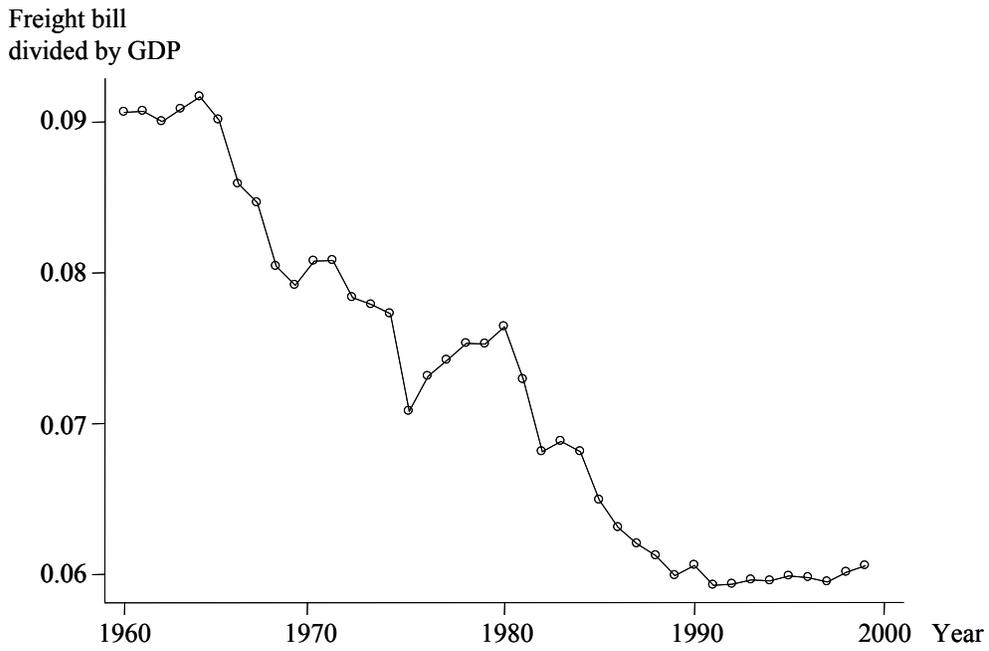
Fig. 2. Transportation bill (freight only) divided by GDP
Source: Bureau of Transportation Statistics Annual Reports.
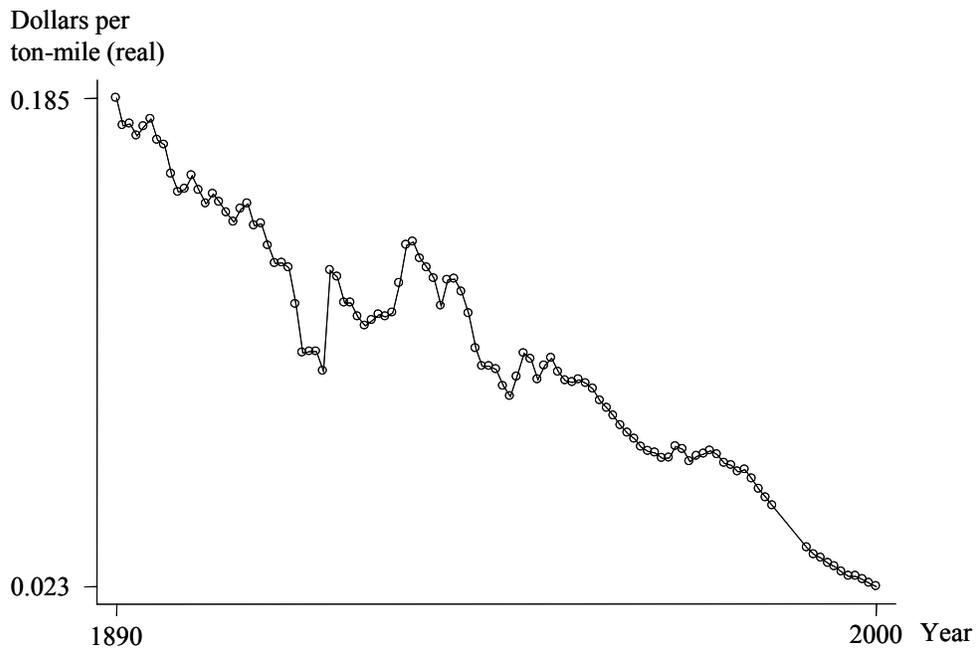
Fig. 3. The costs of railroad transportation over time
Source: Historical Statistics of the US (until 1970), 1994, Bureau of Transportation Statistics Annual Reports 1994 and 2002.

Revenue per ton-mile
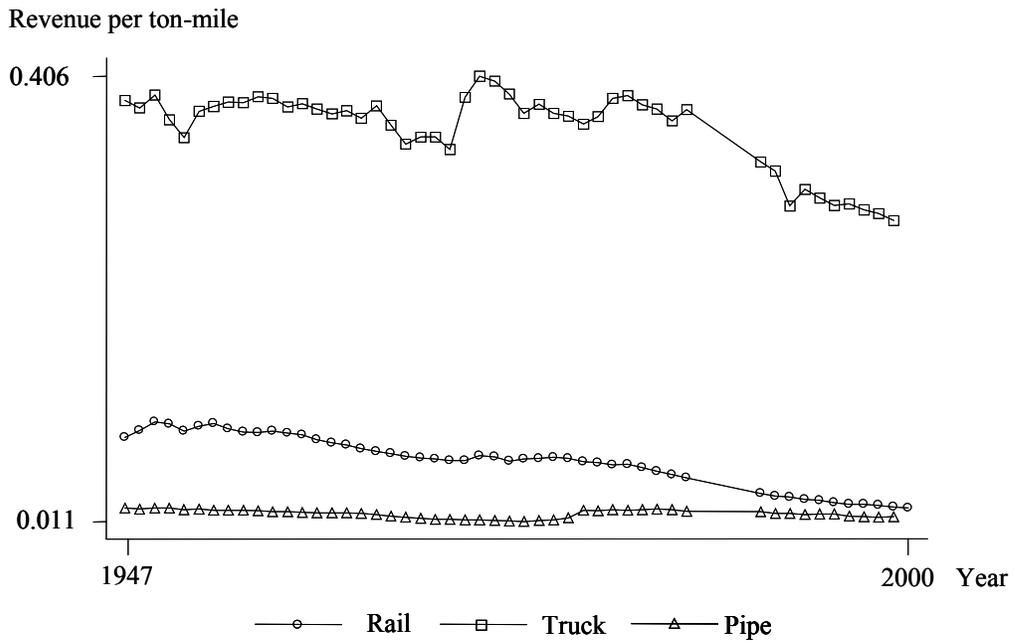
Fig. 4. Revenue per ton-mile, all modes
Source: Bureau of Transportation Statistics Annual Reports
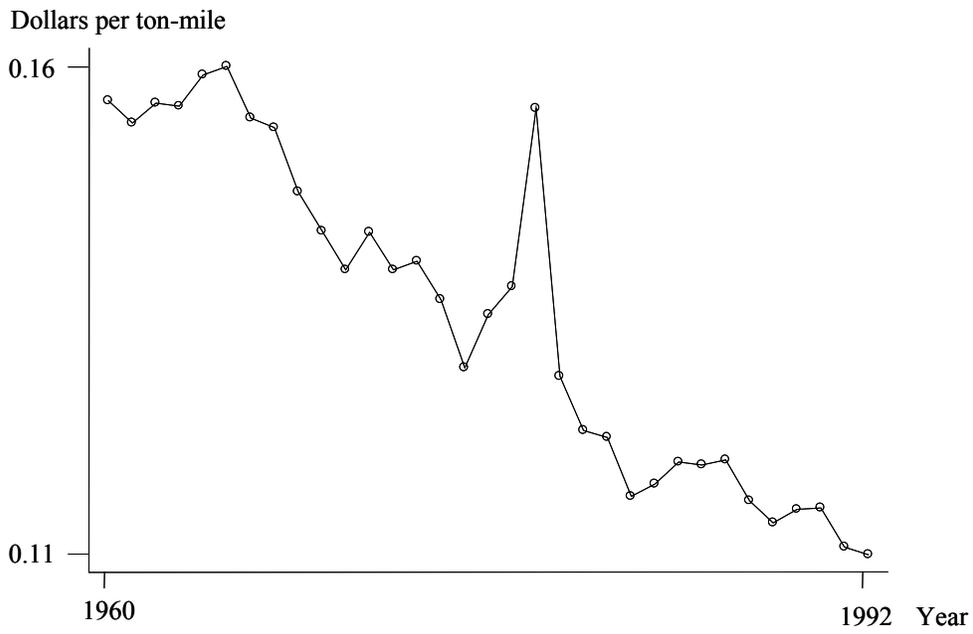
Fig. 5. Revenue per ton-mile, all modes together
Source: Bureau of Transportation Statistics Annual Reports

Ton-miles of freight
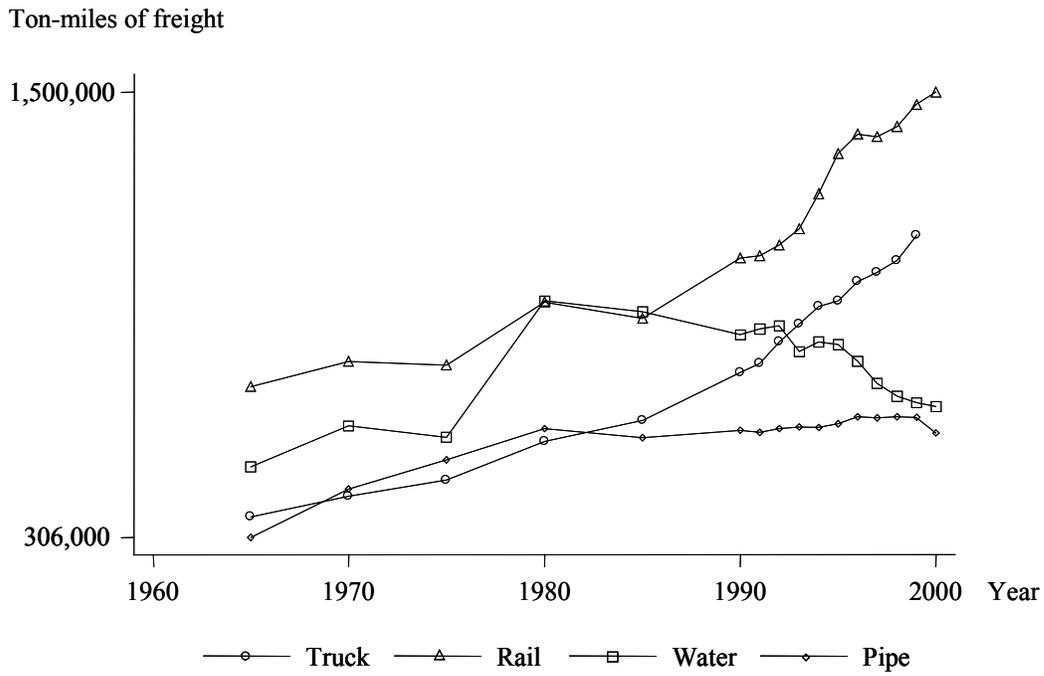
Fig. 6. Ton-miles of freight over time
Source: Bureau of Transportation Statistics Annual Reports
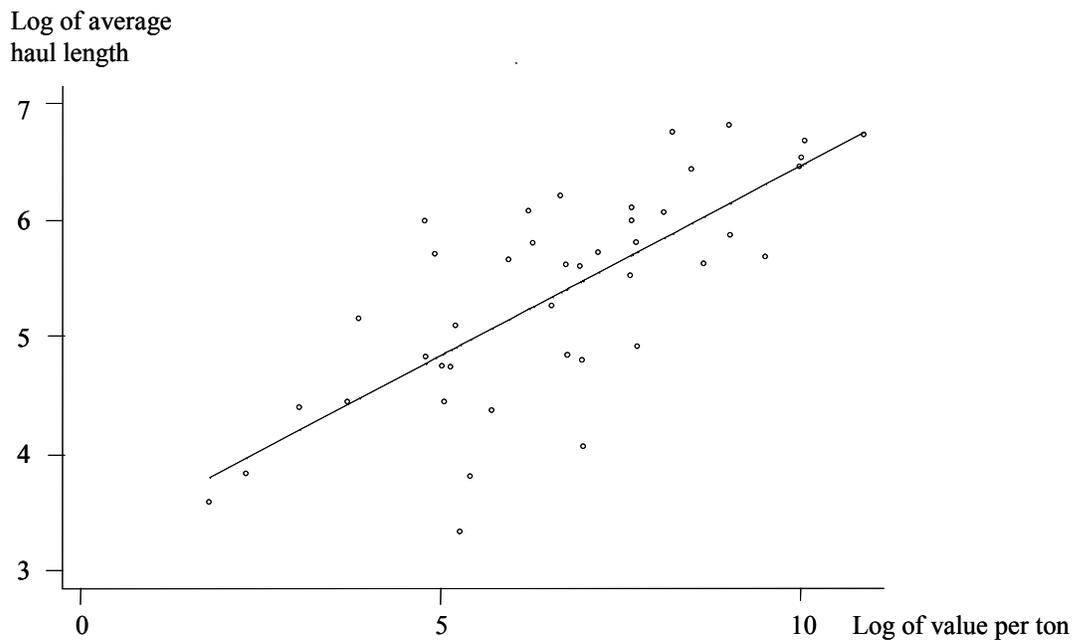
Fig. 7. Distance and value per ton
Source: National Transportation Statistics, 2001, Table 1-52.
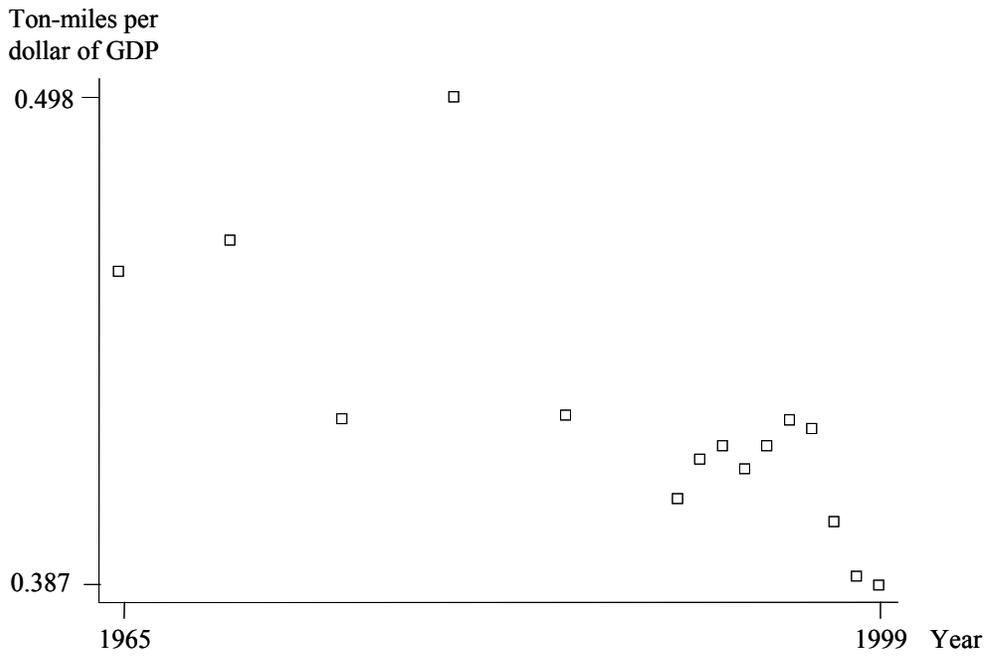
Fig. 8. Ton-miles divided by GDP (in 2001 dollars)
Source: Bureau of Transportation Statistics Annual Reports
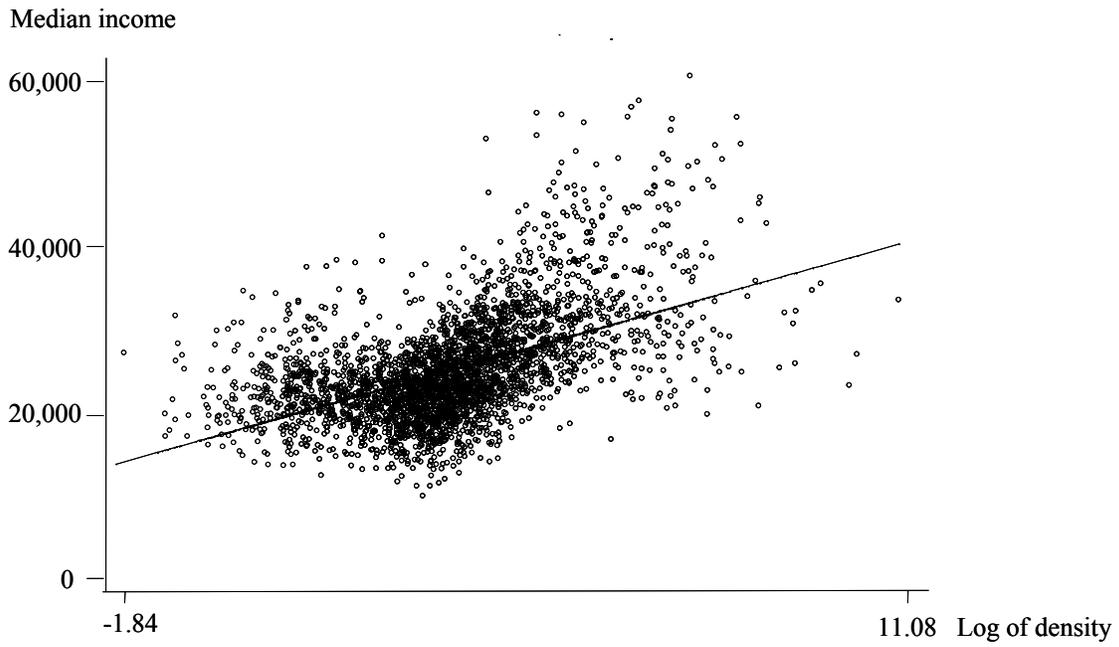
Fig. 9. Income in 1989 and density in 1990
Note: The regression line has a slope of 2055.6 (standard error of 60), and an R^2 of 27.4 percent.
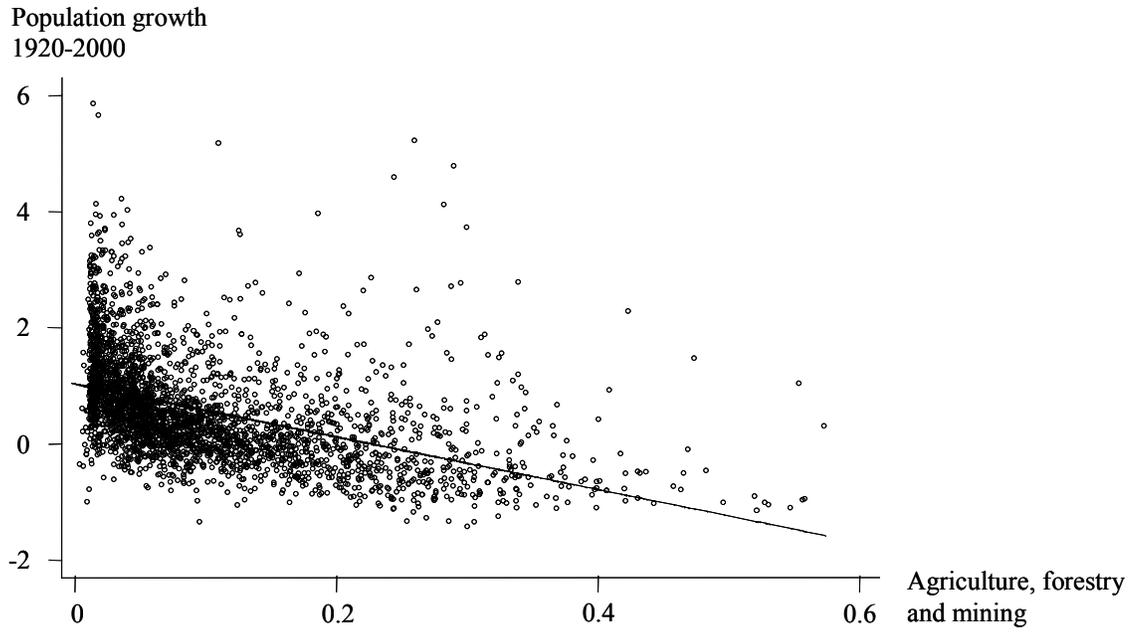
Fig. 10. Population decline and natural resources
Source: US Census, 1920, 1990 and 2000.
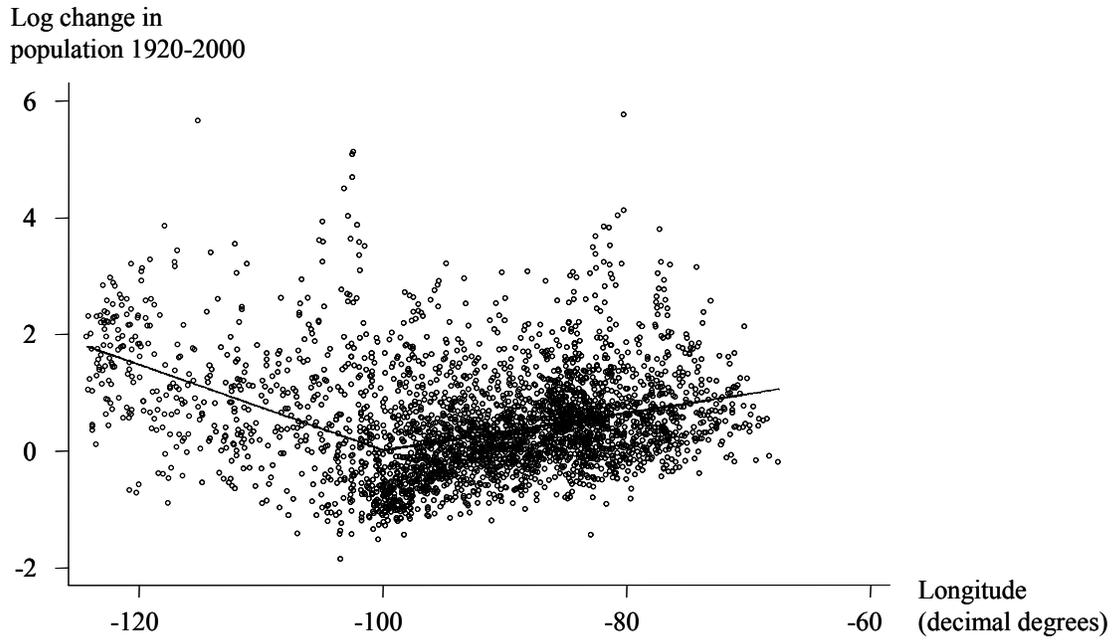
Fig. 11. The emptying of the hinterland, 1920–2000

Population change
1980-2000

0.66

-0.16

2.2

72.6

Avg. daily January
temperature (F)
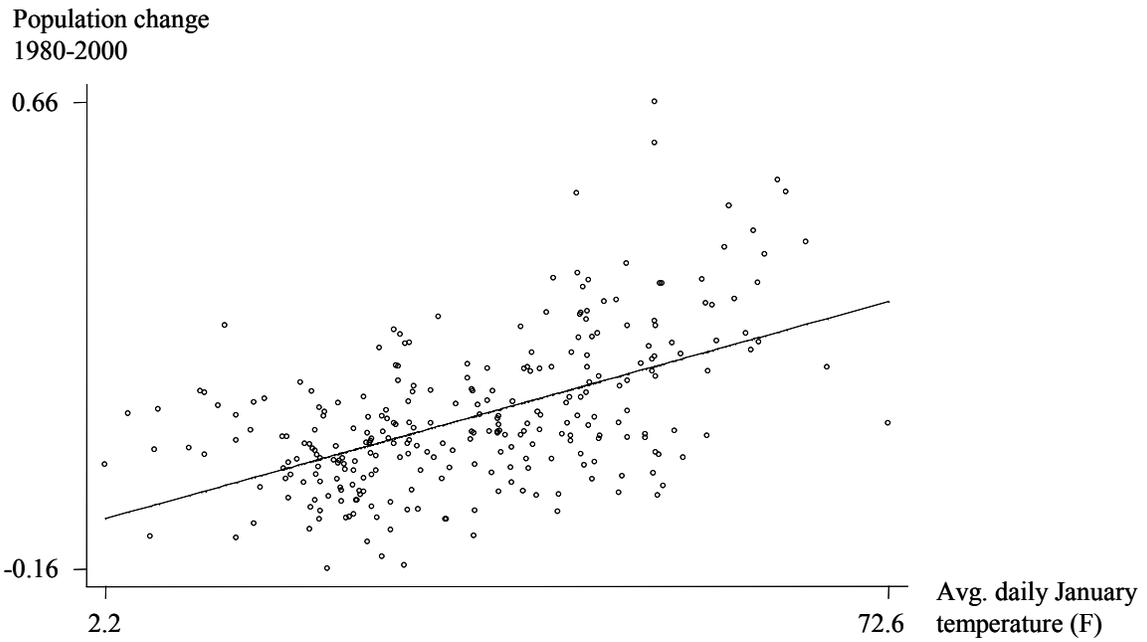
Fig. 12. The growth of temperate places, 1980–2000

Share of employment
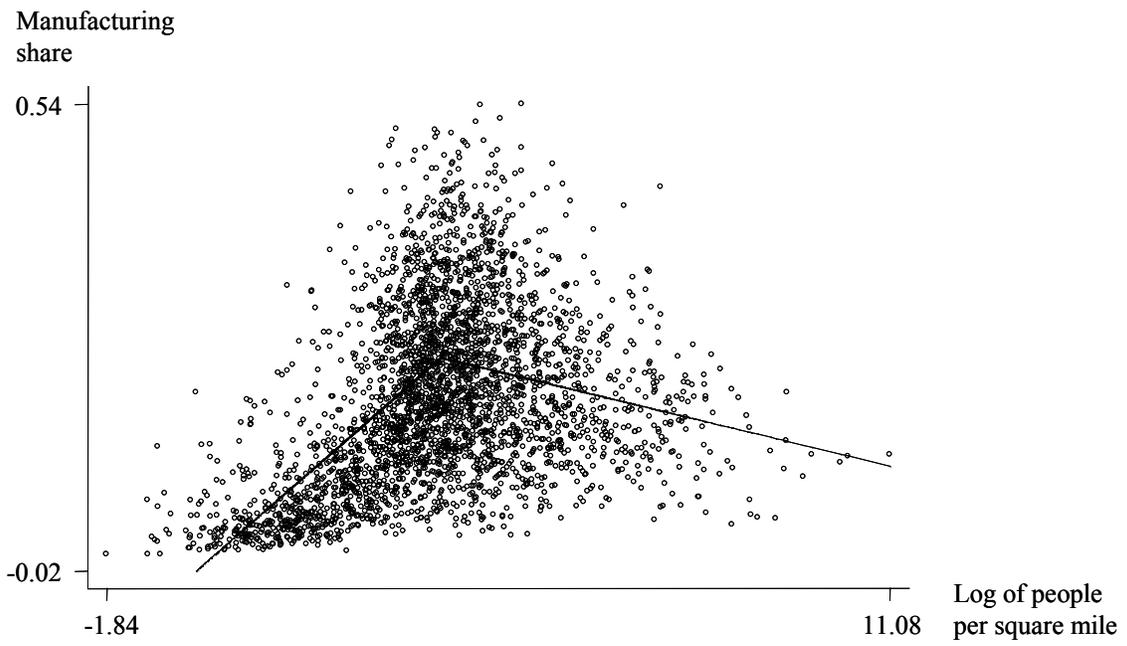in FIRE, 1990
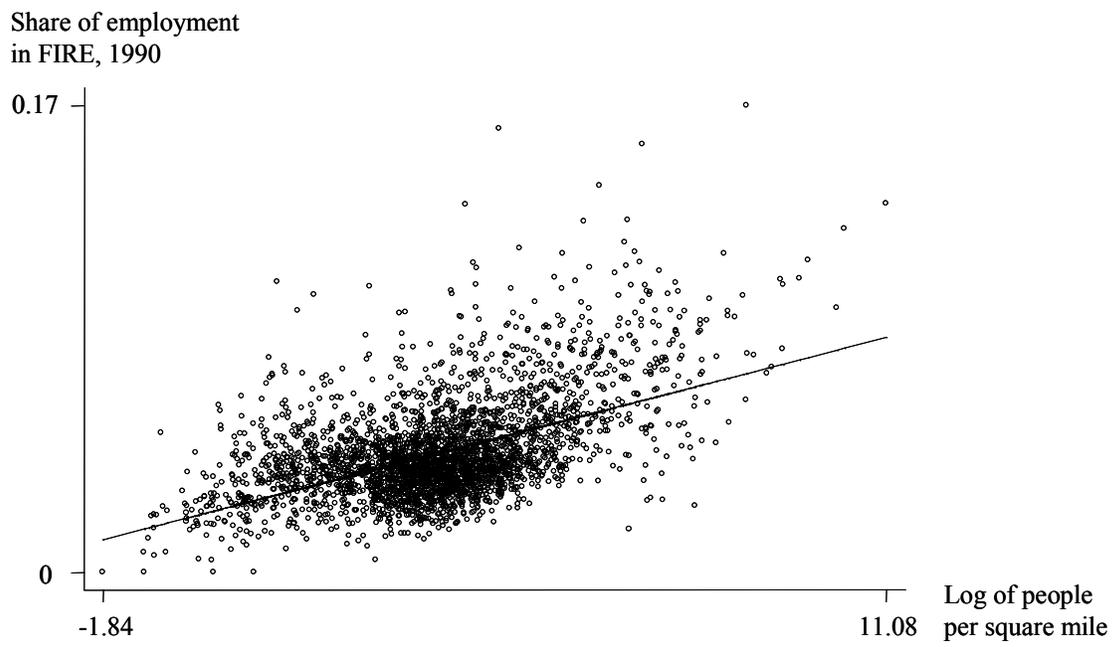
Fig. 13. Services and density
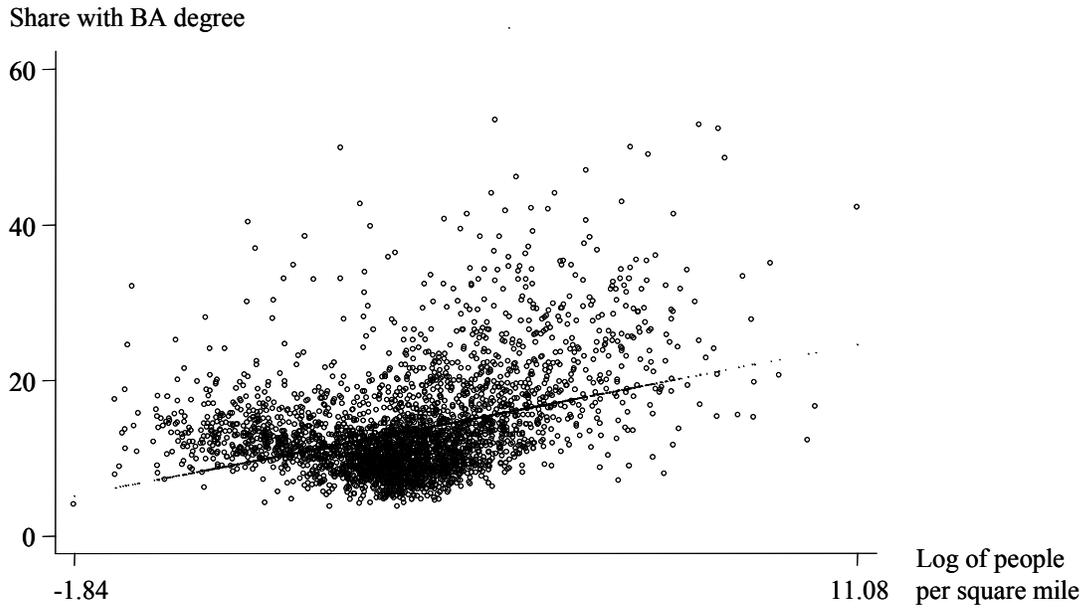
Fig. 14. Manufacturing and density

Fig. 15. Density and the share of the population with college degrees