

Historical Perspectives on U.S. Economic Geography

By

Sukkoo Kim

and

Robert A. Margo

August 2003

We thank Vernon Henderson and participants at the North American Regional Science Association meetings in Puerto Rico for their comments. Prepared for the *Handbook of Regional and Urban Economics*, Volume 4, edited by Vernon Henderson and Jacques-Francois Thisse.

Abstract

We review historical patterns of economic geography for the United States from the colonial period to the present day. The analysis is framed in terms of two geographic scales: regions and cities. The compelling reason for studying geographic area of two different scales is that models that explain location of economic activities at one scale may not apply to other scales. We consider the process of settling the frontier; the development of national markets in goods and factors and, more generally, the convergence (and divergence) of regional economies; the growth of cities and the relationship between urbanization and trends in aggregate economic structure, such as industrialization; and changes in the internal structure of cities.

JEL No. R0

Key Words: Historical Perspectives on U.S. Economic Geography

Index Words:

1. Introduction

2. U.S. Regional Development

2.1 Settlement of the U.S. Continent

2.2 Economic Integration and the Rise of a National Economy

2.3 Divergence and Convergence of U.S. Regional Economies

3. U.S. Urban Development

3.1 Cities in the Colonial Period

3.2 Urbanization and Industrialization

3.3 Urbanization in the Twentieth Century

3.4 Spatial Structures of Cities and Metropolitan Areas

3.5 Size Distribution of Cities

3.6 Urbanization and Inventive Activity

3.7 Urbanization and Productivity

3.8 Urbanization and Health

3.9 Immigration and Urbanization

4. Conclusions

References

1. Introduction

Economic geography is central to economic development. As economies develop, the initial allocation of resources across locations rarely remains fixed. New industries develop and, for technological or other reasons, find it profitable to situate in different locations than old industries. Transportation networks emerge, linking far-flung markets, within and across countries, again potentially altering the spatial distribution of resources. Equally as or even more importantly, the density of economic activity changes with development. In low income societies, much economic activity is concentrated in low-density activities, such as agriculture or cottage industries. As per capita income rises, capital and labor migrate to cities, and average density rises. But within cities, density may also change over time, as people and firms find it to their advantage to spread out over a wider “metropolitan” area.

We explore the economic geography of the U.S. from a historical perspective in two geographic scales: regions and urban areas. The most compelling reason for studying geographic areas of differing scales is that models which explain the location of economic activities at one scale, such as the region, may not apply to smaller scales such as urban areas, or even finer ones like financial and industrial districts.¹ Further, the examination of these two geographic units of observation, regions and urban areas, presents a more coherent understanding of the historical geography of the U.S., than treating each geographic dimension separately.

We have three major goals: to systematically present historical data on U.S. regions and urban areas, to survey the existing literature on the historical geography of the U.S., and to direct scholars to promising avenues of research. Because this subject is a vast one that touches on

¹ Fujita and Thisse (2002) note that geographers have long known that geographic scale matters. According to Fujita and Thisse, to ascribe what is true at one spatial scale to another is to commit ecological fallacy.

many disciplines, our review of the literature is necessarily selective and highly incomplete. However, even an incomplete review, we believe, can offer important insights on the fundamental role of geography in American economic development. It is impossible, in our opinion, to understand long-term development of the U.S. divorced from its geography. Surprisingly, however, many important areas remain relatively unexplored and we try to highlight the most significant of these in our review.

2. U.S. Regional Development

One of the most distinctive characteristics of America is space.² In 1800, the United States was already one of the largest nations in the world. In the nineteenth century, the U.S. became a continental nation of enormous proportions from sea to sea. Moreover, the Americans faced unprecedented challenges in integrating vast unpopulated territories politically and economically. The Northwest Ordinance provided an orderly procedure for incorporating unsettled territories into states, but sectional differences over slavery led to divergence and disunion. However, by the early twentieth century, the states of the union emerged as a unified nation. This section addresses three of the most significant developments in U.S. economic geography from a regional perspective: the westward migration of Americans, the rise of a national economy, and the divergence and convergence of U.S. regional economies.

2.1 Settlement of the U.S. Continent

In the colonial period, most Americans of European and African descent lived along the eastern seaboard. With the Treaty of Versailles in 1783, Americans gained independence and

² “North and South, east and west, city and country, from the 1790s to the 1850s, the most powerful influence in shaping of American society was space. The ideas about space, the uses of space, the projects across space, and the accommodations to space, above all else, charted the basic changes that occurred between these years.” Robert H. Wiebe as quoted in Meinig (1993; p.219).

vast stretches of land through the mid-western U.S. Between 1783 and 1898, through purchase, conquest, or treaty, Americans acquired a continental area of 3 million square miles. Moreover, Americans migrated westward in waves. During the first half of the nineteenth century, people from the Northeast migrated to the East North Central region, and those from the Southeast migrated to the East South Central region. In the second half of the nineteenth century, migration continued westward to the West North Central and, to a lesser extent, the West South Central region. Finally, in the twentieth century, population migrated further west and south to the Mountain, Pacific and West South Central regions (See Figure 1).

Most migrants moved west along the latitude. Meinig (1993) documents the geographic patterns of westward migration and the establishment of western settlements in the nineteenth century. Since most people involved in the westward migration were farmers, people may have migrated along the latitude because they acquired latitude-specific skills in agriculture (see Steckel (1983)). However, the westward migration is not without puzzles. Normally, economists argue that individuals migrate from low to high per capita income areas, all other things equal. In the case of the South, this appears to be consistent with the gradient of per capita incomes (higher in the west than in the east). But in the North, the gradient ran in the opposite direction - incomes were higher in the east than in the west.

Why did people go west? The most famous explanation is the safety valve hypothesis proposed by the historian Frederick Jackson Turner; the economically dispossessed easterners migrated to the west where land was cheap. However, most economic historians dismiss the safety valve hypothesis on the grounds that migration to the west was costly, both in terms of opportunity costs and resources. In particular, farm-making on the western frontier was very

costly, well beyond the means of unemployed industrial workers in eastern cities. Nevertheless, using sophisticated econometric methods, Ferrie (1997) finds some evidence that migration was indeed selective in the sense required by the safety valve hypothesis. In addition, the safety valve hypothesis is consistent with Fogel and Rutner's (1972) analysis that the sale of large tracts of western lands by the federal government induced the westward movement.

The safety valve hypothesis is consistent with the east-west gradient of per capita income distribution in the North, but many scholars believe that income per capita does not always accurately measure gains to migration. It is well known that fertility was high on the western frontier. Adjusting for the higher dependency ratio (children to economically active adults) narrows the difference in per capita incomes between the Midwest and the Northeast (Fogel 1989). More recently, Margo (1999, 2000) has provided real wage indices that are corrected for regional differences in cost of living. Margo (1999, 2000) finds considerable evidence for the hypothesis, initially proposed by Coeholo and Shepherd (1976), that westward migration was prompted by higher marginal product of labor on the frontier than in the east. According to Margo's estimates, real wages of unskilled labor were about 30% higher in the Midwest than in the Northeast ca. 1830. However, as east-west migration began in earnest, the real wage gap in the North diminished to about 15% or so on the eve of the Civil War. By the end of the nineteenth century, a gap remained, but it was smaller (about 10%) than the gap that existed just prior to the War. Margo also shows that the real wage gap for skilled labor was initially much higher, but also converged before the Civil War.

Finally, migration to the antebellum frontier may not be explained by current income, but by future returns - that is, by capital gains. There is evidence that "getting there first" had

economic value on the frontier. Early migrants got the best land and, if and when the railroad came through, got the most capital gains (Galenson and Pope 1992; Craig, Palmquist, and Weiss 1998; Coffman and Gregson 1998).

2.2 *Economic Integration and the Rise of a National Economy*

In the seventeenth and the eighteenth centuries, the colonial economies were integrated with Europe and the West Indies. Beginning in the early nineteenth century, greater investments in internal transportation began to integrate the domestic economy intra-regionally and inter-regionally. Initially, private construction of turnpike roads increased the mileage of roads in New England and Middle Atlantic regions from 4,684 to 11,662 miles between 1810 and 1830.³ However, most of these investments were not profitable. In the following decades, the construction of canals and the use of steamboats on navigable waters lowered the costs of trade markedly. However, the integration of U.S. regional economies really began in earnest with the emergence of the railroads in the 1850s. The growth in the mileage of railroads and telegraphs between the late nineteenth and the turn of the twentieth centuries was spectacular. For example, between 1860 and 1890, the mileage of operation in railroads increased sharply from 30,626 to 166,703 miles whereas that of telegraphs increased from 50,000 to 19,382,000 miles.⁴ These advances in transportation led to dramatic declines in domestic freight rates (see North (1973)).

Regional convergence in prices suggests that integration occurred at different rates for goods and factor markets. Goods market integration seems to have been realized by the latter half of the nineteenth century, capital markets by the early twentieth century, and labor markets

³ See Fishlow (2000) for a summary of developments in internal transportation during the nineteenth and the early twentieth centuries.

⁴ Despite the growing importance of the railroads, Fogel (1964) argues that the direct benefits from lower transportation costs due to the railroads were about two percent of GNP in 1890.

by the mid-twentieth century. North (1961) and Harley (1980) find rapid convergence in the prices of goods such as lard, pork, flour, and wheat across regions between 1820 and 1860 (North 1961; Harley 1980). Davis (1965) finds marked convergence in interest rates by the early 1900s.⁵ Finally, while there is convergence of wages and earnings by the mid-twentieth century, labor markets seem to have remained segmented for a longer period between the North and the South (Rosenbloom 1990; Margo and Villaflor 1987; Goldin and Margo 1992; Margo 2003).

Over the past two centuries or so, the United States has become, in effect, a national market in goods and factors of production. Incomes are now less tied to place than they are to skills, and regions are more homogenous with respect to economic activity than they once were. Nonetheless, it would be a mistake to conclude that region no longer “matters” in the United States. Regions are still subject to economic shocks, and the evidence is that such shocks have consequences for incomes and employment, at least for the medium run, if not for the very long run (Blanchard and Katz 1992).

2.3 *Divergence and Convergence of U.S. Regional Economies*

The most significant long-run trend in U.S. regional economies is the divergence and convergence of U.S. regional economies. In the colonial period, regional differences in industrial structures and incomes were relatively slight. The U.S. regional economies diverged between the nineteenth and the early twentieth centuries. The divergence was first seen in industrial structures as the Northeast region industrialized during the first half of the nineteenth century. In the late nineteenth century, industrial divergence deepened with the formation of the manufacturing belt

⁵ However, Bodenhorn and Rockoff (1992) suggest that the integration of capital markets occurred much earlier. They find that regional interest rates were well integrated in the antebellum period, but that the disruptions of the Civil War caused the dis-integration of capital markets in the postbellum period, especially in the South and the Pacific Coast regions.

in the North and the continued specialization of agriculture in the South. The divergence in income per capita, especially between the North and the South, emerged strongly in the post-bellum period. Southern income per capita fell relatively and absolutely. However, in the second half of the twentieth century, regional industrial structures as well as regional income per capita converged dramatically.

A. Trends in U.S. Regional Economic Structures

The colonial economy was dominated by agriculture, extractive industries and fisheries and was motivated by trade with Europe and the West Indies. Data on exports constructed by Shepherd and Walton (1972) suggest that the American colonies were highly specialized in their export staples.⁶ The New England colonies were specialized in fish, meat, wood products, whales, potash and grain; the Middle Colonies in grain, and to a smaller extent in flaxseed, wood, iron, meat and potash; the Upper South colonies in tobacco, and to a lesser extent in grain, iron and wood products; the Lower South colonies primarily in rice and indigo as well as deerskin, naval store, wood, grain and meat products. Most scholars believe that these patterns of colonial regional economic specialization can be explained by regional comparative advantage based on soil, climate and geography.

The early industrialization of the U.S. economy led to the first divergence of U.S. regional economies. Prior to industrialization, domestic manufactured goods were produced by artisans in urban areas. As the U.S. economy industrialized in the first half of the nineteenth century, manufacturing moved out of artisan shops into non-mechanized and mechanized

⁶ The New England colonies consist of Massachusetts, New Hampshire, Connecticut, and Rhode Island; the Middle colonies consist of New York, New Jersey, Pennsylvania, and Delaware; the Upper South colonies consist of Maryland and Virginia; and Lower South consists of North Carolina, South Carolina, Georgia, and Florida.

factories. Sokoloff (1984, 1986) documents that the rise of factory production in a variety of industries such as boots and shoes, coaches and wagons, furniture, hats, paper, leather tanning, and textiles was associated with significant increases in labor and total factor productivity. The early industrialization led to regional divergence as manufacturing became increasingly concentrated in the Northeast region. In 1840, the share of the labor force in New England and Middle Atlantic regions in non-agricultural activities was 38% and 32% respectively, whereas, for the remaining regions, the figure ranged from 10% to 18%.

Why did industrialization begin in New England? Historians like Hunter (1979) have emphasized the abundance of waterpower sites in New England. However, it is also well known that women and children played an important role in the early industrialization of the U.S. Goldin and Sokoloff (1982) propose that early industrialization proceeded in New England because of the opportunity costs of women and children relative to men in farming were lower in that region than in others. Goldin and Sokoloff find that the relative wages of women and children compared to men were lower in New England than in other regions. Moreover, the relative wage of females rose sharply with industrialization.

Sokoloff (1984) documents that the extension of markets due to the construction of canals led to the re-organization of production from one based on artisanal shops to non-mechanized and mechanized factories in the Northeast. Sokoloff suggests that initial advantages of factory production came from firm reorganization. The greater division of labor caused by the extension of markets in non-mechanized factories led to significant increases in productivity. However, increased inventive activity caused by extension of markets, also documented by Sokoloff (1988), may have contributed to the second phase where increases in productivity are

associated with mechanized factories that adopted new methods of production, power sources, and new machinery. Lindstrom (1978) proposes a slightly different hypothesis for the concentration of manufacturing in the Northeast based on demand. Lindstrom argues that the increase in intra-regional demand brought on by falling transportation costs in the Northeast led to the concentration of manufacturing in that region. Overtime, however, supply-side factors may have become more important.⁷

The rise of a mature industrial economy in the U.S. between the late nineteenth and the early twentieth century was accompanied by a significant divergence in U.S. regional economies, especially between the North and South. Manufacturing spread from the Northeast to the East North Central region to form the manufacturing belt whereas the South remained entrenched in agriculture. In the second half of the twentieth century, as manufacturing spread South and West, U.S. regional economies converged significantly.

The divergence and convergence of U.S. regional economies are documented by Kim (1995, 1998). Kim finds that the patterns of U.S. regional specialization differed considerably by sector (see Figure 2). In agriculture, regions continued to become more specialized over time in various crops such as wheat, barley, corn, oats, rye, cotton, tobacco and vegetables. In manufacturing, data at the two-digit industry level indicate that regions became more specialized between the late nineteenth and the early twentieth centuries; however, since the mid-twentieth century, regions became substantially more similar in industrial structure.⁸ Regional

⁷ See Rosenberg (1963).

⁸ The two-digit industries in manufacturing are food, tobacco, apparel, lumber and wood, furniture and fixtures, paper, printing and publishing, chemicals, petroleum and coal, rubber and plastics, leather, stone, clay and glass, primary metal, fabricated metal, machinery, electrical machinery, transportation, instruments, and miscellaneous.

specialization in the service industries, wholesale trade, retail trade, and other services were low for the second half of the twentieth century when data were available.⁹

Kim (1998) finds that regions became more specialized at the 1-digit sectoral level between 1870 and 1880, fell slightly between 1880 and 1910, and then fell significantly over the twentieth century.¹⁰ However, the data indicate that differences in regional industrial structures between the North and South were particularly high until the second half of the twentieth century. Between the late nineteenth and the early twentieth centuries, regional industrial structures between the North and South differed by more than 50% from that of the overall average. This difference was caused by the concentration of manufacturing in the North and the concentration of agriculture in the South. In the twentieth century, regional industrial structures converged significantly. The convergence was caused by the dispersal of agriculture and the shift in economic structure into services.

For most regional and urban economists, the emergence of the manufacturing belt during the second half of the nineteenth century marked a major turning point in the U.S. economic geography.¹¹ The manufacturing belt, composed mostly of the New England, Middle Atlantic and East North Central regions, contained the majority of manufacturing employment between

⁹ The wholesale industries match the two-digit industries in manufacturing; the retail industries consist of food and general stores, general merchandise, apparel, furniture, automotive, filling stations, eating, drug stores, and other; services industries consist of hotel and lodging, personal, business, auto repair, garages, miscellaneous repair, motion pictures, amusement and recreational, health, legal, educational, social, museum and botanical zoological gardens, membership organizations, and miscellaneous.

¹⁰ For 1870 and 1950, the one-digit industries consist of agriculture, mining, forestry, fishing, manufacturing and services. For the years between 1939 and 1987, construction, transportation, wholesale trade, retail trade, and government industries were also included.

¹¹ See Holmes and Stevens (2004) for an informative map of the manufacturing belt and the location of large manufacturing plants in 1947 and 1999.

the late nineteenth and the first half of the twentieth centuries. Krugman (1991) proposes that the U.S. manufacturing belt emerged when economies of scale in production rose and transportation costs fell. To minimize transportation costs, firms chose to locate in one region - the manufacturing belt. The concentration of firms further increased local demand and locked in the initial advantage. Krugman suggests that the manufacturing belt also benefitted from a denser population and a transportation network that provided a substantially larger local market.

Meyer (1983, 1989) presents a slightly different interpretation of the rise of the manufacturing belt. For Meyer, the manufacturing belt arose as series of regional industrial systems in the antebellum period that merged to form the manufacturing belt in the late nineteenth century. Meyer, in an analysis similar to that of Lindstrom (1978) for early industrialization, argues that local regional demand triggered industrialization in a series of regions stretching from the East to the Midwest over time. Although regions like the Midwest industrialized later than those in the Northeast, Meyer argues that high interregional transportation costs enabled Midwestern manufacturers to develop their own manufacturing sector during the antebellum period. When transportation costs fell in the late nineteenth century, Midwestern manufacturers were able to effectively compete with Eastern producers.

Kim (1999) presents alternative evidence for the rise of the manufacturing belt based on natural advantages. Since data on interregional trade are not available to test the Heckscher-Ohlin model, Kim exploits the Rybczynski theorem that provides a linear relationship between factor endowments and production. Kim finds that a relatively small number of factor endowments is able to explain significant variations in U.S. state production for twenty 2-digit manufacturing industries for 1880 and 1900 as well as later periods in the twentieth century.

While these results do not rule out the importance of increasing returns, they suggest that natural advantages may have played a significant role in the rise of the manufacturing belt.

The emergence and persistence of the manufacturing belt was a significant phenomenon in U.S. economic history. It justly deserves greater attention by scholars. However, the declining importance of the manufacturing belt and the rapid convergence of regional industrial structures during the second half of the twentieth century should receive equal attention. Kim (1995) argues that the rise and decline of regional specialization, or the divergence and convergence of regional industrial structures, can be accounted for by changes in the factor mobility of factors and goods. As transportation costs of final goods relative to inputs fell between the nineteenth and early twentieth centuries, regions became more specialized in order to exploit their comparative advantage. During the second half of the twentieth century, as factors became increasingly more mobile relative to goods, and as new technologies such as electricity reduced the importance of regional differences in resources, regional industrial structures converged.

An alternative interpretation of the convergence of U.S. regional industrial structures is the core-periphery reversal (Krugman and Venables (1995), Puga (1999), Tabuchi (1998), Helpman (1998)) or the bell-shape relationship discussed in Ottaviano and Thisse (2004). Assuming labor mobility throughout, as transportation costs fall low enough with any type of congestion in the core (or with limited mobility throughout), the core-periphery structure will reverse with industrial diffusion. However, at this point, we do not know of any empirical studies based on this theory.

B. Trends in U.S. Regional Incomes

In the colonial period, differences in regional income per capita seem to have been slight.

While we do not have estimates of regional income per capita, we have estimates of regional wealth per capita. Using probate records, Jones (1980) finds that differences in wealth per capita were very small. In 1774, wealth estimates in pounds sterling for New England, Middle Colonies, and the Southern colonies were £36.4, £40.2 and £35.4 respectively. However, the wealth per capita for the free population was significantly higher for the Southern colonies.

In the early nineteenth century, regional incomes diverged somewhat. Yet the divergence was not along North and South. Although income per capita in the Northeast was higher than the national average, it was highest in the West South Central region in 1840 and 1860 (see Table 1 and Fogel (1989)). In the second half of the nineteenth century, Figure 3 shows that regional incomes diverged significantly, especially between the North and the South. Income per capita in the South fell sharply during the Civil War decade, absolutely and relative to other regions. Although there was some recovery after the initial decline, the pace was slow. By 1900, the income per capita of the South was barely half of the national average. However, during the second half of the twentieth century, regional incomes converged substantially (Easterlin (1960, 1961), Mitchener and McLean (1999)). Although there are still regional differences, these are much smaller than before World War II, particularly in the case of the South.

Barro and Sala-i-Martin (1991) find that the convergence of U.S. regional income per capita in the twentieth century is consistent with the predictions of the neoclassical growth model. As predicted by the model, growth in income per capita was inversely correlated with initial incomes in 1880. However, there is no work in the growth literature that examines the divergence of U.S. regional income per capita in the late nineteenth century. While this divergence may be consistent with models of growth based on increasing returns, scholars have

not examined regional income per capita between 1860 and 1900 using this framework.

The divergence and convergence of regional industrial structures can potentially cause divergence and convergence of regional income per capita. To the extent that wages differ across industries, regional specialization in different industries may cause aggregate incomes to differ. Using a simple decomposition procedure, Kim (1998) finds that, while the level of importance varies with regions and time, the divergence and convergence of regional industrial structures are highly correlated with that of regional income per capita. Caselli and Coleman (2001) report similar findings with a more sophisticated decomposition procedure. In addition, these studies also show that differences in regional wages in identical industries also contributed to the divergence and convergence of regional incomes. However, this latter component is better explained by models of economic growth rather than economic geography.

For Krugman, the divergence and convergence of regional incomes are accounted for by transportation costs and increasing returns. When transportation costs fell, increasing returns in manufacturing led to the rise of the manufacturing belt where high-wage jobs were concentrated. However, Krugman and Venables (1995) show that when transportation costs fall even further, then regions eventually converge even when increasing returns are present. On the other hand, Kim (1998) and Caselli and Coleman (2001) present explanations for the divergence and convergence of regional incomes based on comparative advantage.¹² Caselli and Coleman suggest that the shift out of agriculture in the South was aided by the declining costs of acquiring nonagricultural skills.¹³

¹² Kim's (1998) argument is outlined above. Caselli and Coleman's (2001) present an interesting model based on the following assumptions: less than unit income elasticity of farm good demand, faster factor productivity growth in agriculture, and declining costs of acquiring nonfarming skills.

¹³ Caselli and Coleman (2001) argue that the costs of acquiring education fell as reductions in

However, most economic historians argue that the institution of slavery had a major impact on U.S. regional development over time. Slavery and its dissolution greatly influenced the productivity of agriculture in the South before and after the Civil War and, consequently on U.S. regional per capita incomes. The legacy of slavery and the Civil War led to an isolated labor market in the South and it contributed to poor educational infrastructure. Economic historians interpret the Civil War as a major negative shock that took many years to work through. Many economic historians also believe that the federal government played an important role in the integration of the South into the national economy.

The causes of the decline in per capita incomes in the South have been extensively studied by economic historians. There is general agreement that the economic dislocation caused by the Civil War was responsible for at least some of the initial decline. Fogel and Engerman (1974, 1979) argue that slave labor was relatively efficient compared with free labor in the South, due to the use of the so-called “gang system” by slave owners. With the end of the War came emancipation and the end of slavery, and a once-and-for-all loss in productivity in southern agriculture (Fogel and Engerman 1974, 1979; Moen 1992). Other potential causes of the relative decline in southern income per capita incomes include a reduction in the rate of growth of the demand for southern cotton (Temin 1979), as well as the possibility that per capita incomes in the South in 1860 may have been above their long-run trend (Wright 1979).

Recent work by Margo (2003) has examined this issue from the perspective of regional wage gaps. Wright (1986) points out that, at the turn of the twentieth century, the South - particularly the South Atlantic - was a low-wage region in a high-wage country. The South ca.

transportation costs in rural areas allowed schools to exploit economies of scale, as school curricula was transformed in the early twentieth century, as life expectancy rose over time, and as desegregation improved the access to education for blacks.

1900 could have been low-wage due to recent economic shocks. Or it could have been low-wage for some time - that is, predating the Civil War. Margo presents a variety of data on wages in the South relative to the North, before and after the Civil War. His results show that all three explanations have merit. Wages in the South Atlantic regions were below those in the North before the Civil War, although not in the South Central regions. Wages in the South also declined relative to the North in the 1890s, possibly due to a short-run decline in agricultural productivity. However, the data clearly show that the war left an imprint on the regional wage structure - as was true of per capita incomes, wages in the South declined sharply relative to the North between 1860 and 1870.

Why was convergence delayed so long? Economic historians have suggested several answers to this question. One answer emphasizes the “flawed” economic institutions of the South, particularly in agriculture, after the war (Ransom and Sutch 1977). Another explanation, due to Wright, involves network externalities in labor markets. In the North, according to Wright (1986), jobs in manufacturing were allocated not by an impersonal labor market, but by one dominated by personal contacts and references. However, the Southerners, consumed by the turbulence of reconstruction, failed to develop these employment networks. As a result, the jobs in the North went to immigrants rather than to Southerners. Only after immigration was abruptly cut off during and after World War I did northern manufacturers turn to the South for their labor supply (Collins 1997).

Another explanation of the slow pace of Southern recovery emphasizes the poor quality of education in the region. Using census data, Margo (1990) establishes that the probability that a Southern-born individual left the region was a positive function of schooling; that is,

interregional migrants were better educated than those that stayed behind. This factor was particularly important for blacks, who attended racially segregated schools that generally were inferior in quality to that of schools attended by whites.

3. U.S. Urban Development

The U.S. transformed itself from a rural to an urban society between the nineteenth and the twentieth centuries. The onset of industrialization and the expansion of domestic markets significantly increased the number and size of cities. By the turn of the twentieth century, America became an urban nation. The spatial structures of cities changed markedly over time as well. Initially, cities were compact and dense. Over time, boundaries of cities increased as firms and households moved further away from the city center. In this section, we examine the growth of cities and the changes in the spatial structures of cities over time.

In addition, we examine factors that differentiate urban from rural areas. Unlike rural places, population sizes of cities varied systematically at any given point in time. While the relationship has not remained constant over time, the size distribution has been usefully characterized by the rank size rule or by Zipf's law. Moreover, urban places have been centers of inventive activity, culture, entertainment, government and a magnet for immigrants. Urban places fostered higher productivity and higher nominal wages. Unfortunately, for a time in U.S. history, urban places were also extremely unhealthy and mortality rates of urban areas far exceeded those of rural areas.

3.1 Cities in the Colonial Period

The colonial British America was predominantly rural. In 1690, there were only four cities with populations greater than 2,500. Boston was the largest city with a population of 7,000

with Philadelphia and New York trailing behind with populations near 4,000. During the next one hundred years from 1690 to 1790, the number of cities rose from 4 to 24, yet the share of the urban population fell from 8.3 to 5.1% (see Table 2). For the only extended time in American history, the rural population grew faster than the urban population. The rank of cities also shifted over this period; Philadelphia emerged as the largest city during the mid-eighteenth century but then gave way to New York towards the end of the century.

The colonial urban occupational structure differed greatly from the rest of the nation. Whereas the great majority of the general population was engaged in agriculture, most of the urban population was engaged in mercantile, manufacturing and other service activities. Historians have constructed occupational structures of various colonial cities using tax records and directories for the late eighteenth century. According to Price (1974), based on information for Boston, Philadelphia and New York, an average of 23% of the population was employed in mercantile or commercial activities, 24% in manufacturing, 50% in services and 4% in government.¹⁴ Thus, despite the reputation of port cities as centers of mercantile activity, the industrial and service sectors were responsible for the majority of the labor force.

The cities in the colonial period arose along the eastern seaboard. Urban economists believe that port cities arise because of increasing returns and because geography bestowed on them lower transport costs.¹⁵ However, few mainstream urban economists have examined the rise of port cities in the colonial period. Most of what is known about cities in this period derives from the work of historians. While these works are rarely guided by explicit economic theory, most scholars seem to believe that merchants became concentrated in cities to benefit from

¹⁴ Also see Warner (1968).

¹⁵ Fujita and Mori (1996) and Konishi (2000).

information spillovers. Merchants gathered regularly in coffee shops and exchanges to discuss market conditions in Europe and the West Indies.¹⁶

Most historians believe that port cities arose in places endowed with excellent harbors and access to a rich hinterland.¹⁷ Due to high interregional transportation and communications costs, the size of the hinterland determined the size of a city's mercantile activity and the quality of the port determined the costs of transportation overseas. Yet, while this thesis may explain the rise of cities like New York and Philadelphia in the Middle Colonies or even Boston in New England, it does not easily explain why there were no major cities in the Upper South, and why cities in New England were so numerous. The Chesapeake region had excellent navigable waters and a rich hinterland; on the other hand, New England's hinterland was relatively small.

Historians have proposed a variety of theories to explain the absence of cities in the Upper South, but the most attractive theory is based on colonial regulation of trade.¹⁸ According to Price (1974), the lack of cities in the Chesapeake is accounted for by British regulation of tobacco trade. The Navigation Acts restricted exports of tobacco to England whereas similar

¹⁶ See Albion (1939); also see Burrows and Wallace (1999).

¹⁷ Boston, the largest port city in the U.S. until the mid-seventeenth century, possessed an excellent, deep, sheltered harbor, and a hinterland composed of coastal settlements from Cape Cod to Maine; Philadelphia, the premier city between the mid- to the end of the seventeenth century, possessed an excellent agricultural hinterland but a harbor subject to freezing in winter; New York, the foremost city from the early eighteenth century, possessed a harbor unparalleled as well as a rich agricultural hinterland (see Price (1974)).

¹⁸ Two of the more prominent explanations for the lack of cities in the Upper South are the abundance of navigable waters in the Chesapeake and that tobacco required little processing (see Middleton (1953), Carville (1992)). Neither of these explanations seem satisfactory. First, if mercantile activity in tobacco is subject to spillovers, then lower transportation costs should foster rather than deter the concentration of merchants in cities. Second, it is not clear that the processing of wheat or other crops was any more extensive than that of tobacco. In any case, forward linkages associated with wheat processing seems an unlikely explanation for significant agglomeration economies in Philadelphia or New York.

restrictions did not apply for fish, wheat, corn, pork, beef, etc. In addition, regulations and high customs duties greatly restricted entry into the tobacco trade. For both of these reasons, a handful of merchants coordinated the tobacco trade from London and Glasgow rather than from cities in the Chesapeake.¹⁹ By contrast, merchants in New York or Philadelphia exported their products to many markets, and entry into the business was relatively open. For these merchants, regular meetings in coffee houses provided valuable information spillovers on local and overseas market conditions as well as information on trustworthy contacts and potential partners.²⁰ Thus, when Baltimore and Alexandria and other cities arose in the Upper South in the mid-eighteenth century, their activity was based on wheat rather than on tobacco.²¹

Despite the fact that most scholars believe that port cities arose for mercantile reasons, the occupation structures of cities indicate that the number of merchants were equaled by artisans and surpassed by service workers. Urban artisans, working alone or with an apprentice, produced a wide variety of products such as barrels, kegs, bread, candles, soaps, etc. Why did artisans locate in cities? Historians and economists seem to rarely ask this question. The most likely answer is transportation costs. Due to high transportation costs, except for trade in export staples, most cities produced goods for local consumption. As local demand for manufactures grew, artisans emerged in cities to satisfy this demand. Services in cities, such as tavern and inn keeping, can be explained in a similar manner. Thus, the concentration of artisans and service workers added greatly to an urban economy, but these activities alone did not sustain an urban area during this period.

¹⁹ See Price and Clemens (1987).

²⁰ See Doerflinger (1986).

²¹ See Klingaman (1969).

3.2 *Urbanization and Industrialization*

In the United States, there was a strong correlation between industrialization and urbanization. While cities existed in the pre-industrial era, the rapid growth in the number and size of cities coincided with the development of a manufacturing economy. Moreover, cities were concentrated in regions where industrialization began. Table 2 and Figure 4 show that in the early phase of industrialization, urban population grew from 7% in 1820 to 20% in 1860. Between 1820 and 1860, urban population grew by an average of 57% per decade. Moreover, the cities became significantly larger. In 1820, there was only one city with a population above 100,000; in 1860, there were 9. Throughout this period, New York remained the largest city with its population increasing from 123,706 to 805,651. The urban population in the first half of the nineteenth century was concentrated in the Northeast region where industrialization began. For these Northeastern cities, Williamson and Swanson (1966) find that manufacturing accounted for approximately 60% of the labor force during this period.

Did industrialization cause urbanization in the U.S.? The accepted view in the development literature is that urbanization and industrialization go hand in hand. Labor saving technological change combined with innovation and development of new manufactured and service products spur the growth of urban areas where products are most efficiently produced. However, did industrialization lead urbanization or vice versa?

While few studies examine this question directly, the existing literature hints at the possibility that industrialization led urbanization. First, the Goldin-Sokoloff hypothesis on U.S. early industrialization argues that industrialization first began in rural areas where the opportunity cost of women and children relative to men were low. Although the Goldin-Sokoloff

model does not address the issue of industrialization and urbanization, the fact that industrialization is seen to arise in rural areas suggests that industrialization preceded urbanization. Thus, in the Northeast rural regions, cities arose as manufacturing developed in these areas.

Williamson's (1965) work on urban growth during early industrialization seems to confirm the view that industrialization led urbanization. If urbanization caused industrialization, then industrialization should take place in large urban areas which formed prior to industrialization. Moreover, larger cities rather than smaller cities or rural areas should industrialize and grow faster over time. However, contrary to expectations, Williamson (1965) finds that rates of urbanization converged rather than diverged between 1790 and 1890. During this period, rural areas urbanized more rapidly than the developed urban areas. In addition, according to Williamson and Swanson, urban growth was consistently unrelated to the initial size of cities and younger cities grew faster than older cities.

The U.S. became a mature industrial economy between the second half of the nineteenth and the early twentieth centuries. America also became an urban nation during this period. By 1920, more than half of the population lived in cities. In this period, based on a sample of cities, Weiss (1974) finds that slightly less than 50% of the labor force was engaged in manufacturing, 50% in services, and a tiny fraction in agriculture. The rise of an industrial economy was accompanied by the rise of the manufacturing belt as industrialization moved westward but not southward. Once again, as with early industrialization, there was high geographic correlation between later industrialization and urbanization as the urban population was highly concentrated in the industrial belt. In addition, the cities in this region were significantly more oriented toward

manufacturing than those in other regions. The cities in the New England, Middle Atlantic and East North Central regions had twice the labor force allocated to manufacturing than that of cities in other regions.²²

Despite the historical importance of urbanization during this period, the topic has received limited attention from economists and economic historians in recent years. The most widely shared interpretation of the rise of cities in this period is based on pecuniary and Marshallian externalities. Meyer (1983, 1988) emphasizes pecuniary externalities; for Meyer, the manufacturing belt is really an agglomeration of regional urban city systems. Ades and Glaeser (1999) find that initial development proxied by urbanization in 1850 was significantly correlated with later urban development. Thus, increasing returns led to divergent urban development. However, Ades and Glaeser find openness or the extension of markets was negatively correlated with urban growth.

Kim (2000) argues that the location of cities was largely driven by superior access to resources and that cities arose to take advantage of economies in local public goods and in transportation. A variety of factors suggest that the concentration of manufacturing cities in the industrial belt resulted from that region's comparative advantage in resources. First, the cities in the industrial belt were specialized in manufacturing. They had a significantly higher proportion of the labor force engaged in manufacturing than cities in other regions. Second, within the manufacturing sector, the cities were highly specialized in one or two industries and cities within a region tended to specialize in the same set of industries. City specialization in a few industries might signal the importance of localization economies, but it cannot explain why cities within regions tended to specialize in the same set of industries. Third, the rise of the Midwestern cities

²² See Kim (2000).

in the industrial belt was correlated with the growing importance of primary metals, machinery, and transportation industries.

3.3 Urbanization in the Twentieth Century

The pattern of urban development in the second half of the twentieth century differed in nature and scope from the industrial period (see Table 2 and Figure 4). First, the pace of urbanization slowed. The share of population in cities rose modestly from 51.2% to 61.6% between 1920 and 1990. Second, Table 3 and Figure 5 show that the urban population became increasingly more suburbanized in metropolitan areas; in 1940, there were 138 metropolitan areas in the U.S. which contained 51% of the U.S. population. However, in 1990, the number of metropolitan areas in the U.S. rose to 335 as did the share of population in these areas to 77.5%. In this period, the industrial structure of cities shifted dramatically into services. By the end of the twentieth century, the percentage of employment in services reached almost three times that of manufacturing (see Kolko (1999)).

The patterns of urbanization in the second half of the twentieth century exhibited marked regional convergence. The regional distribution of cities became more even as the shares fell in the Northeast and the Midwest and rose in the West. While economic structures of cities continued to differ by size, they converged rapidly across regions. Kim (1995, 2000) suggests that these trends, like those of regions, are roughly consistent with diminished differences in regional comparative advantage. During this period, regional input supplies became more similar as factors became more mobile and as technical advances reduced regional differences in inputs. The shift in industrial structure toward services further reduced the importance of differences in resources. Thus, the convergence of geographic differences in resources accounts for the

convergence of city industrial structure as well as the convergence of the geographic distribution of cities across regions.

It is beyond the scope of this section to summarize the literature on modern twentieth century data. For excellent reviews, we refer the reader to Duranton and Puga (2004) and Rosenthal and Strange (2004). Since some kind of agglomeration economies are needed to generate cities, the literature on cities attempts to identify which sources of agglomeration economies are most important for urban development. Here, we highlight a few papers that possess long-run historical dimension. Two of these studies emphasize the role of Marshallian externalities. Simon and Nardinelli (2002) find that cities with higher levels of human capital in 1880 grew faster over the next one hundred years, suggesting the importance of knowledge spillovers. Beeson, DeJong, and Troesken (2001) find that initial advantages have persistent results for a much longer period in U.S. history. However, Rappaport and Sachs (2001) argues that geography, especially access to coasts of both oceans and Great Lakes play an important role in explaining population distribution and density. Mitchner and McLean find that states with navigable waterways, large mineral endowment, and no slaves in 1860 had higher labor productivity levels between 1880 and 1980.

3.4 Spatial Structures of Cities and Metropolitan Areas

The land use patterns of U.S. urban areas have changed dramatically over time. In the colonial period, cities were small, compact and densely organized near the waterfront. Between the nineteenth and the mid-twentieth centuries, urban boundaries increased as outlying areas were annexed. However, cities generally became more dense as population migrated to cities and building density in downtowns rose. The land-use patterns of urban areas changed dramatically

in the second half of the twentieth century. As population moved further away from city centers, outside traditional municipal boundaries, a new definition of urban areas emerged based on the metropolitan area. In this period, urban densities declined dramatically as density in city centers fell and as urban boundaries increased ever more outwards. In this section, we use the density gradient, which measures changes in density as one moves further away from the city center, and average density, defined simply as persons per square mile, to study the historical trends in urban spatial structures.

A. Density Gradients

Due to data limitations, it is difficult to construct data on spatial structures of cities until the late nineteenth century. However, using federal census documents, tax records and the city directory for Philadelphia in 1790, Schweitzer (1993) was able to identify the locations of residence and occupation of a majority of inhabitants in Philadelphia in 1790. Schweitzer finds that the population density gradient was very steep and population densities were significantly lower just a few city blocks away from the waterfront on the Delaware River. Schweitzer also finds that there was significant spatial differentiation by socioeconomic groups. Most merchants lived near the waterfront, artisans were distributed throughout the city, the wealthy were clustered in an area known as the New Society Hill, and the poor were found everywhere but disproportionately present in the outskirts of the city.

One of the most striking empirical regularities in urban economics is the consistent decline in density gradients of cities over time. Mills (1972) finds that density gradients of four metropolitan areas, Baltimore, Milwaukee, Philadelphia, and Rochester, declined more or less monotonically over time between 1880 and 1963.²³ Macauley (1985) estimates that, for a sample

²³ Mills' (1972) estimate of Philadelphia's density gradient for the late nineteenth century is

of 18 metropolitan areas, density gradients in population and employment for manufacturing, retailing, wholesaling and services all tended to decline over the twentieth century. Similar trends are found by Edmonston (1975) for a larger sample of metropolitan areas as well as for the same sample of cities as Mills.

The Alonso-Muth-Mills monocentric city model provides a number of possible explanations as to why urban population or employment might eventually decentralize or why the density gradient might fall. According to this model, households choose where to reside within an urban area, given their tastes for housing versus other goods. To make this concrete, imagine that households maximize a utility function $U(h, x, t)$ where h = housing services, x = a composite commodity, and t = leisure. The budget constraint is $p(r)h + x = A + w(T - cr - t)$, where $p(r)$ is the price of housing at distance r from the central business district (CBD), A is non-labor income, T is the household's time endowment, w is the wage per unit of time, and c is time spent commuting per unit of distance r from the CBD. Employment in this model is concentrated in the CBD but households may choose to live at various points r from the CBD. The model can readily accommodate fixed or variable monetary costs of commuting but for our purposes it is not necessary to include these.

In this model the variable r serves as the indicator of "suburbanization" because higher values of r correspond to distances further from the urban "core" (the CBD). The first order condition for r is $p'h = -wc$. This equation will hold at non-zero values of r because the slope of the pricing function, p' , is negative – housing price declines as distance r from the CBD increases. In choosing its optimal location, each household trades off a lower price of housing

considerably lower than that of the other three cities. However, Mills believes that the low estimate may have been due to problems with data construction associated with Philadelphia.

versus higher commuting costs.

If housing is a normal good – and all empirical studies suggest that it is – it is straightforward to demonstrate that $dr/dA > 0$. That is, an increase in non-labor income will cause the household to live further away from the CBD – to suburbanize. The demand for housing increases (because A has increased), and the household chooses optimally to locate further away where housing price is lower. However, an increase in w has off-setting income and substitution effects. The income effect of an increase in w is the same as the effect of an increase in A , as long as housing is a normal good. But the substitution effect goes in the opposite direction because a higher wage means a higher time cost of commuting. If the income effect outweighs the substitution effect, an increase in w will imply an increase in r , that is, a distance further from the CBD.

In the Alonso-Muth-Mills model, higher income households choose to suburbanize because they demand more housing relative to other goods, and the relative price of housing of land is cheaper at locations that are distant from the CBD than at locations that are close in. Of course, there may be other reasons why higher income households choose to live in the suburbs. It is possible, for example, that suburban local governments might be more efficient at providing local public goods, such as education, and these goods could be income-elastic. Models of jurisdictional “sorting” – the archetype being the Tiebout model – imply that households sort across local governments on the basis of their demands for local public goods. Such sorting may be reinforced by “peer group” effects – the quality of the local public schools, for example, may be affected by socioeconomic backgrounds of the students that attend them.

The existence of a density gradient for Philadelphia in 1790 presents a puzzle for the

monocentric city model. In that model, the density gradient is generated by the commuting choice of households. However, in this period, very few people commuted to work. Most people worked where they lived. Thus, it is more likely that the density gradient was caused by the fact that businesses found it desirable to locate near the waterfront on the Delaware River (see Fujita and Ogawa (1982), Berliant, Peng and Wang (2002), Lucas and Rossi-Hansberg (2002), and Rossi-Hansberg (2003) for possible explanations.)

The decline in the density gradient or the decentralization of employment and population, especially in the second half of the twentieth century, is often attributed to developments in transportation and rising incomes. Glaeser and Kahn (2004) argue that the adoption of automobiles and trucks fundamentally transformed how Americans live and work in the latter half of the twentieth century. Automobiles and trucks lowered transportation costs and, just as importantly, eliminated fixed costs associated with rail depots and ports. The decline in agglomeration economies associated with these fixed costs enabled firms and households to decentralize and organize their activities around a more dispersed environment.

Margo (1992) finds that rising incomes also played an important role in suburbanization. Margo uses the public use sample of the 1950 census, which provides information on household and personal characteristics, including whether households living in metropolitan areas reside in central cities or suburban rings. Margo estimates a cross-sectional logistic regression of the probability of suburban residence as a function of household characteristics, including household income, which is instrumented. The model is identified by assuming that education and industry of employment affect income, but do not otherwise (that is, independently) affect the likelihood of a suburban residence. The estimated coefficient on income in this regression is positive and

highly significant.

Using this coefficient in conjunction with the mean growth of real household income between 1950 and 1980, Margo predicts that the percent suburbanized should have increased by 7.4 percentage points. The actual increase was 17.5 percentage points (from 42.7 to 60.2 percent). Thus, according to Margo's calculation, changes in income account for about 43 percent of suburbanization over the 1950 to 1980 period. It is likely that this estimate is biased upwards, for several reasons. First, education and industry may not be proper instruments. For example, education may have a direct impact on the likelihood of a suburban residence, if better-educated parents value their children's schooling, and suburban schools, as most studies indicate, are superior in quality to urban schools. Similarly, industry may have a direct effect on the likelihood of suburban residence, if industries differ in their ability to substitute land for other inputs. As a result, some industries locate further away from the CBD than others, and their workers follow suit.

Second, as Margo (1992, p. 308) notes, some central cities expanded their boundaries by annexing suburban areas after 1950. Had jurisdictional boundaries remained constant between 1950 and 1980, the increase in the proportion suburbanizing after 1950 would have greater than 17.5 percentage points. Third, Margo's calculation is partial equilibrium. In particular, it does not take into account induced effects on price of housing in the suburbs relative to central cities. These effects are positive – house values should (and did, see Collins and Margo 2002) rise in the suburbs relative to the central cities, dampening the incentive to suburbanize due to rising incomes.

It is important to recognize that Margo's calculation does not identify precisely why

higher income households suburbanize. This could be because of the causal mechanism identified by the Alonso-Muth-Mills model, or it could be for some other reason. Glaeser, Kahn, and Vigdor (2001) argue, in fact, that the reason emphasized in the Alonso-Mills model – land prices – is not the reason why “[t]he poor live in cities.” Rich people, Glaeser, et. al, agree, do have a higher demand for land than poor people, but the income elasticity of the demand for land is not high enough to explain the positive association between income and suburban residence. Poor people, their argument continues, live in cities because the poor place a high value on public transportation. In Glaeser, et.al.’s modification of the Alonso-Mill’s model, the marginal monetary cost of public transportation is cheap, but the fixed and time costs are large relative to automobile transport.

B. Urban Densities

While the density gradient and other measures of decentralization are informative, we believe that a study of simple average density of cities provides a fuller picture of urban spatial structures. Shammass (2000) finds that colonial cities were compact and dense. In 1800, data on sample of four cities (Philadelphia, New York, Baltimore and Boston) indicate that each city consisted of less than 2 square miles of area. Population densities of these cities ranged from Boston’s 20,781 to Philadelphia’s 45,800 persons per square mile. The differences in population densities in these cities reflected their differences in housing lot sizes. It appears that Philadelphia’s high density relative to other cities was due to its relatively small house lot sizes. While Philadelphians averaged 7.1 compared to Boston’s 8.7 persons per dwelling, Philadelphia’s house lot sizes only averaged 1,392 square feet as compared to Boston’s 3,441.

Unfortunately, systematic information on urban land areas is available only from 1890

onwards with the publication of the *Social Statistics of Cities*. Kim (2002) finds that urban densities rose and fell between the late nineteenth and the twentieth centuries. Between 1890 and 1950, average population density rose from 7,230 to 8,876 persons per square mile for a consistent sample of cities whose population was greater than 25,000 (see Table 4 and Figure 6).²⁴ During this period, the cities also annexed considerable amounts of land. In 1890, the cities averaged approximately 19 square miles of land; by 1950, they averaged 40 square miles. Yet, despite the significant increases in the boundaries of cities, population density rose as urban population growth outpaced annexation. However, in the second half of the twentieth century, the average population density of cities declined substantially. By 1990, average population density fell to 5,647 persons per square mile. In this period, cities continued to annex nearby areas, but urban population growth did not keep pace with annexation. For metropolitan areas, Kim (2002) finds that average density rose between 1940 and 1960 but then fell sharply between 1960 and 1990.

The combined information from urban densities and density gradients presents a more coherent picture of the changes in urban spatial structures between the late nineteenth and the twentieth centuries. The average densities of urban areas rose and fell over time; however, density gradients of urban areas declined monotonically over time. These two trends can be reconciled accordingly. Between 1890 and 1950, the density gradient curve shifted upward but its slope fell, causing urban density to rise even as the density gradient declined. However, between 1950 and 1990, the density gradient curve shifted downward as its slope continued to fall, causing urban density to decline sharply.

²⁴ Also see Mills (1972) and Macauley (1985) for estimates of population density and gradients for a smaller subset of metropolitan areas.

By concentrating most of their empirical analysis on the density gradient, largely motivated by the monocentric city model, Kim (2002) argues that urban economists have underemphasized the location decisions of firms and its impact on urban spatial structures. Although falling transportation costs and rising incomes tended to disperse households outwards, Kim suggests that firm agglomeration economies in manufacturing and business services may have contributed to the rise in urban density during the first half of the twentieth century. Moreover, advances in skyscraper technologies lowered the costs of dense employment, especially for sectors that used office space. However, in the second half of the twentieth century, the decline in the importance of agglomeration economies in employment, re-enforced by household's preference for living in larger housing away from the city center, is likely to have contributed to the decline in average urban density as well as the continuing decline in the density gradient.²⁵

C. Urban Land Rents

With some notable exceptions, twentieth century evidence on land price implications of the monocentric city model is extremely limited. This is simply because, in most urban areas, vacant land parcels are relatively few in number and are vacant (rather than developed) for non-random reasons. Historical studies offer more promise, but until recently, most such studies have been based on a single data set, that for Chicago originally compiled by Homer Hoyt (1933); exceptions include Edel and Sclar's (1979) study of Boston land prices, Gin and Sonstelie's

²⁵ Kain (1968) argues that the relative and absolute decline of central cities is the result of several important technological changes. First, developments in transportation have made land much more substitutable. With trucks and automobiles, firms no longer need to locate near ports, freight and passenger terminals and rail lines. Second, changes in production methods that require spacious, single-story plants have made it more attractive for firms to build plants in outlying areas where land is cheaper. Finally, technological changes in communications and data processing have reduced the need for armies of white collar workers employed in central city offices. On the other hand, Chatterjee and Carlino (2001) argue that increases in congestion costs due to increases in population have led to the deconcentration of MSA employment in the second half of the twentieth century.

(1992) analysis of Philadelphia, and Smith's (2001) study of land values in Cleveland).

Representative studies based on the Hoyt data include Mills (1969), McDonald and Bowman (1979), Kau and Sirmans (1979), Kau, Lee, and Sirmans (1986), and McMillan, Jarmin, and Thorsnes (1992).

A few recent studies, however, have examined archival data for cities other than Chicago. Using data culled from newspaper advertisements, Margo (1996) studies the relationship between rental price of housing and distance from the CBD in antebellum New York. Although Margo's study is not, strictly speaking, about the price of land, he nevertheless finds a relatively steep rental price gradient. Atack and Margo (1998) also use newspaper listings to study sale prices of vacant land in New York City over the period 1835 to 1900. Atack and Margo also find a steep negative gradient between the price of land and distance, but one which was flattening over time. The steep rental gradient is consistent with the high cost of intra-urban transport before the Civil War and the flattening of the vacant land price gradient is also consistent with what is known about postbellum technological improvements in urban transportation. Atack and Margo's other important finding is a marked increase in the value of vacant land in New York after the Civil War. The increase in value between 1860 and 1870 is well in excess of the general rise in the price level at the time, and the increase was sustained after the price level began to fall in the 1870s. This rise in land values coincides, of course, with the economic ascendance of New York after the Civil War as a premier financial and manufacturing center.

3.5 *Size Distribution of Cities*

From the earliest period in American history, cities varied considerably in their populations at any given point in time. Madden (1956) finds that the size distributions of cities

for every decade between 1790 and 1950 were remarkably stable and that they can be usefully summarized by the rank-size rule. Dobkins and Ioannides (2000) and Black and Henderson (2002) provide estimates for the Pareto distribution or the rank-size rule for metropolitan areas between 1900 and 1990. We refer the reader to Gabaix and Ioannides (2004) for an extremely useful summary of the literature on the size distribution of cities.

While there is considerable debate as to what forces are responsible for the size distribution of cities, most economic historians have emphasized the role of markets and transportation nodes.²⁶ Weiman (1988) argues that the initial growth of Atlanta was due to its strategic location at the intersection of three railroads, but that its emergence as a regional metropolis was accounted for its development as a mercantile and financial center that cleared regional transactions. Odell and Weiman (1998) suggest that the founding of the Federal Reserve Banks in Dallas and Atlanta was due to their positions as regional centers of commerce and finance in the lower South.

Why do cities lower the costs of trade? Every trade must overcome two types of costs: transportation and information. Since goods are traded over space, they must incur transportation costs. In addition, specialization reduces the information set of agents. As compared to a more self-sufficient economy, information on demand and supply are geographically more dispersed. Specialization also exacerbates the problems of asymmetric information where sellers are more informed about the quality of their product than the buyers. The concentration of merchants, exchanges, insurance firms, lawyers, judges, etc. in larger cities lowers the costs of trade by providing better market information on supply and demand for goods and by solving the

²⁶ See Cain (1985).

asymmetric information problem.²⁷ Kim (2000) finds that the size distribution of cities is partly explained by the concentration of transaction services in the larger cities.

William Cronon's (1991) *Nature's Metropolis* provides an excellent illustration on the role of markets in causing the rise of Chicago. While Cronon's work is loosely based on Christaller's central place theory, scholars find his distinction between "first nature" (natural landscape) and "second nature" (human actions) particularly useful. Cronon argues that the growth of Chicago was due as much to "second nature" than to "first nature."²⁸ Although Cronon's definition of "second nature" is rather all encompassing, many urban scholars like Krugman (1996) have interpreted "second nature" as indicating the importance of increasing returns that is often self-reinforcing. Cronon's various case studies of wheat, lumber, meat, capital and reapers provide us with a good sense of what he means by "second nature."

While each case study is somewhat idiosyncratic, the case of wheat is illustrative. There is no doubt that Chicago was endowed with a rich hinterland west of the Great Lakes for growing wheat. Without first nature, no wheat would flow to Chicago.²⁹ However, St. Louis, not

²⁷ Ogawa and Fujita (1980) and Imai (1982) provide models of agglomeration based on the exchange of information. In their model, firms own bits of information and due to its public goods nature, the benefits of face to face communications rise with the number of firms. In their model, communication is costly as the exchange of information requires organization and time (see Fujita and Thisse, 2000.)

²⁸ Cronon (1991, 56-57) writes: "A kind of 'second nature,' designed by people and 'improved' toward human ends, gradually emerged atop the original landscape that nature - 'first nature' - had created such an inconvenient jumble. Despite the subtly differing logic that lay behind each, the geography of second nature was in its own way as compelling as the geography of first nature, so boosters and others often forgot the distinction between them. Both seemed quite 'natural.' Nowhere was this more true than in the new artificial transportation technologies that changed the ways people and commodities moved back and forth between city and country... Second nature defined the corridors of commerce at least as much as first nature."

²⁹ Lumber illustrates the primacy of first nature. When the supply of white pines from its hinterland diminished, Chicago's wholesale market in lumber declined as well. In addition, unlike grain, Chicago's advantages in the lumber trade were linked to transporting timber via the Great Lakes rather than railroads.

Chicago, was the major destination of wheat when transportation was based on water. The “second nature” advantages of Chicago were tied to the coming of the railroads. The railroads and the steam-powered grain elevator fundamentally changed how wheat was transported. Under the water-based marketing system, wheat was transported as sacks of grain; under the railroad-based system, grain was shipped freely in carloads. The railroad, as compared to water transportation, enjoyed significant economies of scale and labor-saving benefits.³⁰

Yet, while the increase in productivity of transportation and storage may explain why grain was shipped to Chicago rather than to St. Louis, it does not explain why grain was shipped to these cities in the first place. Why did farmers not simply sell their products to merchants or consumers in the cities of final destination? Why did wheat move through Chicago or St. Louis? Cronon’s answer is “second nature.” “By imposing their own order and vocabulary on the world of first nature, the city’s traders invented a world of second nature in which they could buy and sell grain as commodity almost independently from grain as crop (Cronon, 1991, 146).” Once quality was assured with the adoption of a standard grading system by the Chicago Board of Trade, buyers and sellers across the country could trade grain without sampling it in advance. In addition, the concentration of grain exchanges in Chicago meant that the posted trading prices provided instant information on grain market conditions of the entire nation.

3.6 *Urbanization and Inventive Activity*

The industrialization of the U.S. economy was accompanied by a significant increase in

³⁰ Cronon (1991, 113-114) writes: “The economic benefits of such efficient handling were so great that moving a bushel of grain from railroad car to lake vessel cost only half a cent, giving Chicago a more than tenfold advantage over St. Louis.” According to Cronon, St. Louis did not construct grain elevators until the end of the Civil War because, unlike the railroad owners, river interests faced high coordination costs and because the changing height of the Mississippi River made it impractical to build one on the levee.

inventive activity. Sokoloff (1988) finds that the expansion of markets brought on by canals in the early nineteenth century significantly increased inventive activity. Sokoloff and Kahn (1990) show that the growth of inventive activity in this period was due to inventions of ordinary citizens with common skills and knowledge. However, Kahn and Sokoloff and Lamoreaux and Sokoloff (1999) find that, over time, more and more inventors became career specialists who produced patents that were sold or leased. Thus, unlike during the second half of the twentieth century when research and development was mostly done by firms in-house, Lamoreaux and Sokoloff find that most patented inventions in the late nineteenth and the early twentieth centuries were traded in the marketplace.

Inventive activity was concentrated in urban areas. In the early nineteenth century, Pred (1973) finds that the three largest cities, New York, Philadelphia, and Boston were responsible for over 20% of all patents granted (also see Feller (1971)). However, while Sokoloff (1988) also finds that per capita patenting rates were significantly higher in urban areas during this period, he also notes that patenting was concentrated in the Southern New England and New York regions, and that both urban and rural places in these regions had disproportionately higher rates of patenting than elsewhere. In the late nineteenth and the early twentieth centuries, Lamoreaux and Sokoloff (1999) show that patents were concentrated in the New England, Middle Atlantic and the East North Central regions and that per capita patenting rates were positively correlated with city size.

The concentration of inventive and innovative activities in urban areas is often interpreted as providing evidence for the importance of information spillovers in generating new ideas. Lamoreaux and Sokoloff (1999) write that the concentration of patent agents and attorneys and

the availability of a wide array of publications such as the Scientific American promoted the diffusion of technological knowledge in the Northeast and in urban centers. However, an alternative explanation is also possible. Patent agents and patent attorneys were intermediaries who facilitated trade in inventions, and like other intermediaries, it is not surprising that they were concentrated in urban areas. However, why should inventors reside in cities if not for information spillovers? Lamoreaux and Sokoloff note that it was not a simple matter for inventors to obtain property rights to their inventions. Potential ideas had to be shepherded through the Patent Office application process and then protected in court in cases of infringement. Thus, inventors may have chosen to locate near patent agents and lawyers because access to these intermediaries lowered the costs of securing property rights to their inventions as well as lowering the transactions costs of marketing their inventions.

3.7 Urbanization and Productivity

Direct historical information linking urban locations to productivity are minimal. In his study of economies of scale in ante-bellum manufacturing, Sokoloff (1984) finds that firms located in more urbanized counties had higher levels of total factor productivity. Using a sample of manufacturing firms from the 1880 census, Atack, Bateman and Margo (2003) find that urban establishments were more productive, all other factors held constant, than rural establishments.

There is a considerable body of evidence showing that, throughout most of U.S. history, urban wages were higher - sometimes markedly so - in nominal terms than rural wages. Sokoloff and Villaflor (1992) find that in the New England and Middle Atlantic regions, wages were higher in urban than rural areas for various years between 1820 and 1860. However, the urban-rural wage gap declined over this period. In 1820 and 1842, nominal wages were 15-20% higher

in urban rather than in rural counties. In 1850, the urban-rural wage gap declined to 10% and, by 1860, the gap was insignificant. Furthermore, Goldin and Sokoloff (1984) find that in the early nineteenth century male wages were positively correlated with urbanization but that female wages were negatively correlated. It appears that female workers who were employed in large textile mills near waterpower sites earned relatively higher wages. For the late nineteenth century, Atack, Bateman and Margo (2000) find that nominal urban wages, based on data from samples from the decennial census manuscripts of manufacturing, were significantly higher than non-urban areas in 1850 and 1880. Moreover, the data indicate that the urban-rural wage gap increased over this period.

For the twentieth century, using data on a sample of cities from every decade between 1900 and 1990, Dobkins and Ioannides (2000) find that population size was significantly correlated with nominal wages. More specifically, they find that after controlling for individual, time and regional effects, a 10% increase in population was associated with a 1% increase in wages. Glaeser and Mare (2001) find that wages in large cities are 33% higher than outside metropolitan areas in 1990.

While economic historians have documented the urban-rural wage gap, few have attempted to explain the causes of this gap for the nineteenth and the early twentieth centuries. While most scholars believe that the urban wage premium is due to the higher productivity of urban workers, it is important to establish the sources of these productivity gains. It is also important to rule out the alternative hypothesis that the urban-rural wage gap is simply caused by the fact the more able workers choose to live in cities. In the early nineteenth century, Sokoloff (1984) finds that employees in urban firms may have been more productive because larger

markets led to greater division of labor. In the twentieth century, Glaeser and Mare (2001) suggest that urban workers were more productive because cities enabled workers to accumulate more human capital.

3.8 *Urbanization and Health*

The industrialization and urbanization of the U.S. economy between the early nineteenth and turn of the twentieth centuries were associated with significant increases in productivity and real wages. Yet, despite the fact that income per capita rose over this period, the average quality of health for Americans declined markedly over this period. The heights of native-born males, as well as life expectancy, fell between 1830 and 1890 and recovered to 1830 levels only by 1930.³¹ Since health is an important component of human welfare, Costa and Steckel (1995) estimate that the overall welfare of Americans may have stagnated or fallen between 1830 and 1890.

Economic historians suspect that much of the decline in health was associated with urbanization. Unfortunately, the U.S. economy industrialized and urbanized prior to the invention of modern medicine and public health. People living in high density areas were subject to greater exposure to infectious and parasitic diseases, both respiratory (air-borne) and gastrointestinal (water-borne). Consequently, the health of urban Americans was much worse than their rural counterparts. In 1900, Haines (2001) finds that life expectancy of urban Americans was 10 years less than rural Americans. In 1880, Higgs (1979) finds that the urban mortality rate was 50% higher than rural mortality; in 1890 and 1900, Condran and Crimmins (1980) finds that the urban-rural differential was 27% and 18% respectively. The urban-rural mortality differential was even more significant for infants and young children. For infants, excess urban mortality was 63% and 49% in 1890 and 1900, respectively and for young children

³¹ See Fogel (1986) and Costa and Steckel (1995).

aged one to four, the figures were 107% and 97%, respectively. Furthermore, Costa (2000) finds that exposure to diseases at early ages may have had long-term health consequences. Costa finds that Union Army veterans who grew up in large cities faced much higher mortality rates at older ages.

The urban-rural differentials in health and mortality began to decline and reverse over the twentieth century. By around 1940, it was sometimes healthier to reside in urban rather than in rural areas. Scholars attribute the decline and elimination of excess urban mortality to improved public works such as sewers and water systems and advances in public health and medicine (Haines (2001)). However, the transition to a healthier urban environment did not occur without complications. In 1897, half of all municipalities installed lead pipes to deliver water. Troesken (2003) finds that municipalities which used lead water pipe systems experienced higher rates of infant mortality and stillbirths by 25% and 50% respectively.

3.9 Immigration and Urbanization

Since the early nineteenth century, immigrants have been concentrated in cities, especially in the North (Ward (1972)). By 1890, Easterlin (1972) shows that 53% of the urban population was foreign born and the presence of foreign born seems to be positively correlated with city sizes (see Thompson and Whelpton (1933)). In contrast, 75% of native whites lived in rural areas in 1890. Moreover, within cities, immigrants tended to live in close geographic proximity, forming distinctive ethnic neighborhoods, especially in the larger cities.

Why are the foreign born more attracted to urban areas than natives? Why do immigrants form ethnic enclaves? Borjas (1995) argues that ethnic neighborhoods in cities arise because of ethnic specific externalities that are captured by residents who live in close proximity. While the

process by which ethnic externalities are transmitted is not well understood, Borjas suggests that ethnic spillovers foster the accumulation of human capital that raises the productivity of workers.

4. Conclusion

The most important lesson that emerges from this review of U.S. economic geography from a historical perspective is that economic geography is fundamentally linked to the process of economic development. While no overwhelming consensus emerges on the causal relationship between economic geography and development, the review highlights the enormous advances made on this subject by economic historians in recent decades.

From a regional perspective, while there is continued debate on the exact causes, it is now widely believed that the westward migration was caused by higher productivity of labor in the West. Moreover, the work on early industrialization and slavery enhance our understanding of when and why U.S. regions diverged and converged. It appears that regional divergence began in the early nineteenth century as the Northeast began to industrialize. Yet, in the antebellum period, due to the relatively high productivity of Southern agriculture under slavery, the divergence was not fully reflected in the regional income per capita. In the post-bellum period, regional divergence was felt even more keenly due to the elimination of slavery in the South. However, despite the important impact of slavery on regional incomes, the root cause of the divergence and convergence of U.S. regional economies is likely to be related to the development of an industrial and post-industrial society.

From an urban perspective, the review indicates that cities have come full circle. In the colonial period, as is today, urban activities are largely based on services. However, the rise of an urban society was due to the rise of industrial cities. Urbanization was significantly correlated

with industrialization over time and across space. The number and size of cities rose with industrialization and cities were geographically concentrated in areas where industrialization began and spread. The spatial structures of cities changed markedly over time. Scholars find that the changes in land-use patterns are explained by falling commuting costs and rising incomes. We also have a much better understanding of what factors differentiate urban from rural areas. Inventive activity, productivity, nominal wages, immigrants, government, and entertainment all favor urban areas. However, urbanization was not always associated with positive developments. Because urbanization occurred prior to advances in modern medicine, urbanization was associated with a significant decline in the health of Americans.

The review also highlights a number of interesting puzzles and many important areas that remain relatively unexplored. In the colonial period, even prior to industrialization, why were cities and towns more abundant in New England and relatively absent in the Chesapeake? Why is there a rent gradient in 1790 Philadelphia when people did not commute to work? If industrialization and urbanization were associated with a significant deterioration of health, why did rural Americans migrate to urban areas in significant numbers? What are the causal relationships between industrialization and urbanization? If industrialization preceded urbanization, what kinds of agglomeration economies in manufacturing led to urbanization? In later industrialization, what forces led to the rise of the manufacturing belt and why did its importance decline in the twentieth century? Additionally, are cities, as argued by Glaeser, Kolko and Saiz (2001), becoming consumer cities?

Finally, we point to two neglected areas of research. First, few works have systematically examined the relationship between demography and economic geography (see Costa and Kahn

(2000)). The long-run changes in population, fertility, mortality, life expectancy, family size and income are likely to have major impact on how Americans organize their spatial environment. Second, the role of political economy on the location of economic activities is not well understood. Recent work by Holmes (1998) suggests that state policies have an important impact on the location of industries. However, we know very little how the long-run changes in the political economy of the federal, state and local governments influenced the evolution of the economic geography of the United States.

References

- Ades, Alberto and Edward Glaeser. "Evidence on Growth, Increasing Returns, and the Extent of the Market." *Quarterly Journal of Economics* (1999): 1025-1045.
- Albion, Robert. *The Rise of New York Port, 1815-1860*. Charles Scribner's Sons, 1939.
- Alexanderson, G., *The Industrial Structure of American Cities*. University of Nebraska Press, 1956.
- Atack, Jeremy and Peter Passell. *A New Economic View of American History*. 2nd Ed. W.W. Norton & Co., 1994.
- Atack, Jeremy and Robert Margo. "'Location, Location, Location!' The Price Gradient for Vacant Urban Land: New York, 1835 to 1900." *Journal of Real Estate Finance and Economics* 16 (1998): 151-172.
- Atack, Jeremy, Fred Bateman and Robert Margo. "Rising Wage Dispersion Across American Manufacturing Establishments, 1850-1880." NBER working paper #7932, 2000.
- Atack, Jeremy, Fred Bateman and Robert Margo. "Productivity in Manufacturing and the Length of the Working Day: Evidence from the 1880 Census of Manufactures." *Explorations in Economic History*, 2003, forthcoming.
- Barro, Robert and Xavier Sala-I-Martin. "Convergence across States and Regions." *Brookings Papers in Economic Activities*, no. 1 (1991): 107-182.
- Beeson, Patricia, David DeJong, and Werner Troesken. "Population Growth in U.S. Counties, 1840-1990." *Regional Science and Urban Economics* 31 (2001): 669-699.
- Berry, Brian. *Geography of Market Centers and Retail Distribution*. Prentice-Hall, 1968.
- Berliant, Marcus, Shin-Kun Peng, and Ping Wang. "Production Externalities and Urban Configurations." *Journal of Economic Theory*, 2002, forthcoming.
- Blanchard, Olivier and Lawrence Katz. "Regional Evolutions." *Brookings Papers on Economic Activity* 1 (1992): 1-75.
- Bodenhorn, Howard and Hugh Rockoff. "Regional Interest Rates in Antebellum America." in *Strategic Factors in Nineteenth Century American Economic Growth*, edited by Claudia Goldin and Hugh Rockoff, Cambridge University Press, 1992.
- Bogue, Donald. *The Population of the United States*. The Free Press, 1985.
- Borjas, George. "Ethnicity, Neighborhoods, and Human-Capital Externalities," *American Economic Review* (1995): 365-390.
- Breen, T.H. *Tobacco Culture*. Princeton University Press, 1985.
- Bridenbaugh, Carl. *Cities in the Wilderness*. The Ronald Press, 1938.
- _____. *Cities in Revolt*. Alfred Knopf, 1955.
- Burrows, Edwin and Mike Wallace. *Gotham: A History of New York City to 1898*. Oxford University Press, 1999.
- Cain, Louis, "William Dean's Theory of Urban Growth: Chicago's Commerce and Industry, 1854-1871," *Journal of Economic History*, 45 (1985).
- _____. "Historical Perspectives on Infrastructure and U.S. Economic Development," *Regional Science and Urban Economics* 27 (1997): 117-138.
- Carville, Earle. *Geographical Inquiry and American Historical Problems*. Stanford University Press, 1992.
- Caselli, Francesco and Wilbur J. Coleman. "The U.S. Structural Transformation and Regional Convergence: A Reinterpretation." *Journal of Political Economy* 109 (2001): 584-616.

- Chatterjee, Satyajit and Gerald A. Carlino. "Aggregate Metropolitan Employment Growth and the Deconcentration of Metropolitan Employment." *Journal of Monetary Economics* 48 (2001): 549-583.
- Christaller, W., *The Central Places of Southern Germany*. Prentice-Hall, 1966.
- Coehlo, Phillip and James Shepherd. "Regional Differences in Real Wages: The United States, 1851-1880." *Explorations in Economic History* 13 (1976): 203-230.
- Coffman, Chad and Mary E. Gregson. "Railroad Development and Land Values." *The Journal of Real Estate Finance and Economics* 16 (1998): 191-204.
- Condran, Gretchen and Eileen Crimmins. "Mortality Differentials between Rural and Urban Areas of States in Northeastern United States, 1890-1900," *Journal of Historical Geography* 6 (1980): 179-202.
- Costa, Dora. "Understanding Mid-life and Older Age Mortality Declines: Evidence From Union Army Veterans," NBER Working Paper #8000, 2000.
- Costa, Dora and Matthew Kahn. "Power Couples: Changes in the Locational Choice of the College Educated 1940-1990," *Quarterly Journal of Economics* 115 (2000): 1287-1315.
- Costa, Dora and Richard Steckel. "Long-Term Trends in Health, Welfare, and Economic Growth in the United States." NBER Historical Paper #76, 1995.
- Craig Lee A., Raymond B. Palmquist, and Thomas Weiss. "Transportation Improvements and Land Values in the Antebellum United States: A Hedonic Approach." *The Journal of Real Estate Finance and Economics* 16 (1998): 173-189.
- Cronon, William. *Nature's Metropolis*. W.W. Norton & Co., 1991.
- Crowther, Simeon. "Urban Growth in the Middle Atlantic States, 1785-1850," *Journal of Urban Economics* 36 (1976): 624-643.
- Cutler, David, Edward Glaeser, and Jacob Vigdor. "The Rise and Decline of the American Ghetto," NBER working paper #5881 (1997).
- Davis, Lance. "The Investment Market, 1870-1914: The Emergence of a National Market," *Journal of Economic History* 25 (1965): 355-393.
- Dobkins, Linda and Yannis Ioannides. "Dynamic Evolution of the Size Distribution of U.S. Cities." In *Economics of Cities* edited by Jean-Marie Huriot and Jacques-François Thisse. Cambridge University Press, 2000.
- Doerflinger, Thomas. *A Vigorous Spirit of Enterprise: Merchants and Economic Development in Revolutionary Philadelphia*. University of North Carolina Press, 1986.
- Duranton, Gilles and Diego Puga. "Micro-foundations of Urban Agglomeration Economies," in *Handbook of Regional and Urban Economics*, Vol. 4, edited by Vernon Henderson and Jacques-François Thisse, North-Holland, 2004.
- Earle, Carville. *The Evolution of a Tidewater Settlement System*. University of Chicago Department of Geography Research Paper No. 170, 1975.
- _____. *Geographical Inquiry and American Historical Problems*. Stanford University Press, 1992.
- Easterlin, Richard. "Interregional Differences in Per Capita Income, Population, and Total Income, 1840-1950." In *Trends in the American Economy in the Nineteenth Century*, edited by William Parker. Princeton University Press, 1960.
- _____. "Regional Income Trends, 1840-1950." In *American Economic History*, edited by Seymour Harris. McGraw-Hill, 1961.
- _____. "Population," in *American Economic Growth*. edited by Lance Davis et. al., Harper and

- Row, 1972.
- Edmonston, Barry. *Population Distribution in American Cities*. Lexington Books, 1975.
- Ellison, Glenn and Edward Glaeser. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy* (1997).
- Feller, Irwin. "The Urban Location of United States Invention, 1860-1910." *Explorations in Economic History* 8 (1971): 285-304
- Ferrie, Joseph. "Migration to the Frontier in Mid-Nineteenth Century America: A Re-Examination of Turner's 'Safety-Valve'." Department of Economics, Northwestern University Manuscript, 1997.
- Fishlow, Albert. "Internal Transportation in the Nineteenth and Early Twentieth Centuries," in *Cambridge Economic History of the United States*, Vol. II, edited by Stanley Engerman and Robert Gallman. Cambridge University Press, 2000.
- Fogel, Robert. *Railroads and American Economic Growth*. Johns Hopkins University Press, 1964.
- _____. "Nutrition and the Decline in Mortality since 1700: Some Preliminary Findings." In *Long-Term Factors in American Economic Growth* edited by Stanley Engerman and Robert Gallman. University of Chicago Press, 1986.
- _____. *Without Consent or Contract*. W.W. Norton, 1989.
- Fogel, Robert and Stanley Engerman. *Time on the Cross*. Little, Brown, 1974.
- Fujita, Masahisa. *Urban Economic Theory*. Cambridge University Press, 1989.
- Fujita, Masahisa and H. Ogawa. "Multiple Equilibria and Structural Transition of Non-monocentric Urban Configurations." *Regional Science and Urban Economics* 12 (1982): 161-196.
- Fujita, Masahisa and Jacques-François Thisse. *Economics of Agglomeration*. Cambridge University Press, 2002.
- Fujita, Masahisa, Paul Krugman, and Anthony Venables. *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press, 1999.
- Fujita, Masahisa and Tomoya Mori. "The Role of Port in the Making of Major Cities: Self-Organization and Hub-effects." *Journal of Development Economics* 49 (1996): 93-120.
- Gabaix, Xavier and Yannis Ioannides. "Evolution of City Size Distribution," in *Handbook of Regional and Urban Economics*, Vol. 4, edited by Vernon Henderson and Jacques-Francois Thisse, North-Holland, 2004.
- Galenson, David W. and Clayne Pope. "Precedence and Wealth: Evidence from Nineteenth-Century Utah." In C. Goldin and H. Rockoff, eds. *Strategic Factors in Nineteenth Century American Economic History*. University of Chicago Press, 1992.
- Galloway, Lowell and Richard Vedder. "The Increasing Urbanization Thesis: Did 'New Immigrants' to the United States have a Particular Fondness for Urban Life?" *Explorations in Economic History* 8 (1971): 305-320.
- Glaeser, Edward, Kallal, J. Scheinkman and A. Shleifer. "Growth in Cities." *Journal of Political Economy* (1992).
- Glaeser, Edward and David Maré. "Cities and Skill." *Journal of Labor Economics* 19 (2001): 316-342.
- Glaeser, Edward, Jed Kolko, and Albert Saiz. "Consumer City," *Journal of Economic Geography* 1 (2001): 27-50.
- Glaeser, Edward and Matthew Kahn. "Sprawl and Urban Growth," in *Handbook of Regional and*

- Urban Economics*, Vol. 4, edited by Vernon Henderson and Jacques-Francois Thisse, North-Holland, 2004.
- Goldin, Claudia and Kenneth Sokoloff. "The Relative Productivity Hypothesis of Industrialization: The American Case, 1820-1850." *Quarterly Journal of Economics* 99 (1984): 461-488.
- Goldin, Claudia and Robert A. Margo. "Wages, Prices, and Labor Markets Before the Civil War," in *Strategic Factors in Nineteenth Century American Economic Growth*, edited by Claudia Goldin and Hugh Rockoff, Cambridge University Press, 1992.
- Haines, Michael. "The Population of the United States, 1790-1920," in *Cambridge Economic History of the United States*, Vol. II, edited by Stanley Engerman and Robert Gallman. Cambridge University Press, 2000.
- Haines, Michael. "The Urban Mortality Transition in the United States, 1800-1940." NBER Historical Paper #134, 2001.
- Harley, C. Knick. "Transportation, the World Wheat Trade, and the Kuznets Cycle 1850-1913," *Explorations in Economic History* 17 (1980): 218-250.
- Helpman, Elhanan. "The Size of Regions," in *Topics in Public Economics*, edited by D. Pines, E. Sadka and I. Zilcha, Cambridge University Press, 1998.
- Henderson, J. Vernon. *Urban Development*. Oxford University Press, 1988.
- Higgs, Robert. "Cycles and Trends of Mortality in 18 Large American Cities, 1871-1900." *Explorations in Economic History*. 16 (1979): 381-408.
- Holmes, Thomas. "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders," *Journal of Political Economy* 106 (1998): 667-705.
- Holmes, Thomas and John Stevens. "Spatial Distribution of Economic Activities in North America," in *Handbook of Regional and Urban Economics*, Vol. 4, edited by Vernon Henderson and Jacques-Francois Thisse, North-Holland, 2004.
- Hunter, Louis. *History of Industrial Power in the United States, 1780-1930*, Volume 1: Water Power in the Century of the Steam Engine. University Press of Virginia, 1979.
- Imai, H. "CBD hypothesis and Economics of Agglomeration." *Journal of Economic Theory* 28 (1980): 275-299.
- Jones, Alice Hanson. *Wealth of a Nation to Be*. Columbia University Press, 1980.
- Kain, John. "The Distribution and Movement of Jobs and Industry," in *Metropolitan Enigma*, edited by James Wilson. Harvard University Press, 1968.
- Kau, James B. and C. F. Sirmans. "Urban Land Value Functions and the Price Elasticity of Demand for Housing." *Journal of Urban Economics* 6 (1979): 112-121.
- Kau, James B., C. F. Lee, and C. F. Sirmans. "Urban Econometrics." *Research in Urban Economics* 6 (1986): 388-406.
- Kim, Sukkoo. "Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U.S. Regional Manufacturing Structure, 1860-1987." *Quarterly Journal of Economics* (1995): 881-908.
- _____. "Economic Integration and Convergence: U.S. Regions, 1840-1987." *Journal of Economic History*, (1998): 659-683.
- _____. "Regions, Resources, and Economic Geography: Sources of U.S. Regional Comparative Advantage, 1880-1987." *Regional Science and Urban Economics* 29 (1999): 1-32.
- _____. "Urban Development in the United States." *Southern Economic Journal* 66 (2000): 855-880.

- _____, "The Reconstruction of the American Urban Landscape in the Twentieth Century." NBER working paper #8857 (2002).
- Klingaman, David. "The Significance of Grain in the Development of the Tobacco Colonies." *Journal of Economic History* 24 (1969): 268-278.
- Kolko, Jed. "Can I Get Some Service Here? Information Technology, Service Industries, and the Future of Cities," mimeo, 1999.
- Konishi, Hideo. "Hub Cities." *Journal of Urban Economics* (2000):
- Krugman, Paul. *Geography and Trade*. MIT Press, 1991.
- _____. *The Self Organizing Economy*. Blackwell Publishers, 1996.
- Krugman, Paul and Anthony Venables. "Globalization and the Inequality of Nations." *Quarterly Journal of Economics* 110 (1995): 857-880.
- Lamoreaux, Naomi and Kenneth Sokoloff. "Inventive Activity and the Market for Technology in the United States, 1840-1920." NBER working paper #7107, 1999.
- _____. "Intermediaries in the U.S. Market for Technology, 1870-1920." NBER working paper #9017, 2002.
- Lebergott, Stanley. *Manpower in Economic Growth*. McGraw-Hill, 1964.
- Lindstrom, Diane. *Economic Development in the Philadelphia Region, 1810-1850*. Columbia University Press, 1978.
- Lucas, Robert and Esteban Rossi-Hansberg. "On the Internal Structure of Cities." mimeo, 2002.
- Macauley, Molly. "Estimation and Recent Behavior of Urban Population and Employment Density Gradients." *Journal of Urban Economics* 18 (1985): 251-260.
- Madden, C., "On Some Indicators of Stability in the Growth of Cities in the United States," *Economic Development and Cultural Change* (1956).
- Majewski, John. *A House Dividing*. Cambridge University Press, 2000.
- Margo, Robert A. *Race and Schooling in the South, 1880-1950: An Economic History*. University of Chicago Press, 1990.
- _____. "The South as an Economic Problem: Fact or Fiction?" In D. Doyle and L. Griffin, eds. *The South as an American Problem*. University of Georgia Press, 1975.
- _____. "Explaining the Postwar Suburbanization of Population in the United States: The Role of Income." *Journal of Urban Economics* 31 (1992): 301-310.
- _____. "The Rental Price of Housing in New York City, 1830-1860." *Journal of Economic History* 56 (1996): 605-625.
- _____. "Regional Wage Gaps and the Settlement of the Midwest." *Explorations in Economic History* 36 (1999): 128-143.
- _____. *Wages and Labor Markets in the United States, 1820-1860*. University of Chicago Press, 2000.
- _____. "The North-South Wage Gap: Before and After the Civil War." In D. Eltis, F. Lewis, and K. Sokoloff, eds. *Slavery in the Development of the Americas*. Cambridge University Press, 2003, forthcoming.
- Margo, Robert A. and Georgia Villaflor. "The Growth of Wages in Antebellum America: New Evidence," *Journal of Economic History* 47 (1987): 873-896.
- McDonald, John and H. Woods Bowman. "Land Value Functions: A Re-Evaluation." *Journal of Urban Economics* 6 (1979): 25-41.
- McCusker, John and Russel Menard. *The Economy of British America, 1606-1789*. University of North Carolina Press, 1985.

- McMillan, Daniel P., R. Jarmin, and P. Thorsnes. "Selection Bias and Land Development in the Monocentric Model." *Journal of Urban Economics* 22 (1992): 273-284.
- Meinig, D.W. *The Shaping of America*, Vol. 2. Yale University Press, 1993.
- Meyer, David. "Emergence of the American Manufacturing Belt: An Interpretation." *Journal of Historical Geography* 9 (1983): 145-174.
- _____. "Industrial Retardation of Southern Cities, 1860-1880," *Explorations in Economic History*, 25 (1988): 366-386.
- _____. "Midwestern Industrialization and the American Manufacturing Belt in the Nineteenth Century," *Journal of Economic History* 49 (1989): 921-937.
- Middleton, Arthur. *Tobacco Coast*. Johns Hopkins University Press, 1953.
- Mills, Edwin. "An Aggregation Model of Resource Allocation in a Metropolitan Area," *American Economic Review* (1967).
- _____. "The Value of Urban Land". In H. Perloff, ed. *The Quality of the Urban Environment*. Johns Hopkins Press, 1969, pp. 231-253.
- _____. *Studies in the Structure of the Urban Economy*. Johns Hopkins Press, 1972.
- _____. *Urban Economics*, 2nd Edition. Scott, Foresman and Co., 1980.
- Mitchener, Kris and Ian McLean. "U.S. Regional Growth and Convergence, 1880-1980," *Journal of Economic History* 59 (1999): 1016-1042.
- Mitchener, Kris and Ian McLean. "The Productivity of U.S. States Since 1880," NBER Working Paper #9445, 2003.
- North, Douglass. *The Economic Growth of the United States 1790-1860*. Prentice-Hall, 1961.
- Odell, Kerry and David Weiman. "Metropolitan Development, Regional Financial Centers, and the Founding of the Fed in the Lower South," *Journal of Economic History* 58 (1988): 103-125.
- Ogawa, H. And Masahisa Fujita. "Equilibrium Land Use Patterns in a Non-monocentric City." *Journal of Regional Science* 20 (1980): 455-475.
- Ottaviano, Gianmarco and Jacques-Francois Thisse. "Agglomeration and Economic Geography," in *Handbook of Regional and Urban Economics*, Vol. 4, edited by Vernon Henderson and Jacques-Francois Thisse, North-Holland, 2004.
- Pred, Allan. "Manufacturing in the American Mercantile City, 1800-1840," *Annals of the American Association of Geographers* 56 (1966): 307-325.
- _____. *Urban Growth and the Circulation of Information: The United States System of Cities, 1790-1840*. Harvard University Press, 1973.
- _____. *The Spatial Dynamics of U.S. Urban-Industrial Growth, 1800-1914*. MIT Press, 1977.
- Price, Jacob, "Economic Function of the Growth of American Port Towns in the Eighteenth Century." *Perspectives in American History* 8 (1974), 121-186.
- Price, Jacob and Paul Clemens. "A Revolution of Scale in Overseas Trade: British Firms in the Chesapeake Trade, 1675-1775." *Journal of Economic History* 47 (1987): 1-43.
- Prude, Jonathan. *The Coming of Industrial Order: Town and Factory Life in Rural Massachusetts 1810-1860*. Cambridge University Press, 1983.
- Puga, Diego. "The Rise and Fall of Regional Inequalities," *European Economic Review* 43 (1999): 303-334.
- Ransom, Roger and Richard Sutch. *One Kind of Freedom: The Economics of Emancipation*. Cambridge University Press, 1975.
- Rappaport, Jordan and Jeffrey Sachs. "The U.S. As a Coastal Nation," mimeo, 2001.

- Rosenberg, Nathan. "Technological Change in the Machine Tool Industry, 1840-1910," *Journal of Economic History* (1963).
- Rosenbloom, Joshua. "One Market or Many? Labor Market Integration in the Nineteenth Century United States," *Journal of Economic History* 50 (1990): 85-108.
- Rosenthal, Stuart and William Strange. "Evidence on the Nature and Sources of Agglomeration Economies," in *Handbook of Regional and Urban Economics*, Vol. 4, edited by Vernon Henderson and Jacques-Francois Thisse, North-Holland, 2004.
- Rossi-Hansberg, Esteban. "Optimal Urban Land Use and Zoning," mimeo, 2002.
- Schweitzer, Mary. "The Spatial Organization of Federalist Philadelphia, 1790." *Journal of Interdisciplinary History* 24 (1993): 31-57.
- Shammas, Carole. "The Space Problem in Early United States Cities," *William and Mary Quarterly* 57 (2000): 505-542.
- Shepherd, James and Gary Walton. *Shipping, Maritime Trade, and the Economic Development of North America*. Cambridge University Press, 1972.
- Simon, Curtis and Clark Nardinelli. "Human Capital and the Rise of American Cities, 1900-1990." *Regional Science and Urban Economics* (2002):
- Smith, Frederic Homer. *Real Estate Values and the Location of Economic Activity: A Case Study of the Urban Form of Cleveland, 1915-1960*. Ph.D. dissertation, Department of Economics, Vanderbilt University, 2001.
- Smith, Thomas R. *The Cotton Textile Industry of Fall River Massachusetts*. King's Crown Press, 1944.
- Sokoloff, Kenneth. "Was the Transition from the Artisanal Shop to the Non-mechanized Factory Associated with Gains in Efficiency? Evidence from the U.S. Manufactures Censuses of 1820 and 1850," *Explorations in Economic History* 21 (1984): 351-382.
- _____. "Productivity Growth in Manufacturing during Early Industrialization: Evidence from the American Northeast, 1820-1860," in *Long-Term Factors in American Economic Growth*, Studies in Income and Wealth, Vol. 51, NBER, edited by Stanley Engerman and Robert Gallman, University of Chicago Press, 1986.
- _____. "Inventive Activity in Early Industrial America: Evidence from Patent Records, 1790-1846." *Journal of Economic History* 48 (1988): 813-850.
- Sokoloff, Kenneth and Georgia Villaflor. "The Market for Manufacturing Workers during Early Industrialization: The American Northeast, 1820 to 1860," in *Strategic Factors in Nineteenth Century American Economic Growth*, edited by Claudia Goldin and Hugh Rockoff, Cambridge University Press, 1992.
- Sokoloff, Kenneth and Zorina Kahn. "The Democratization of Invention During Early Industrialization: Evidence from the United States, 1790-1846." *Journal of Economic History* 50 (1990): 363-378.
- Steckel, Richard. "The Economic Foundations of East-West Migration during the Nineteenth Century." *Explorations in Economic History* 20 (1983): 14-36.
- Tabuchi, T. "Agglomeration and Dispersion: A Synthesis of Alonso and Krugman," *Journal of Urban Economics* 44 (1998): 333-351.
- Thompson, Warren and P.K. Whelpton. *Population Trends in the United States*. McGraw-Hill Book Co., 1933.
- Troesken, Werner. "Lead Water Pipes and Infant Mortality in Turn-of-the-Century Massachusetts," NBER Working Paper #9549, 2003.

- Ward, David. *Cities and Immigrants*. Oxford University Press, 1971.
- Ware, Caroline. *The Early New England Cotton Manufacture*. Boston, 1931.
- Warner, Sam, *Streetcar Suburbs: The Process of Growth in Boston 1870-1900*. Harvard University Press, 1962.
- _____. *The Private City: Philadelphia in Three Periods of its Growth*. University of Pennsylvania, 1968.
- Weiman, David. "Urban Growth on the Periphery of the Antebellum Cotton Belt: Atlanta, 1847-1860," *Journal of Economic History* 48 (1988): 259-272.
- Weiss, Thomas. "The Industrial Distribution of the Urban and Rural Workforces: Estimates for the United States, 1870-1910." *Journal of Economic History* 32 (1972): 919-937.
- _____. "Urbanization and the Growth of the Service Workforce." *Explorations in Economic History* 8 (1971): 241-259.
- Williamson, Jeffrey. "Antebellum Urbanization in the American Northeast." *Journal of Economic History* (1965): 592-608.
- Williamson, Jeffrey and J. Swanson, "The Growth of Cities in the American Northeast, 1820-1870," *Explorations in Entrepreneurial History* (1966): 3-101.
- Wright, Gavin. *Old South, New South*. Basic Books, 1986.

Table 1

Per Capita Income by Region, 1840 and 1860 (in 1860 Prices)

	Total Population			Free Population	
	1840	1860	1880	1840	1860
National Average	\$96	\$128	\$173	\$109	\$144
North	109	141	205	110	142
Northeast	129	181	244	130	183
North Central	65	89	170	66	90
South	74	103	88	105	150
South Atlantic	66	84	78	96	124
East S. Central	69	89	88	92	124
West S. Central	151	184	104	238	274

Source: Fogel (1989)

Table 2

Number and Population of Cities in the United States, 1690-1990

City Size	Number of Cities									
	1690	1790	1820	1860	1880	1920	1940	1960	1960*	1990*
2500-5000	3	12	26	163	467	1255	1422	1777	-	-
5000-10000	1	7	22	136	249	715	965	1320	16434	16929
10000-25000		3	8	58	146	465	665	1142	978	1290
25000-50000		2	2	19	42	143	213	424	366	567
50000-100000			2	7	15	76	107	203	180	309
100000-250000			1	6	12	43	55	80	79	131
250000-500000				1	4	13	23	29	30	40
500000-1,000,000				2	3	9	9	16	16	15
1,000,000 +					1	3	5	5	5	8
Number	4	24	61	392	939	2722	3464	4996	18088	19289
City Size	Distribution of Population (Percent)									
	1690	1790	1820	1860	1880	1920	1940	1960	1960*	1990*
2500-5000	5.0%	1.1%	1.0%	1.9%	3.2%	4.1%	3.8%	3.5%	-	-
5000-10000	3.3	1.2	1.6	3.1	3.4	4.7	5.1	5.2	13.9	11.3
10000-25000	0	1.2	1.3	2.8	4.4	6.7	7.6	9.9	8.4	8.2
25000-50000	0	1.6	0.7	2.1	2.9	4.8	5.6	8.2	7.1	8.0
50000-100000	0	0	1.3	1.4	1.9	5.0	5.6	7.8	7.0	8.5
100000-250000	0	0	1.3	3.2	3.6	6.2	5.9	6.5	6.4	7.7
250000-500000	0	0	0	0.8	2.6	4.3	5.9	5.9	6.0	5.7
500000-1,000,000	0	0	0	4.4	3.8	5.9	4.9	6.2	6.2	4.1
1,000,000 +	0	0	0	0	2.4	9.6	12.1	9.8	9.8	8.0
Total Urban	8.3%	5.1%	7.2%	19.8%	28.2%	51.2%	56.5%	63.1%	64.7%	61.6%
(millions)										
Urban Population	0.02	0.2	0.7	6.2	14.1	54.2	74.4	112.5	116.0	153.1
U.S. Population	0.21	3.9	9.6	31.4	50.2	105.7	131.7	178.5	178.5	248.7

* Uses the new urban areas defined by the 1960 Census of Population.

Sources: Data for 1690 are from Bridenbaugh (1938) and the Historical Statistics of the United States. All

other data are from the Censuses of Population, 1960, Vol. 1, Characteristics of the Population, U.S. Summary, 1-14-15 and the Statistical Abstract, 1992.

Table 3

Number and Population of Metropolitan Areas in the United States, 1940-1990

	Number of Cities		
	1940	1960	1990
Size Categories			
50,000-100,000	15	24	26
100,000-250,000	69	91	143
250,000-500,000	30	52	75
500,000-1,000,000	13	31	45
1,000,000 +	11	24	46
Number	138	222	335
	Distribution of Population (Percent)		
	1940	1960	1990
Size Categories			
50,000-100,000	1.0%	1.1%	0.9%
100,000-250,000	16.9	8.2	9.1
250,000-500,000	8.3	9.7	10.6
500,000-1,000,000	6.9	12.0	13.1
1,000,000 +	26.2	34.3	43.9
Total Metropolitan	51.0%	65.3%	77.5%
(millions)			
Metropolitan Population	67.1	116.6	192.9
U.S. Population	131.7	178.5	248.7

Sources: *County Data Book*, 1947; *Statistical Abstract of the U.S.*, 1965; *Census of Population*, 1990.

Note: In general, the metropolitan area is a county or a group of contiguous counties (except in New England) which contains at least one central city of 50,000 inhabitants or more or "twin cities" with a combined population of at least 50,000. In New England, towns and cities are used in defining metropolitan areas.

Table 4

Population and Employment Densities of Cities, 1890-1990

	Number of Cities	Average Population (sq. miles)	Average Area	Average Population Density
1890	122	113835	16.7	7648
1900	160	123243	20.2	7377
1910*	184	148442	23.1	7176
1920*	252	145966	20.9	7597
1930*	310	152890	21.9	7366
1940	412	128051	19.2	6742
1950	481	128811	19.5	6536
1960	673	112400	22.8	5340
1970	835	104785	28.5	4673
1980	944	97756	32.8	3998
1990	1068	98108	34.9	3783

Note: The data, except for years 1910-1930, are for cities with population over 25,000. In 1890, two cities were omitted due to lack of data on land area. Cities in Alaska and Hawaii are excluded.

*Data for 1910-1930 are for cities with population over 30,000.

Sources: Social Statistics of Cities, 1890; Census of Population, 1900; Financial Statistics of Cities, 1910, 1920, 1930; County and City Data Book, 1949, 1952, 1962, 1972, 1982, 1988, 1994.

Figure 1
U.S. Population Distribution, 1790-1990
(Percent)

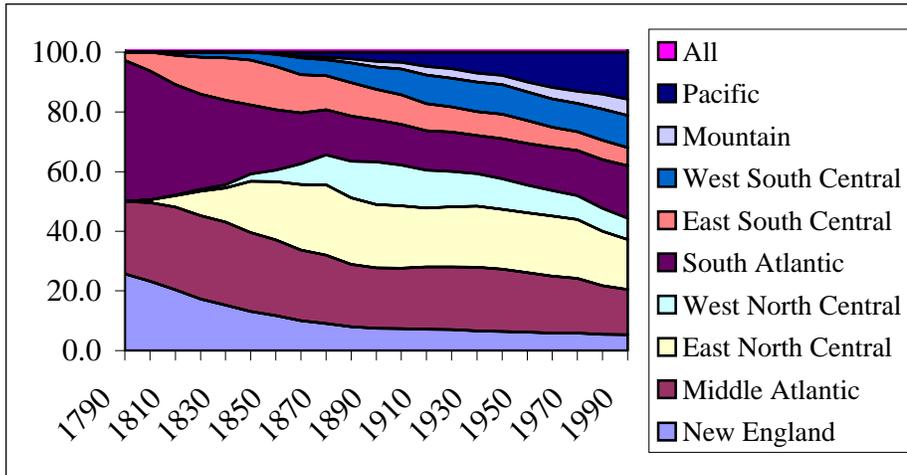
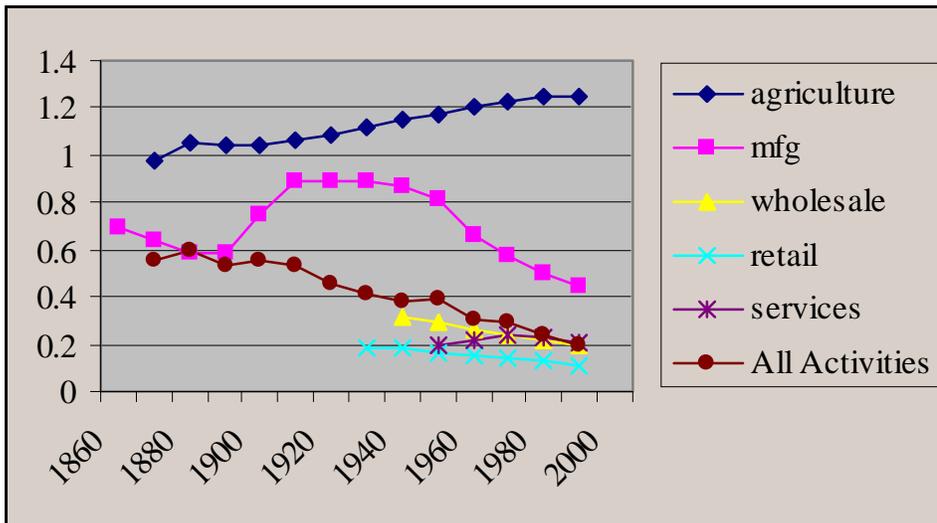
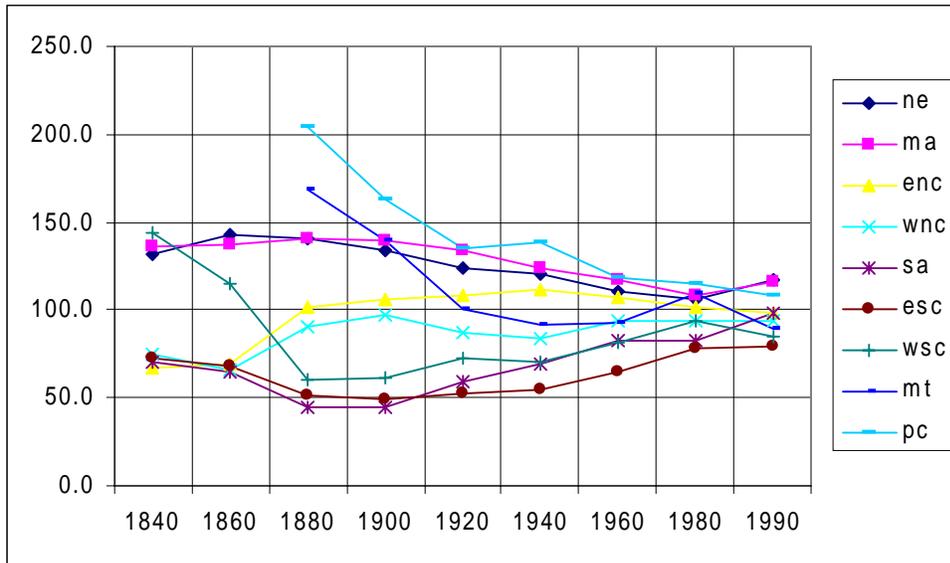


Figure 2
U.S. Regional Specialization, 1860-1990



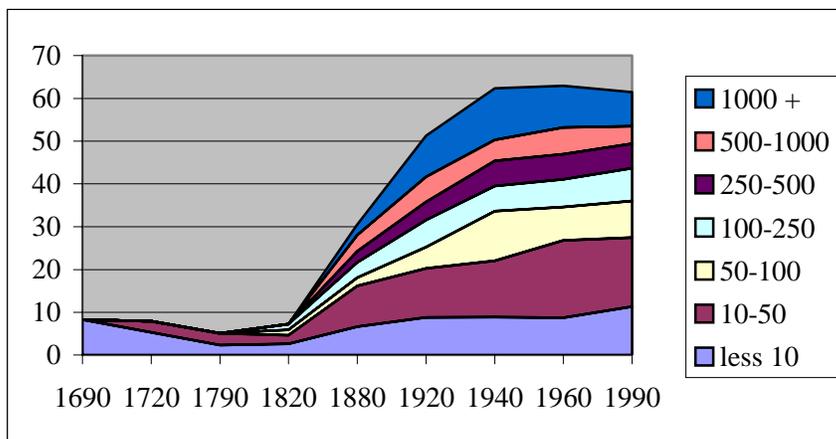
Source: Kim (1995, 1998). Index of regional specialization is based on Krugman (1991).

Figure 3
Regional Income Per Capita, 1840-1990
(U.S. = 100)



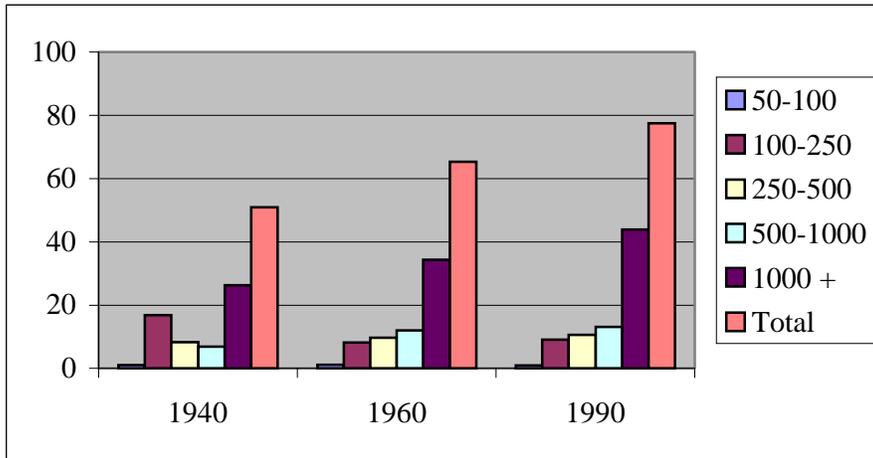
Sources: Easterlin (1960, 1961), Statistical Abstract of the U.S, various years.

Figure 4
U.S. Urban Population by Size, 1690-1990
(Percent)



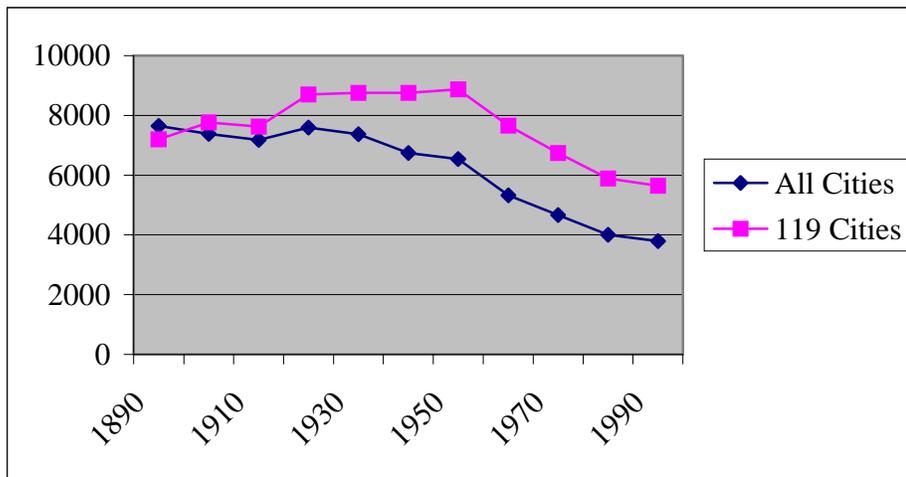
Sources: See table 2. City size is in thousands.

Figure 5
 U.S. Metropolitan Population by Size, 1940-1990
 (Percent)



Sources: See table 3. Metropolitan size is in thousands.

Figure 6
 Population Density of Cities, 1890-1990
 (Persons per square mile)



Sources: See Kim (2002)