Division of labor and the rise of cities: evidence from US industrialization, 1850–1880

Sukkoo Kim*

Abstract

Industrial revolution in the USA first took hold in rural New England as factories arose and grew in a handful of industries such as textiles and shoes. However, as factory scale economies rose and factory production techniques were adopted by an ever-growing number of industries, industrialization became concentrated in cities throughout the Northeastern region which came to be known as the manufacturing belt. While it is extremely difficult to rule out other types of agglomeration economies such as spillovers, this paper suggests that these geographic developments associated with industrial revolution in the USA are most consistent with explanations based on division of labor, job search, and matching costs.

1. Introduction

Economic historians and economists generally believe that modern economic growth was triggered by the process of industrial revolution. In the USA, industrial revolution occurred in two distinct phases between the nineteenth and the twentieth centuries. Between 1820 and 1840, early industrialization began in New England as manufacturing re-organized from artisanal shops to non-mechanized factories in a relatively small number of industries such as textiles, leather, and shoes. In the second phase of industrialization, which occurred between 1850 and 1920, factory production rose in scale, became more mechanized, and spread to numerous industries and to the North-eastern region known as the manufacturing belt.

One of the more distinctive differences between the two phases of industrialization was location. Whereas early industrialization began in rural locations, late industrialization was predominantly urban. While the number of urban places between 1820 and 1840 grew from 61 to 131, and the share of urban population grew from 7.2 to 10.8%, the dominant economic activities of the largest cities were mercantile and artisanal manufacturing rather than factory industrial production. Between 1860 and

* Department of Economics, Washington University in St Louis, Campus Box 1208, One Brookings Drive, St Louis, MO 63130-4899, USA
email <soks@artsci.wustl.edu>

1920, as agriculture declined and manufacturing became the dominant economic activity, the number of cities rose from 392 to 2722 and the share of urban population rose from 25.8 to 51.2%.[1] Thus, by the end of the industrial revolution, a majority of Americans lived in urban areas.

Why early factories arose in rural New England and then moved to cities throughout the Northeastern region as industrialization matured is still open to some debate. One of the more popular explanations is based on the shift in primary power sources for manufacturing. For Hunter (1985) and Rosenberg and Tratjtenberg (2004), early factories arose in rural New England because it possessed abundant water-power, but as the steam-engine released firms from the locational constraints of topography and climate, factories relocated to cities. For Goldin and Sokoloff (1984) and Kim (2005), however, the more likely explanation is based on the re-organization of production from artisans to factories. Goldin and Sokoloff argue that industrialization began in New England because early factories used the labor of women and children intensely and that the wages of women and children relative to men were lower in that region compared with those in other regions. Kim finds that factory organization, regardless of the type of primary power sources, rather than the adoption of steam power, was the most important reason why firms chose urban locations.[2]

Pred (1966) presents an alternative but influential explanation for the rise of industrial cities in the USA. According to Pred, early industrialization began in rural places between 1800 and 1840 because most of the population was simply rural.[3] Moreover, Pred argues that manufacturing did not locate in large commercial cities like New York, Baltimore, Philadelphia, and Boston during this period because of shortages of capital and labor as well as restricted access to markets.[4] However,

---

1 The US census defines a city as an area having a population greater than 2500. See Kim and Margo (2005).

2 Kim (2005) finds that while steam-powered firms were five to six times more likely to locate in cities than water-powered firms, they were 0.55 times less likely to locate in cities as compared with hand-powered firms. Calculations suggest that the shift in primary power sources from water to steam may have contributed at most 8–10% of the increase in the rate of urbanization between 1850 and 1880. Rather than the adoption of steam power, the shift from artisanal to factory organization of production seems to have been more important for urbanization. Factory workers who did not use inanimate power, as well as steam-powered and water-powered workers, were on average 2–3 times more likely to locate in cities than artisans. At any given point in time, factory production may have contributed to a 27% increase in the rate of urbanization. A decomposition of industry fixed effects show that resource-intensive industries were more likely to locate in rural areas whereas labor-intensive industries were more likely to locate in urban areas.

3 Pred (1966, p. 143):

> During the initial decades of the last century, American manufacturing was characterized predominantly by an emphasis on consumer rather than capital goods, by handicraft rather than machine techniques, by household rather than factory organization, and by rural dispersion rather than concentration in major urban centers. Even in the textile industries, where factories were the most important production units by the 1830's, activity was largely confined to rural waterfall sites and mill towns recently superimposed upon the rural landscape. The factory and industrial capitalism had not as yet become the cornerstones of metropolitan growth. In other words, at a time when the economy of the USA was agricultural, the industrial as well as the agrarian population was predominantly rural.

4 Pred (1966, p. 152): What factors militated against the location of additional manufacturing activities in the larger commercial cities? It is reasonable to contend that the limited dimensions of manufacturing in

between 1860 and 1914, changes in technology and the rapid decline in transportation costs with the introduction of the railroads, unleashed Marshallian agglomeration economies which led industries to locate in cities.[5]

Although it is extremely difficult to disentangle the causal linkages between industrialization and urbanization, this paper suggests that the most consistent explanation for why industrialization arose in rural areas and then shifted to urban areas is likely to be based on division of labor and labor-matching costs. Early industrialization began in rural New England because early factories made extensive use of women and child labor from rural farms. Because division of labor, though important, was relatively limited, the early industrial labor force was relatively homogenous. In addition, because early industrialization was confined to a handful of industries, labor heterogeneity across industries was also limited. As a result of limited division of labor in society, labor-matching costs remained low even in rural locations.

As industrial revolution deepened, however, division of labor and labor-matching costs rose significantly and manufacturing became concentrated in urban areas. The rise in scale economies at the firm level correlated with a growing division of labor between skilled, managerial workers who supervised an increasingly mechanized factory based on increasing sub-division of tasks that utilized relatively unskilled labor. Moreover, as the industrial method of production spread to hundreds of industries, the heterogeneity of the manufacturing labor force rose significantly. As labor-matching costs rose over the second half of the industrial revolution, workers and firms found it more advantageous to locate in cities.

To establish the existence of urban agglomeration economies during the second phase of the industrial revolution, this paper uses firm-level data constructed from the decadal manuscript censuses of manufactures between 1850 and 1880 to estimate whether there were urban wage and productivity premiums. Because a firm's decision to locate in an urban area is endogenous, the paper uses several instruments to correct for this potential endogeneity bias. The instruments used were lagged urban variable, access to water transportation, date of incorporation, and the distance from the eastern seaboard. In addition, the regressions include various firm-level characteristics such as capital intensity, 3-digit industry fixed effects, and location fixed effects at the county or the state level.

The regression estimates for each of the four decades between 1850 and 1880 indicate that workers in urban firms were paid significantly higher wages relative to those in rural firms in every decade during period. Moreover, the data show that the urban wage-premium applied across the board for different skilled categories. Evidence suggests that the wages of urban workers were higher because urban firms were more productive. The data indicate that labor productivity of urban workers and total factor productivity of urban firms were both significantly higher than those of rural

the American mercantile city were attributable to shortages of capital and labor, the state of technology, an expensive and inadequate transport network, and the restricted size of the accessible market.

5   Glaeser (2005) suggests that other factors such as economies in transportation hubs may also explain why industries became concentrated in cities. Glaeser argues that the rise of industrial Boston was aided by its comparative advantage in seafaring human capital as well as its position as a railroad hub of New England. Also see Kim (2000).

firms. Pooled regression involving all the years of the data indicate that the urban wage-premium rose over time.

Despite the fact that there is a wealth of anecdotal evidence on the rise of division of labor and labor-matching costs during the second half of the industrial revolution, it is extremely difficult to systematically measure division of labor in firms and society. To the extent that worker skills and requirements might differ across industries, however, division of labor is likely to be correlated with industrial diversity. This paper examines whether larger urban areas were associated with a greater diversity in manufacturing industries. Data on industrial diversity for large cities in 1860 and 1880 for which detailed industrial data are available suggest that larger cities were considerably more diverse than smaller cities. Thus, evidence suggests that larger cities fostered division of labor, and lowered the search and matching costs of firms and workers.

2. Empirical framework

Why did industrialization cause urbanization in the second half of the nineteenth century? Duranton and Puga (2004) provide a useful summary of the literature on the models of micro-foundations of urban agglomeration economies. These authors categorize the models of city formation into three essential categories: sharing, matching, and learning. Sharing economies come from indivisibilities in the provision of goods or facilities such as local public goods, marketplaces, intermediate input, and labor tasks.[6] Matching economies exist when search costs of firms and workers decline with agglomeration.[7] Finally, learning economies occur when geographic proximity of agents facilitate learning.[8]

In order to determine whether there were benefits to locating in cities for workers and firms, this paper examines whether workers in cities earned higher wages and whether firms in cities were characterized by high labor and total factor productivities. If cities fostered sharing, matching or learning, then workers in cities should be more productive and earn higher wages. In recent years, a number of studies such as Glaeser and Mare (2001), Wheeler (2001, 2004), and Wheaton and Lewis (2002) have found evidence for an urban wage-premium for the twentieth century whereas Ciccone and Hall (1996) find that employment density is positively correlated with average labor productivity.

Unlike most of these previous studies which estimate wage regressions using individual level data from the Census of Population, this paper uses firm-level data from the Census of Manufactures.[9] This paper estimates the following wage, labor, and total

---

6 See Buchanan (1965) for models based on indivisibilities of facilities, Berliant and Wang (1993) for those based on marketplaces, and Abdel-Rahman and Fujita (1990) for those based on intermediate inputs.

7 See Kim (1989, 1990) and Helsley and Strange (1990).

8 See Duranton and Puga (2001), Jovanovic and Rob (1989), Glaeser (1999), Palivos and Wang (1996).

9 Most studies estimate a version of the following equation: $\log(w_{ict}) = \alpha + X_{ict}\beta + Z_{ct}\gamma + d_c + d_t + u_{ict}$ where w_{ict} is wage of individual i living in city c in period t, X_{it} is a vector of individual characteristics, Z_{ct} is a vector of city characteristics; d_c is city fixed effect; d_t is year effect. Typical proxies for spillovers are average human capital level of a city or whether the location is urban or metropolitan. See Moretti (2004) for a useful review of the literature.

factor productivity equations:

$$\ln(w_{ic}) = \alpha_0 + \alpha_1 \text{Urban}_{ic} + \alpha_2 F_{ic} + d_l + d_j + u_{ic} \tag{1}$$

$$\ln(\text{LP}_{ic}) = \beta_0 + \beta_1 \text{Urban}_{ic} + \beta_2 F_{ic} + d_l + d_j + u_{ic} \tag{2}$$

$$\ln(\text{TFP}_{ic}) = \beta_0 + \beta_1 \text{Urban}_{ic} + \beta_2 F_{ic} + d_l + d_j + u_{ic} \tag{3}$$

where w_{ic} is wages, LP_{ic} and TFP_{ic} are labor and total factor productivity of firm i in county c, respectively; Urban_{ic} is a dummy variable for whether firm i was located in an urban location, F_{ic} is firm-level characteristics, d_l is location (county or state) fixed effect, and d_j is industry fixed effect.[10]

The regressions estimate whether firms in urban locations are characterized by higher wages, labor, and total factor productivity than those in rural locations. The regressions control for firm-level characteristics, such as capital intensity, sex composition (skill composition) of workers, power type, firm size, and industrial category, which may influence worker productivity and wages. Workers in capital-intensive firms may earn higher wages if there are capital-skill complementarities or if workers in capital-intensive firms have greater bargaining power. Male workers also generally earned higher wages as compared with women and children. As the regression estimates may be subject to omitted variable bias, the regressions include appropriate location (county or state) fixed effects and industry fixed effects at the 3-digit industry level. These fixed effects are likely to control for unobserved locational and industry characteristics that might influence productivity and wages.

The paper uses a combination of several instruments to correct for the potential endogeneity of the urban variable (see Ciccone and Hall, 1996). One instrument used is whether a county was urban in previous decades. For the 1850 regressions, the lagged urban variable was the share of the county population that was urban in 1840; for the 1860–1880 regressions, it was the share of the county population that was urban in 1850. In both cases, the urban threshold was population exceeding 2500. The lagged urban variable is most likely to be exogenous for the 1880 regressions as the variable is lagged three decades.

In addition, the paper utilizes three additional instruments which are likely to be correlated with the urban variable but much less so with the structural error term. Two instruments, access to water transportation and distance to the eastern seaboard are functions of physical geography. The final instrument, county date of incorporation, is likely to be correlated with urbanization as counties that were founded earlier were likely to have some advantageous characteristics for settlement. However, as most counties at their time of incorporation were likely to have been rural, the date of incorporation is likely to be exogenous.

This paper also examines the impact of urbanization on the distribution of wages as job matching efficiencies or spillovers may affect skilled and unskilled workers

---

10   Atack et al. (2004) estimate a version of equation (1). However, their interest is in examining whether wages are decreasing with establishment size.

differently. Thus, we estimate the following equations:

$$\ln(w_{sic}) = \alpha_0 + \alpha_1 \text{Urban}_{ic} + \alpha_2 F_{ic} + d_1 + d_j + u_{ic} \tag{4}$$

$$\ln(wu_{ic}) = \alpha_0 + \alpha_1 \text{Urban}_{ic} + \alpha_2 F_{ic} + d_1 + d_j + u_{ic} \tag{5}$$

$$\ln(w_{sic}) - \log(wu_{ic}) = \alpha_0 + \alpha_1 \text{Urban}_{ic} + \alpha_2 F_{ic} + d_1 + d_j + u_{ic} \tag{6}$$

where w_{sic} is wage paid by a firm i to skilled labor in county c and w_{uic} is wage paid by a firm i to unskilled labor in county c, and the remaining variables defined as above.

In order to examine whether urban agglomeration economies gained force over time, the paper estimates the following pooled difference-in-difference regression of the form:

$$\ln(w_{ict}) = \alpha_0 + \alpha_1 \text{Urban}_{ict} + \alpha_2 F_{ict} + \alpha_3 d_t \text{Urban}_{ct} + d_t + d_l + d_j + u_{ict} \tag{7}$$

where the parameter of interest is α_3. If α_3 is positive, then agglomeration economies are likely to have increased over time. The regression includes year fixed effects as wage data might not be comparable over different years because of changes in census methodology or price levels.

3. Data and empirical evidence

This paper uses the Atack–Bateman–Weiss (ABW) sample of manufacturing firms drawn from the manuscripts of the decennial censuses for 1850, 1860, 1870, and 1880 (see Atack and Bateman, 1999).[11] The data contain the typical information reported in the census of manufactures such as output, capital, labor, raw materials, wages, primary power source among others. The establishments are categorized by the standard industrial code (sic) at the 3-digit industry level. The firm-level information contained in the ABW data was supplemented with county-level information from the published decennial censuses. Consequently, the data set contains a rich array of county-level information such as population, land area, and various economic and demographic characteristics.

The dependent variables are defined as the logarithms of the average annual wage, labor, and total factor productivities of a given establishment or firm.[12] In this period, employees consisted of men, women, and children. For 1850 and 1860, the

---

[11] Because the 1850–1870 data were collected under constraints of punch cards and mainframe computers, the data were drawn at random from firms in each state with the goal of reaching a sample size of 200–300 firms in each state. Thus, the 1850–1870 data required a post-sample weighting procedure to construct a nationally representative sample. The 1880 data, however, drawn more recently using personal computers, were constructed to reflect both the state and national samples that would not require post-sample weighting. The data used in this paper were not weighted to reflect a nationally representative sample. However, the descriptive statistics of this data set match closely those of Atack et al. (2004). The sample sizes of the data used in this paper were 7.1, 6.3, 4.0, and 5.6% of total establishments in 1850, 1860, 1870, and 1880, respectively. See Atack et al. (2004) and Kim (2005) for a fuller discussion of the data.

[12] The estimates on total factor productivity are based on a neoclassical production function: $Q_i = F(K_i, L_i, M_i)$ where Q_i is the output in value added of establishment i and K_i, L_i and M_i are capital, labor and raw materials, respectively. Total factor productivity in logarithmic values can be expressed as: $\ln TFP_i = \ln Q_i - \alpha_K \ln K_i - \alpha_L \ln L_i - \alpha_M \ln M_i$ where α_K, α_L, α_M, are income shares from

Table 1. Summary statistics: manuscript censuses, 1850–1880 (mean values)

All Firms	1850	1860	1870	1880
Annual wage (log)	5.15	5.27	5.52	5.23
Rural	5.07	5.19	5.35	4.94
Urban	5.40	5.48	5.89	5.57
Labor productivity (Log)	5.76	6.00	6.38	6.11
Rural	5.67	5.91	6.25	5.89
Urban	6.04	6.24	6.65	6.36
Urban	0.25	0.28	0.31	0.46
Log K/L ratio	5.62	5.97	6.20	5.99
Share of male labor	0.67	0.68	0.68	0.69
Factory	0.10	0.10	0.15	0.16
Steam power	0.09	0.20	0.29	0.25
Water power	0.30	0.18	0.15	0.16
Steam*factory	0.03	0.03	0.09	0.09
Water*factory	0.02	0.01	0.01	0.01
Number of firms	4,333	3,550	2,289	8,658

Note: Urban = 1 if establishment was located in an incorporated town or city of population of 2500 or more. Factory = 1 if an establishment employed more than 15 employees. Sample selection criteria was same as that of Atack et al. (2004): Gross output greater than $500; for 1850 and 1860, average monthly wage greater than $4.76 but less than $190.5; and, for 1870 and 1880, average monthly wage greater than $5.20 but less than $208.
Source: See Atack and Bateman (1999).

average annual wage for each establishment was constructed from the reported average monthly wages by assuming that workers worked through the entire year; for 1870 and 1880, the constructed average annual wage was based on reported annual wage costs. As not all firms operated throughout the year, the estimates of average wages for non-fulltime establishments may be biased downwards in 1870 and 1880. The potential bias caused by part-time operation can be investigated for 1880 as that census reported information on the actual months of operation.

The urban variable used for the regression is a dummy variable that indicates whether an establishment was located in an urban or rural location. The census officials for the period 1850–1880 defined an urban area based on whether it was an incorporated town or city which contained a population of at least 2500. Unlike modern definitions which use the county as the smallest unit of analysis to determine whether a place is urban or rural, the typical geographic unit of observation was the minor civil divisions. As cities during this period were geographically compact, often no larger than 3 square miles, it made little sense to use the entire county as the unit of analysis to determine whether a location was urban or rural. This paper also examines whether the urban wage relationship is robust to other indicators of labor market density such as total county population, population density, shares of county population in cities of 2500 or higher, and shares of county population in cities of 25 000 and higher.

the respective factors of production. The income shares are estimated using ordinary least squares and the standard Cobb–Douglass production function specification for each year.

Table 2. OLS wage equations, 1850–1880

OLS	1850 (1)	1860 (2)	1870 (3)	1880 (4)	1880+ (5)
Urban	0.10*	0.07*	0.20*	0.38*	0.27*
	(6.2)	(3.2)	(5.1)	(24.9)	(19.4)
Ln(K/L)	0.06*	0.06*	0.20*	0.19*	0.12*
	(10.2)	(8.7)	(15.4)	(29.3)	(20.1)
Men/L	0.91*	1.1*	0.78*	0.95*	0.92*
	(25.9)	(26.9)	(10.6)	(25.2)	(26.4)
Factory	−0.00	0.05	0.17*	0.05	0.10*
	(0.2)	(1.7)	(3.1)	(1.7)	(3.9)
Steam	0.04	−0.02	−0.00	−0.08*	0.06*
	(1.5)	(−0.8)	(−0.0)	(−3.3)	(2.7)
Water	−0.05*	−0.05*	−0.19*	−0.18*	−0.07*
	(−2.4)	(−2.0)	(−3.3)	(−6.4)	(−2.6)
Steam*factory	0.10*	0.07	−0.04	0.18*	0.01
	(2.1)	(1.5)	(0.5)	(4.8)	(0.2)
Water*factory	0.12*	0.11	0.16	0.22*	0.03
	(2.6)	(1.5)	(1.2)	(3.0)	(0.5)
Constant	3.6*	3.7*	3.2*	2.8*	3.4*
	(8.2)	(8.0)	(3.8)	(5.0)	(6.6)
R-squared	0.63	0.65	0.60	0.45	0.33
Observations	4,285	3,536	2,280	8,658	8,658

Note: *t*-statistics in parentheses. *Significant at the 5% level. Regressions include county and 3-digit industry fixed effects. +Wages for equation (5) were adjusted for months of operation.

To eliminate potential outliers in the data, the samples were restricted to establishments with positive values of output, employment, and capital. In addition, following Atack et al. (2004), the sample was further restricted to establishments with gross output of greater than $500 and those with extremely low and high average wages were also excluded. The number of establishments that remained in the sample were 4333, 3550, 2289, and 8658 for 1850, 1860, 1870, and 1880, respectively and they represent 3.5, 2.5, 0.9, and 3.4% of total establishments for their respective years. It is also important to point out that the data for 1850–1870, unlike the 1880 data, do not represent a nationally representative sample; data for these years were collected with the goal of having at least 200 establishments per state.

Table 1 reports the regression sample means. The data show that urbanization increased significantly for manufacturing firms. Over the period between 1850 and 1880, the share of establishments in urban locations almost doubled from 25 to 46%. The logarithm of average annual wages of all firms ranged from 5.2 to 5.5 over this period. The logarithm of capital-labor ratios remained relatively constant between 5.6 and 6.2, as did the share of male labor force at around 68%. The share of factory establishments, defined as those who employed more than 15 workers, rose from 10 to 16% over this period. The percentage of steam-powered establishments rose from 9 to 25% whereas the share of water-powered firms fell from 30 to 16%. The share of steam-powered factories rose from 3 to 9% whereas that of water-powered factories fell from 2 to 1%. Wages of urban establishments were 30% higher than rural wages

Table 3. First-stage regressions

Urban	1850 (1)	1860 (2)	1870 (3)	1880 (4)
Lagged urban	0.60*	0.94*	0.54*	0.68*
	(29.0)	(36.0)	(18.6)	(39.4)
Water transportation	0.08*	0.05*	0.11*	0.08*
	(5.0)	(3.2)	(5.3)	(6.5)
R-squared	0.44	0.49	0.45	0.49
F-Statistics	30.3	34.4	19.2	50.1
Observations	3991	3249	2162	6870

Urban	1850 (5)	1860 (6)	1870 (7)	1880 (8)
Ln(distance)	−0.02*	−0.06*	−0.02*	−0.02*
	(−3.3)	(−2.5)	(−2.4)	(−5.1)
Ln(incorporation date)	−3.4*	−5.5*	−3.4*	−3.4*
	(−13.8)	(−17.2)	(−7.6)	(−15.0)
R-squared	0.31	0.34	0.36	0.38
F-statistic	19.3	18.8	13.9	37.8
Observations	4240	3433	2288	8445

Note: *t*-statistics in parentheses. *Significant at the 5% level. Regressions include state and 3-digit industry fixed effects. For 1850, the lagged urban variable is the share of county population that was urban in 1840; for 1860, 1870, and 1880, the lagged urban variable is the share of county population that was urban in 1850. Distance measures square miles from Kings County, New York and is constructed using data on county longitude and latitude from http://www.census.gov/tiger/tms/gazetteer. Distance between two points, A and B, is calculated using the following approximate formula: $[(69.1*(\text{longitude}A - \text{longitude}B))^2 + (53.0*(\text{latitude}A - \text{latitude}B))^2]^{1/2}$. County dates of incorporation are obtained from Kane (1972).

between 1850 and 1860, and about 50–60% higher in 1870 and 1880; labor productivities of urban firms were also 33–47% higher than rural firms.

The OLS regression estimates on wages reported in Table 2 indicate that workers in urban firms consistently earned higher wages on average than those of rural firms. In 1850 and 1860, urban wages were 10.5 and 7.3% higher than rural wages; in 1870 and 1880, they were 22.1 and 46.2% higher.[13] Because the regressions include county fixed effects, the variations in the data come from establishments within counties that are in rural and urban locations. As many firms did not operate throughout the year, the wage estimates of non-fulltime establishments may be biased downwards. For 1880, it is possible to adjust wages for months of operation. The OLS regression estimate of the urban wage-premium declined from 46.2 to 31% when wages are adjusted for months of operation. The decline in this estimate suggests that rural firms were more likely to operate less than full-time than urban firms. An alternative interpretation of this result may be that urban firms and workers were more likely to be employed on a full-time basis which may be worth as much as 33% of one's wages.

13　The data censoring restrictions had the largest impact for the data sample in 1880. In that year, a large number of small firms located in rural areas were excluded from the sample. When the regressions were run without any censoring restrictions, the urban wage-premium was higher than that reported for 1880. Note that as the regression is in semilogarithmic form, the percentage effect of the dummy variable is equal to $100(\exp(\alpha_1) - 1)$.

Table 4. IV Wage equations, 1850–1880

IV	1850 (1)	1860 (2)	1870 (3)	1880 (4)	1880+ (5)
Urban	0.47*	0.22*	0.83*	0.78*	0.52*
	(5.0)	(4.7)	(10.0)	(19.8)	(14.6)
Ln(*K*/*L*)	0.06*	0.06*	0.19*	0.18*	0.11*
	(9.0)	(8.6)	(14.5)	(24.0)	(16.7)
Men/*L*	0.83*	1.2*	0.82*	0.85*	0.82*
	(17.4)	(27.7)	(11.0)	(19.1)	(20.6)
Factory	−0.04	0.03	0.15*	0.04	0.10*
	(1.1)	(1.0)	(2.5)	(1.1)	(3.5)
Steam	0.10*	−0.00	−0.01	−0.05	0.08*
	(3.5)	(−0.4)	(−0.2)	(−1.7)	(3.3)
Water	0.0	−0.05*	−0.18*	−0.12*	−0.03
	(0.03)	(−2.2)	(−3.2)	(−3.8)	(−1.1)
Steam*factory	−0.02	0.06	−0.07	0.13*	−0.04
	(−0.4)	(1.3)	(−0.9)	(2.9)	(−0.9)
Water*factory	0.07	0.09	0.19	0.21*	0.02
	(1.4)	(1.2)	(1.5)	(2.7)	(0.3)
Constant	4. 1*	4.8*	2.9*	3.5*	4.1*
	(11.8)	(14.6)	(5.0)	(5.6)	(7.2)
Overid test Sargan *P*-value	0.61	0.32	0.77	0.39	0.27
R-squared	0.42	0.51	0.37	0.40	0.30
Observations	3850	3419	2153	6873	6873

Note: *t*-statistics in parentheses. *Significant at the 5% level. Regressions include state and 3-digit industry fixed effects. + Wages for equation (5) were adjusted for months of operation.

Table 3 reports the first-stage regressions of whether a firm resided in an urban or a rural location on the instruments lagged urban variable, a dummy variable for access to water transportation, logarithm of distance from the eastern seaboard, and logarithm of dates of incorporation.[14] All of these variables are significantly correlated with the urban variable in the expected manner. The lagged urban variable and access to water transportation were both positively correlated with whether a firm was likely to locate in cities; firms in younger counties and counties further from the eastern seaboard were less likely to locate in cities.

The IV regression estimates reported in Table 4 indicate that the OLS estimates of the urban wage-premium are likely to be biased downwards. The IV regression estimates show that workers of urban firms earned 60, 25, 129, and 118% higher wages than those of rural firms for 1850, 1860, 1870, and 1880, respectively.[15] In general,

---

14 Distance from the eastern seaboard measures distance in square miles from Kings County, New York; the measure is constructed using data on county longitude and latitude from http://www.census.gov/tiger/tms/gazetteer. The distance between two points, A and B, is calculated using the following approximate formula: $[(69.1*(longitudeA - longitudeB))^2 + (53.0*(latitudeA - latitudeB))^2]^{1/2}$. County dates of incorporation are obtained from Kane (1972).

15 The IV regressions use state rather than county fixed effects. When county fixed effects were used, the IV regressions were not identified. When the OLS regressions were run separately for county and state fixed effects, the urban premium with county fixed effects were generally smaller. Thus, IV estimates of the urban wage-premium are likely to decline somewhat if finer geographic controls are used.

Table 5. IV Labor productivity equations, 1850–1880

IV	1850 (1)	1860 (2)	1870 (3)	1880 (4)
Urban	0.77*	0.39*	0.74*	0.85*
	(9.8)	(6.1)	(6.5)	(18.0)
Ln(K/L)	0.21*	0.27*	0.26*	0.26*
	(16.1)	(17.8)	(14.6)	(28.9)
Men/L	0.57*	1.1*	0.42*	0.27*
	(7.0)	(11.2)	(4.1)	(5.1)
Factory	−0.14*	−0.07	−0.00	−0.11*
	(−2.5)	(−1.1)	(−0.2)	(−2.8)
Steam	0.23*	0.08	0.13*	−0.05
	(4.1)	(1.6)	(2.0)	(−1.5)
Water	−0.01	0.02	−0.13	−0.11*
	(−0.2)	(0.4)	(−1.6)	(−2.9)
Steam*factory	−0.06	0.03	−0.04	0.15*
	(−0.6)	(0.3)	(0.4)	(2.8)
Water*factory	−0.02	0.19	0.16	0.34*
	(−0.2)	(1.2)	(0.9)	(3.5)
Constant	4.2*	3.1*	4.4*	3.8*
	(6.0)	(4.3)	(5.5)	(5.1)
Overid test Sargan *P*-value	0.82	0.81	0.38	0.16
R-squared	0.24	0.29	0.24	0.28
Observations	3,765	3,190	2,113	6,848

Note: *t*-statistics in parentheses. * Significant at the 5% level. Regressions include state and 3-digit industry fixed effects.

the IV estimates were significantly higher than those of the OLS estimates. As there is more than one instrument, it is possible to test the overidentifying restrictions. In general, different combinations of any two of the instruments, such as lagged urban and access to water, or distance and date of incorporation, exhibited the highest Sargan *P*-value indicating that the instruments were uncorrelated with the structural error term. The IV regressions were least successful from the perspective of the overidentifying restriction test when all four instruments were used. The IV regression estimates reported were chosen for their highest Sargan *P*-values but the coefficients were generally robust to different sets of instruments.

The data show that capital intensity was positively correlated with wages as expected. For the capital intensity variable, the OLS and IV regression estimates are very similar. In 1850 and 1860, a percentage increase in capital intensity increased wages by 6%, but the figures rose to 18 and 19% in 1870 and 1880, respectively. However, the estimates for 1880 decline markedly when wages are adjusted for months of operation. Thus, capital-intensive firms were more likely to operate full-time than labor-intensive firms. The regressions also indicate that firms with a higher share of male workers paid higher average wages.

The IV regressions indicate that firms which utilized male workers more intensely were likely to exhibit higher average wages. Workers in water-powered artisanal firms earned less wages than the omitted category of hand-powered, artisanal firms. On the other hand, water-powered factory workers earned between 7 and 23% higher wages than workers in the omitted category. The human-powered factory variable was positively significant only for 1870; the steam and steam-powered factory variable exhibited little systematic patterns over time.

Table 6. IV Total factor productivity equations, 1850–1880

IV	1850 (1)	1860 (2)	1870 (3)	1880 (4)
Urban	0.59*	0.28*	0.69*	0.60*
	(8.0)	(4.8)	(6.4)	(14.3)
Ln(K/L)	0.02	0.04*	0.08*	0.04*
	(1.8)	(2.6)	(4.5)	(4.4)
Men/L	0.31*	0.47*	0.35*	0.10*
	(4.0)	(5.4)	(3.6)	(2.2)
Factory	−0.32*	−0.39*	−0.04	−0.19*
	(−6.3)	(−6.5)	(−0.6)	(−5.5)
Steam	0.11*	−0.02	0.10	−0.06*
	(2.1)	(0.3)	(1.8)	(−2.0)
Water	−0.05	−0.01	−0.15*	−0.09*
	(−1.2)	(−0.3)	(−2.0)	(−2.8)
Steam*factory	−0.01	0.09	−0.03	0.13*
	(−0.1)	(0.9)	(−0.3)	(2.7)
Water*factory	0.01	0.07	0.19	0.29*
	(0.1)	(0.5)	(1.1)	(3.4)
Constant	3.5*	2.7*	4.0*	2.8*
	(5.2)	(4.0)	(5.3)	(4.5)
Overid test Sargan *P*-value	0.69	0.66	0.42	0.09
R-squared	0.12	0.15	0.13	0.18
Observations	3765	3190	2113	6848

Note: *t*-statistics in parentheses. *Significant at the 5% level. Regressions include state and 3-digit industry fixed effects.

The IV regression estimates on labor productivity reported in Table 5 show that urban workers were significantly more productive than rural workers. The labor productivity of urban firms were 48–134% higher than those of rural firms between 1850 and 1880. As with wages, labor productivity was also positively correlated with capital and male labor intensities. The remaining variables did not exhibit systematic patterns over time. The IV regression estimates on total factor productivity reported in Table 6 show that urban firms were 32–180% more productive that rural firms.[16] Like the wage and labor productivity regressions, total factor productivity was positively correlated with capital and male intensities. However, unlike labor

---

16 Total factor productivity estimates must be interpreted with caution. There are two potential reasons for why total factor productivity estimates may be unreliable. One major reason is that the neoclassical assumptions used to estimate total factor productivity is unlikely to hold if spatial externalities are important. In the neoclassical model, labor productivity in a simple two factor Cobb–Douglas production function ($Y = AK^{\alpha} L^{1-\alpha}$) can be expressed as follows: $Y/L = A (K/L)^{\alpha}$. Thus, labor productivity is equal to total factor productivity (A) times some proportion of capital intensity. However, if there are spatial external economies, then the relationship between labor and total factor productivity is more complex. As shown in Ciccone (2002), average labor productivity of a given region can be expressed as total factor productivity multiplied by some proportion of capital intensity and the factor responsible for spatial externalities. The second important reason may be that the OLS estimates of the neoclassical production function used to calculate total factor productivity are biased as firms with higher productivity shocks may use more inputs (see Olley and Pakes, 1996 and Levinsohn and Petrin, 2003). To the extent that firms in urban areas are more likely to experience productivity shocks, the estimates of total factor productivity of urban firms may be biased upwards.

Table 7. IV Wage equations for skilled and unskilled workers, 1850–1880

	1850	1860	1870	1880
	(1)	(2)	(3)	(4)
Skilled workers IV				
Urban	0.38*	0.30*	—	0.21*
	(8.1)	(9.0)	—	(9.7)
Overid test Sargan *P*-value	0.67	0.51	—	0.53
R-squared	0.28	0.27	—	0.27
Observations	3920	3210	—	6296
Unskilled workers IV	1850	1860	1870	1880
	(5)	(6)	(7)	(8)
Urban	0.59*	0.28	—	0.24*
	(3.9)	(1.3)	—	(9.9)
Overid test Sargan *P*-value	0.37	0.81	—	0.41
R-squared	0.20	0.40	—	0.36
Observations	513	353	—	5406
Skilled–unskilled Wage gap IV	1850	1860	1870	1880
	(9)	(10)	(11)	(12)
Urban	−0.31	−0.09	—	0.16*
	(−1.8)	(−0.9)	—	(2.6)
Overid test Sargan *P*-value	0.97	0.97	—	0.24
R-squared	0.25	0.31	—	0.12
Observations	513	333	—	6006

Note: *t*-statistics in parentheses. *Significant at the 5% level. Regressions include the same set of independent variables as well as state and 3-digit industry fixed effects as in previous tables. For 1850 and 1860, unskilled and unskilled wages are proxied by male and female wages, respectively; for 1880, skilled and unskilled wages are daily wages for skilled mechanics and common laborers, respectively.

productivity, total factor productivity was also consistently positively correlated with human-powered factory and steam variables.

Regression estimates on the urban wage premia of skilled and unskilled workers reported in Table 7 show that the urban wage-premium appears to apply across the board for different skilled categories. Skilled and unskilled wages, for the years 1850 and 1860, are proxied by male and female annual wages, respectively; for 1880, the skilled and unskilled wages are daily wages for a skilled mechanic and common laborer, respectively. Sample sizes of these regressions are smaller because not all firms reported data on these variables. Table 7 shows that urban male wages were 46 and 35% higher than rural male wages in 1850 and 1860, respectively; in 1880, the estimated urban wage-premium for a skilled mechanic was 23%. Table 8 shows that urban female wages were 80 and 32% higher than rural female wages, respectively; in 1880, urban common laborers earned 27% higher wages than rural common laborers.

Table 7 also reports regression results on the male–female wage gap for 1850 and 1860 and the skilled–unskilled wage gap for 1880. Because the samples for these regressions were restricted to firms which reported both skilled and unskilled wages, the regressions control for firm-specific effects which might influence the skilled–unskilled wage gap. Urbanization generally seems to have reduced the skilled–unskilled wage

Table 8. Wage-IV, labor productivity, and total factor productivity regressions by different urban definitions, 1850–1880

Wages-IV	1850	1860	1870	1880
% Urban >= 2500	0.23*	0.25*	0.51*	0.28*
% Urban >= 25 000	0.23*	0.26*	0.45*	0.37*
Ln(population)	0.03*	0.06*	0.11*	0.14*
Ln(density)	0.03*	0.05*	0.06*	0.09*
LP-IV	**1850**	**1860**	**1870**	**1880**
%urban >= 2500	0.45*	0.35*	0.47*	0.31*
%urban >=25 000	0.45*	0.36*	0.42*	0.38*
ln(population)	0.10*	0.09*	0.13*	0.15*
ln(density)	0.06*	0.07*	0.07*	0.09*
TFP-IV	**1850**	**1860**	**1870**	**1880**
% Urban >= 2500	0.35*	0.25*	0.44*	0.23*
% Urban >=25 000	0.35*	0.26*	0.38*	0.33*
Ln(population)	0.07*	0.06*	0.12*	0.10*
Ln(density)	0.05*	0.05*	0.07*	0.07*

Note: *Significant at the 5% level. Regressions include the same set of independent variables as well as state and 3-digit industry fixed effects as in previous tables. These coefficients were not reported due to space limitations.

gap in 1850 and 1860, but the coefficient was only weakly significant for 1860. In 1860, the female–male wage gap in urban areas were 36% lower than in rural places. However, the skilled–unskilled wage gap seems to have widened in urban areas in 1880. In that year, urban location increased the wage gap of skilled mechanics and unskilled common laborers by 17%. Other factors that seem to have reduced the wage gap was the intensive use of male workers and the use of water-power in factories.

Table 8 reports urban wage, labor and total factor productivities premiums for four alternative measures of urban agglomeration: (i) the share of population in cities with populations of 2500 or more; (ii) the share of population in cities with populations of 8,500 or more; (iii) the total county population, and (iv) the county population density. The IV regression estimates indicate that there was a significant correlation between wages and productivity for all these different measures of urban agglomeration. While not reported, the regressions all contain the same set of independent variables and instruments.

The pooled difference-in-difference regression estimates reported in Table 9, based on pooled data for all years between 1850 and 1880, indicate that urban wage and labor productivity premiums increased significantly over this period. According to the IV estimates, the urban wage-premium remained constant between 1850 and 1860, but rose significantly in the two succeeding decades. Relative to 1850, the omitted year, wages of workers in urban firms rose from 28 to 35% and labor productivity urban firms rose from 6 to 11% in 1870 and 1880. For the wage regression, year dummies were positively significant for 1860 and 1870, but negatively so for 1880. For the labor productivity, year dummies were positively significant for all years. For the pooled regression, capital and male intensities, steam and water-powered factory

Table 9. Pooled difference-in-difference wage and labor productivity regressions, 1860 and 1880

	Ln(Wage) OLS	Ln(LP) OLS
Urban	0.11*	0.26*
	(6.2)	(9.8)
D1860	0.10*	0.11*
	(7.3)	(5.8)
D1870	0.12*	0.34*
	(7.5)	(14.9)
D1880	−0.20*	0.06*
	(−15.2)	(3.1)
D1860*urban	−0.03	−0.01
	(−1.0)	(−0.2)
D1870*urban	0.25*	0.06
	(8.8)	(1.6)
D1880*urban	0.30*	0.10*
	(15.0)	(3.4)
Ln(K/L)	0.15*	0.27*
	(38.8)	(48.3)
Men/L	0.94*	0.53*
	(41.2)	(16.1)
Factory	0.08*	−0.04
	(4.8)	(−1.8)
Steam	−0.04*	0.02
	(−2.9)	(0.6)
Water	−0.12*	−0.09*
	(−7.9)	(−4.4)
Steam*factory	0.11*	0.10*
	(4.6)	(3.0)
Water*factory	0.09*	0.08
	(2.3)	(1.3)
R-squared	0.44	0.32
Observations	18 759	18 509

Note: *t*-statistics in parentheses. *Significant at the 5% level. Regressions include state and 3-digit industry fixed effects.

variables were positively correlated with wages and labor productivity; steam and water-powered non-factory variables were negatively correlated.

4. Division of labor and matching economies

The analysis of the firm-level data from the manuscript censuses for the period between 1850 and 1880 demonstrates that urban firms enjoyed significant urban agglomeration economies and that these economies are likely to have increased over time. Evidence indicates that urban firms were significantly more productive than rural firms and, as a result, compensated their workers accordingly. However, the fact that wages, labor, and total factor productivities were higher for urban firms provides little guidance as to which types of urban agglomeration economies, sharing, matching or learning economies, caused firms and workers to choose urban locations during this phase of the industrial revolution. This

section provides a discussion of the potential causal linkages of industrialization and urbanization.

The fact that early industrialization began in rural areas of New England casts some doubt on the importance of certain types of agglomeration economies such as sharing or learning. If these economies were at work during early industrialization, then the early factories should have located in major cities like New York and Boston rather than in rural townships of Dudley and Oxford, MA.[17] In particular, explanations based on variety and market sizes, indivisibility of facilities, and access to intermediate inputs seem inconsistent with the rural beginnings of industrialization. Moreover, geographic evidence on patents for the early and late industrial period suggests that technological learning or spillovers may have caused patent activity to concentrate in cities, but may not have had a similar impact on production. In general, there seems to be little evidence that inventive and production activities were spatially correlated during this period.[18]

While it is extremely difficult to empirically distinguish between the various causes of urban agglomeration during late industrialization, a variety of evidence points toward the importance of division of labor and matching economies.[19] Indeed, the two salient features of industrialization are the rise of the division of labor and the rise of a labor market. Before industrialization, self-employed master craftsmen or artisans made the entire product using their hand tools and learned all production techniques in apprenticeship. With industrialization, factory owners recruited workers for a wage in the labor market and reduced their jobs into a myriad of tasks that needed modest training to master. Because most jobs were obtained through a network of family, friends, and people of similar ethnicity, job search costs were likely to be significantly lower with the urban concentration of workers and firms.

The fact that industrialization first began in rural locations and then moved to cities seems more consistent with explanations based on division of labor and matching. In early industrialization, firms located in rural New England because early factories made extensive use of rural native women and children.[20] Because early industrialization was concentrated in only a handful of industries, and because the division of

---

17 As supplies of capital and labor are both likely to be more favorable in cities rather than in the rural countryside, Pred's (1966) thesis that early industrialization did not start in large mercantile cities because these places lacked capital and labor seems rather suspect. Prude (1983), Hirsch (1978), Faler (1981) and Ware (1931) provide useful historical accounts of early industrialization in rural New England.

18 Lamoreaux and Sokoloff (2000) find that inventive and production activities in the glass industry were geographically concentrated in different areas. Sutthiphisal (2005) finds similar geographic patterns for a number of low- and high-tech industries between 1870 and 1910.

19 Duranton (1988) demonstrates that the agglomeration of the labor force in cities deepens division of labor between workers and raises their productivity. However, the rise of division of labor within and across firms and industries is likely to significantly increase the heterogeneity of the labor market. When specialized firms and workers must search and match for production, Kim (1989, 1990), and Helseley and Strange (1990) show that urban agglomeration raises productivity by improving the average quality of matches of firms and workers. Thus, division of labor is likely to be limited by the extent of labor market matching economies. The importance of the division of labor for growth and development has also been explored by Baumgardner (1988), Becker and Murphy (1992), and Acemoglu (1996) among others.

20 See Goldin and Sokoloff (1984).

labor was only modest, labor market transaction costs were also likely to have been modest.[21] Anecdotal evidence suggests that labor market information was easily communicated throughout rural New England because factory towns and cities were very specialized. For example, according to Ware (1931), the well-known Slater Mill or mills in Lowell rarely needed to advertise for help as sufficient numbers of families and single women came to their factories seeking employment.[22]

In the second half of the nineteenth century, as division of labor within firms and across industries became more extensive, and as factory production spread to numerous industries, industrialization became a distinctly urban phenomenon. With industrialization, the occupational categories of workers as well as the number of industries in manufacturing rose dramatically. In 1820, US census officials only reported occupations by broad sectoral categories such as agriculture, manufacturing, trade, etc. However, as occupations proliferated with industrialization, the censuses began reporting these categories in great detail. In 1850 and 1860, the censuses reported occupation data for 323 and 584 categories, respectively (Edwards, 1943). In addition, the census of manufactures indicates that the number of manufacturing industries rose markedly. Although it is difficult to compare industries over time as the level of industry aggregation differed over time, there may have been as many as 40–50 very finely defined industries in 1820 whereas the number of industries rose to 261 and 631 in 1850 and 1860, respectively.

There are many reasons to believe that labor market transaction costs rose significantly in the second half of the nineteenth century. As the scale of factories rose with greater utilization of the division of labor, and as firm and labor specialization increased, it became increasingly more difficult to recruit and efficiently match workers and firms in rural locations. As immigration increased significantly over the second half of the twentieth century, manufacturers shifted their labor force from native women and children to immigrant workers who favored urban locations. As the most important source of labor market information came from a close network of family and friends, firms and workers found it more effective to locate in

---

21 Sokoloff (1984) finds that the shift in production from artisans to non-mechanized factories, with only a modest increase in the size of the work force, was associated with significant gains in productivity. Sokoloff attributes these gains to the division of labor, especially among the relatively unskilled workers composed of women and children.

22 Ware (1931, p. 212):

The use of advertising to secure help does not appear to have been wholly successful or necessary, to judge from the experience of the only mill whose advertisements could be compared with additions to the labor force. Between 1818 and 1825 only one of the Slater's six advertisements appears to have brought desired response while another secured a quarter of the help which it called for. Twenty-three of the twenty-six families who entered his employ during this period seem to have come unsolicited. When he opened a new thread mill in 1828, however, he was successful in using this means for additional hands. Except when a sudden increase in labor force was needed, most workers probably drifted in with the growth of towns around the mills or came through chance knowledge that there might be employment available.... The boarding house mills, on the other hand, did not have to let people know that work was available. If we are to believe the tales in the *Lowell Offerings*, there never was any lack of employment at Lowell. Girls came from long distances unheralded, selected a boarding house, perhaps where a friend from the same village lived, and the next morning found work in the mill, usually in the same room with her friend.

cities where such networks could be extended.[23] Because there were considerable turnover and instability in the job market during the period of the second industrial revolution, cities lowered the search costs of workers in a variety of ways.[24]

There seems to be substantial anecdotal evidence that division of labor and labor-matching costs rose as factories increased in scale and became more mechanized. One such example comes from the meat-packing industry. While meat-packing factory workers generally possessed less skill than a traditional artisan butcher, division of labor based on skill, ability, and physical attributes led to significant improvements in productivity (Barrett, 1987). The factory disassembly was organized into cattle-killing gangs of 230 men composed of 139 common laborers, 98 skilled, and 11 highly skilled workers.[25] Division of labor and supervision increased productivity by increasing the speed and intensity of work. Barrett (1987) writes that the great advantage of cities like Chicago over country towns was the availability of labor. Because few workers were hired on a permanent basis, workers generally made their rounds to various firms early in the morning where the foreman hired the required workers for the day. As most workers walked to work, they lived in close proximity to the packing houses to make their morning rounds. Workers obtained job market information from friends, neighbors, and the gatherings at neighborhood saloons.

While there may be considerable anecdotal evidence that division of labor arose with factory production and increased with mechanization, it is extremely difficult to measure it directly. However, to the extent that skill levels and occupational requirements are likely to differ across industries, one potential measure of division of labor at the economy-wide level may be the level of industrial diversification. As firms and workers are likely to be more heterogeneous across industries, matching firms, and workers of appropriate skills across industries are likely to be more costly. If cities foster division of labor and lower matching costs, then cities should also foster industrial diversity.

Following Duranton and Jayet (2005), industrial diversity is measured using the following index: $DI = 1/\Sigma_i (E_{ij}/E_j)^2$ where E_{ij} is employment in industry i in city j, E_j is total employment in city j.[26] As data at the industry level are available only for the largest cities, the diversity measure is constructed for 40 and 99 cities in 1860 and

---

23   See Nelson (1995) and Rosenbloom (2002).

24   Prude (1983) finds that the mean turnover rate at the Slater textile mill between 1813 to the mid-1830s was about 163%. He also finds that over half of the workers chose to depart voluntarily. Slichter (1919) provides additional evidence for the extremely high turnover rate of industrial workers.

25   Barrett (1987, 26): "The packers achieved three important and interrelated accomplishments through this subdivision of labor. First, by grossly reducing the quantity and quality of skill required to do the job, they destroyed the control which the all-around butcher had exercised over the slaughtering and cutting processes. A few highly skilled positions remained, but these were very specialized. In fact, mass pro-duction created a new, more narrowly defined notion of skill. Splitting the backbone of a steer, for example, required great dexterity as well as strength, and only a few men could do the job. Thus, the splitter and a few other trades earned high wages under the new system." See Hirsch (1978) for evidence of the rise of task differentiation in shoe and hatting factories. Thomson (1989) provides a complex analysis of the changes in the technology of shoe production and how these changes altered the division of labor between women, children, and skilled workers.

26   Duranton and Jayet (2005) examine whether division of labor is related to city size for French cities in 1990. They find that scarce occupations, as predicted by their model, were over-represented in larger cities suggesting that division of labor increased with the extent of the market.

Table 10. Industrial diversity of cities, 1860–1880 summary statistics

	1860		1880	
	Mean	SD	Mean	SD
Industrial diversity	15.08	(8.52)	9.84	(6.82)
Population	96 382	(153 272)	91 717	(167 595)
Manufacturing employment	14 945	(20 890)	12 972	(30 409)
Number of cities	40		99	

Urban-industrial diversity is measured using the following index: $DI = 1/\Sigma_i (E_{ij}/E_j)^2$ where E_{ij} is employment in industry i in city j, E_j is total employment in city j. The industry is defined at the product level and contains about 390 manufacturing industries.
Sources: Census of Manufactures, 1860, 1880; Census of Population, 1860, 1880.

Table 11. Estimates of urban-industrial diversity, 1860

Log(diversity)	(1)	(2)	(3)	(4)
Ln(population)	0.14	—	0.52*	0.46*
	(1.42)	(4.34)	(3.66)	
Ln(Mfg Emp.)	—	−0.13	−0.47*	−0.40*
		(−1.41)	(−4.33)	(−3.45)
Regional dummies				
New England	−0.42	−0.51*	−0.03	0.02
	(−1.70)	(−2.16)	(0.12)	(0.08)
South	−0.57*	−0.81*	−1.16*	−0.97*
	(−2.33)	(−2.91)	(−4.81)	(−3.58)
Midwest	0.11	−0.02	−0.26	−0.10
	(0.45)	(−0.09)	(−1.18)	(−0.42)
NY Metro	—	—	—	0.46
				(1.41)
Constant	1.17	4.00*	1.34	1.29
	(1.02)	(4.58)	(1.43)	(1.40)
R-squared	0.29	0.29	0.54	0.57
Observations	40	40	40	40

Note: *t*-statistics in parentheses. *Significant at the 5% level. Urban-industrial diversity is measured using the following index: $DI = 1/\Sigma_i (E_{ij}/E_j)^2$ where E_{ij} is employment in industry i in city j, E_j is total employment in city j. The industry is defined at the product level and contains about 390 manufacturing industries.

1880, respectively. Because the standard industrial codes were not devised until the mid-twentieth century, the data reported by the census officials were at the level of product categories that perhaps resemble the current three- or four-digit categories. In both of these years, the census officials reported information for close to 400 industries. Table 10 reports the descriptive information on these cities.

Tables 11 and 12 report regressions of the log of urban-industrial diversity on the log of total urban population and manufacturing employment; the regressions also include regional dummy variables where the omitted region is the Middle Atlantic. The regressions show that industrial diversity was positively and significantly

Table 12. Estimates of urban-industrial diversity, 1880

Log(diversity)	(1)	(2)	(3)	(4)
Ln(population)	0.44*	—	0.77*	0.72*
	(6.88)	(6.31)	(5.76)	
Ln(Mfg Emp.)	—	0.19*	−0.26*	−0.24*
		(3.90)	(−3.10)	(−2.86)
Regional dummies				
New England	−0.28	−0.44*	−0.16	−0.12
	(−1.64)	(−2.28)	(−0.96)	(−0.72)
South	0.05	0.08	−0.06	−0.01
	(0.31)	(0.39)	(−0.39)	(−0.04)
Midwest	0.47*	0.45*	0.46*	0.51*
	(3.05)	(2.55)	(3.05)	(3.36)
West	0.52	0.60	0.32	0.38
	(1.87)	(1.86)	(1.16)	(1.39)
NY Metro	—	—	—	0.56
				(1.61)
Constant	−2.85*	0.34	−4.19*	−3.86*
	(−3.85)	(0.75)	(−5.17)	(−4.67)
R-squared	0.44	0.27	0.49	0.51
Observations	99	99	99	99

Note: *t*-statistics in parentheses. *Significant at the 5% level. Urban-industrial diversity is measured using the following index: $DI = 1/\Sigma_i (E_{ij}/E_j)^2$ where E_{ij} is employment in industry i in city j, E_j is total employment in city j. The industry is defined at the product level and contains about 390 manufacturing industries.

correlated with urban population for both 1860 and 1880. According to the specification (4), which also includes total manufacturing employment, a percentage increase in population increased industrial diversity by 46 and 72% in 1860 and 1880, respectively.

The regressions show that manufacturing employment was negatively correlated with industrial diversification, especially when the regressions control for population. One interpretation is that manufacturing employment and population are collinear. However, it is instructive to note that large manufacturing cities which were highly specialized were concentrated in New England. It appears that early factory towns in New England which grew into large cities over time remained very specialized in those early industries. As a labor force specialized in few industries is unlikely to possess diverse occupational skills, it is not surprising that large industrial cities in New England were incapable of fostering industrial diversity.

5. Conclusion

Industrial revolution in the USA, as in England and other parts of Europe, occurred in two stages. Industrialization first began in rural New England between 1820 and 1840 as the artisanal method gave way to factory organization of production in a handful of industries. However, between 1850 and 1920, as industrial revolution deepened, scale economies rose, factory organization spread to numerous industries and regions, and industrialization became significantly concentrated in urban areas throughout the

northern region of the US. This paper examines the potential causal linkages in US industrialization and urbanization.

This paper finds that, at least for the period between 1850 and 1880 for which firm-level data from the manuscript censuses of manufactures are available, industrialization caused urbanization because there were significant urban agglomeration economies. The data show that urban firms were more productive, paid higher wages to their workers, and that urban premiums in wages and productivity increased over time. The urban wage-premium accrued to both skilled (male) and unskilled (female) workers throughout the entire period. The urban skilled–unskilled wage gap seems to have narrowed in 1850 and 1860, but may have widened in 1880. As expected, other firm-level characteristics such as capital intensity and skill intensity (male labor intensity) were also associated with higher wages and productivity. When the data were pooled over the entire period, those workers who worked in water and steam-powered factories also tended to earn higher wages when controlling for all other factors.

While it is difficult to rule out the importance of other types of agglomeration economies such as spillovers, this paper suggests that the geographic patterns of the two phases of industrialization are most consistent with explanations based on division of labor and labor market transaction costs. Because division of labor and labor-matching costs were relatively limited in early industrialization, early factories who located near rural townships in New England did not face high costs of recruiting labor. However, in the second phase of industrialization, as scale economies rose and as factory organization of production spread to hundreds of industries, division of labor and matching costs rose, and firms and workers became concentrated in cities. Evidence suggests that cities with higher population levels fostered greater industrial diversity.

Whereas workers of early factories in rural New England were composed of native women and children, the industrial workforce of the second industrial revolution was composed of immigrant workers. Over the century of the industrial revolution, immigrants and 'birds of passage' came to the USA in extraordinary numbers from varying European nations and became the dominant industrial labor force. Because the extent of division of labor of the native labor force was likely to have been limited during this period, immigrants from numerous different nations with varying skills and physical attributes are likely to have significantly increased the extent of division of labor available in society. Thus, immigrants who came during the industrial period are likely to have contributed significantly to sustaining the industrial revolution. Moreover, because division of labor is likely to have operated along ethnic lines, it is not surprising that ethnic communities formed in cities to lower labor market transaction costs.

Acknowledgments

References

Abdel-Rahman, H., Fujita, M. (1990) Product variety, Marshallian externalities, and city sizes. *Journal of Regional Science*, 30(2): 165–183.

Acemoglu, D. (1996) A microfoundation for social increasing returns in human capital accumulation. *Quarterly Journal of Economics*, 111: 779–804.

Atack, J., Bateman, F. (1999). U.S. historical statistics: nineteenth-century U.S. industrial development through the eyes of the census of manufactures. *Historical Methods*, 32(4): 177–188.

Atack, J., Bateman, F., Margo, R. (2004). Skill intensity and rising wage dispersion in nineteenth-century American manufacturing. *Journal of Economic History*, 64(1): 172–192.

Barrett, J. R. (1987) *Work and Community in the Jungle: Chicago's Packinghouse Workers 1894–1922*. University of Illinois Press.

Baumgardner, J. R. (1988) The division of labor, local markets and worker organization. *Journal of Political Economy*, 96(3): 509–527.

Becker, G. S., Murphy, K. M. (1992) Division of labor, coordination costs, and knowledge. *Quarterly Journal of Economics*, 107(4): 1137–1160.

Berliant, M., Wang, P. (1993) Endogenous formation of a city without agglomerative externalities or market imperfections: marketplaces in a regional economy. *Regional Science and Urban Economics*, 23(1): 121–144.

Buchanan, J. M. (1965) An economic theory of clubs. *Economica*, 32(125): 1–14.

Ciccone, A. (2002) Agglomeration-effects in Europe. *European Economic Review*, 46(2): 213–227.

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

Duranton, G. (1998) Labor specialization, transport costs and city size. *Journal of Regional Science*, 38: 553–573.

Duranton, G., Puga, D. (2001) Nursery cities: urban diversity, process innovation, and the life cycle of products. *American Economic Review*, 91(5): 1454–1477.

Duranton, G., Puga, D. (2004) Micro-foundations of urban agglomeration economies. In J. Thisse and J. V. Henderson (eds) *Handbook of Regional and Urban Economics*, Vol. 4, and, Elsevier.

Duranton, G., Jayet, J. (2005) Is division of labour limited by the extent of the cmMarket? Evidence from French Cities. Mimeo.

Edwards, A. M. (1943) Comparative occupation statistics for the United States, 1870 to 1940. US Bureau of the Census.

Faler, P. G. (1981) *Mechanics and Manufacturers in the Early Industrial Revolution: Lynn, Massachusetts, 1780–1860*. Albany, NY: SUNY Press.

Glaeser, E. L. (1999) Learning in cities. *Journal of Urban Economics*, 46(2): 254–277.

Glaeser, E. L. (2005) Reinventing Boston: 1630–2003. *Journal of Economic Geography* 5: 119–153.

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

Goldin, C., Sokoloff, K. L. (1984) The relative productivity hypothesis of industrialization: the American case, 1820–1850. *Quarterly Journal of Economics*, 99: 461–488.

Helsley, R. W., Strange, W. C. (1990) Matching and agglomeration economies in a system of cities. *Regional Science and Urban Economics* 20: 189–212.

Hirsch, S. E. (1978) *Roots of the American Working Class: The Industrialization of Crafts in Newark, 1800–1860*. Philadelphia: University of Pennsylvania Press.

Hunter, L. C. (1985). *A History of Industrial Power in the United States: Steam Power*, Vol. 2. Charlottesville, VA: University of Virginia Press.

Jovanovic, B. Rob, R. (1989) The Growth and Diffusion of Knowledge. *Review of Economic Studies*, 56(4): 569–582.

Kane, J. N. (1972) *The American Counties*, 3rd edn. Lanham, MD: Scarecrow Press.

Kim, S. (1989) Labor market specialization and the extent of the market. *Journal of Political Economy*, 97(3): 692–705.

Kim, S. (1990) Labor heterogeneity, wage bargaining, and agglomeration economies. *Journal of Urban Economics*, 28: 160–177.

Kim, S. (2000) 'Urban Development in the United States, 1690-1990,' *Southern Economic Journal*, 66(4): 855–880.

Kim, S. (2005) Industrialization and urbanization: did the steam engine contribute to the growth of cities in the United States? *Explorations in Economic History*, 42(4) 586–598.

Kim, S., Margo, R. (2004) Historical perspectives on U.S. economic geography. In V. Henderson and J. Thisse (eds) *Handbook of Regional and Urban Economics*, Vol. 4, North-Holland: Elsevier.

Lamoreaux, N. R., Sokoloff, K. L. (2000) Location and technological change in the American glass industry during the late nineteenth and early twentieth centuries. *Journal of Economic History*, 60: 700–729.

Levinsohn, J., Petrin, A. (2003) Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, 70(2): 317–342.

Moretti, E. (2004) Human capital externalities in cities. In J. V. Henderson and J. F. Thisse (eds.) *Handbook of Regional and Urban Economics*, Vol. 4. North-Holland: Elsevier.

Nelson, D. (1995) *Managers and Workers: Origins of the Twentieth Century Factory System in the United States 1880–1920*. Madison, WI: University of Wisconsin Press.

Olley, G. S., Pakes, A. (1996) The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(4): 1263–1297.

Palivos, T., Wang, P. (1996) Spatial agglomeration and endogenous growth. *Reg Sci Urban Econ* 26 (6): 645–669.

Pred, A. R. (1966) *The Spatial Dynamics of U.S. Urban-Industrial Growth, 1800–1914*. Cambridge, MA: MIT Press.

Prude, J. (1983) *The Coming of Industrial Order: Town and Factory in Rural Massachusetts 1810–1860*. Cambridge: Cambridge University Press.

Rosenberg, N., Trajtenberg, M. (2004) A general purpose technology at work: the corliss steam engine in the late-nineteenth-century United States. *J Econ History*, 64(1): 61–99.

Rosenbloom, J. L. (2002) *Looking for Work, Searching for Workers*. Cambridge: Cambridge University Press.

Slichter, S. H. (1919) *The Turnover of Factory Labor*. New York: D. Appleton and Co.

Sokoloff, K. L. (1984) Was the transition from the artisanal shop to the nonmechanized factory associated with gains in efficiency?: evidence from the U.S. manufacturing censuses of 1820 and 1850. *Explor Econ Hist*, 21: 351–382.

Sutthiphisal, D. (2005) The Geography of Invention in High- and Low-Technology Industries: Evidence from the Second Industrial Revolution. Mimeo.

Thomson, R. (1989) *The Path to Mechanized Shoe Production in the United States*. North Carolina, NC: University of North Carolina Press.

Ware, C. F. (1931) *The Early New England Cotton Manufacture*. Boston, MA: Houghton Mifflin Company.

Wheaton, W. C., Lewis, M. J. (2002) Urban wages and labor market agglomeration. *J Urban Econ*, 51: 542–562.

Wheeler, C. (2001) Search, sorting, and urban agglomeration. *J Labor Econ*, 19(4): 879–899.

Wheeler, C. (2004) Wage inequality and urban density. *J Econ Geogr*, 4: 421–437.