Sep 2017 No.339

A Long-Run Perspective on the Spatial Concentration of Manufacturing Industries in the United States.

Nicholas Crafts and Alexander Klein

WORKING PAPER SERIES

Centre for Competitive Advantage in the Global Economy

Department of Economics

A Long-Run Perspective on the Spatial Concentration of Manufacturing Industries in the United States.[1]

Nicholas Crafts
(University of Warwick)

and

Alexander Klein
(University of Kent)

August 2017

Abstract

We construct spatially-weighted indices of the geographic concentration of U.S. manufacturing industries during the period 1880 to 1997 using data from the Census of Manufactures. Several important new results emerge from this exercise. 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, the persistent tendency to greater spatial dispersion was characteristic of most manufacturing industries. Third, 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 manufacturing industries have changed markedly over time. Kim and Margo (2004) describe a trajectory where regional divergence developed in the context of industrialization during the late 19th and early 20th centuries but was then superseded by regional convergence in the second half of the 20th century. Holmes and Stevens (2004) stress that this latter phase is notable for the decline of the manufacturing belt in which 70 per cent of manufacturing employment was located in1947 but only 40 per cent by 1999.

Kim (1995) developed the point that industries first became more localized as regions became more specialized and then industries became more dispersed as regions became less specialized. He calculated Hoover's coefficient of localization for 2-digit industries through time an found that the weighted average rose from 0.242 in 1900 to 0.316 in 1927 before falling to 0.197 in 1987. While not all industries experienced this pattern a significant number, notably including transportation equipment, did.

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. Ellison and Glaeser (1997) explained that it is important to control for differences in the size distribution of plants when measuring spatial concentration and developed an index in which a measure of raw geographic concentration is modified by taking account of the plant Herfindahl index. An important refinement to the basic EG index is to take account of the geographical position of regions through allowing for 'neighborhood 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.

In this paper we re-examine long-run trends in the spatial concentration of U.S. manufacturing industries over the long run. In particular, 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 0.184 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 at all times 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.

Fourth, changes in the distribution of spatially-weighted EG indices over the long run are statistically significant. For both 1940 compared with 1880 and 1940 compared with 1997, this is the case. However, there is a difference between the two sub-periods. Prior to 1940, distributions compared at decadal intervals are generally not statistically significantly different whereas post 1940 they usually are. The correlation coefficient at the SIC3 industry level between the 1880 and 1997 scores is 0.17, between 1947 and 1997 is 0.13, and between 1967 and 1997 only 0.17. Taking a long-term perspective, we do not find the stability in the distribution of geographic concentration that was emphasized by Dumais et al. (2002).

2. HISTORICAL PERSPECTIVE AND LITERATURE REVIEW

Ellison and Glaeser (1997) introduced an index of spatial concentration adjusted to allow for the distribution of employment across plants recognizing that if output in an industry was produced by only a few plants it would be present in only a few states. They suggested values for their index, 0.02 and 0.05, respectively, below which a sector would be described as not very concentrated or as highly concentrated. They obtained results for 1987 which showed that for almost all SIC4 industries (446/459) there was greater spatial concentration than would be expected to arise randomly but using their criteria 43% would be regarded as not very concentrated and just over a quarter as highly concentrated.

In a sister paper, Dumais et al. (2002) reported that the mean level for the EG index across 134 SIC3 manufacturing industries was 0.036 in 1987 having fallen slightly from 0.039 in 1972. This paper emphasized stability in geographic concentration at the industry level over this 20-year period such that the correlation between the values of the index in 1972 and 1992 was 0.92. This was the case even though analysis of entry and exit showed that there was a considerable degree of mobility but with a pattern that tended to sustain a very similar degree of spatial concentration as if sustaining an equilibrium based on fundamental industry characteristics.

A long-term perspective on localization of industry is provided by Kim (1995). He calculated Hoover's coefficient of localization for SIC2 industries for selected years from 1860 through 1987. He found that the weighted average value of this statistic was rising up to 1927 but then fell steadily to 1987. Average localization was considerably lower at the end of the 20th century than in the late 19th century. This rise and fall of localization was characteristic of some but not by any means all sectors (8/20). In explaining these results Kim highlighted the role of plant size (measured in terms of employment), which rose significantly on average until after World War II, and changes in energy sources. Dumais et al. (2002) noted that Kim's data also exhibit considerable stability in spatial concentration with a correlation coefficient of 0.64 for Hoover's coefficient of localization at the SIC2 industry level between 1860 and 1997.

Important aspects of the economic development of the United States since the late 19th century include steadily declining transport costs for manufactured goods driven by improvements in technology and rising size of plants through the 1970s. 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.

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.[2] 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. We calculate that the ratio of the average wage in manufacturing 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.[3]

This combination of changes over time is reminiscent of the stylized core-periphery model presented by Krugman and Venables (1995). This 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 which raises productivity but also tends 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.

Recent research has produced empirical results which are broadly consistent with this 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

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

[3] 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.

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.

3. METHODOLOGY

We present estimates based both on the original index of spatial concentration developed by Ellison and Glaeser (1997) and also a spatially-adjusted version proposed by Guimarães et al. (2011) which we believe is more appropriate for our purposes. The EG index has a welcome property, namely, that it takes industrial structure into account, thus measuring geographical concentration in excess of what would be expected given industrial concentration. A well-known example illustrating the importance of this is the vacuum-cleaner industry. In this industry in the U.S., four plants account for over 70% of employment. The raw data seem to indicate high spatial concentration as almost three-quarters of the industry is located in at most four regions but this reflects industrial structure rather than locational factors. The EG index deals with this issue by incorporating a measure of industrial structure. The index for industry *i* in a country with *j* regions can be expressed using vectors as

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

where H_i is a Herfindahl index measuring the industry concentration at plant level, G_i is the index of geographical concentration defined as $G_i=(S-X)'(S-X)$ 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*.

A limitation of the EG index is that it does not take into account the geographical position of regions – not even adjacent regions – even though the construction of the index requires spatial data. As was noted by Duranton and Overman (2005), this means that the EG index is potentially downward biased due to the ‘checkboard problem’ since any agglomeration of an industry that crosses state borders will be split into two or more pieces. Spatial weighting corrects for this problem and the values of spatially weighted indices are usually higher than those for spatially un-weighted indices.[4]

The checkboard problem has been addressed by Guimarães et al (2011) who developed a spatially weighted version of the index by introducing the idea of a ‘neighborhood effect’. This adjusts the EG index as follows

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

where H_i and X' are defined as in the equation (1), $G_i^S=(S-X)'\Psi(S-X)$ is the spatially weighted version of the geographical index 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. Here we follow closely Guimarães et al (2011) who, for the case of the US, define the $\boldsymbol{W}$ matrix such that each element takes one for contiguous US states and zero otherwise.

A problem in using the EG index to study long-run development of spatial concentration is that it requires plant-level employment data which are not available throughout 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 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

[4] This is not always the case as is noted by Guimarães et al. (2011).

provide a spatially-weighted version of MS index (henceforth SMS) which is defined as follows:

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

The formula for the SMS index in equation (3) is the main focus of our analysis although we also report estimates based on the MS index.

4. DATA SOURCES

We analyze 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.[5] 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 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

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

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 (2017) for the years 1880, 1890, 1900, 1910, 1920, 1930, and 1940. 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 Manufacturers 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 means that we can use across-state and across-industry adding-up constraints to recover the missing data.[6]

5. RESULTS

We report estimates for the average of MS Index and the average SMS index for all SIC3 industries over the long run in Table 1. The MS estimates are not spatially weighted and are the equivalent of the original EG index. Compared with the EG averages reported by Dumais et al. (2002) for the years 1972 to 1992 our MS estimates are somewhat larger but show a similar decrease in this period. The highlight of the longer-term account that we provide is that the levels of spatial concentration reflected in an EG-type index were considerably higher

[6] We use a procedure described in Golan et al. (1994) which allows recovery of missing information from incomplete multi-sectoral data.

(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.

We regard the spatially-weighted SMS estimates as more informative. As expected, these show higher levels of spatial concentration than the MS estimates. Looking through this lens, our key finding is that 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. Our estimates are that spatial concentration of industries was much more prevalent in the late 19th- than in the late 20th-century. Unlike Kim (1995), who reported the weighted average of Hoover's coefficient of localization for SIC2 industries rising from 0.242 in 1900 to 0.316 in 1927, we do not find an episode of increasing spatial concentration in the early decades of the 20th century.

SMS estimates are reported for all SIC2 industries in Table 2. 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 3. 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.

Our results do not lend support to the hypothesis of stability in geographic concentration. In Table 4 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).

A useful way to further investigate changes in geographical concentration over time is to use the method developed by Midelfart-Knarvik et al. (2000). This entails grouping industries at the start and end of a period ranked from most- to least-spatially concentrated into thirds. Then identify five categories of industry: (i) those in the top third in both years (CC); (ii) those in the bottom third in both years (DD); (iii) those in the top third at the start and bottom third at the end of the period (CD); (iv) those in the bottom third at the start and top third at the end of the period (DC); (v) all others.

The results of this analysis for SIC3 industries for 1880-1997, 1880-1940, and 1940-1997 are shown in Table 5. The overriding impression is that there was not much persistence in ranking in terms of spatial concentration as measured by the SMS index. Only 6 industries, Blast Furnace and Basic Steel Products, Carpets and Rugs, Chewing and Smoking Tobacco, Miscellaneous Converted Paper Products, Paperboard Containers and Boxes, and Women's and Children's Undergarments, were classified as CC and only 8 industries, Beverages, Cement (Hydraulic), Fats and Oils, Household Furniture, Miscellaneous Food and Kindred Products, Preserved Fruits and Vegetables, Rubber and Plastics Footwear, and Ship & Boat Building and Repairing, as DD in all three periods.

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 1: a relatively stable distribution of the spatial concentration of manufacturing activities before World War II and quite rapid changes after that. In Table 6, 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 means that on both tests the distribution in 1940 was significantly different from 1880.

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. Table 7 lists all the cases where the SMS index is not statistically significantly above zero. There are only 20 such instances and none after 1940.

Figure 2 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 level described as 'highly concentrated' by Ellison and Glaeser (1997), and, as we saw in Table 1, the mean SMS at SIC3 level is way above 0.05 throughout the period. The criterion of 0.05 was chosen because it is consistent with the existence of substantial local cost advantages. Therefore, our results imply that economically significant spatial concentration was very highly prevalent across industry continuously from 1880 through 1997.

6. DISCUSSION

An obvious 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. Over the long run, industrial location continually evolved as fundamentals changed.

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 1 to 4 show an evolving pattern of 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 center 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 Unites 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 labor costs.

The semiconductor industry, which belongs to SIC 367, is another example 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 tubes 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).[7]

We have already underlined the importance of spatial weighting in obtaining an appropriate measure of geographic concentration. For example, it is the SMS index rather than the MS index which contradicts the finding in Kim (1995) that spatial concentration of manufacturing industries was increasing during the early 20th century. It is apparent from Table 1 that, on average, spatial weighting indicates higher levels of geographic concentration. Our estimates give means for SMS that are always at least 50 per cent above those for MS while in the late 19th century they are more than double. This compares with the result reported by Guimarães et al. (2011) for 2000 of a 39 per cent difference between their spatially-weighted and not spatially-weighted indices.

It should, however, be remembered that sometimes the SMS for a sector is smaller than the MS index and it can move in an opposite direction over time. A case in point concerns the three major ICT industries, namely, Computer and Office Equipment (SIC 357), Household Audio and Video Equipment (SIC 365), and Electronic Components and Accessories (SIC 367),

[7] Ketelhöhn (2006) calculates the EG index for the semiconductor industry to have risen from 0.061 in 1967 to a peak of 0.093 in 1982 before declining to 0.065 in 1997. We calculate the SMS index for SIC 367 as 0.154 in 1958 and 0.075 in 1997.

between 1958 and 1997. The geographical index G_i for these industries is shown in Maps 5-10 for the years 1958 and 1997.[8]

As we noted in the section on methodology, the limitation of G_i is that it does not take into account the geographical position of regions. Looking at the maps in Maps 5-10, we observe the same patterns in all three industries. In 1958, there were at most two pockets of concentration, and none of them spatially connected. In 1997, we have more pockets of concentration, but they are all scattered across the country, similarly as in 1958. Between these two years, SMS fell from 0.106 to 0.069, from 0.114 to 0.066, and from 0.154 to 0.075 for SIC 357, SIC 365, and SIC 367, respectively. By contrast, between the same two years MS rose from 0.052 to 0.126, from 0.056 to 0.116, and from 0.100 to 0.102 for these three industries, respectively. In 1997 but not in 1958 the MS are greater than the SMS scores.

The main point of this example is that spatial weighting is a crucial aspect of measuring geographic concentration. The spatially-weighted index shows a decrease in geographical concentration as it accounts for the fact that the pockets of concentration in 1997 are disjointed while the not spatially-weighted index interprets the changes between 1958 and 1997 as an increase in geographical concentration. The visual displays in Maps 5-8, especially, suggest that the spatially-weighted index is superior.

A theme that underpins this paper is the value of having a long-run perspective on today's industrial geography. It is useful to know that spatial concentration at the end was much lower than at the start of the 20th century and that it declined steadily over most of the intervening decades. In the context of this move towards greater spatial dispersion, it is noteworthy how low correlations of localization at the industry level were over time. Even so, it is striking how

[8] $G_i=(S-X)'(S-X)$ 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.

pervasive significant excess spatial concentration has been throughout our period. Nevertheless, it seems quite possible that the underlying reasons for this have changed over time and this emerges as an important topic 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.

The first striking implication 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. This experience is characteristic of the vast majority of SIC2 industries. It is also important to note that correlations over time of our index of geographic concentration are quite low. The second major point that comes from our estimates is that 'excess' spatial concentration is pervasive at the SIC3 level throughout the whole period. Across almost all industries and all years, spatial concentration is significant both statistically and economically.

[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).

These findings differ from the conventional wisdom quite considerably. In particular, we reject the picture of an inverted-U shape in long-run geographic concentration suggested by Kim (1995), the hypothesis of continuity in the extent of spatial concentration at the industry level proposed by Dumais et al. (2002), and we believe that economically significant spatial concentration has been much more prevalent in American economic history than would be expected by a reader of Ellison and Glaeser (1997). The added value of a long-run perspective is apparent.

REFERENCES

Carter, Susan B., Scott S. Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright (eds.). 2006. *Historical Statistics of the United States: Earliest Times to the Present*. Cambridge: Cambridge University Press.

Crafts, Nicholas and Alexander Klein. 2015. "Geography and Intra-National Home Bias: U.S. Domestic Trade in 1949 and 2007," *Journal of Economic Geography*, 15, 477-497.

Dumais, Guy, Glenn Ellison, and Edward L. Glaeser. 2002. "Geographic Concentration as a Dynamic Process," *Review of Economics and Statistics*, 84, 193-204.

Duranton, Gilles and Henry G. Overman. 2005. "Testing for Localization using Micro-Geographic Data," *Review of Economic Studies*, 72, 1077-1106.

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

Glaeser, Edward L. and Janet E. Kohlhase. 2004. "Cities, Regions and the Decline of Transport Costs," *Papers in Regional Science*, 83, 197-228.

Golan, Amos, George Judge, and Sherman Robinson. 1994. "Recovering Information from Incomplete or Partial Multi-Sectoral Data," *Review of Economics and Statistics*, 74, 541-549.

Guimarães, Paulo, Octavávio Figueiredo, and Douglas Woodward. 2011. "Accounting for Neighboring Effects in Measures of Spatial Concentration," *Journal of Regional Science*, 51, 678-693.

Holmes, Thomas J. and John J. Stevens. 2004. "Spatial Distribution of Economic Activities in North America," in J.V. Henderson and J. F. Thisse (eds.), *Handbook of Regional and Urban Economics*, vol. 4. Amsterdam: Elsevier, pp. 2797-2843.

Ketelhöhn, Niels W. 2006. "The Role of Clusters as Sources of Dynamic Externalities in the US Semiconductor Industry," *Journal of Economic Geography*, 6, 679-699.

Kim, Sukkoo. 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.

Kim, Sukkoo and Robert A. Margo. 2004. "Historical Perspectives on U. S. Economic Geography," in J. V. Henderson and J-F. Thisse (eds.), *Handbook of Regional and Urban Economics*, vol. 4. Amsterdam: Elsevier, pp. 2981-3019.

Klein, Alexander and Nicholas Crafts. 2012. "Making Sense of the Manufacturing Belt: Determinants of U.S. Industrial Location, 1880-1920," *Journal of Economic Geography*, 12, 775-807.

Klein, Alexander and Nicholas Crafts. 2017. "Agglomeration Externalities and Productivity Growth: U.S. Cities in the Railroad Era, 1880-1930," University of Warwick CAGE Working Paper No. 235.

Klier, Thomas and James Rubenstein. 2008. *Who Really Made Your Car?* Kalamazoo: W.E. Upjohn Institute for Employment Research.

Krugman, P. 2011."The New Economic Geography, Now Middle-Aged," *Regional Studies*, 45, 1-7.

Krugman, P. and Anthony J. Venables. 1995. “Globalization and the Inequality of Nations,” *Quarterly Journal of Economics*, 110, 857-880.

Maurel, Francoise and Beatrice Sedillot. 1999. “A Measure of the Geographic Concentration in French Manufacturing Industries,” *Regional Science and Urban Economics*, 29, 575-604.

Midelfart-Knarvik, Karen H., Henry G. Overman, Stephen J. Redding, and Anthony J. Venables. 2000. “The Location of Industry,” *European Economy Economic Papers* No. 142.

Table 1. MS and SMS Indices, SIC 3-Digit Industries, 1880-1997

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

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

Table 2. 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 3. Percentage of SIC3 Industries in each SIC2 Group which became More Localized and Dispersed, 1880-1997.

SIC 2	Industry	1880-1940		1940-1997		1880-1997	
		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 4. 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 5. Persistence in Spatial Concentration

1880-1997

CC

Asphalt Paving and Roofing Materials
Blast Furnace and Basic Steel Products
Carpets and Rugs
Chewing and Smoking Tobacco
Knitting Mills
Miscellaneous Converted Paper Products
Miscellaneous Fabricated Metal Products
Miscellaneous Textile Goods
Nonferrous Foundries (Castings)
Paints and Allied Products
Paperboard Containers and Boxes
Pulp Mills
Screw Machine Products, Bolts etc
Women's and Children's Undergarments

DC

Dairy Products
Farm and Garden Machinery
Grain Mill Products
Meat Products
Metal Forgings and Stampings
Sawmills and Planing Mills
Secondary Nonferrous Metals

CD

Costume Jewellery and Notions
Hats, Caps, and Millinery
Measuring and Controlling Devices
Miscellaneous Fabricated Textile Products
Miscellaneous Furniture and Fixtures
Musical Instruments
Plumbing and Heating, except Electric
Toys and Sporting Goods

DD

Beverages
Books
Broadwoven Fabric Mills, Wool
Cement, Hydraulic
Fats and Oils
Hose and Belting and Gaskets and Packing
Household Furniture
Leather Goods nec
Medical Instruments and Supplies
Miscellaneous Apparel and Accessories
Miscellaneous Food and Kindred Products
Office Furniture
Ordnance and Accessories nec
Pens, Pencils, Office and Art Supplies
Preserved Fruits and Vegetables
Rubber and Plastics Footwear
Ship and Boat Building and Repairing
Watches, Clocks, Watchcases and Parts

1880-1940

CC

Chewing and Smoking Tobacco
Carpets and Rugs
Women's and Children's Undergarments
Miscellaneous Fabricated Textile Products
Paperboard Containers and Boxes
Miscellaneous Converted Paper Products
Greeting Cards
Industrial Organic Chemicals
Handbags and Personal Leather Goods
Blast Furnace and Basic Steel Products
Cutlery, Hand Tools and Hardware
Electrical Industrial Apparatus
Electric Lighting and Wiring Equipment
Measuring and Controlling Devices

CD

Bakery Products
Leather Gloves and Mittens
Miscellaneous Furniture and Fixtures
Paints and Allied Products
Pulp Mills

DD

Beverages
Cement, Hydraulic
Fabricated Structural Metal Products
Fats and Oils
Grain Mill Products
Household Furniture
Meat Products
Miscellaneous Food and Kindred Products

Costume Jewellery and Notions	Miscellaneous Transportation Equipment
DC	Photographic Equipment and Supplies
Broadwoven Fabric Mills, Wool	Preserved Fruit and Vegetables
Footwear Cut Stock	Rubber and Plastic Footwear
Luggage	Sawmills and Planing Mills
Miscellaneous Apparel and Accessories	Ship & Boat Building and Repairing
Miscellaneous Plastics Products nec	
Ordnance and Accessories nec	
Watches, Clocks, Watchcases and Parts	
1940-1997	
CC	***CD***
Blast Furnace and Basic Steel Products	Broadwoven Fabric Mills, Wool
Broadwoven Fabric Mills, Cotton	Costume Jewellery and Notions
Broadwoven Fabric Mills, Manmade	Flat Glass
Carpets and Rugs	Footwear, except Rubber
Chewing and Smoking Tobacco	Glass and Glassware, Pressed or Blown
Leather tanning and Finishing	Measuring and Controlling Devices
Men’s and Boys’ Furnishings	Miscellaneous Apparel and Accessories
Men’s and Boys’ Suits and Coats	Miscellaneous Fabricated Textile Products
Metalworking Machinery	Ophthalmic Goods
Miscellaneous Converted Paper Products	Ordnance and Accessories, nec
Narrow Fabric Mills	Tires and Inner Tubes
Nonferrous Rolling and Drawing	Watches, Clocks, Watchcases and Parts
Paperboard Containers and Boxes	Women’s and Misses’ Outerwear
Textile Finishing, except wool	***DD***
Women’s and Children’s Undergarments	Aircraft and Parts
DC	Beverages
Concrete, Gypsum and Plaster Products	Cement, Hydraulic
Construction and Related Machinery	Commercial Printing
Fur Goods	Fats and Oils
Grain Mill Products	Household Furniture
Meat Products	Metal Cans and Shipping Containers
Paints and Allied Products	Miscellaneous Food and Kindred Products
Printing Trade Services	Miscellaneous Furniture and Fixtures
Pulp Mills	Petroleum Refining
Railroad Equipment	Preserved Fruits and Vegetables
Sawmills and Planing Mills	Refrigeration and Service Industry
Wood Containers	Rubber and Plastics Footwear
	Ship and Boat Building and Repairing
	Structural Clay Products

Notes:

CC: concentrated industry at both start and end of period

CD: concentrated industry at start and dispersed at end of period.

DC: dispersed industry at start and concentrated at end of period.

DD: dispersed industry at both start and end of period.

In each case ‘concentrated’ (‘dispersed’) indicates a ranking in the top (bottom) third of industries in terms of spatial concentration.

Sources: own calculations, see the text.

Table 6. 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.

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.

Figures:

Figure 1:

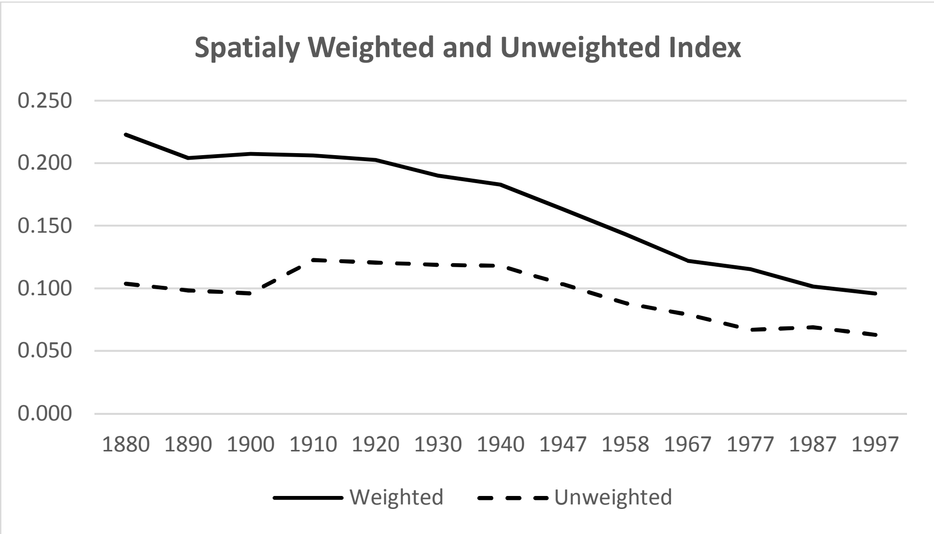
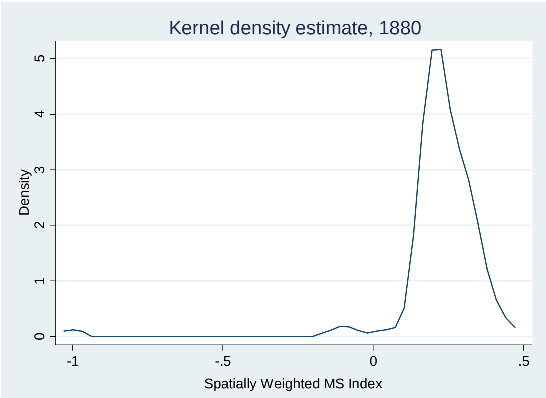
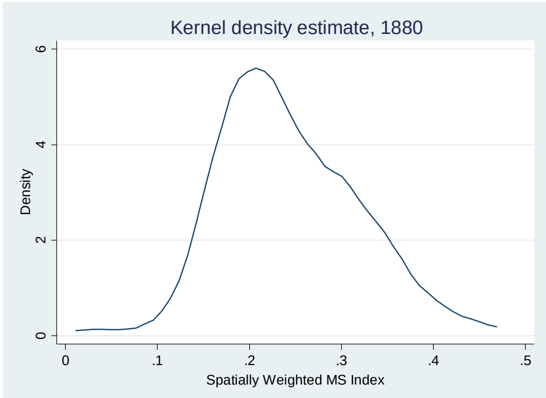
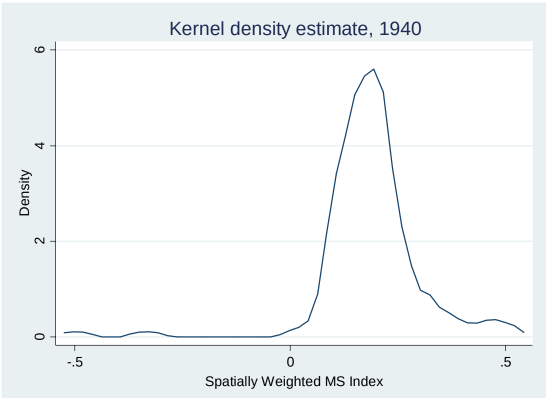
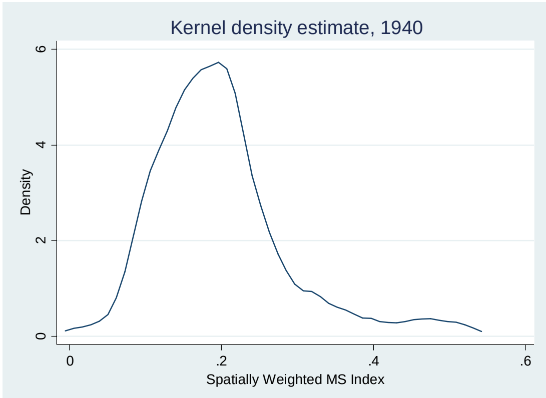


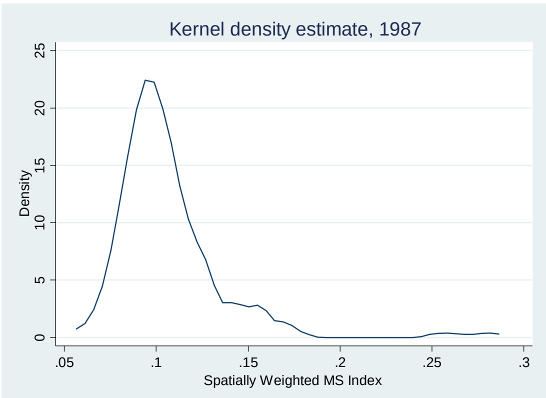
Figure 2:Kernel Density of SMS Index.

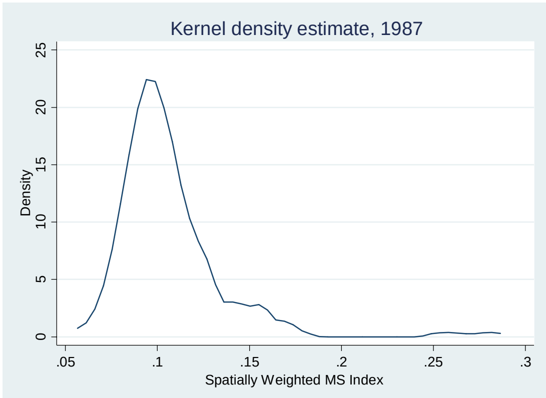












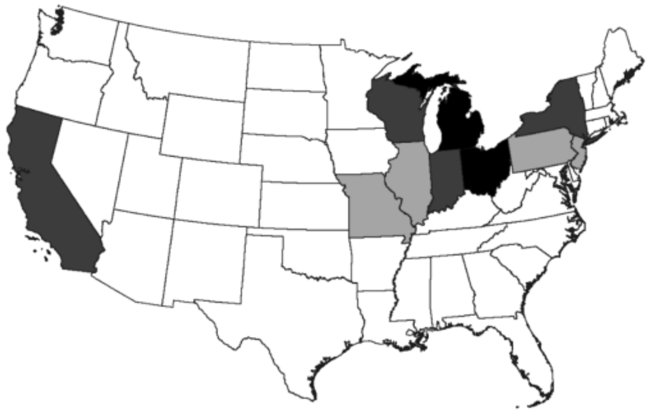
Maps 1-4:

SIC 371 – Motor Vehicles & Motor Vehicle Equipment

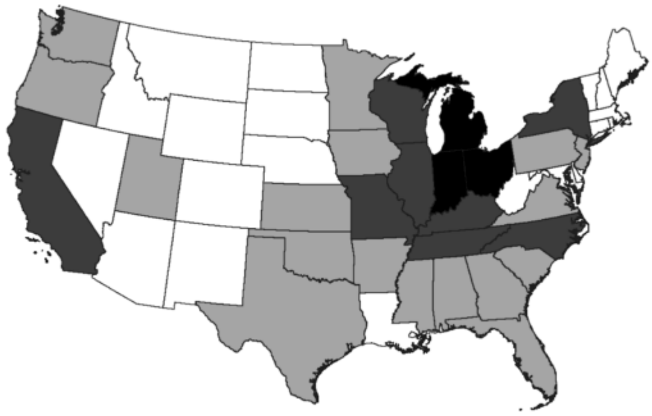
1940

1977

1958



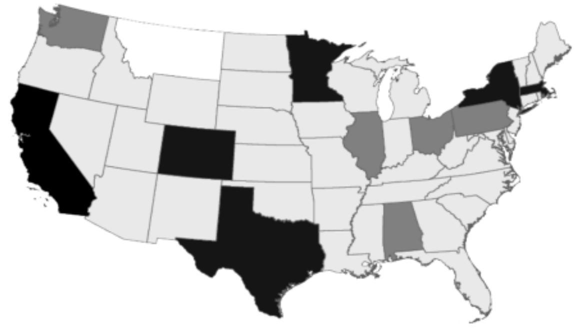
1997



Maps 5-10:

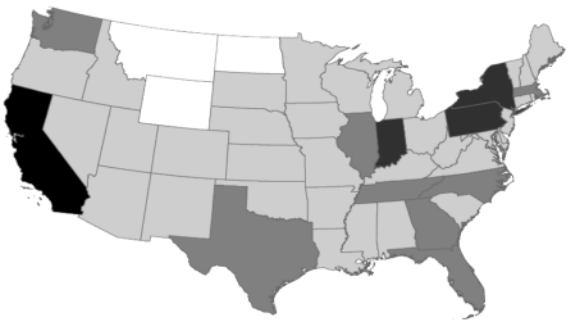
1958: SIC 357 – Computer and Office Equipment

1997: SIC 357– Computer and Office Equipment



1958: SIC 365 – Household Audio & Video

1997: SIC 365– Household Audio & Video



1958: SIC 367 – Electronic Components

1997: SIC 367 – Electronic Components