University of Kent

School of Economics Discussion Papers

The Rise and Fall of US Manufacturing: Re-Examination of Long-Run Spatial Trends

Nicholas Crafts and Alexander Klein

December 2019

KDPE 2009

The Rise and Fall of US Manufacturing: Re-Examination of Long-Run Spatial Trends[1]

Nicholas Crafts
(University of Sussex)

and

Alexander Klein
(University of Kent)

Abstract

We re-examine the long-run geographical development of U.S. manufacturing industries using recent advances in spatial concentration measures. We construct spatially-weighted indices of the geographical concentration of U.S. manufacturing industries during the period 1880 to 1997 using data from the Census of Manufactures and Bureau of Labor Statistics. Doing so we improve upon the existing indices by taking into account industrial structure and checkerboard problem. Several important new results emerge. First, we find that average spatial concentration was much lower in the late 20th- than in the late 19th-century and that this was the outcome of a continuing reduction over time. Second, spatial concentration of industries did not increase in early twentieth century as shown by traditional indices but rather declined, implying that we do not find an inverted-U shape pattern of long-run spatial concentration. Third, the persistent tendency to greater spatial dispersion was characteristic of most manufacturing industries. Fourth, even so, economically and statistically significant spatial concentration was pervasive throughout this period.

Keywords: manufacturing belt; spatial concentration; transport costs.

JEL Classification: N62; N92; R12.

[1] We are grateful to Sylvain Barde for his help with the estimation procedure to recover missing observations.

1. Introduction

It is well-known that patterns of regional specialization and the spatial concentration of American manufacturing industries have changed markedly over time. A standard narrative concerns the rise and fall of the manufacturing belt in the mid-19th century and second half of the 20th century, respectively. This is seen as a key aspect of a pattern of divergence followed by convergence of U.S. regions.

Kim (1995) provided a much-cited quantitative account of these trends. He calculated Hoover's coefficient of localization for 2-digit industries through time and found that the weighted average rose from 0.243 in 1880 to 0.316 in 1927 before falling to 0.197 in 1987. Krugman (2009) saw this experience in terms of new economic geography with the success of the manufacturing belt based on a phase of increasing returns in manufacturing but with the applicability of this model evaporating in the late 20th century.

In this paper we seek to re-examine and improve on these accounts. First, we take advantage of improved measurement techniques to estimate the extent of spatial concentration allowing for industrial structure and the *checkerboard problem*. To do this, we use an approximation to a spatially-weighted Ellison-Glaeser index. Second, we highlight the importance of changing locations patterns *within* the manufacturing belt, and the propensity of manufacturing to move *outside* the manufacturing belt already before World War II. We describe a clear tendency to spatial dispersion even during the heyday of a rising size of plants.

In order to re-examine long-run trends in the spatial concentration of U.S. manufacturing industries we construct a new dataset which permits the calculation of a spatially-adjusted version of the EG index at both SIC2 and SIC3 levels for selected census years from 1880 through 1997. To circumvent data limitations we use the spatially-weighted version of the Maurel and Sedillot (1999) adaptation of the EG index which does not require plant-level employment data. Construction of the index required assignment of industries into SIC categories and a procedure to deal with problems posed by withholding of data to prevent identification of individual firms.

Our main findings are as follows. First, the weighted average of the spatially-weighted EG index for SIC3 industries is at its maximum in 1880 at 0.223 after which it declines slowly to 0.183 in 1940 and then more rapidly to a low of 0.096 in 1997. Unlike Kim (1995), we do not find an episode of increasing spatial concentration in the early 20th century. Spatial-weighting is important in arriving at this conclusion. Second, increasing spatial dispersion over time is a general experience across American manufacturing industries over the long run and especially after 1940. At SIC2 level, all sectors have lower spatial concentration in 1997 than either in 1880 or in 1940 while 17 out of 20 industries were already more dispersed in 1940 than in 1880. At SIC3 level, in 14/20 SIC2 categories at least 2/3rds of the constituent SIC3 industries were more dispersed in 1997 than in 1880 while in 12/20 SIC2 categories the same was true for 1940 compared with 1880.

Third, even so, it is important to recognize that almost all SIC3 industries always exhibit spatial concentration in the sense that their spatially-weighted EG index score is positive and significantly different from zero. This is the case even at the end of the period when spatial concentration has generally declined. In fact, all 20 exceptions out of 1300 observations occur before 1947. The average of 0.096 in 1997 is at a level where it can be thought of as economically significant according to the criterion proposed by Ellison and Glaeser (1997). It would be incorrect to suppose that spatial concentration of manufacturing industry was no longer an important phenomenon in the late 20th century.

2. Literature Review

The relative decline of the manufacturing belt in the second half of the 20th century is well-known and features prominently in economists' reviews of the evolution of American industrial geography. Krugman (2009) in his Nobel Prize Lecture highlights that the manufacturing belt began to dissolve after World War II while Holmes and Stevens (2004) in their handbook chapter stress that as late as the 1950s manufacturing activity was still heavily concentrated in the North East and Upper Midwest around the Great Lakes in the manufacturing belt after which time it moved out and into other parts of the country. The data reported in Table 1 are consistent with these accounts in that they show 72.5 per cent of manufacturing employment was in the manufacturing belt in 1947 but this share fell to only 45.3 per cent in 1997.

That said, the economic geography literature has always recognised that the spatial distribution of manufacturing had evolved considerably before World War II. Already in the 1930s and 1940s geographers were discussing the 'decentralization of industrial activity'. Smith (1947) in a quantitative analysis of manufacturing employment commented on a steady movement in the direction of decentralization. Hoover (1948) noted a trend toward more equal inter-regional distribution of manufacturing for many decades prior to 1940 and pointed out that the locational histories of many industries involved an early stage of increasing concentration followed by a later stage of re-dispersion. Easterlin (1960) found that there had been a substantial shift in the location of manufacturing between 1869 and 1947 and calculated that a minimum of 30 per cent of wage earners in 1947 would need to be relocated to restore the 1869 percentage distribution by state.[2] (Ericksson et al. (2019) documented the spread of manufacturing

[2] Of the 30 per cent, 8.3 percentage points accrued between 1889 and 1909, 10.8 percentage points between 1909 and 1929 and 5.3 percentage points between 1929 and 1947.

between 1910 and 1940 noting the decline of New England and the Northern Great Lakes region and the expansion of the Southern Great Lakes region and most of the Appalachians.

This also is reflected in Table 1 where it is seen that New England declined from 24.0 per cent of manufacturing employment in 1880 to 12.5 per cent in 1940 while the East North Central region rose from 19.1 per cent to 27.9 per cent. Table 1 also shows that there had been a notable decrease in the share of the manufacturing belt between these two dates from 87.2 per cent to 73.6 per cent. The point to note is that while the manufacturing belt still accounted for much of American manufacturing employment in the 1940s it was already a good deal less dominant than in the 1880s.

These developments in shares of manufacturing employment were related to the pattern of trade within the United States. By 1949, the earliest date for which railroad freight data are available, as is reported in Table 2, the East North Central region was responsible for more inter-state trade than New England and Middle Atlantic combined while West North Central and West South Central together exceeded Middle Atlantic while accounting for 21.7 per cent of trade despite having only 9.4 per cent of employment, and California was the source of nearly as much inter-state trade (3.3%) as New England (3.5%).

A staple finding of the literature on the location of manufacturing is that industries with larger plant sizes tend to have higher levels of geographic concentration (Holmes and Stevens, 2004). The basic new economic geography model reviewed by Krugman (2009) predicts that industry will concentrate in the core region with the best market access if economies of scale are large relative to transport costs. Kim (1995) pointed to a rise in the scale of production as reflected by the size of plants measured in terms of employment as a key factor in first rising and then declining spatial concentration over the course of the 20th century. Table 3 reports average plant size at the SIC 2-digit level and this confirms that plant sizes were generally rising until at least the 1940s but were generally falling in the decades towards the end of the century. However, a quite important point to note is that even prior to World War II an increasing number of locations had market sizes which could support large scale production. For example, Rhode (2001) stresses that this was true of California by the 1920s and 1930s where the automobile and tire industries constructed plants at that time. In 1949, shipments of cars from California to Oregon and Washington were 20 times those from Michigan while within California shipments by rail were 10 times those from Michigan to California (ICC, 1951).

Nevertheless, Kim (1995) found that spatial concentration was increasing in the early decades of the 20th century. He calculated Hoover's coefficient of localization for 2-digit industries and found that the unweighted and weighted average figures rose from 0.243 in 1880 to 0.327 in 1947 and from 0.242 in 1900 to 0.316 in 1927, respectively, before subsequently declining. However, since Kim wrote his paper, which has become the standard reference on the topic, there have been important developments in the measurement of spatial concentration which suggest that a new look is required.

Ellison and Glaeser (1997) explained that it is important to control for differences in the size distribution of plants to obtain a meaningful measure of spatial concentration and developed an index in which raw geographic concentration is modified by taking account of the plant Herfindahl index.[3] An important refinement to the basic EG index is to take account of the geographical position of regions through allowing for 'neighbourhood effects'. This leads to the spatially-weighted version of the EG index proposed by Guimarães et al. (2011) which represents a significant advance on Hoover's localization coefficient.

Guimaraes et al. (2011) pointed out that the Ellison-Glaeser index does not consider the geographic position of regions. They illustrated the problem using the diagram reproduced

[3] An industry in which production comes from very few plants will appear as spatially concentrated even if it is randomly located.

here as Figure 1. It is intuitively obvious that spatial concentration is greater in Figure 1a than in Figure 1b – spillovers across regional boundaries would seem much more likely in the former case - but they will be recorded as having the same spatial concentration by the Ellison-Glaeser method. This is an example of what has become known as the 'checkerboard problem'. To address this, spatial weighting is required.

In fact, the checkerboard problem is quite important in the early decades of the 20th century notably in the context of movement within the manufacturing belt from the north-east to the mid-west. For example, in the case of SIC 359 (industrial machinery nec), Map 1 shows that, in 1900, the industry is concentrated in several states most of which are adjacent. Map 2 shows that this changed by 1920 when three separate pockets of employment emerged: New York, Illinois, and Michigan. This is a typical checkerboard problem. Since the EG index does not take the geographical position of individual states into account, it misinterprets the concentration of employment into fewer states as a sign of higher geographical concentration even though those states are geographically disjointed, an error which is corrected by spatial weighting.

3. Methodology

The spatially-adjusted version of the EG index proposed by Guimarães et al. (2011) addresses the checkerboard problem by taking neighbourhood effects into account. The index is defined as follows

$$\gamma_i^{SEG} = \frac{G_i^S - H_i(1 - X'\Psi X)}{(1 - H_i)(1 - X'\Psi X)} \qquad (1)$$

where H_i is a Herfindahl index measuring the industry concentration at plant level, $G_i^S=(S-X)'\Psi(S-X)$ is the spatially weighted version of the geographical index where the vector S is the fraction of employment in industry i across geographical areas j, $X'=[x_1, x_2,\ldots, x_j]$ is the vector of the aggregate employment across geographical areas *j* and Ψ is a spatial weight matrix. Ψ is defined as $\Psi=W+I$ where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{W}$ is a weight matrix for adjacent regions.

We implement variants of this approach. Our main results are derived using a first-order contiguity matrix $\boldsymbol{W}$ defined such that each element takes one for contiguous US states and zero otherwise. As a robustness check, we also use an alternative spatial weighting also suggested by Guimarães et al (2011). In particular, we consider spatial matrices in which neighbours are identified using a pre-defined bandwidth: a spatial unit *j* is considered a neighbour of a spatial unit *i* if the distance between their centroids is less than the pre-defined bandwidth *b*. We discuss this in detail later.

A problem in using the EG index to study the long-run development of spatial concentration is that it requires plant-level employment data. These are not available for the entire period under study. Fortunately, Maurel and Sedillot (1999) (henceforth MS) developed a version of the EG index where the Herfindahl index H_i is replaced by $1/N_i$ (N_i is the number of plants in industry *i*), and where the vector *S* is defined as the fraction of *plants* in industry *i* across geographical areas *j*. They show that their index is an unbiased estimator of the EG index. This allows us to circumvent the problem of the lack of plant-level employment data and we can calculate the MS index for the entire period 1880-1997. Guimarães et al (2011) also provide a spatially-weighted version of the MS index (henceforth SMS) which is defined as follows:

$$\gamma_i^{SMS} = \frac{N_i G_i^S - (1 - X'\Psi X)}{(N_i - 1)(1 - X'\Psi X)} \qquad (2)$$

The formula for the SMS index in equation (2) is the main focus of our analysis. When $\Psi=I$, the index collapses into a standard spatially unweighted EG index. To facilitate a comparison with the spatially weighted index, we also present results for the MS index which is defined as

$$\gamma_i^{MS} = \frac{N_i G_i - (1 - X'X)}{(N_i - 1)(1 - X'X)} \tag{3}$$

where $G_i=(S-X)'(S-X)$ with S and X defined as above.

We will perform a statistical analysis using a one-sided statistical test assuming that γ_i is asymptotically normally distributed. Ellison and Glaeser (1997) derive a formula for the variance of EG index under the null hypothesis that γ_i=0 and the spatially-weighted version is provided by Guimarães et al (2011):

$$V\left(\hat{\gamma}_i^{SMS}\right) = \frac{2H_i^2 tr\left[\Psi[\mathrm{diag(X)}-XX']\Psi[\mathrm{diag(X)}-XX']\right]}{[(1-H_i)(1-X'\Psi X)]^2} \tag{4}.$$

4. Data Sources

We analyse the evolution of the spatial concentration of SIC 2- and SIC 3-digit level industries across 48 U.S. states in every decade between 1880 and 1997, specifically for the following years: 1880, 1890, 1900, 1910, 1920, 1930, 1940, 1947, 1958, 1967, 1977, 1987, 1997. The construction of the indices requires data on employment and on the number of plants by U.S. states at SIC 2- and SIC 3-digit level industries, and also a spatial weight matrix. The spatial weight matrix for 48 U.S. contiguous states was obtained from the REPEC data repository.[4] The data on U.S. state-industry employment and number of plants were collected from the U.S. Census of Manufactures for the period 1880-1967 and from the Bureau of Labor Statistics for the years 1977-1997.

The construction of the EG index over the period of 120 years presents three challenges. First, we need to harmonize SIC 2- and SIC 3-digit level industries across time. Harmonization of the data for the post World War II period is straightforward as the Census of Manufactures

[4] Following Guimarães et al (2011) we used the usswm package developed by Scott Merryman; the original spatial weight matrix was created by Luc Anselin.

reports the SIC industrial categories and a great deal of information was published about changes in SIC classifications between 1947 and 1997. There are no SIC codes reported in the Censuses before 1947. Here we use the assignment of industries into SIC 2- and 3-digit categories created by Klein and Crafts (2012) and by Klein and Crafts (2019) for the years 1880, 1890, 1900, 1910, 1920, 1930, and 1940. Details of the harmonization of SIC 3-digit industries are in the Appendix 3. Second, construction of the Herfindahl index requires data on employment in plants. Ellison and Glaeser (1997) used data from the 1987 Census of Manufactures which reports employment in plants belonging to 10 employment size categories. Unfortunately, the Census of Manufactures does not report plant employment data before 1947. Therefore, we use the MS index and the spatially-adjusted version of it (SMS) which require only the number of plants, making it feasible to construct the indices all the way back to 1880. Third, when there are issues about disclosure of information on individual companies, the Census either withholds the data or reports the data in employment classes. Similarly, the Bureau of Labor withholds information in order to protect the identity or identifiable information of individual firms. Hence we have incomplete state-industry employment and plant data. Fortunately, the data are in the form of matrices with rows being totals for U.S. states and columns totals for U.S. industries. This allowed us to take advantage of a methodology developed in Golan et al. (1994). They use a maximum entropy procedure to recover missing data in multi-sectoral matrices with information about row and column sums as well as information contained in the multi-sectoral matrices. In our case, we used across-state and across-industry adding-up constraints to recover missing information on state-industry employment and plant data.

5. Results

We first report the results of the weighted average SMS index for all SIC3 industries over the long run where the weights are the shares of employment in SIC3 industry, robustness checks with respect to the spatial matrix, and a comparison with the original, spatially EG unweighted index. The weighted average SMS Index is reported in Table 3, column I and we plot it in Figure 2 as well. The highlight of this longer-term account is that the levels of spatial concentration were considerably higher (almost twice as large as in 1997) in the early decades of the 20th century through to 1940 and then fell quite rapidly after World War II. Furthermore, mean spatial concentration for SIC3 industries was distinctly lower in 1930 and 1940 than in 1880. Although the rate of decrease of mean SMS accelerated after 1940, about 30 per cent of the total fall between 1880 and 1997 had already occurred by 1940. Overall, our estimates show that spatial concentration of industries was much more prevalent in the late 19th- than in the late 20th-century.[5]

It is interesting to compare these results with the (spatially unweighted) MS index, other EG-type indices in the literature, and Kim (1995). The MS index is presented in Table 4, column II, and in Figure 2: this shows a similar proportionate decline in geographical dispersion between 1910 and 1997. Unlike the SMS index, however, the MS index shows an increase in spatial concentration in the early 20th century. A comparison with the EG averages reported by Dumais et al. (2002) for the years 1972 to 1992 reveals that our estimates are somewhat larger but show a similar decrease in this period. Contrary to Kim (1995), who reported the weighted average of Hoover's coefficient of localization for SIC2 industries which is presented in Figure 3, we do not find an episode of increasing spatial concentration in the 1910s and 1920s when looking at the SMS index. However, Figure 3 reveals that the MS index shows an

[5] We also explored alternative methods of spatial weighting as a robustness test, see Appendix 1.

increase in spatial concentration in the first quarter of the century similar to that reported by Kim.

SMS estimates for all SIC2 industries are reported in Table 5. A general tendency to greatly increased spatial dispersion over time is clear; in every case, the SMS index was lower in 1997 than in either 1880 or 1940 and in all but one sector the reduction was at least 40 per cent. The highest SMS score in 1997 (0.17) would have been the second lowest in 1880. In the vast majority of sectors (17/20), there was already dispersion between 1880 and 1940. The smallest percentage decrease in the SMS index between 1880 and 1997 is in SIC 22, textile mill products, while the largest reductions are in SIC 30, rubber and plastic products, SIC 35, machinery, SIC 36, electrical equipment, and SIC 37, transportation equipment.

The experience of changing spatial concentration at SIC3 level is summarized in Table 6. In 14/20 SIC2 categories at least 67 per cent of the constituent SIC3 industries were more dispersed in 1997 than in 1880 while in 12/20 SIC2 categories the same was true for 1940 compared with 1880. So, there was quite a high incidence of spatial dispersion but it was by no means universal.[6]

The evolution of spatial concentration in three groups of industries, those whose origins were in the first industrial revolution, those from the second industrial revolution and those from the ICT revolution, is displayed in Figure 4. In each case, spatial concentration starts out quite high and then decreases much as Hoover (1948) suggested. Interestingly, the second industrial revolution industries are dispersing continuously from 1910 onwards and the ICT industries are the least spatially concentrated of the three groups in the late 20th century. The latter fact is due to the fact that ICT industries were developing in three geographically disjoined states of

[6] Our results do not lend support to the hypothesis of stability in geographic concentration advanced by Dumais et al. (2002). See Appendix 2.

Texas, California, and later Washington, which is controlled for in our spatially weighted version of Ellison and Glaeser index.

Although we have stressed that there was a strong tendency for spatial concentration of industries to decline over time, especially after 1940, it is important to recognize that even at the end of our period there was a very high incidence of localization at the SIC3 level. Spatial concentration was almost always present to an extent which was both statistically and economically significant. We have tested the statistical significance using equation 4 under the null hypothesis that SMS index is equal to zero. We can reject the null hypothesis at the 1 percent significance level in all but 20 instances (none after 1940). Table 7 lists all the cases where the SMS index is not statistically significantly above zero.

Furthermore, Figure 5 displays kernel distributions for SMS for selected years with the charts on the right truncated at zero for 1880 and 1940. It is apparent that, with spatial weighting, there are very few observations below 0.05, the conventional level described as 'highly concentrated' and, as we saw in Table 3, the mean SMS at SIC3 level is way above 0.05 throughout the period. The criterion of 0.05 was originally chosen by Ellison and Glaeser (1997) because it is consistent with the existence of substantial local cost advantages. Therefore, our results imply that economically significant spatial concentration was the norm across industry continuously from 1880 through 1997.

We see that spatial weighting makes a considerable difference to the estimated evolution of spatial concentration in the early decades of the 20th century. Even though the SMS index is methodologically superior to the MS index, it is informative to explore what drives the differences between MS and SMS in these years. Figure 2 shows that unlike the SMS index, the MS index increases appreciably between 1900 and 1910 and then plateaus throughout the interwar period. The SMS index, on the other hand, decreases gradually between 1920 and

1940. Examining the sectoral composition of the MS index reveals that the increase of the index in the period 1910-1940 is, on average, driven by half of all SIC3 industries: 55, 49, 50 and 44 percent of all SIC3 industries in 1910, 1920, 1930, and 1940 respectively. In these industries the MS index *doubled* relative to 1900; more precisely, the percentage increase of the MS index in 1910, 1920, 1930, and 1940 relative to 1900 is 105, 92, 102, 103 percent respectively. These industries come from the industrial groups of SIC25 (furniture), SIC 28 (chemicals), SIC 35 (machinery), SIC 36 (electrical), and SIC 39 (miscellaneous). The SMS index, however, does not show that those industries experienced such an increase. The reason is that while the SMS index corrects the checkerboard problem which looms quite large at this point, the MS index does not.

As an illustration, Maps 1-9 show an example of three such industries. They plot the geographical index G_i of SIC 359 (industrial machinery nec), SIC 364 (electrical lighting and wiring equipment), and SIC 393 (musical instruments) for the years 1900, 1920, and 1940. Let us remind ourselves that $G_i = (S-X)'(S-X)$ where the vector S is the fraction of employment in industry i across geographical areas j and $X' = [x_1, x_2, \ldots, x_j]$ is the vector of the aggregate employment across geographical areas j. As we noted in the section on methodology, the limitation of G_i is that it does not consider the geographical position of regions, thus creating the checkerboard problem. Maps 1-9 show prevalent checkerboard problems throughout the entire period 1900-1940. There are two distinct patterns: one between 1900 and 1920, the other between 1920 and 1940. As for 1900-1920, we see that even though the employment in these industries increased in a few individual states in the Middle Atlantic and Midwest regions between 1900 and 1920, those states are geographically *disjointed*. As for 1920-1940, we see that while the increase in employment led to create a contiguous area of Middle Atlantic and Midwest regions, we see a substantial increase of employment in California and in some case Texas and Washington, cause the checkerboard problem again because these two dominant regions of employment are, *again*, geographically disjointed: one in the East and Midwest, the other in the West and South.

For example, in the case of SIC 359 (industrial machinery nec), Map 1 shows that in 1900, the industry is concentrated in a few states which form two disjointed pockets of employment: one consisting of the states of New York, Ohio, and Pennsylvania, the other of Illinois. Map 2 shows that this changed by 1920 when three disjoined pockets of employment emerged: New York, Illinois, and Michigan. This is a typical checkerboard problem. By 1940, these three disjointed pockets of employment formed a contiguous area, but since Texas, Washington, and California experienced a rapid rise of the industrial machinery industry as well, the checkerboard problem re-emerged. Since the MS index does not take the geographical position of individual states into account, it misinterprets the concentration of employment into fewer states as a sign of high geographical concentration even though those states are geographically disjointed. This is reminiscent of the problem illustrated by Figure 1.

The cases of SIC 364 (electric lighting and wiring equipment) in Maps 4-6, and SIC 393 (musical instruments) in Maps 7-9 are similar. Indeed, Map 4 shows that in 1900, there were two disjointed clusters of a few states with employment in electric lighting concentrated around Illinois and the state of New York, respectively. By 1920, as Map 5 shows, this feature becomes even more pronounced and employment concentrates mostly in New York, Ohio, and Illinois – states without adjacent borders. The year 1940 then shows the checkerboard problem being driven by California. Lastly, the musical instruments industry shows even more profoundly how employment in a handful of states which are geographically far apart can be misinterpreted by the MS index as high geographical concentration. Indeed, Maps 7 and 8 show that the industry is concentrated predominantly in two states – New York and Illinois – but those states are geographically far apart without adjacent state borders. Map 9 then shows a similar development as in case of SIC 364 – increase in employment concentration in

California. As a consequence, the MS indices consider these industries as highly geographically concentrated. Spatial weighting, however, takes into account the checkboard problem, correcting thus their bias.

6. Discussion

A notable implication of our results is that forces promoting the spatial dispersion of American manufacturing were present throughout the 20th century. The most important of these was surely the continuing long-run decline of transport costs first in the railroad era and then sustained by trucking. Lower shipping costs for goods meant that manufacturing could move out of the large industrial cities in which it concentrated at the start of the 20th century (Glaeser and Kolhase, 2004). Market potential would matter less and high wage costs in production would matter more and this eroded the advantages of the manufacturing belt. The ratio of the average wage in states in the manufacturing belt compared with other nearby states followed an inverted-U shape with its peak in 1940. Over the long run, industrial location continually evolved as fundamentals changed.

Glaeser and Kohlhase (2004) noted that the costs of moving manufactured goods declined by over 90 per cent in real terms between 1890 and 2000 from 18.5 cents per ton-mile to 2.3 cents (at 2001 prices). In fact, much of this decrease occurred by 1967 when the cost was only 5.6 cents (at 2001 prices) and by 1891 the railroad revolution had cut transport costs to about 10 per cent of the 1820s' level.[7] We calculate that the ratio of the average wage in manufacturing

[7] These estimates of transport costs are based on Carter et al. (2006), volume 4, pages 781 and 932-934.

in East North Central and Mid-Atlantic states relative to East and West South Central states rose from 1.22 in 1890 to 1.52 in 1940 before falling to 1.15 in 1987.[8]

Average plant size according to our estimates from the Census of Manufactures rose from 11.0 in 1880 to 60.6 in 1947, after which it stayed on a plateau until 1977 when it was 62.7 before falling to 46.1 in 1997. As many writers including Kim (1995) have noted, the decrease in plant size in the later 20th century was conducive to lower spatial concentration. In the period of rising plant size combined with spatial dispersion prior to World War II, the point to note is that the rise of the mid-west relative to the north-east which tended to lower SMS scores was associated with establishment of larger plants. By 1940, 14 SIC 2-digit industries out of 20 had a larger average plant size in the mid-west than in New York whereas in 1880 that was true of only 3 of the 20.

So, in the long run the locational advantages of agglomeration in the manufacturing belt were undermined by rising wage costs, falling transport costs and a reduction in average plant size. In some respects, this combination of changes over time is reminiscent of the later phase of the stylized core-periphery model presented by Krugman and Venables (1995). This model would see a move from very high to intermediate to very low transport costs driving a move from dispersed to spatially concentrated then back to dispersed locations for manufacturing. In the spatially concentrated (manufacturing belt) phase the core benefits from economies of scale and proximity to markets and suppliers raise productivity but also tend to raise wages; subsequently, however, in the context of much lower transport costs, the wage gap becomes too high and moves to the periphery promote a convergence of wage rates.

[8] The former group of states comprises Illinois, Indiana, Michigan, New Jersey, New York, Ohio, Pennsylvania, and Wisconsin while the latter comprises Alabama, Arkansas, Kentucky, Louisiana, Mississippi, Oklahoma, Tennessee and Texas. The average wage rates are obtained by dividing the wage bill by the number of workers in the Census of Manufactures.

Recent research has produced empirical results which are broadly consistent with a core-periphery model. Klein and Crafts (2012) found that the location of manufacturing in the early 20th century was strongly influenced by the attraction of market potential to industries with large plants and strong linkages with industrial customers and suppliers. This pattern underpinned the existence of the manufacturing belt. Crafts and Klein (2015) found that home bias in U.S. domestic trade was much lower in 1949 than in 2007. In 1949, some commodities actually exhibited negative home bias at a time when the ratio of inter- to intra-state trade was much higher and much production in the manufacturing belt was still exported to the rest of the United States. They showed that in 1949 home bias was inversely correlated with geographic concentration of industries. This configuration had, however, evaporated by 2007.

Reality was often more complex but reflected similar issues. An excellent example of this is Motor Vehicles and Equipment (SIC 371) where overall geographic concentration fell in the second half of the 20th century but where significant localization persisted in a new configuration. The SMS index for SIC 371 was 0.191 in 1940, 0.120 in 1958, 0.106 in 1977 and 0.094 in 1997. Maps 10 to 13 show an evolving pattern of its spatial concentration over time such that by 1997 the move away from the 1940 situation of a dominant position for Michigan and an east-west corridor in the southern Great Lakes region has been superseded by one in which Michigan is still a major centre but clusters within 'Auto Alley' extend as far south as Alabama (Klier and Rubenstein, 2008). Two key developments that underlay these changes were the switch of assembly plants in the 1960s away from the coasts to central areas to reduce the costs of transporting cars to customers once these plants became specialized in models for sale throughout the United States and the advent of Japanese producers in the 1980s and 1990s who chose to locate further south – initially Kentucky and Tennessee and then in the deep south. Throughout, parts suppliers wanted to locate close to auto producers. Transport costs were instrumental in some of these decisions but the move to the south by the Japanese was encouraged by a quest for lower labour costs.

Besides contributing to the checkerboard problem, the ascent of California as a manufacturing location adds to the richness of the historical picture. Initially, Californian manufacturing was based mainly on resource-processing industries but already by the late-1930s it was developing a significant presence in knowledge-based industries and a comparative advantage based on human capital and localized technological spillovers, first in aircraft followed by electronics and information technology (Rhode, 2001). A good post-war example can be found in the semi-conductor industry where spatial dispersion took place over the long run in the context of a reconfiguration of the sector driven by technological change. The key development was the advent of the integrated circuit in 1959 which was discovered in California and Texas. This triggered a long-term move to those states and away from Massachusetts and New York where, in the 1950s, semiconductors were produced by vacuum tube manufacturers. \nevertheless, the industry continued to experience a significant level of localization in which knowledge spillovers and proximity to buyers played a big part (Ketelhöhn, 2006).

In the context of a general move towards greater spatial dispersion, it is noteworthy how weak correlations of localization at the industry level were over time. Even so, it is striking how pervasive significant excess spatial concentration has been throughout our period. As the manufacturing belt lost its manufacturing dominance the new locational patterns saw new pockets of spatial concentration emerge rather than a scattering of plants across the rest of the country. Nevertheless, it seems quite possible that the underlying reasons for concentration have changed over time and that individual-industry experiences provide many variations on this theme. These are important topics for future research.[9]

7. Conclusions

We have constructed spatially-weighted indices of geographic concentration of SIC2 and SIC3 manufacturing industries in the United States over the period 1880 to 1997 and have shown that this is possible notwithstanding data constraints. These estimates embody recent methodological innovations. We offer a new and improved perspective on long-run trends in spatial concentration of American manufacturing. We show that it is very important to use spatial-weighting in order to achieve this. This leads us to a very different picture of long-run trends in spatial concentration from that which was found by Kim (1995); we do not find an inverted-U shape.

The first striking feature of our estimates is that by the end of the 20th century average levels of spatial concentration in manufacturing were much lower than in the late 19th century. The weighted average for SIC3 industries for the SMS index was 0.096 in 1997 compared with 0.223 in 1880. Although spatial concentration fell more rapidly after World War II, a significant decrease had already taken place by 1940 in the context of an early decline in the importance of the manufacturing belt and a switch towards the mid-west within the manufacturing belt. A second important point is that this experience is characteristic of the vast majority of SIC2 industries. It is also notable that correlations over time of our index of geographic concentration are quite low. The third major finding that comes from our estimates is that ‘excess’ spatial concentration is pervasive at the SIC3 level throughout the whole period.

[9] For example, as one of the founding fathers of the ‘new economic geography’ reflected, its models may have more salience to the era of the manufacturing belt than the present day (Krugman, 2011).

Across almost all industries and all years, spatial concentration is significant both statistically and economically.

References

Carter, S.B., Gartner, S.S, Haines, M.R., Olmstead, A.L., Sutch, R., Wright, G. (eds.) (2006) *Historical Statistics of the United States: Earliest Times to the Present*. Cambridge: Cambridge University Press.

Crafts, N., Klein, A. (2015) Geography and intra-national home bias: U.S. domestic trade in 1949 and 2007. *Journal of Economic Geography*, 15: 477-497.

Dumais, G. Ellison, G., Glaeser, E.L. (2002) Geographic concentration as a dynamic process. *Review of Economics and Statistics*, 84: 193-204.

Easterlin, R.A. (1960) Redistribution of manufacturing. In S. Kuznets, A.R. Miller and R.A. Easterlin (eds.), *Population Redistribution and Economic Growth*, vol. 2, pp. 103-139. Philadelphia: American Philosophical Society.

Ellison, G., Glaeser, E.L. (1997) Geographic concentration in U.S. manufacturing industries: a dartboard approach. *Journal of Political Economy*, 105: 889-927.

Eriksson, K., Russ, K.N., Shambaugh, J.C., Xu, M. (2019) Trade shocks and the shifting landscape of U.S. manufacturing. NBER Working Paper No. 25646.

Glaeser, E. L., Kohlhase, J. E. (2004) Cities, regions and the decline of transport costs. *Papers in Regional Science*, 83: 197-228.

Golan, A., Judge, J., Robinson, S. (1994) Recovering information from incomplete or partial multi-sectoral data. *Review of Economics and Statistics*, 74: 541-549.

Guimarães, P., Figueiredo, O., Woodward, D. (2011) Accounting for neighboring effects in measures of spatial concentration. *Journal of Regional Science*, 51: 678-693.

Holmes, T. J., Stevens, J.J. (2004) Spatial distribution of economic activities in North America. In J.V. Henderson, J. F. Thisse (eds.), *Handbook of Regional and Urban Economics*, vol. 4, pp. 2797-2843. Amsterdam: Elsevier.

Hoover, E.M. (1948) *The Location of Economic Activity*. New York: McGraw-Hill.

Interstate Commerce Commission (1951), *Carload Waybill Statistics, 1949*. Washington, DC.: Bureau of Transport Economics and Statistics.

Ketelhöhn, N.W. (2006) The role of clusters as sources of dynamic externalities in the US semiconductor industry. *Journal of Economic Geography*, 6: 679-699.

Kim, S. (1995) Expansion of markets and the geographic distribution of economic activities: the trends in U.S. regional manufacturing structure, 1860-1987. *Quarterly Journal of Economics*, 110: 881-908.

Klein, A., Crafts, N. (2012) Making sense of the manufacturing belt: determinants of U.S. industrial location, 1880-1920. *Journal of Economic Geography*, 12: 775-807.

Klein, A., Crafts, N. (2019) Agglomeration externalities and productivity growth: U.S. cities, 1880-1930. *Economic History Review*, forthcoming.

Klier, T. and Rubenstein, J. (2008) *Who Really Made Your Car?* Kalamazoo: W.E. Upjohn Institute for Employment Research.

Krugman, P. (2009) The increasing returns revolution in trade and geography. *American Economic Review*, 99: 561-571.

Krugman, P. (2011) The new economic geography, now middle-aged," *Regional Studies*, 45: 1-7.

Krugman, P., Venables, A. J. (1995) Globalization and the inequality of nations. *Quarterly Journal of Economics*, 110: 857-880.

Maurel, F., Sedillot, B. (1999) A measure of the geographic concentration in French manufacturing industries. *Regional Science and Urban Economics*, 29: 575-604.

Rhode, P. (2001) *The Evolution of California Manufacturing*. San Francisco: Public Policy Institute of California.

Smith, G.C. Jr. (1947) Lorenz curve analysis of industrial decentralization. *Journal of the American Statistical Association*, 42: 591-596.

Table 1. Regional Shares of Manufacturing Employment (%).

	1880	1890	1900	1910	1920	1930	1940	1947	1958	1967	1977	1987	1997
Manufacturing Belt	**87.2**	**81.2**	**80.3**	**78.3**	**78.1**	**75.1**	**73.6**	**72.5**	**63.8**	**61.7**	**54.5**	**48.8**	**45.3**
New England	24	19.5	18.3	17.3	15.7	12.9	12.5	10.3	8.6	8.2	7.2	7.2	5.6
Middle Atlantic	37.5	34.2	34.2	33.8	32.1	29.2	27.9	27.7	24.4	22.3	18.2	15.5	12.6
East North Central	19.1	23.3	23.4	22.8	26.6	29	27.9	30.3	26.6	26.9	24.9	21.8	23.2
South Atlantic (part)	6.6	4.2	4.4	4.4	3.7	4	5.3	4.2	4.2	4.3	4.2	4.3	3.9
Non-Manufacturing Belt	**12.6**	**18.8**	**18.7**	**21.8**	**21.9**	**24.8**	**26.4**	**27.5**	**36.1**	**38.2**	**45.3**	**51.2**	**54.7**
South Atlantic (part)	2.3	3.4	4.7	5.5	4.9	6.2	7.4	6.3	8.3	8.5	10.4	12.2	12.2
West North Central	4.5	6.8	5.8	5.4	5	5.1	4.9	5.5	7.2	6.3	6.7	7	8
East South Central	2.7	3.6	3.6	3.8	3.3	4.1	4.7	4.4	4.8	5.6	7	7.1	7.8
West South Central	1	1.8	2.2	3	3	3.2	3.3	3.9	4.8	5.7	7.5	7.8	9.2
Mountain	0.4	0.6	0.8	1	1	1	0.9	1	1.4	1.6	2.4	3.2	4
Pacific	1.7	2.6	2.6	3.1	4.7	5.2	5.2	6.4	9.6	10.5	11.3	13.9	13.5

Notes: South Atlantic states inside the manufacturing belt are Delaware, Maryland, Virginia and West Virginia.

Source: US Census of Manufactures.

Table 2. Inter-State Trade in Manufactures in 1949.

US Region	% carloads originating in each region
Manufacturing Belt	*62.3*
New England	3.5
Middle Atlantic	20
East North Central	32.4
South Atlantic (part)	6.4
Non-Manufacturing Belt	*37.7*
South Atlantic (part)	3.8
West North Central	9
East South Central	5.8
West South Central	12.7
Mountain	1.8
Pacific	4.6

Source: Interstate Commerce Commission (1951)

Table 3. Average Plant Size in SIC 2-Digit Industries: Number of Production Workers.

	1880	1920	1947	1967	1997
Food & Kindred Products	5.09	10.92	35.52	50.63	72.44
Tobacco & Tobacco Products	13.03	14.9	105.23	246.23	243.98
Textile Mill Products	76.4	156.02	151.03	131.59	90.89
Apparel & Related Products	32.67	29.21	35.21	51.7	32.9
Lumber & Wood Products	6.79	18.6	24.24	15.12	20.21
Furniture & Fixtures	10.31	30.8	42.55	42.88	45.47
Paper & Allied Products	30.23	78.31	110.56	108.97	92.56
Printing & Publishing	17.43	8.84	24.54	27.05	23.1
Chemicals & Allied Products	15.75	35.37	62.84	71.63	73.28
Petroleum & Coal Products	12.78	46.65	155.03	76.83	57.49
Rubber & Plastic Products	113.88	329.79	352.49	80.53	56.99
Leather & Leather Products	6.32	55.12	76.91	90.5	46.13
Stone, Clay & Glass Products	12.9	29.68	43.99	37.07	32.7
Primary Metal Products	44.47	57.79	237.69	187.53	96.07
Fabricated Metal Products	3	56.23	65.04	49.04	39.51
Machinery	20.41	94.98	91.69	49.44	35.13
Electrical Engineering	24.03	119	202.77	175.55	87.42
Transportation Equipment	13.28	172.94	319.24	247.18	121.54
Instruments & Related Products	5.52	19.68	90.78	89.21	64.18
Miscellaneous Manufacturing	16.91	40.07	33.12	30.27	21.98

Source: US Census of Manufactures.

Table 4. MS and SMS Indices, SIC 3-Digit Industries, 1880-1997

Year	SMS mean (standard deviation)	MS mean (standard deviation)
1880	0.223 (0.150)	0.104 (0.093)
1890	0.204 (0.129)	0.098 (0.159)
1900	0.207 (0.117)	0.096 (0.136)
1910	0.206 (0.156)	0.123 (0.218)
1920	0.203 (0.094)	0.121 (0.139)
1930	0.190 (0.089)	0.119 (0.142)
1940	0.183 (0.116)	0.118 (0.150)
1947	0.163 (0.056)	0.103 (0.109)
1958	0.143 (0.046)	0.088 (0.084)
1967	0.122 (0.059)	0.079 (0.073)
1977	0.115 (0.030)	0.067 (0.072)
1987	0.102 (0.029)	0.069 (0.059)
1997	0.096 (0.024)	0.063 (0.043)

Note: mean values are weighted averages using employment shares as weights.
Source: own calculations, see the text.

Table 5. SMS Index Estimates, SIC2-Level Industries, 1880-1997

Sic 2 industry code	SIC 2 Industry	1880	1890	1900	1910	1920	1930	1940	1947	1958	1967	1977	1987	1997
20	Food and kindred product	0.16	0.17	0.16	0.15	0.14	0.12	0.12	0.11	0.11	0.1	0.1	0.09	0.05
21	Tobacco and tobacco product	0.23	0.24	0.22	0.21	0.21	0.25	0.24	0.19	0.17	0.17	0.13	0.1	0.13
22	Textile mill product	0.22	0.15	0.16	0.21	0.24	0.2	0.21	0.26	0.23	0.21	0.2	0.18	0.17
23	Apparel and related products	0.25	0.2	0.19	0.26	0.29	0.26	0.25	0.24	0.24	0.22	0.17	0.11	0.07
24	Lumber and wood products	0.17	0.15	0.15	0.14	0.14	0.13	0.11	0.13	0.11	0.12	0.11	0.1	0.1
25	Furniture and fixtures	0.19	0.21	0.19	0.17	0.17	0.14	0.14	0.12	0.11	0.1	0.09	0.09	0.08
26	Paper and allied products	0.32	0.33	0.3	0.28	0.27	0.22	0.21	0.19	0.16	0.14	0.12	0.11	0.1
27	Printing and publishing	0.19	0.15	0.14	0.13	0.13	0.13	0.13	0.12	0.12	0.11	0.1	0.09	0.06
28	Chemicals and allied products	0.21	0.23	0.22	0.22	0.17	0.16	0.13	0.13	0.12	0.11	0.11	0.09	0.08
29	Petroleum and coal products	0.18	0.18	0.16	0.37	0.18	0.19	0.12	0.13	0.11	0.11	0.11	0.1	0.09
30	Rubber and plastic products	0.39	0.34	0.34	0.3	0.19	0.17	0.19	0.2	0.14	0.12	0.1	0.1	0.08
31	Leather and leather products	0.18	0.2	0.17	0.19	0.2	0.27	0.25	0.28	0.23	0.21	0.14	0.11	0.09
32	Stone, clay, and glass products	0.19	0.18	0.17	0.16	0.17	0.16	0.15	0.14	0.11	0.11	0.1	0.1	0.09
33	Primary metal products	0.23	0.22	0.2	0.18	0.16	0.16	0.19	0.19	0.15	0.14	0.13	0.12	0.11
34	Fabricated metal products	0.18	0.17	0.16	0.2	0.22	0.19	0.17	0.19	0.13	0.12	0.11	0.1	0.06
35	Machinery	0.2	0.2	0.18	0.18	0.17	0.18	0.15	0.19	0.13	0.12	0.11	0.1	0.03
36	Electrical equipment	0.26	0.24	0.23	0.23	0.21	0.18	0.18	0.17	0.15	0.12	0.1	0.08	0.05
37	Transportation equipment	0.18	0.17	0.17	0.16	0.14	0.15	0.15	0.11	0.08	0.08	0.08	0.08	0.03
38	Instruments and related products	0.18	0.19	0.16	0.18	0.17	0.16	0.17	0.17	0.15	0.13	0.11	0.09	0.07
39	Miscellaneous manufacturing	0.25	0.18	0.16	0.16	0.15	0.14	0.13	0.17	0.14	0.13	0.11	0.09	0.08

*Source*s: see text

Table 6. Percentage of SIC3 Industries in each SIC2 Group which became More Localized and Dispersed, 1880-1997.

		1880-1940		1940-1997		1880-1997	
SIC 2	Industry	more dispersed in 1940 than 1880	more localized in 1940 than in 1880	more dispersed in 1940 than 1997	more localized in 1940 than in 1997	more dispersed in 1997 than 1880	more localized in 1997 than in 1880
20	Food and kindred product	89	11	22	78	67	33
21	Tobacco and tobacco product	50	50	50	50	0	100
22	Textile mill product	50	50	33	67	50	50
23	Apparel and related products	43	57	67	33	43	57
24	Lumber and wood products	100	0	17	83	33	67
25	Furniture and fixtures	67	33	60	40	100	0
26	Paper and allied products	50	50	60	40	100	0
27	Printing and publishing	80	20	67	33	80	20
28	Chemicals and allied products	100	0	63	38	83	17
29	Petroleum and coal products	100	0	33	67	100	0
30	Rubber and plastic products	50	50	80	20	75	25
31	Leather and leather products	17	83	71	29	83	17
32	Stone, clay, and glass products	71	29	78	22	71	29
33	Primary metal products	43	57	33	67	29	71
34	Fabricated metal products	75	25	33	67	88	13
35	Machinery	60	40	25	75	40	60
36	Electrical equipment	67	33	75	25	67	33
37	Transportation equipment	100	0	43	57	67	33
38	Instruments & related prod	75	25	83	17	75	25
39	Miscellaneous manufacturing	67	33	83	17	100	0

Sources: see text

Table 7. Not Significantly Spatially Concentrated SIC3 Industries.

SIC 3	Industry
1880	
305	Hose and Belting and Gaskets and Packing
323	Products of Purchased Glass
334	Secondary Nonferrous Metals
1890	
302	Rubber and Plastics Footwear
308	Miscellaneous Plastics Products nec
358	Refrigeration and Service Industry
1900	
261	Pulp Mills
305	Hose and Belting and Gaskets and Packing
365	Household Audio and Video Equipment
1910	
261	Pulp Mills
302	Rubber and Plastics Footwear
354	Metalworking Machinery
364	Electric Lighting and Wiring Equipment
365	Household Audio and Video Equipment
1920	
305	Hose and Belting and Gaskets and Packing
372	Aircraft and Parts
1930	
302	Rubber and Plastics Footwear
358	Refrigeration and Service Industry
1940	
302	Rubber and Plastics Footwear
374	Railroad Equipment

Note: these industries in every case have a negative SMS index.
Source: own calculations, see the text.

Figure 1. The Checkerboard Problem.

3	3		
3	3		

Figure 1a: Hypothetical Distribution of Firms

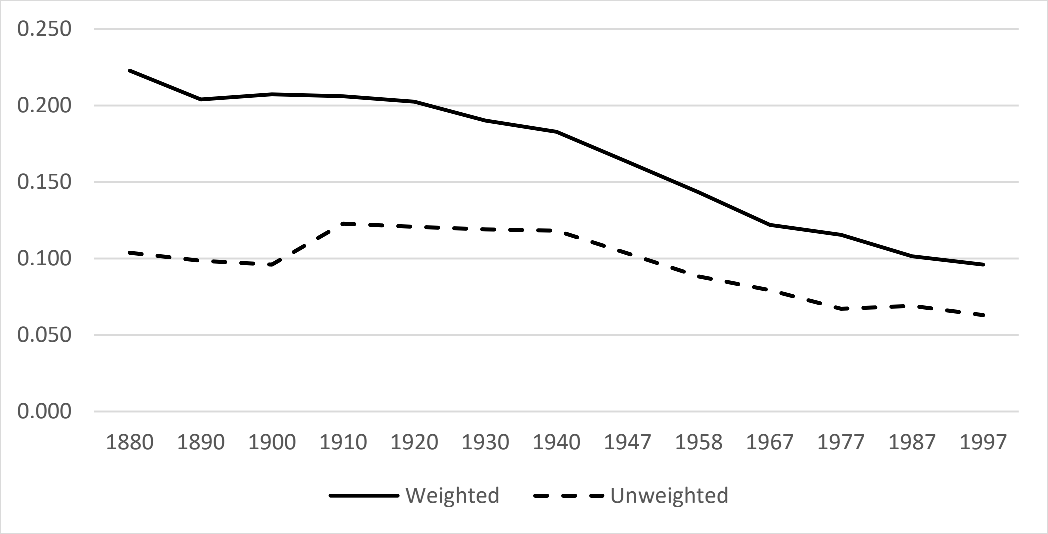
	3		3
	3		3

Figure 1b: Hypothetical Distribution of Firms

Note: these are hypothetical distributions of 12 equally sized firms in 16 equally sized regions and each firm in is the centre of its region. The Ellison-Glaeser measure gives the same score for spatial concentration in each case (0.1273). A spatially- weighted version of the Ellison-Glaeser index will result in distribution a getting a much higher score: (0.2857 vs. 0.0649) if the weighting scheme suggested by Guimaraes et al. (2011) is used.

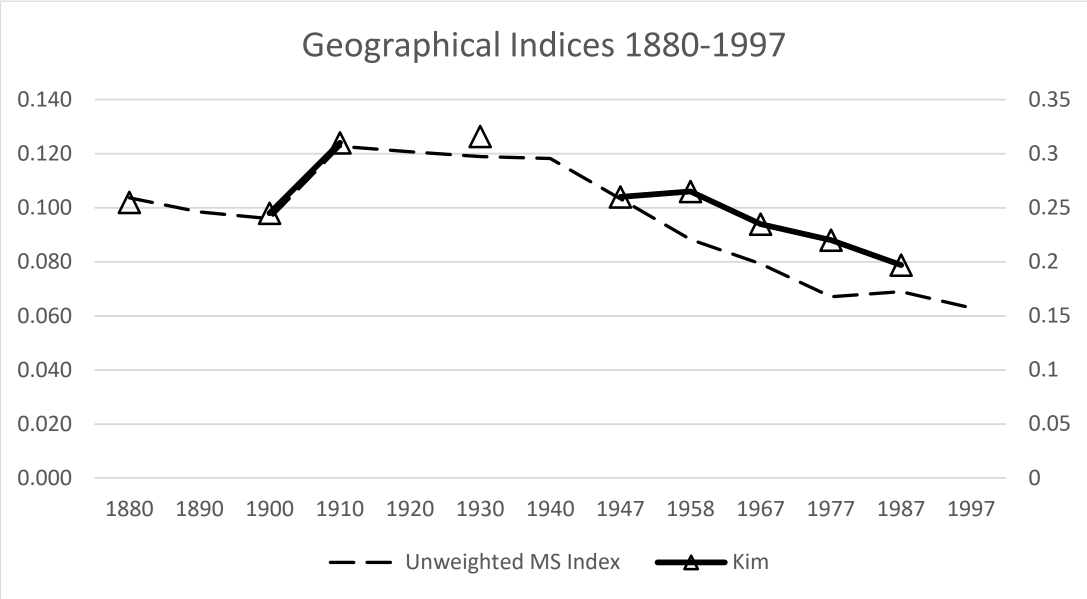
Source: Guimaraes et al. (2011).

Figure 2: Spatially Weighted and Unweighted Index, 1880-1997.



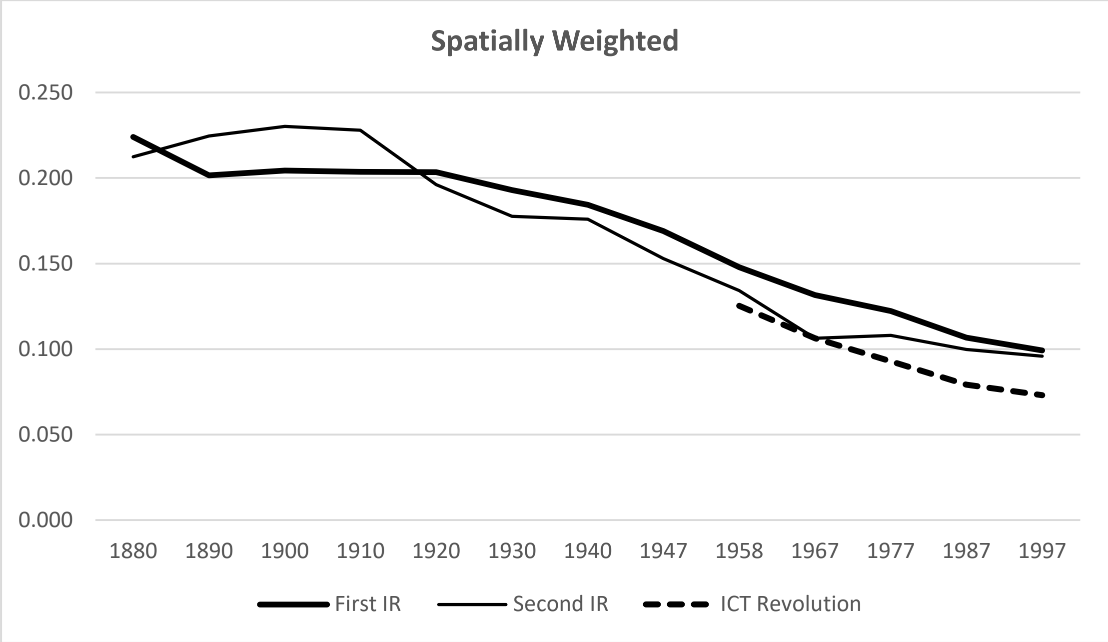
Sources: see the text.

Figure 3: Unweighted MS Index 1880-1997, Kim (1995) Index 1880-1987



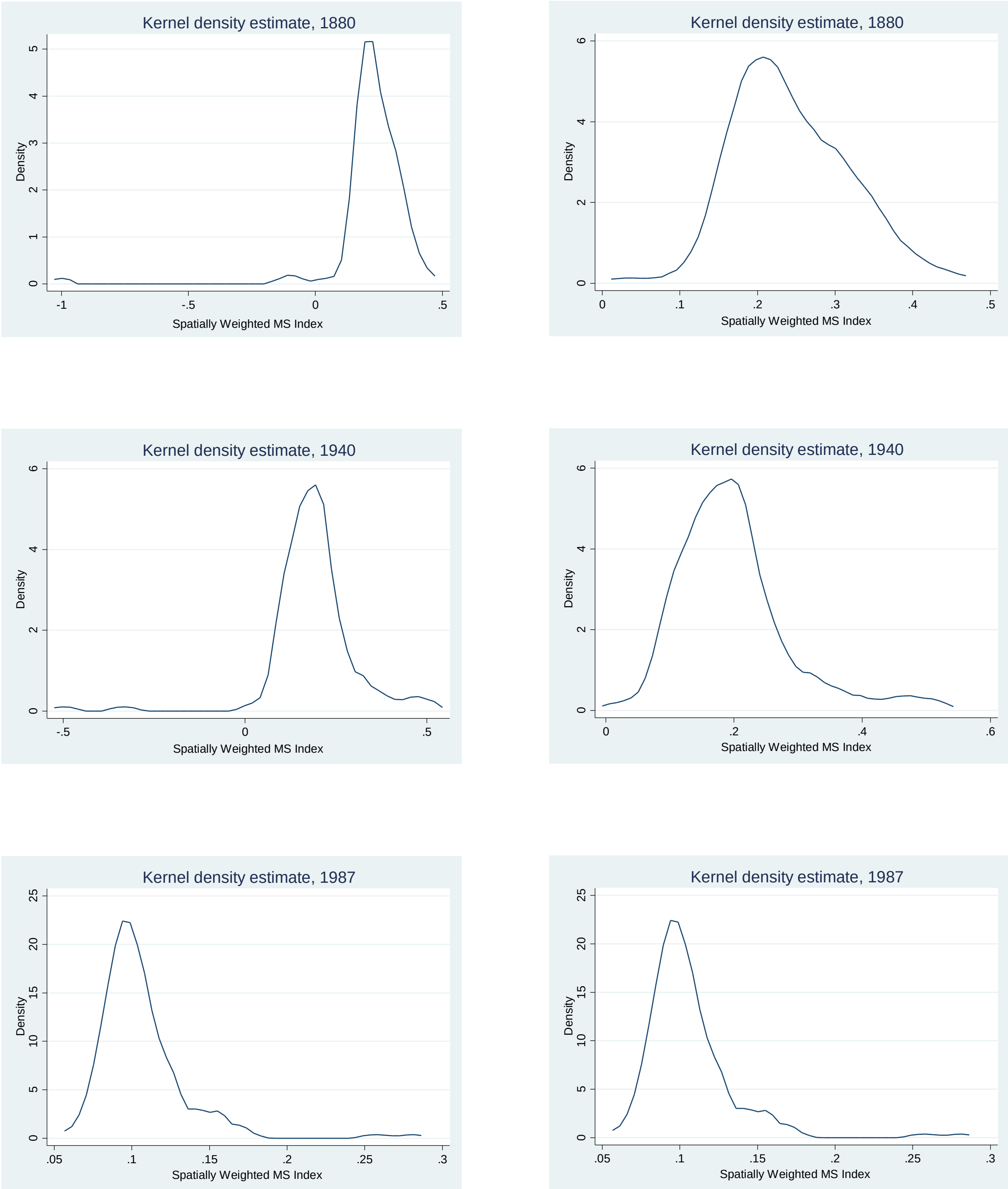
Sources: Kim's index: Kim (1995), unweighted MS index: Figure 2.
Notes: the right x-axis refer to Kim's index, the left x-axis to the unweighted MS index.

Figure 4: Spatially Weighted MS Index by Industries, 1880-1997.



Sources: see the text

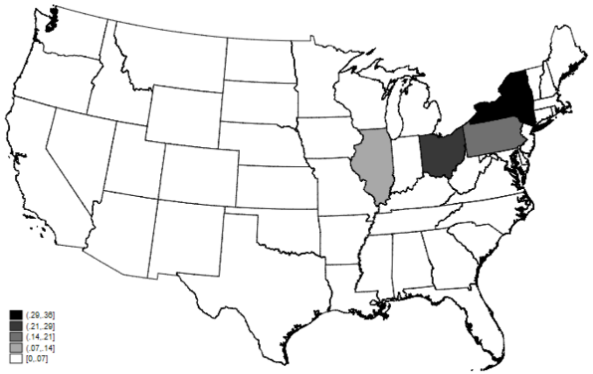
Figure 5:Kernel Density of SMS Index.



Maps 1-9

Map 1: 1900

SIC359: Industrial Machinery nec

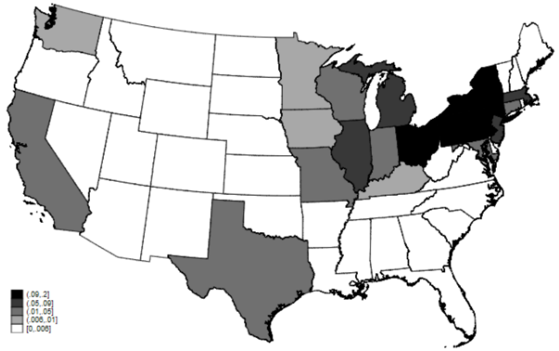


Map 2: 1920

SIC359: Industrial Machinery nec

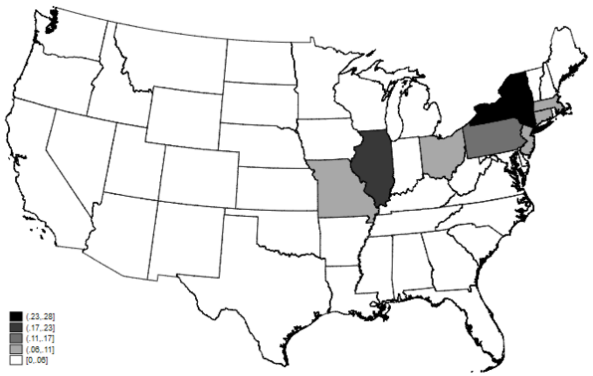
Map 3:1940

SIC359: Industrial Machinery nec



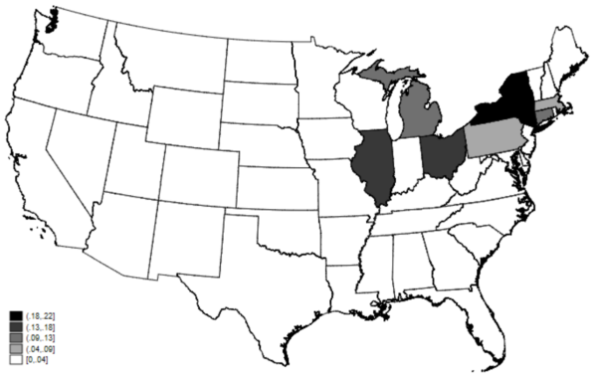
Map 4: 1900

SIC364: Electric Lighting & Wiring Equip.



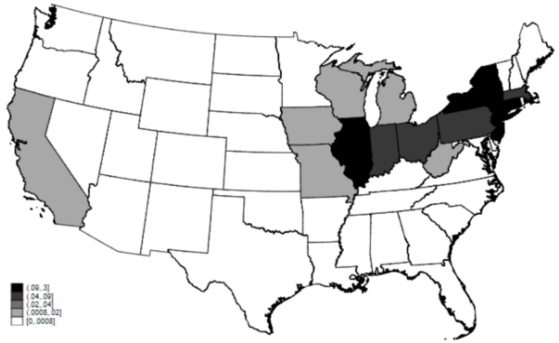
Map 5: 1920

SIC364: Electric Lighting & Wiring Equip.



Map 6: 1940

SIC364: Electric Lighting & Wiring Equip

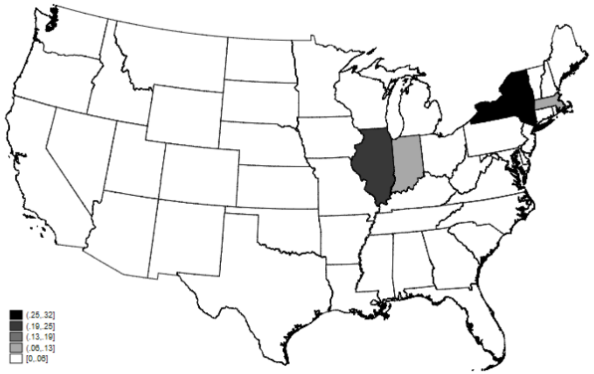


Map 7: 1900

SIC393: Musical Instruments

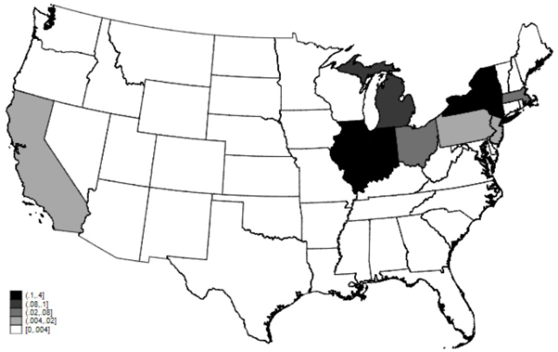
Map 8: 1920

SIC393: Musical Instruments



Map 9: 1940

SIC393: Musical Instruments



Maps 10-13

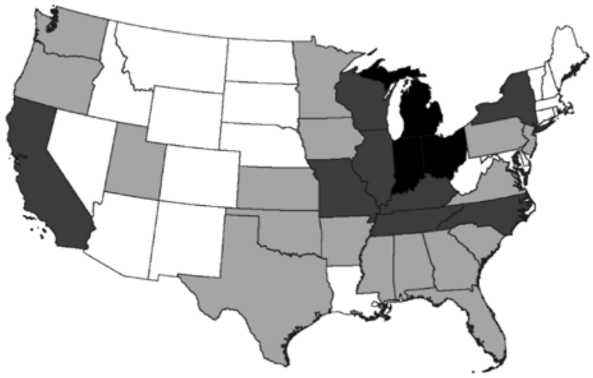
SIC 371 – Motor Vehicles & Motor Vehicle Equipment

Map 10: 1940

Map 11: 1958

Map 12: 1977

Map 13: 1997



Appendix 1

The results in Table 1 and Figure 2 were derived using the first-order contiguity spatial matrix. As was discussed in the methodology section, Guimarães et al (2011) point out that it might be desirable to explore other spatial matrices, especially where there is considerable heterogeneity in the size and shape of spatial units. Therefore, we also use spatial matrices in which neighbours are identified using a pre-defined bandwidth. For example, it could be that the effect of the nascent rubber industry in Ohio in 1900 spilled over not only to neighbouring states such as Pennsylvania or Indiana but also to the state of New York or Illinois which are not adjacent to Ohio.

Ideally, the bandwidth should be large enough to capture spillovers that extend beyond areal boundaries, but not too large to dilute spillover effects. There is little theoretical research on how large the bandwidth should be. Therefore we follow a pragmatic approach. Specifically, we calculated SMS indices with bandwidths starting at 100 miles and going up to 1000 miles by a 50-mile increment, creating thus spatial matrices with neighbours defined over the mile distance of 100-150, 100-200, 100-250, 100-300, etc. This ensures that the spatial matrix captures potential spillovers which extend beyond the borders of individual states.

Figure A1 presents the average of these SMS indices. Specifically, it presents four different versions: first, the average of SMS indices starting at 100 miles and going up to 1000 miles, second, the average of the bandwidth starting at 100 miles and going up to 600 miles, third, the average of the bandwidth starting at 300 miles and going up to 600 miles, and fourth, the average of the bandwidth starting at 400 miles and going up to 500 miles. The reason we included the averages of the bandwidths ranging from 100 to 600 miles and from 300 to 600 miles respectively is to see how sensitive they are to the bandwidths which are either too small or too large. The inclusion of the average of the bandwidths ranging from 400 to 500 miles is due to a suggestion of Guimarães et al (2011) that a bandwidth of approximately 450 miles was preferable when analysing modern state-level data.

We see the overall pattern in all four cases remains basically the same as in Figure 2: large geographical concentration in 1880 relative to 1997, distinctly lower spatial concentration in 1930 and 1940 and a considerable decline of the values of SMS index after 1940. We also see the importance of exploring the role of very large bandwidths. In particular, Figure A1 shows that the plot for the averages of SMS with bandwidths between 100 and 1000 miles crosses the plot for the averages of SMS with the bandwidths between 400 and 500 miles in 1967 and then declines more. The reason is the inclusion of the bandwidths over 600 miles in 100-1000 plot and once excluded, the plot of averages over the bandwidths between 100 and 600 miles no longer crosses the plot for the bandwidths of 400-500 miles. This is not surprising: the spatially weighted indices with the bandwidths over 600 miles boil down to the unweighted MS indices since the bandwidths are too large to capture the neighbourhood effects of US states.

Appendix 2

In Table A1 we report the correlation matrix for the SMS index at SIC3 industry level between different years. On the whole, the correlation coefficients are quite low and a different picture emerges. We find a correlation coefficient of 0.17 between 1967 and 1997 and of 0.17 for 1880 and 1987 compared with 0.92 for the EG index between 1972 and 1992 and 0.64 (at SIC2 level) for Hoover's coefficient of localization between 1860 and 1987 in Dumais et al. (2002).

To test whether the distributions of the SMS indices across decades are similar or not, we use two non-parametric tests, namely, the median test and the Mann-Whitney test. The former is based on the position of each observation relative to the overall median of the distribution, while the latter also takes into account the rank of the observation. As a result, the median test makes fewer assumptions than the Mann-Whitney test. Both tests confirm the pattern emerging from Figure 2: a relatively stable distribution of the spatial concentration of manufacturing activities before World War II and quite rapid changes after that. In Table A2, we see that the distributions of the SMS indices decade-by-decade are mostly *not* statistically significantly different from each other before 1940 while that picture changes after 1940. Even so, the cumulative effect of the pre-1940 changes has the implication that on both tests the distribution in 1940 was significantly different from 1880.

Appendix 3

Manufacturing employment at the two-digit and three-digit SIC level in the U.S. state 1880-1940: The data are taken from the U.S. Census of Manufactures 1880, 1890, 1900, 1910, 1920, 1930, and 1940. We aggregated them into the two-digit SIC level using Niemi (1974) classification and into three-digit SIC level using 1987 SIC classification. To assign an industry listed in the U.S. Census of Manufactures into to the relevant SIC three-digit category, we used detailed descriptions of activities to make the correct matching. The 1910 Census of Manufactures excluded so-called hand trades which are the industries providing repair work or work based on individual orders, e.g. bicycle repairing, furniture repairing, blacksmithing, jewelry engraving. To make the data comparable, we have excluded the hand trades in other years as well. Furthermore, we have excluded repair shops in car manufacturers from 1890 onwards since they did not conduct manufacturing activities. The Census of Manufactures reports a special industry category called 'All Other'. This industry category contains less than one percent of the state's total manufacturing employment and includes the industries with a small number of firms to prevent the identification of those firms. As a result, this category contains a heterogeneous set of industries which makes it difficult to assign it to any of the SIC categories. We have decided to perform the analysis with this industry category assigned to SIC 39, miscellaneous, as well as without that industry. The results are virtually unchanged.

Manufacturing employment at the two-digit and three-digit SIC level in the U.S. state 1947-1997: The data are taken from the U.S. Census of Manufactures 1947, 1958, and 1967, and from Bureau of Labor Statistics for the years 1977, 1987, and 1997. Harmonization of SIC three-digit industries was needed, especially for the year 1947 which presented us with the biggest challenge. Several industries which were coded as SIC four-digit industries became three-digit industries in later census, hence they needed to be recoded. For 1947, Industry

SIC261 had to be adjusted by subtracting SIC2611, SIC286 by subtracting SIC2823 and 2825, and SIC346 by subtracting SIC347. For 1958, SIC264 was added to SIC267 to make it consistent with the subsequent censuses. For that year, as well as for the years 1967 and 1977, SIC303 and SIC307 were reclassified to SIC305 and SIC308 respectively. For 1977 and 1987, SIC264 was recoded as SIC267, and SIC383 was added to SIC382. Furthermore, SIC303 and 304 were added up to create SIC305.

Table A1 . SMS Correlation Matrix, 1880-1997.

	1880	1890	1900	1910	1920	1930	1940	1947	1958	1967	1977	1987
1890	0.56											
1900	0.33	0.18										
1910	0.19	0.54	0.23									
1920	0.29	0.20	0.52	0.37								
1930	0.29	0.65	0.07	0.58	0.55							
1940	0.29	0.63	-0.09	0.46	0.24	0.77						
1947	0.25	0.32	0.16	0.25	0.43	0.50	0.54					
1958	0.14	0.27	0.16	0.32	0.44	0.48	0.53	0.80				
1967	0.20	0.15	0.13	0.01	0.33	0.23	0.24	0.49	0.64			
1977	0.19	0.18	0.23	0.26	0.49	0.44	0.47	0.67	0.82	0.64		
1987	0.17	0.13	0.14	0.31	0.36	0.39	0.43	0.49	0.59	0.42	0.83	
1997	0.17	0.29	-0.05	0.22	0.05	0.28	0.32	0.13	0.16	0.17	0.35	0.67

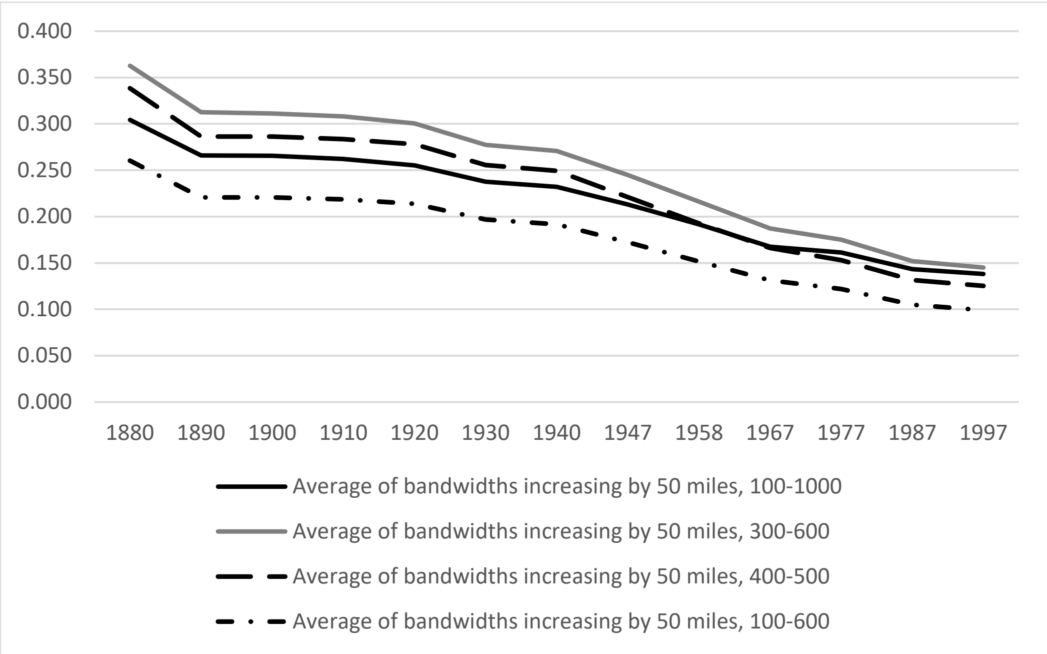
Sources: see text.

Table A2. Non-Parametric Tests on the Similarity of the Distributions of SMS Indices.

Decades	Median test	Mann-Whitney test
1880-1890	0.393	0.314
1890-1900	0.472	1.491
1900-1910	0.123	0.472
1910-1920	0.119	0.577
1920-1930	4.03*	1.883*
1930-1940	0.003	0.063
1940-1947	17.932***	4.596***
1947-1958	7.124*	2.902**
1958-1967	6.3*	3.735***
1967-1977	0.357	-0.997
1977-1987	12.857***	4.082***
1987-1997	6.914**	2.774**
1880-1940	14.966***	4.595***
1940-1997	116.522***	11.428***

Notes: the Median test tests a hypothesis that two samples come from distributions with the same median. The reported statistic is Pearson's chi-square statistic.
The Mann-Whitney test tests a hypothesis that two samples come from the same distribution.
*, **, *** denote statistical significance at 10%, 5%, and 1% respectively.
Source: own calculations, see the text.

Figure A1: Spatially Weighted Index: Robustness of Spatial Weighting Matrix



Sources: see the text

Recent Kent Discussion Papers in Economics