Journal of Economic Geography Advance Access published August 25, 2011

Journal of Economic Geography (2011) pp. 1–33 doi:10.1093/jeg/lbr023

Making sense of the manufacturing belt: determinants of U.S. industrial location, 1880–1920

Alexander Klein and Nicholas Crafts**,†*

*School of Economics, University of Kent, Giles Lane, Canterbury CT2 7NZ, UK
**Department of Economics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK
†Corresponding author: Professor Nicholas Crafts, Department of Economics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK. *email* <n.crafts@warwick.ac.uk>

Abstract

This article investigates industrial location in the USA around the turn of the 20th century using a model which subsumes both market-potential and factor-endowment arguments. The results show that market potential was central to the existence of the manufacturing belt, that it mattered more than factor endowments, and that its impact came through interactions both with scale economies and with linkage effects. Market potential was generally much higher for states in the manufacturing belt. Natural advantage played a role in industrial location decisions in the late 19th century but its importance then faded away.

Keywords: Factor endowments, linkage effects, manufacturing belt, market potential, new economic geography
JEL classifications: N61, N91, R12
Date submitted: 2 November 2010 **Date accepted:** 20 July 2011

1. Introduction

Traditional accounts of industrial location decisions in the USA during the early 20th century pointed to a number of key factors and stressed that there were differences between industrial sectors. Manufacturing industries were seen in the detailed descriptions given by sources like National Resources Committee (1939) as in some cases natural resource oriented (e.g. blast furnaces), in other cases tied to local consumers (e.g. manufactured ice) or seeking to minimize transport costs while exploiting economies of scale (e.g. automobiles). These accounts have clear similarities to hypotheses that might be derived from Heckscher–Ohlin theories based on factor endowments and from new economic geography focusing on market access.

Beyond this, descriptions of American industrial geography also sought to understand the manufacturing belt. The term 'manufacturing belt' has long been used to describe the remarkable spatial concentration of industry in the USA that prevailed from the third quarter of the 19th century to the third quarter of the 20th century. The area was an approximate parallelogram with corners at Green Bay, St Louis, Baltimore and Portland (Maine). In 1900, about 4/5th of American manufacturing output was produced in this part of the country, which comprised only 1/6th of its land area and a little over half its population.[1] A remarkable feature of this manufacturing belt was its long persistence for a century or so from the Civil War.

The advantages of being located in the manufacturing belt were partly seen as high market accessibility, which was particularly advantageous in the context of realizing scale economies (Harris, 1954). In addition to this, however, stress was also placed on proximity to suppliers and purchasers of intermediate goods (forward and backward linkages), while noting the importance of manufactured intermediates in the production of manufactures (Perloff et al., 1960). A large market for intermediates was seen as making the manufacturing belt a very attractive place to produce such goods and, in turn, better access to intermediates made production of final goods cheaper. These ideas would later be formalized in Krugman and Venables (1995). There are no internal trade data with which to quantify flows of manufactured goods at the turn of the 20th century but the maps derived from railroad freight data for the late 1940s by Ullman (1957) show quite clearly that states in this area bought and sold their manufactured goods not only within the manufacturing belt but predominantly within state or to their neighbours.

The data source used by Ullman illustrates this very well in the case of automobiles, as Table 1 reports. Upper panel shows that 70% of passenger vehicles were exported from the manufacturing-belt states with >50% exported from Michigan alone. Except for a very small percentage coming from Tennessee and Missouri, the intermediates for the vehicle producers of Michigan came from manufacturing-belt states, and especially from nearby ones.

There is relatively little modern empirical work on determinants of industrial location in the late 19th and early 20th centuries. The most important papers are Kim (1995, 1999). Kim (1999) estimated the following equation based on the Rybczynski theorem for production of two-digit manufacturing industries across U.S. states for snapshot years between 1880 and 1987:

$$Y = \alpha_0 + \alpha_1 \text{Labour} + \alpha_2 \text{Capital} + \alpha 3 \text{Natural Resources}.$$

He found that factor endowments were the fundamental explanation for the geographic distribution of U.S. manufacturing from 1880 through 1987. High R^2 for these equations were interpreted by Kim to mean that once factor endowments had been taken into account, there was little left to be explained. Kim (1995) considered the relationship between spatial concentration of an industry (localization) and plant size (scale) and raw-material intensity (resource) by estimating the following equation for panel data for U.S. manufacturing industries for selected years between 1860 and 1987:

$$\text{Localization} = \beta_0 + \beta_1 \text{Scale} + \beta_2 \text{Resource}.$$

He found that U.S. regional specialization in the late 19th and early 20th centuries was positively related to both variables. Thus, the manufacturing belt was based on the rise

1 At a disaggregated level, it is appropriate to demarcate the manufacturing belt in terms of counties. Our analysis is at the state level; states whose territory is wholly or predominantly in the manufacturing belt are Connecticut, Delaware, Illinois, Indiana, Maine, Maryland, Massachusetts, Michigan, New Hampshire, New Jersey, New York, Ohio, Pennsylvania, Rhode Island, Vermont, Virginia, West Virginia and Wisconsin.

Table 1. U.S. passenger vehicle and vehicle parts trade in 1949

State	Carloads (%)
U.S. states exporting passenger vehicles to other U.S. states	
California	491 (30.29)
Illinois	42 (2.59)
Indiana	129 (7.96)
Michigan	901 (55.58)
Ohio	57 (3.52)
Pennsylvania	1 (0.06)
Total	1621 (100.00)
Imports of vehicle parts to Michigan	
Illinois	10 (1.73)
Indiana	54 (9.33)
Massachusetts	2 (0.35)
Michigan	204 (35.23)
Minnesota	1 (0.17)
Missouri	5 (0.86)
New Jersey	6 (1.04)
New York	40 (6.91)
Ohio	160 (27.63)
Tennessee	12 (2.07)
West Virginia	9 (1.55)
Wisconsin	76 (13.13)
Total	579 (100.00)

Sources: see the text.

of large-scale production methods that were intensive in the use of raw materials and energy sources that were relatively immobile.

However, these papers by Kim are not fully convincing.[2] Most obviously, there are likely to be problems of omitted variables. In particular, no account is taken of market access or linkage effects, which are taken to be important in the traditional literature. In this article, we address this issue by using a version of a model originally proposed by Midelfart-Knarvik et al. (2000), which incorporates both factor-endowment and market-access determinants of location. This is estimated at the state level for U.S. manufacturing for the earliest feasible period, 1880–1920. We operationalize the notion of market access by the use of 'market potential', the concept introduced by Harris (1954). Our framework allows an explicit analysis of the roles of each of scale economies, backward linkages and forward linkages.

We model industrial location taking explicit account of interactions between industrial characteristics and regional characteristics. The approach that we use is grounded in a model of production and trade that takes account both of input price variations resulting from factor endowments and from costs of intermediate inputs and also of the spatial pattern of demand. We try to explain the existence and persistence of the manufacturing belt around the turn of the 20th century in an analysis of the shares

2 See also the critique in Combes et al. (2008).

of employment in two-digit manufacturing industries across 48 U.S. states using a newly constructed data set.[3]

In particular, we address the following questions relating to U.S. manufacturing at the two-digit level:

1. Which factor endowments affected the location of manufacturing?
2. Did market potential influence the location of manufacturing through linkage effects and/or scale effects? and
3. How important were factor endowments and market potential, respectively, as determinants of industrial location?

2. An empirical framework

We wish to examine the relative importance of natural advantages and market potential in explaining the existence of the manufacturing belt around the turn of the 20th century. Theory offers competing, but not mutually exclusive, hypotheses. The Heckscher–Ohlin model implies that the distribution of economic activity is determined by comparative advantage, that is, by the factor endowments available at each location. New Economic Geography models argue that firms tend to locate so as to minimize transport and communication costs related to the supply of their inputs and demand for their outputs and so wish to be near to large markets. These ideas evidently have their counterparts in the literature on the history of American industrial geography.

Recent empirical studies such as Davis and Weinstein (1999, 2003), Ellison and Glaeser (1999), Midelfart-Knarvik et al. (2000) and Wolf (2007) all take the following stance. First, H–O and NEG models are not mutually exclusive but have different views of the trade-offs that firms face in their choice of location. Second, location theories work on the basis of the interaction of the characteristics of places (states) with those of economic activities (industries). H–O theories predict that industries that are intensive in the use of a factor of production will be attracted to states, which are relatively abundant in that factor; NEG theories predict that a state's market potential will have a greater impact on location when industries have larger scale or stronger linkage effects. Hence, explanation of the pattern of industrial location can be sought in terms of a set of H–O- and NEG-type interactions between industry and state characteristics.[4]

Our methodology follows these earlier studies, in particular, that of Wolf (2007). We rely on an estimating equation that is best viewed as a reduced form, which has both H–O- and NEG-type interactions between state and industry characteristics on the right-hand side.[5] We consider interactions of an H–O-type between the share of farmland and agricultural input use, the share of educated population and use of white-collar workers and coal prices and the use of steam power together with

3 We do not seek to explain the emergence of the manufacturing belt which happened in the decades before the Civil War. Our data do not permit analysis earlier than 1880.

4 There is a well-established tradition of explaining patterns of international trade using estimating equations which contain variables that are interactions between commodity characteristics and country characteristics, for example, Nunn (2007) and Romalis (2004).

5 A formal model of this type was derived by Midelfart-Knarvik et al. (2000). We present a simplified version of their model which delivers an estimating equation similar to the one that we use in the Supplementary Material.

NEG-type interactions between market potential and each of plant size, intermediate input use and sales to industry. Clearly, endogeneity is a big issue and our estimation strategy relies heavily on an instrumental-variables approach.

Formally, the basic model can be written as follows:

$$\ln\left(s_{i,t}^{k}\right) = c + \sum_{j} \beta^{j}\left(y_{i,t}^{j} - \gamma^{j}\right)\left(x_{t}^{j,k} - \chi^{j}\right) + \varepsilon_{i,t}^{k} \tag{1}$$

where ${s_{i,t}}^{k}$ is the share of the output of industry k in state i and time t, ${y^{j}}_{i,t}$ is the level of jth state characteristic in state i and time t; x_{t}^{jk} is the industry k value of the industry characteristic paired with state characteristic j at time t, c is a constant and ${\varepsilon_{i,t}}^{k}$ is the error term. The interaction forces between the characteristics of states and the characteristics of industries are represented by the terms in the summation and β^{j}, γ^{j} and χ^{j} are coefficients to be estimated.

To understand this specification, consider one particular characteristic, say j = skilled labour.[6] So x[skilled labour]$_{t}^{k}$ is white-collar worker intensity of industry k at time t, and y[skilled labour]$_{i,t}$ is educated population abundance of state i at time t. The model can be interpreted as follows. First, there exists an industry with a level of skilled-labour intensity χ[skilled labour] such that its location is independent of state skilled-labour abundance. Second, there exists a level of skilled-labour abundance γ[skilled labour] such that the state's share of any industry is independent of the skilled-labour intensity of the industry. Third, if β[skilled labour] > 0, then industries with skilled labour intensities greater than χ[skilled labour] will be induced to locate in states with skilled-labour abundance greater than γ[skilled labour]. Estimation of the model will produce the following key parameters for each interaction variable: $\beta[j]$, $\gamma[j]$ and $\chi[j]$ with j running over the interactions. If, for example, skilled labour is an important determinant of location patterns, we should see a high value of β[skilled labour].

Expanding the relationships in Equation (1), we obtain the estimating equation

$$\ln\left(s_{i,t}^{k}\right) = c + \sum_{j} \left(\beta^{j} y_{i,t}^{j} x_{t}^{j,k} - \beta^{j}\gamma^{j} x_{t}^{j,k} - \beta^{j}\chi^{j} y_{t}^{j,k}\right) + \varepsilon_{i,t}^{k}. \tag{2}$$

This gives a list of independent variables that comprises scaling terms, state characteristics, industrial characteristics and interactions between state and industrial characteristics. The estimated coefficients of the state characteristics, y^{j}, and industry characteristics, x^{j}, are estimates of $-\beta^{j}\gamma^{j}$ and $-\beta^{j}\chi^{j}$, respectively, and so are expected to have negative signs. The estimated coefficients of the interaction variables—the main variables of interest—$y^{j}x^{j}$ are estimates of β^{j}, which are expected to be positive and comprise the crucial set of parameters in the model. The relative magnitude and statistical significance of this coefficient on, for example, educated population × white-collar workers provides us with a measure of how important this factor endowment was in influencing the location of industries in the USA.

The intuition behind the formulation of our estimating equation is that the share of an industry's output produced in each state will depend on input prices and the spatial distribution of demand. Inputs include both primary factors and intermediates, the prices of which reflect factor endowments and proximity to suppliers, respectively.

6 The discussion follows Crafts and Mulatu (2006).

The spatial distribution of demand has its effect through the attraction of market access that is driven by the geography of GDP and transport costs.

3. Data and implementation of the empirical framework

In this section, we describe the implementation of our model and the data used in the article; a detailed description of the variables can be found in Appendix A. The approach is quite similar to that of Wolf (2007) but is somewhat different from that of Midelfart-Knarvik et al. (2000). Our dependent variable is measured in terms of shares of employment rather than shares of output. This suggests estimation using region and industry dummies to control for the effects that productivity differences might have on the employment-based location quotient. In the implementation of the model, we estimate a version of Equation (2) with state and industry dummies replacing the list of state and industry characteristics. This is acceptable since, in any case, we are interested only in the interaction terms and the impact that they have on the spatial distribution of industry.

3.1. Regression equation

We use four state characteristics (share of farm land, share of educated population, coal prices and market potential), six industry characteristics (the share of white-collar workers, steam power use, plant size, agricultural input use, intermediate input use and sales to industry), six interactions and add state and industry dummies. The estimated Equation (2) can be expressed as follows:

$$\begin{aligned}\ln\left(s_{i,t}^{k}\right) = {} & \beta_1(\text{FARM LAND} \times \text{AGRICULTURE INPUT USE})_{\text{i,k,t}} + \\ & + \beta_2(\text{EDUCATED POPULATION} \times \text{WHITE COLLAR WORKERS})_{\text{i,k,t}} + \\ & + \beta_3(\text{COAL ABUNDANCE} \times \text{STEAM POWER USE})_{\text{i,k,t}} + \\ & + \beta_4(\text{MARKET POTENTIAL} \times \text{INTERMEDIATE INPUT USE})_{\text{i,k,t}} + \\ & + \beta_5(\text{MARKET POTENTIAL} \times \text{SALES TO INDUSTRY})_{\text{i,k,t}} + \\ & + \beta_6(\text{MARKET POTENTIAL} \times \text{SIZE OF ESTABLISHMENT})_{\text{i,k,t}} + \\ & + \Sigma_{\text{i}}\gamma_{\text{i}}\text{STATE}_{\text{i,t}} + \Sigma_{\text{k}}\theta_{\text{k}}\text{INDUSTRY}_{\text{k,t}} + \varepsilon_{i,t}k.\end{aligned} \quad (3)$$

The first three of these interactions are predicted by the Heckscher–Ohlin (H–O) theory based on factor endowments.[7] The relative magnitude and statistical significance of β_1 shows the importance of farmland in influencing the location of industry and so on. The last three are predicted by NEG models. The first market potential interaction says that industries which use relatively large amounts of intermediate goods would prefer locations of high market potential. Here, the importance of forward linkages is the key but how strongly firms value centrality will depend on transport costs; cheaper inputs have to be traded off against a higher costs of sending goods to the final consumer. The second market potential interaction is based on backward linkages and

7 Coal abundance is measured by coal prices. We also used an alternative measure—the share of mining in the labour force—and discuss the implications for the results in the section on robustness.

presumes that industries that sell a relatively large fraction of their output to other firms rather than to the final consumer tend to locate relatively close to other producers. The third market potential interaction hypothesizes that industries operating at relatively large scale will value locations relatively close to market demand (at least at some levels of transportation costs). The coefficients β_4, β_5 and β_6 show the importance of market potential as a determinant of industrial location. In the original work by Midelfart-Knarvik et al. (2000), the authors estimated their version of Equation (2) using OLS, and took account of heteroskedasticity and also country and industry fixed effects. We address additional estimation issues including endogeneity and clustered-sample methods.

3.2. Data set

We created a unique data set of the employment shares for 48 U.S. states and 19 two-digit level industries, 6 industry and 4 state characteristics, including market potential for each census year during 1880–1920.[8] The data on the shares of two-digit level industrial employment in the U.S. states are drawn from the U.S. Census of Manufactures. The aggregation of individual industries at the two-digit level follows the standard industrial classification provided by Niemi (1974). The population data are from Carter et al. (2006). The data on labour force in each U.S. state are from Perloff et al. (1960), the share of farm land is calculated from Carter et al. (2006), coal prices are taken from various U.S. government sources and the data on educated population by states come from the U.S. occupation censuses and Goldin (1998).[9] The share of white-collar workers as well as of steam power use is extracted from the U.S. Censuses of Manufactures 1880–1920. Average plant size is from O'Brien (1988). Forward and backward linkages are evaluated using an input–output table for the U.S. economy in 1899 (Whitney, 1968).[10] Summary statistics of the variables and the units in which they are expressed are shown in Table 2.

Left panel in Table 3 reports industrial characteristics obtained from the 1899 input–output table, which relate to key aspects highlighted by locational hypotheses based either on market potential (Columns 1 and 2) or on natural advantages (Column 3). It is clear that there are big differences across industries. For example, SIC 33, primary metal products, has high use of intermediates and sales to industry relative to gross output whereas for SIC 21, tobacco products, these proportions are negligible. Conversely, tobacco uses agricultural inputs quite heavily but primary metal products do not. Overall, it is noticeable that many sectors have substantial linkages (medians in Columns1 and 2 are both 26%) whereas few sectors rely heavily on agricultural inputs (median in Column 3 is 0.4%).

Right panel in Table 3 shows the distribution of two-digit manufacturing employment between the manufacturing belt states and the states outside the belt. We see that

8 There are 46 states in 1880 since Oklahoma did not exist then, and North and South Dakota was considered a single territory. Alaska is excluded throughout the whole period. More details of data sources and methods are given in Appendix A.

9 We thank Claudia Goldin for providing the data.

10 Leontief (1941) constructed an input–output table for 1919. However, this is not suitable for our purposes because it does not include service-sector activities and does not distinguish between metal production and machinery.

Table 2. Summary statistics, 1880–1920

Variable	*N*	Min	Max	Mean	SD
Share of manufacturing labour force (19 two-digit SIC industries)	4560	0.00	55.21	2.08	4.68
Share of farm land	4560	0.00	97.19	51.08	29.06
Agricultural input use	4560	0.00	23.64	5.22	7.57
Share of educated population	4560	0.00	55.50	12.41	10.82
Share of white-collar workers	4560	2.39	36.85	12.46	6.84
Coal price	4560	2.80	46.85	10.92	6.53
Steam power use	4560	0.03	2.00	0.52	0.46
Intermediate input use	4560	1.70	51.60	31.65	12.67
Sales to industry	4560	0.00	62.99	30.03	18.71
Size of establishment	4560	6.94	967.53	74.27	139.87
Market potential	4560	6.66	212.41	42.54	40.63

Notes: Share of manufacturing labour force is the percentage of a state's 2-digit industry from the state's total manufacturing labour force. Share of farm land is the percentage of a state's farm land from the state's total land area. Agricultural input use, intermediate input use and sales to industry are expressed as percentage of the gross value of output. Steam power use is steam horse power per $1000 gross output. Plant size is calculated as the average number of wage earners per establishment. Coal price is $ per ton of 2200 pounds. Market potential is in millions of current $ U.S.
Source: see the text.

industries having substantial linkages but little use of agricultural inputs are highly concentrated in the manufacturing belt (for example SIC 33, primary metals or SIC 35&36, machinery,) while industries that rely on agricultural inputs (for example SIC 28, chemicals and allied products) are less so. The differences are even more profound in 1920 when, for example, SIC 24, lumber and wood products, employs more people outside the manufacturing belt than inside it. Lower panel also shows that there is a slight decrease of the share of manufacturing employment in the manufacturing belt for some industries between 1880 and 1920. Those industries largely produce final consumer products and since the population living outside the manufacturing belt increased by 1920 it is not surprising that those industries increased their shares outside the belt too. Despite this, the overall pattern of the industries with substantial linkages being located in the manufacturing belt is preserved, with the primary metal products, machinery and chemical industry even increasing their presence in the belt.

The only variable that needs to be estimated is market potential. The estimation of market potential goes back to Harris's (1954) seminal paper, which calculates market potential as the inverse distance-weighted sum of incomes. In recent years, several studies have linked market potential rigorously to theory (e.g. Krugman, 1992; Head and Mayer, 2004) with the implication that a gravity equation framework should be used to estimate market potential. However, the resulting methodology requires internal trade flows data, which are unavailable for the USA for the period 1880–1920. Therefore, we use Harris's original approach and calculate the market potential of a U.S. state i using the formula $M_i = \sum_j \varphi_{ij}\, \mathrm{GDP}j$ where φ_{ij} is the accessibility of market j for goods from the U.S. state i defined as $\varphi_{ij} = d_{ij}^{\delta}$ with $\delta = -1$. The market j consists of nominal GDP in foreign countries, in other U.S. states and in the home state i.

Table 3. Industry characteristics in 1899 and manufacturing employment in 1880 and 1920

	Industry characteristics, 1899 (%gross output)				Manufacturing employment and population (%U.S. total)			
					1880		1920	
	SIC	Intermediate input use	Sales to industry	Agricultural input use	MB	Outside MB	MB	Outside MB
Food and kindred product	20	18.2	11.7	23.6	75.25	24.75	61.05	38.95
Tobacco and tobacco product	21	1.7	0	18.9	78.97	21.03	71.27	28.73
Textile mill product	22	24.6	57.8	19.9	94.63	5.37	75.79	24.21
Apparel and related products	23	46.2	9.0	1.7	93.73	6.27	88.97	11.03
Lumber and wood products	24	38.9	54.2	7.1	77.00	23.00	40.69	59.31
Furniture and fixtures	25	43.2	5.9	0.0	87.58	12.42	81.62	18.38
Paper and allied products	26	38.5	63.0	6.7	95.76	4.24	92.61	7.39
Printing and publishing	27	23.9	14.3	0.0	83.15	16.85	74.08	25.92
Chemicals and allied products	28	37.3	42.8	11.2	69.25	30.75	72.48	27.52
Petroleum and coal products	29	23.4	33.1	0.0	91.31	8.69	54.25	45.75
Rubber and plastic products	30	22.4	30.3	0.0	99.97	0.03	98.35	1.65
Leather and leather products	31	51.1	37.4	8.2	84.88	15.12	88.87	11.13
Stone, clay and glass products	32	21.0	23.5	0.0	81.09	18.91	80.72	19.28
Primary metal products	33	47.8	58.4	0.0	90.22	9.78	92.31	7.69
Fabricated metal products	34	10.4	25.6	0.0	89.68	10.32	88.22	11.78
Machinery	35, 36	32.3	22.6	0.0	89.35	10.65	93.00	7.00
Transportation equipment	37	25.9	35.7	0.4	86.16	13.84	73.03	26.97
Instruments and related products	38	51.6	15	0.0	94.36	5.64	95.07	4.93
Miscellaneous manufacturing	39	26.8	15.7	1.3	96.46	3.54	90.92	9.08
Total manufacturing					86.83	13.17	76.96	23.04
Population					57.55	42.45	53.37	46.63

MB stands for the manufacturing belt.
Sources: Whitney (1968); SIC 21, 25 and 34 are from Leontief (1941); SIC 38 is from Thomas (1984).

The market accessibility of own U.S. states is calculated as

$$\varphi_{ii} = d_{ii}^{-\delta} = [2/3\,.\,(\text{area}_i/\pi)^{0.5}]^{-\delta} \tag{4}$$

Nominal GDP of U.S. states in 1880–1910 is taken from Klein (2009), which provides new estimates of 1890 and 1910 nominal GDP for each U.S. state based on the methodology developed by Easterlin (1957), and re-estimates Easterlin's original 1880 and 1900 estimates.[11] Data for 1920 are from Easterlin (1957). The sources of nominal

11 Easterlin's (1957) study provides estimates of nominal GDP from the income side for each U.S. state in 1880, 1900, 1919–1921 and 1949–1951. Estimation involves two steps. First, the ratio of the state total personal income per capita relative to the U.S. total personal income per capita for each U.S. state is constructed from the census publications. These ratios are then used to allocate the U.S. total personal income per capita among the states. The calculation of the ratios involves the calculation and the weighting of the sectoral ratios for agriculture and six non-agriculture sectors. Total personal income

GDP for foreign countries and the corresponding exchange rate are in Appendix A2.2. The area of U.S. states is taken from Carter et al. (2006), the distance between the U.S. states and the foreign countries is the kilometre distance between the corresponding capitals and the distance between the U.S. states is calculated as the kilometre distance between their largest cities.[12]

Although there are no U.S. internal trade flows data for the period 1880–1920, we can justify the assumption of $\delta = -1$ in two ways. First, our estimates of market potential are for the railroad era and we believe that by this time physical distances are a reasonable approximation to economic distances inside the USA. Our choice of -1 for δ is consistent with estimates for modern internal U.S. trade (Wolf, 2000; Hillberry and Hummels, 2003; Knaap, 2006). Second, we can analyse U.S. internal railroad commodity trade in 1949. This is the earliest date for which internal trade data exist and it is suitable for our purposes because the manufacturing belt was still intact at that time, and the railroads were still the most important transportation mode.[13] The data come from the Interstate Commerce Commission Carload Waybill Statistics, which report commodity flows between the U.S. states at three-digit level. We estimate the following gravity regression:

$$\ln X_{ij} = EX_i + IM_j + \delta \ln d_{ij} + \beta_j B_{ij} + \varepsilon_{ij}, \tag{5}$$

where X_{ij} is the aggregate value of the state's i export to country j, EX_i and IM_j are exporter and importer fixed effects, B_{ij} is a dummy variable that is 1 if i and j share a border. We estimate this equation using Poisson pseudo-maximum-likelihood estimator, following the suggestion of Santos Silva and Tenreyro (2006). The estimated coefficient of δ is statistically significant at 1% with the magnitude of -1.03187 and a standard error of 0.04906, justifying the use of $\delta = -1$ in the calculation of the market potential.

Table 4 displays our estimates of market potential by state for 1880 and 1920. Two points stand out. First, the rank order of market potential is very stable during this period. Second, the 'manufacturing-belt' states tend to have the highest levels of market potential in both years. It should be noted that states with similar GDP inside and outside the manufacturing belt generally have quite different levels of market potential; for example, Delaware and Arizona have very similar GDP but, as Table 4 shows, market potential of the former was about seven times that of the latter.

includes wages, salaries and proprietor's income in agriculture and six non-agriculture sectors; property income includes rental income, personal interest income and dividends, in agriculture and six non-agriculture industries. The non-agriculture sectors consist of manufacturing, mining, construction, transportation and communication and public utilities, private households including domestic service performed in private households, and 'all other' which includes finance, trade, government and other services than domestic services. The re-estimated 1880 and 1900 figures in Klein (2009) are very close to Easterlin's original estimates.

12 In the case of the pair New Jersey and New York, we use the capital cities, Trenton and Albany, rather than the largest cities, Newark and New York City, because the latter are very close to each other leading to a distorted measure of market potential.

13 In 1949, the interstate highway network was still in the future.

Table 4. Market potential and the rank of states based on market potential in 1880 and 1920 Market potential estimates based on $\delta = -1$, in millions of current $U.S.

	1880		1920			1880		1920	
	Market potential	Rank	Market potential	Rank		Market potential	Rank	Market potential	Rank
Delaware	45.47	1	307.22	1	Arkansas	12.02	29	80.99	30
Connecticut	36.16	2	246.35	2	Alabama	11.77	30	87.57	26
Pennsylvania	34.80	3	240.58	3	Mississippi	11.74	31	77.88	32
New York	33.35	4	223.86	4	Oklahoma	11.51	32	80.65	31
New Jersey	32.28	5	223.36	5	Louisiana	11.22	33	68.59	35
Rhode Island	32.08	6	210.25	6	Florida	10.80	34	68.66	34
Massachusetts	29.57	7	189.62	7	North Dakota	10.10	35	65.58	36
Maryland	27.22	8	181.75	8	Texas	8.97	36	56.38	38
New Hampshire	26.14	9	164.52	9	South Dakota	8.70	37	75.09	33
Ohio	22.24	10	146.17	10	Colorado	8.50	38	56.08	39
Indiana	22.06	11	146.12	11	Wyoming	8.29	39	57.08	37
Kentucky	20.71	12	134.71	13	New Mexico	7.53	40	49.16	40
Maine	20.04	13	119.83	18	Nevada	7.48	41	40.23	48
Illinois	20.04	14	138.81	12	Utah	7.25	42	46.06	43
Vermont	19.24	15	118.27	19	Montana	7.25	43	45.28	44
Wisconsin	19.21	16	133.96	14	California	7.03	44	47.28	41
Virginia	18.95	17	125.85	17	Idaho	6.93	45	43.48	46
West Virginia	18.87	18	128.29	15	Oregon	6.62	46	42.14	47
Michigan	18.71	19	126.53	16	Arizona	6.56	47	44.56	45
Tennessee	15.84	20	90.51	24	Washington	6.47	48	46.11	42
Missouri	15.79	21	105.05	20					
Iowa	14.56	22	95.48	22					
North Carolina	14.21	23	104.28	21					
Georgia	13.75	24	91.32	23					
South Carolina	13.07	25	83.06	29					
Minnesota	12.69	26	83.81	27					
Nebraska	12.66	27	83.76	28					
Kansas	12.43	28	88.67	25					

Source: see the text.

4. Empirical Results

4.1. Estimation issues

This section discusses the statistical properties and robustness of the results while their historical interpretation is left to the following section. Estimation of Equation (3) raises the following issues: heteroskedasticity, endogeneity of some of the regressors and the use of panel data techniques. Our data, as seen from the specification of the regression equation, have three dimensions: industry k, state i and time t. Leaving aside the time dimension for a moment, state and industry dimensions are potential sources of heteroskedasticity. Furthermore, having 19 industries in each of 48 U.S. states suggests that we might face an unobserved cluster effect coming from the U.S. states. Failure to take this into account could have a dramatic effect on t-statistics (Pepper, 2002), which would then invalidate our statistical inference. Indeed, cluster-robust standard errors place no restriction on heteroskedasticity and correlation within clusters. A standard way of accounting for the clustering is to use cluster-robust standard errors (White, 1984; Arellano, 1987). Recently, a two-way cluster-robust standard error estimation technique has been proposed, which would call for clustering not only at the level of U.S. states but also at the level of industries (Cameron et al., 2011).[14]

The issue of endogeneity arises for two reasons. First, there is a direct implication of the unobserved cluster effect discussed in the previous paragraph. Using cluster-robust standard errors assumes that the unobserved cluster effect is not correlated with the regressors. However, if this assumption were invalid, then the estimators would be inconsistent. In this case, a 'within' estimator that would sweep away the unobserved within-cluster effect is attractive (Wooldridge, 2003; Cameron and Trivedi, 2005, J. M. Wooldridge, unpublished data). Second, market potential and hence its corresponding interactions may be endogenous. This calls for instrumental variable estimation. In our setting, we have to rely on an exogenous geographical determinant, as used in several recent studies (e.g. Redding and Venables, 2004; Head and Mayer, 2006; Knaap, 2006). We consider three instruments: distance to New York City, the inverse of the distances to New England and Middle Atlantic, and the sum of inverse distances between the U.S. states.[15]

Recent research has shown that instrumental variable estimation has its pitfalls. Although it provides consistent estimates, it is much less efficient than the OLS estimator (Wooldridge, 2002). This is exacerbated when the correlation between instruments and instrumented variables is weak, leaving us with IV estimation of low precision (Staiger and Stock, 1997; Kleibergen, 2002; Hahn and Hausman, 2003). Another profound implication of weak instruments is that even mild instrument endogeneity can lead to IV being even more inconsistent than OLS (Bound et al., 1995). To account for this, we perform weak instrument tests to justify the appropriateness of using instrumental variables estimation. In addition, we follow the suggestion of Wooldridge (2002, 104) and perform endogeneity tests on the suspect regressors.

Returning to the time dimension, its presence naturally calls for the use of panel data techniques. However, panel data estimation is done on pooled data, which assumes the same parameters over time and across regions. In our case, pooling the data across time might not be that innocent. Indeed, the period 1880–1920 is known for dramatic changes in the U.S economy, which suggests a cautious approach to pooling the data across time. Consequently, a testing of poolability is carried out to see whether panel data techniques should be used or not (Baltagi, 2005).

As was mentioned above, Equation (2) can be estimated either with all the industry and state characteristics or with state and industry dummies, as in Wolf (2007). All the regressions discussed in the next section were estimated both ways and the differences in the magnitude and the statistical significance of the six interaction variables, which are the main variables of interest, are miniscule. We present the results using Wolf's (2007)

14 Existing studies using models similar to ours by Midelfart-Knarvik et al. (2000), Crafts and Mulatu (2006) and Wolf (2007) do not take into account the clustering of standard errors.

15 We have also used lagged variables as the instruments and the results of the regression analysis conducted below were confirmed. These additional results are available from the authors upon request.

specification; the results with the full set of industry and state characteristics are available from the authors upon request.

4.2. The basic results

For each year 1880, 1890, 1900, 1910 and 1920, we have estimated Equation (3) with OLS using cluster-robust standard errors and cluster-specific fixed effects.[16] The reason for using cluster-robust standard errors is the possibility that there is an unobserved cluster effect, which needs to be taken into account. We use both one-way as well as two-way cluster-robust standard errors, although it should be noted that the latter is implemented using only a small number of clusters in one dimension, as there are only 19 industries.

The cluster-robust standard errors estimator assumes, however, that the unobserved cluster effect is not correlated with the regressors and puts it into the composite error term ε_i^k. If the unobserved cluster effect actually happens to be correlated with the regressors, the OLS estimator becomes inconsistent. Therefore, we have also estimated a cluster-specific fixed effect, to allow for the possibility of that correlation.[17] The results are presented in Tables 5 and 6 using one- and two-way clustering, respectively.

A general overview of the estimation results in Table 5 suggests that market potential interaction variables matter in each of the years, though some variation exists before 1900. The H–O interactions are less prevalent except for agriculture in 1890 and 1900. Of the market potential interactions, the plant-size interaction is always statistically significant, usually at the 1% significance level. The backward-linkages interaction is significant, except for 1890 and 1910. Forward linkages are first significant in 1890, after that, they remain significant until 1920. The H–O interactions are very different in terms of significance. The skilled-labour interaction changes signs and is insignificant for most of the time. The coal interaction is significant with a correct sign in 1890 only.[18] The agriculture interaction, on the other hand, is statistically significant in 1890 and 1900, before and after which it is insignificant.[19] The estimation results in Table 6 show that, despite some loss of the statistical significance, the overall picture that emerged from Table 5 can still be seen: H–O interactions are jointly significant, the plant-size interaction is always statistically significant and the significance of the forward-linkages interaction increases by 1920.

16 As was argued earlier, pooling data across time might pose a problem. Bearing in mind that the U.S. economy was undergoing dramatic changes in 1880–1920, the assumption of the same parameters across time could be too strong. Therefore, we carried out a Chow test to determine whether the data should be pooled or not. The calculated F-statistic F(23, 4465) is 27.2265 which enables us to reject the null hypothesis that $\beta[j]_t = \beta[j]\ \forall t$ at the 1% significance level.

17 Even in the case of cluster-specific fixed effect estimation, we use cluster-robust standard errors to estimate a fully robust variance matrix, as shown in Wooldridge (2003; J. M. Wooldridge, unpublished data). We have also estimated the cluster-specific random effect model, and the results remain qualitatively unchanged; they are available from the authors upon request.

18 The availability of coal in a U.S. state is captured by coal prices. This implies that the correct sign of the coal interaction is negative—a low price of coal makes a U.S. state attractive for the manufacturing firms.

19 An *F*-test for joint significance of the H-O interactions shows that the null hypothesis cannot be rejected for 1910 and 1920.

Table 5. OLS and cluster-specific fixed effect, estimations year by year

	1880		1890		1900		1910		1920	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	H–O forces									
Agricultural farm land × Agricultural input use	0.0007	0.0007	0.001*	0.001*	0.0009*	0.0009*	0.0007	0.0007	0.00007	0.00007
	(0.0005)	(0.0005)	(0.00065)	(0.0006)	(0.0005)	(0.0005)	(0.0005)	(0.00053)	(0.0005)	(0.0004)
Educated population × White-collar workers	0.006	0.006	−0.01	−0.01	−0.02**	−0.02**	−0.0009	−0.0009	0.001	0.001
	(0.004)	(0.004)	(0.008)	(0.0078)	(0.0056)	(0.005)	(0.0042)	(0.004)	(0.001)	(0.0012)
Coal abundance × Steam power use	0.15*	0.15*	−0.18**	−0.18**	0.02	0.02	0.06	0.06	0.006	0.006
	(0.08)	(0.076)	(0.08)	(0.079)	(0.049)	(0.048)	(0.069)	(0.068)	(0.035)	(0.034)
	Market potential forces									
Market potential × Intermediate input use	0.001	0.001*	0.002*	0.002*	0.001*	0.001*	0.0015**	0.0015**	0.0005***	0.0005***
	(0.0007)	(0.0006)	(0.001)	(0.00108)	(0.00058)	(0.00056)	(0.00072)	(0.00070)	(0.00017)	(0.00016)
Market potential × Industry sale	0.0014***	0.0014***	0.0008	0.0008	0.0012**	0.0012**	0.0003	0.0003	0.0001**	0.0001**
	(0.0005)	(0.0005)	(0.0006)	(0.0005)	(0.0005)	(0.0005)	(0.00042)	(0.00041)	(0.00007)	(0.00006)
Market potential × Size of establishment	0.0008***	0.0008***	0.0007***	0.0007***	0.0005***	0.0005***	0.0003***	0.0003***	0.00004**	0.00004**
	(0.00015)	(0.00014)	(0.0002)	(0.0002)	(0.0002)	(0.00018)	(0.00009)	(0.00009)	(0.00002)	(0.00002)
Constant	−3.5***	−3.4***	−0.3	−1.7	0.9	1.7	−2.8**	−2.6*	−1.9**	−2.3*
	(0.6)	(0.6)	(1.3)	(1.5)	(1.2)	(1.5)	(1.2)	(1.5)	(0.9)	(1.2)
State dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	912	912	912	912	912	912	912	912	912	912
R^2	0.75	0.46	0.7	0.45	0.67	0.48	0.64	0.46	0.61	0.36
Breusch–Pagan chi square (1)	299.83***		218.15***		289.87***		227.12***		259.35***	
F-test joint significance H–O	2.75*	2.9**	3.03**	3.2**	6.39***	6.75***	1.29	1.37	0.27	0.29

Notes: *Significant at 10%; **significant at 5%; ***significant at 1% OLS—cluster-robust SE, FE—cluster-specific fixed effect with cluster-robust se, clusters at the U.S. state level Degrees of Freedom in *F*-test are (3, 47) for OLS and FE.
Sources: see the text.

Table 6. OLS and cluster-specific fixed effect with two-way clustering, estimations year by year

	1880		1890		1900		1910		1920	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
	H–O forces									
Agricultural farm land × Agricultural input use	0.0007	0.0007	0.001	0.001	0.0009	0.0009	0.0007	0.0007	0.00007	0.00007
	(0.001)	(0.001)	(0.001)	(0.001)	(0.0012)	(0.0012)	(0.0014)	(0.0013)	(0.001)	(0.001)
Educated population × White-collar workers	0.006	0.006	−0.01	−0.01	−0.02	−0.02	−0.0009	−0.0009	0.001	0.001
	(0.005)	(0.008)	(0.014)	(0.013)	(0.013)	(0.012)	(0.01)	(0.009)	(0.0016)	(0.0015)
Coal abundance × Steam power use	0.15**	0.15**	−0.18	−0.18	0.02	0.02	0.06	0.06	0.006	0.006
	(0.06)	(0.08)	(0.12)	(0.13)	(0.05)	(0.05)	(0.06)	(0.04)	(0.03)	(0.07)
	Market potential forces									
Market potential × Intermediate input use	0.001	0.001	0.002****	0.002	0.001	0.001	0.0015*****	0.0015	0.0005**	0.00052**
	(0.002)	(0.0019)	(0.001)	(0.001)	(0.0012)	(0.001)	(0.0009)	(0.001)	(0.0002)	(0.0002)
Market potential × Industry sale	0.0014	0.0014	0.0008	0.0008	0.0012	0.0012	0.0003	0.0003	0.0001	0.0001
	(0.0015)	(0.0014)	(0.0009)	(0.001)	(0.001)	(0.001)	(0.00076)	(0.00077)	(0.00025)	(0.00026)
Market potential × Size of establishment	0.0008***	0.0008***	0.0007***	0.0007***	0.0005***	0.0005***	0.0003***	0.0003***	0.00004**	0.00004**
	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.00017)	(0.0002)	(0.00008)	(0.00009)	(0.00001)	(0.00002)
Constant	−3.5***	−6.8***	−0.3	−3.8**	0.9	−4.8***	−2.8	−7.5***	−1.9**	−5.9***
	(1.3)	(1.2)	(2.03)	(1.7)	(2.4)	(1.6)	(2.9)	(2.3)	(0.9)	(1.6)
State dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	912	912	912	912	912	912	912	912	912	912
R^2	0.75		0.7		0.67		0.64		0.61	
Breusch–Pagan chi-square (1)	299.83***		218.15***		289.87***		227.12***		259.35***	
F-test joint significance H–O	2.9**	5.03*	1.11	3.24	2.05*	7.61*	0.42	2.18	0.19	0.53

Notes: *Significant at 10%; **significant at 5%; ***significant at 1%, ****significant at 16%, *****significant at 11%.
OLS—cluster-robust se, FE—cluster-specific fixed effect with cluster-robust SE, clusters at the U.S. state and industry level. Degrees of Freedom in *F*-test are (3,911) for OLS and FE.
Sources: see the text.

4.3. Instrumental variable estimation

The endogeneity issue regarding market potential interactions is addressed by instrumental-variable estimation. As was noted earlier, we use three different instruments. The first one is the distance to an eastern seaport—New York City. The measure of physical distance between the U.S. state's largest city and New York City is highly correlated with the state's market potential and is considered to be a strong instrument as will be discussed below. However, it might be that there are some unobservable factors in $\varepsilon_{i,t}{}^{k}$, which are still correlated with that distance measure. Indeed, New York City was a major economic centre long before our time period and this might have influenced the geographical distribution of manufacturing labour force across U.S. states. Also, this measure might be correlated with the stock of educated workers or the locations of farms, thus confounding the instrument.[20] Therefore, we also use two other instruments to address these problems, namely, the sum of inverse distances to New England and Middle Atlantic States and the sum of inverse distances between the U.S. states.[21]

Instrumental variable estimation does not perform well in the presence of weak instruments. Therefore, we check whether our instruments are 'weak' or not using Shea's (1997) partial R^2 and the weak instrument test as suggested by Stock and Yogo (2005). In additional, we perform an endogeneity C-test (Hayashi, 2000, 233–234). Instrumental variable estimation is carried out using IV/2SLS as well as two-step GMM, which is more efficient than IV/2SLS. The differences in the magnitude and the statistical significance of the estimated coefficients are very small. For each year, we re-estimate Equation (3) and we use both one-way and two-way cluster-robust standard errors. The results of IV/2SLS estimation with one-way and two-way clustering are in Tables 7–9 and Tables 10–12, respectively. The results of two-step GMM are available from the authors upon request.[22]

First, we check the correlation between our instruments and instrumented market potential and the corresponding interactions. Shea's partial R^2 in Tables 7–12 show a very strong correlation between the instruments and the instrumented variables, ranging from 0.65 to 0.92. For the weak-instrument test, the relevant F-statistics are well above the critical values reported by Stock and Yogo (2005). Finally, the endogeneity tests in Tables 7–12 suggest that market potential and its interactions might be endogenous.[23]

The results in Tables 7–9 show that the market potential interaction variables are generally significant and have the correct sign. The plant-size interaction is statistically significant except in 1920 in Table 7 and usually at the 1% significance level, except for Table 7. The forward-linkage interaction is usually significant from 1890, and the significance rises by 1920. The estimated backward-linkages coefficients are significant,

20 We would like to thank a referee for pointing out this issue.
21 We follow the suggestion of Head and Mayer (2006) to use instruments which do not explicitly impose a centre.
22 We have also instrumented for the potentially endogenous H-O interaction with the educated population using average monthly mean temperature. The instrument, however, is weak and thus we decided not to include these results in Tables 7–12. These results are in the Supplementary data.
23 We have also performed the Sargant test of over-identifying restrictions and the results further confirm the usefulness of all instruments. The results are available from authors upon request.

Table 7. 2SLS instrumental variable estimation: instrument is distance to New York City

	1880	1890	1900	1910	1920
	H–O forces				
Agricultural farm land × Agricultural input use	0.0004	0.001*	0.00088*	0.0005	-0.00001
	(0.0005)	(0.0006)	(0.0005)	(0.0005)	(0.0004)
Educated population × White-collar workers	0.005	−0.01	−0.02***	−0.001	0.0009
	(0.004)	(0.008)	(0.006)	(0.004)	(0.001)
Coal abundance × Steam power use	0.16**	−0.2**	0.02	0.06	0.02
	(0.08)	(0.08)	(0.04)	(0.07)	(0.04)
	Market potential forces				
Market potential × Intermediate input use	0.0006	0.002**	0.001**	0.001*	0.0006**
	(0.001)	(0.001)	(0.0006)	(0.0008)	(0.0002)
Market potential × Industry sale	0.002***	0.0009	0.0015***	0.0006	0.0003**
	(0.0008)	(0.0008)	(0.0005)	(0.0004)	(0.0001)
Market potential × Size of establishment	0.0004**	0.0004**	0.0004***	0.0002**	0.00002
	(0.0002)	(0.0002)	(0.0001)	(0.00008)	(0.00002)
Constant	−3.2***	−0.1	0.99	−2.6**	−1.9**
	(0.7)	(1.2)	(1.3)	(1.3)	(0.9)
State dummy	Yes	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes	Yes
Observations	912	912	912	912	912
R^2	0.75	0.70	0.67	0.63	0.61
Shea partial R^2					
mp1vs2_intermed	0.65	0.66	0.65	0.65	0.66
mp1vs2_sale	0.64	0.65	0.64	0.68	0.67
mp1vs2_plant	0.68	0.68	0.68	0.67	0.67
Endog. C test [chisq (3)]	28.6***	9.09**	6.69*	14.90***	11.2**
Joint significance					
Heckscher–Ohlin, chi-square (3)	6.55*	8.35**	20.55***	3.12	0.77

Notes: Regression with cluster-robust SE, Cluster at the U.S. state.
*Significant at 10%; **significant at 5%; ***significant at 1%.
Sources: see the text.

except for 1890 and 1910. The results in Table 10–12 show that H–O interactions and the backward linkages interaction lose statistical significance. Despite this, the general picture is the same: H–O interactions are jointly significant except for 1910 and 1920, the plant-size interactions are almost always significant and the forward-linkages interaction is significant in 1890, 1910 and 1920.

4.4. Robustness, standardized coefficients and the economic significance of the results

We have also undertaken additional robustness checks.[24] First, as an alternative way to address endogeneity, we also re-estimated Equation (3) with a revised market potential variable, which was calculated summing distance-deflated GDP as usual except for

24 Results are available from authors on request.

Table 8. 2SLS instrumental variable estimation: instrument is sum of inverse distances to New England and Middle Atlantic states

	1880	1890	1900	1910	1920
	H–O forces				
Agricultural farm land × Agricultural input use	0.0007	0.001**	0.0009*	0.0007	−0.00005
	(0.0004)	(0.0006)	(0.0005)	(0.0005)	(0.0004)
Educated population × White-collar workers	0.006	−0.008	−0.02***	−0.0008	0.001
	(0.004)	(0.008)	(0.005)	(0.004)	(0.001)
Coal abundance × Steam power use	0.2**	−0.2**	0.02	0.06	0.008
	(0.07)	(0.08)	(0.05)	(0.06)	(0.03)
	Market potential forces				
Market potential × Intermediate input use	0.0006	0.001*	0.0008	0.001**	0.0004***
	(0.0006)	(0.0008)	(0.0005)	(0.0005)	(0.0001)
Market potential × Industry sale	0.001***	0.0007	0.0012***	0.0002	0.0002***
	(0.0005)	(0.0005)	(0.0004)	(0.0004)	(0.00007)
Market potential × Size of establishment	0.001***	0.0008***	0.0006***	0.0003***	0.00003*
	(0.0002)	(0.0003)	(0.0002)	(0.00008)	(0.00002)
Constant	−3.5***	−0.4	1.05	−2.5**	−1.7**
	(0.6)	(1.2)	(1.2)	(1.1)	(0.8)
State dummy	Yes	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes	Yes
Observations	912	912	912	912	912
R^2	0.75	0.77	0.67	0.63	0.61
Shea partial R^2					
mp1vs2_intermed	0.71	0.68	0.71	0.68	0.71
mp1vs2_sale	0.70	0.66	0.69	0.67	0.69
mp1vs2_plant	0.72	0.68	0.71	0.68	0.70
Endog. C test [chi-square (3)]	18.02***	4.3	1.66	5.48	2.28
Joint significance					
Heckscher–Ohlin, chi-square (3)	10.65**	9.54**	20.11***	3.44	0.89

Notes: regression with cluster-robust se, Cluster at the U.S. state level.
*Significant at 10%; **significant at 5%; ***significant at 1%.
Sources: see the text.

omitting own GDP. The results that were obtained are again very similar. The market potential interactions are generally significant while over time the linkage interactions become stronger; the agriculture factor endowment interaction is significant initially but not after 1900.

We have also checked the robustness of the H–O interaction variables and market potential interactions. As for the H–O interactions, we have used the share of agriculture in the labour force instead of the share of farm land in the agricultural-interaction variable, the share of mining in the labour force instead of coal prices and the share of coal inputs in gross product instead of the ratio of horse power to gross output in the coal-interaction variable.[25] In both cases, the qualitative results

25 The share of agricultural and mining employment in each U.S. state is calculated from Perloff et al. (1960); the share of coal inputs in gross product comes from Whitney (1968) and Leontief (1941).

Table 9. 2SLS instrumental variable estimation: instrument is sum of inverse distances between the U.S. states

	1880	1890	1900	1910	1920
	H–O forces				
Agricultural farm land × Agricultural input use	0.0007	0.001**	0.0009*	0.0007	0.00004
	(0.0004)	(0.0006)	(0.0005)	(0.0005)	(0.0004)
Educated population × White-collar workers	0.005	−0.01	−0.02***	−0.001	0.001
	(0.004)	(0.008)	(0.005)	(0.004)	(0.001)
Coal abundance × Steam power use	0.1**	−0.2**	0.02	0.06	0.006
	(0.07)	(0.08)	(0.05)	(0.07)	(0.03)
	Market potential forces				
Market potential × Intermediate input use	0.001*	0.002*	0.001*	0.0014**	0.0005***
	(0.0007)	(0.001)	(0.0006)	(0.0007)	(0.0002)
Market potential × Industry sale	0.001***	0.0008	0.001***	0.0003	0.0002**
	(0.0005)	(0.0005)	(0.0004)	(0.0004)	(0.00006)
Market potential × size of establishment	0.0008***	0.0006***	0.0005***	0.0003***	0.00003**
	(0.0002)	(0.0002)	(0.0002)	(0.00009)	(0.00002)
Constant	−3.5***	−0.3	0.99	−2.6**	−1.9**
	(0.6)	(1.2)	(1.2)	(1.1)	(0.9)
State dummy	Yes	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes	Yes
Observations	912	912	912	912	912
R^2	0.75	0.70	0.67	0.63	0.61
Shea partial R^2					
mp1vs2_intermed	0.91	0.90	0.91	0.89	0.89
mp1vs2_sale	0.91	0.90	0.90	0.90	0.89
mp1vs2_plant	0.92	0.91	0.91	0.90	0.90
Endog. C test [chi-square (3)]	0.25	0.40	3.10	3.40	4.26
Joint significance					
Heckscher–Ohlin, chi-square (3)	8.95**	9.85**	20.80***	3.58	0.87

Notes: regression with cluster-robust se, Cluster at the U.S. state level.
*Significant at 10%; **significant at 5%; ***significant at 1%.
Sources: see the text.

are similar to the results in Tables 5 and 6 and 7–12, with agriculture being the most prevalent among all H–O interaction variables.[26] As for the market potential interactions, we checked the robustness of the results doing the following. First, we used the distances between the capitals of the U.S. states, not their largest cities. Second, we have explicitly taken into account the possible impact of navigable waterways on our distance assumption by estimating all the regressions with a dummy being 1 if a state has an access to navigable waterways for inter-state trade and 0 otherwise, as well as interacting that dummy with the market potential interactions. Third, we have also estimated the regressions with market potential based on foreign

26 The shares of agriculture and mining in the labour force are also potentially endogenous. Therefore, we ran regressions in which they were instrumented with their lagged values. The sign and the statistical significance are the same as when they are treated as exogenous.

Table 10. 2SLS instrumental variable estimation with two-way clustering: instrument is sum of inverse distances between the U.S. states: instrument is distance to New York City

	1880	1890	1900	1910	1920
	H–O forces				
Agricultural farm land × Agricultural input use	0.0004	0.001	0.001	0.0006	−0.00003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.0009)
Educated population × White-collar workers	0.006	−0.007	−0.02*	−0.002	0.001
	(0.007)	(0.006)	(0.01)	(0.009)	(0.001)
Coal abundance × Steam power use	0.15**	−0.19	0.02	0.07	0.01
	(0.08)	(0.14)	(0.05)	(0.05)	(0.06)
	Market potential forces				
Market potential × Intermediate input use	0.0008	0.003	0.0019	0.002*	0.0007*
	(0.003)	(0.002)	(0.0019)	(0.001)	(0.0005)
Market potential × Industry sale	0.003	0.001	0.0017	0.001	0.0003
	(0.002)	(0.002)	(0.0015)	(0.001)	(0.0004)
Market potential × Size of establishment	0.0005	0.001*	0.0005**	0.0002**	0.00002
	(0.0003)	(0.0003)	(0.0002)	(0.0001)	(0.00002)
Constant	−3.41***	−1.39	0.7	−2.7	−2.1***
	(0.73)	(1.71)	(2.5)	(2.6)	(0.8)
State dummy	Yes	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes	Yes
Observations	912	912	912	912	912
R^2	0.74	0.70	0.67	0.63	0.61
Shea partial R^2					
mp1vs2_intermed	0.75	0.76	0.76	0.76	0.75
mp1vs2_sale	0.74	0.75	0.75	0.78	0.76
mp1vs2_plant	0.77	0.77	0.78	0.77	0.76
Endog. C test [chi-square (3)]	35.5***	16.85***	17.6***	33.01***	20.8***
Joint significance					
Heckscher–Ohlin, chi-square (3)	4.65*	8.22**	6.82*	1.74	0.56

Notes: regression with cluster-robust se, Cluster at the U.S. state and industry level.
*Significant at 10%; **significant at 5%; ***significant at 1%.
Sources: see the text.

countries only. Again, in all cases, the qualitative results are similar to the results in Tables 5, 6 and 7–12.[27]

Overall, these results show the statistical importance of two of the market potential interactions and the H–O forces jointly during the period 1880–1920, irrespective of the estimation technique. The factor endowment interactions are always found to be weaker than the market potential interactions by the early 20th century and this reflects a tendency in all estimations for forward-linkage effects to strengthen over time while the agricultural endowments effects weaken or are statistically insignificant depending on the method of estimation.

27 The regression results with the navigable waterway dummy and market potential based on foreign countries, respectively, only are in the Supplementary Material. Other regression results are available from authors upon request.

Table 11. 2SLS instrumental variable estimation with two-way clustering: instrument is sum of inverse distances to New England and Middle Atlantic states

	1880	1890	1900	1910	1920
	H–O forces				
Agricultural farm land × Agricultural input use	0.0007	0.0013	0.0009	0.0007	−0.00008
	(0.001)	(0.001)	(0.001)	(0.0013)	(0.0008)
Educated population × White-collar workers	0.006***	−0.009	−0.021*	−0.0016	0.001
	(0.002)	(0.012)	(0.012)	(0.009)	(0.0014)
Coal abundance × Steam power use	0.16**	−0.18	0.014	0.05	0.006
	(0.07)	(0.1)	(0.04)	(0.035)	(0.06)
	Market potential forces				
Market potential × Intermediate input use	0.0005	0.0017***	0.0009	0.0014*	0.0004*
	(0.002)	(0.0006)	(0.0008)	(0.0009)	(0.0003)
Market potential × Industry sale	0.0017	0.0007	0.001	0.0002	0.0002
	(0.001)	(0.001)	(0.0013)	(0.0009)	(0.0003)
Market potential × Size of establishment	0.001***	0.0009***	0.0007***	0.0004***	0.00004**
	(0.0002)	(0.0003)	(0.0002)	(0.00008)	(0.00002)
Constant	−3.7***	−0.58	0.93	−2.5	−1.7***
	(1.4)	(2.1)	(2.44)	(2.9)	(0.6)
State dummy	Yes	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes	Yes
Observations	912	912	912	912	912
R^2	0.75	0.70	0.67	0.64	0.61
Shea partial R^2					
mp1vs2_intermed	0.76	0.75	0.75	0.73	0.76
mp1vs2_sale	0.75	0.73	0.73	0.71	0.74
mp1vs2_plant	0.76	0.75	0.74	0.72	0.75
Endog. C test [chi-square (3)]	20.1***	7.4*	5.4	9.3**	6.4*
Joint significance					
Heckscher–Ohlin, chi-square (3)	19.6***	4.5	16.05***	2.6	0.69

Notes: regression with cluster-robust se, Cluster at the U.S. State and industry level.
*Significant at 10%; **significant at 5%; ***significant at 1%.
Sources: see the text.

This suggests that industrial location was driven both by agglomeration mechanisms related to market potential and natural advantages, although the former seem to be the stronger. We can support this inference by calculating standardized or so-called β-coefficients of all the interaction variables. The β-coefficients provide a comparison of the relative importance of the interaction variables in determining state shares of manufacturing employment by industry. The results reported in Table 13 show that throughout 1880–1920 the sum of the contributions of the market potential interactions exceeds that of the H–O interactions and this is increasingly the case over time. Among the market potential interactions, scale economies always have a substantial impact but it is noticeable that forward linkages become more important over time and that, by 1920, the contribution of linkages outweighs everything else.

To evaluate the economic significance of the market potential interaction variables and the interaction variables capturing the states' natural advantages, we follow Redding and Venables (2004) and examine their effect on the predicted share of

Table 12. 2SLS instrumental variable estimation with two-way clustering: Instrument is sum of inverse distances between the U.S. states

	1880	1890	1900	1910	1920
	H–O forces				
Agricultural farm land × Agricultural input use	0.0007	0.001	0.0009	0.0007	0.000003
	(0.001)	(0.001)	(0.001)	(0.001)	(0.00091)
Educated population × White-collar workers	0.006	-0.01	−0.02*	−0.0015	0.001
	(0.007)	(0.012)	(0.01)	(0.009)	(0.0013)
Coal abundance × Steam power use	0.16**	−0.18	0.03	0.06	0.008
	(0.07)	(0.12)	(0.05)	(0.05)	(0.064)
	Market potential forces				
Market potential × Intermediate input use	0.00147	0.002*	0.001	0.0018*	0.0007**
	(0.00234)	(0.001)	(0.0014)	(0.0009)	(0.0003)
Market potential × Industry sale	0.00178	0.001	0.0018	0.0005	0.0002
	(0.00168)	(0.0012)	(0.001)	(0.0009)	(0.0003)
Market potential × Size of establishment	0.00098***	0.0007***	0.0006**	0.0003***	0.00003*
	(0.00023)	(0.0003)	(0.0002)	(0.00009)	(0.00002)
Constant	−3.8***	−0.72	0.55	−2.77	−2.05***
	(1.3)	(1.8)	(2.30)	(2.7)	(0.72)
State dummy	Yes	Yes	Yes	Yes	Yes
Industry dummy	Yes	Yes	Yes	Yes	Yes
Observations	912	912	912	912	912
R^2	0.75	0.70	0.67	0.64	0.61
Shea partial R^2					
mp1vs2_intermed	0.84	0.85	0.85	0.85	0.85
mp1vs2_sale	0.84	0.85	0.85	0.85	0.85
mp1vs2_plant	0.85	0.86	0.86	0.86	0.86
Endog. C test [chi-square (3)]	0.87	4.01	11.9***	12.1***	13.9***
Joint significance					
Heckscher–Ohlin, chi-square (3)	5.4*	3.82	7.54*	2.00	0.65

Notes: Regression with cluster-robust se, Cluster at the U.S. State and industry level.
*Significant at 10%; **significant at 5%; ***significant at 1%.
Sources: see the text.

manufacturing employment using counterfactual analysis. Our illustrative example uses the estimated coefficients reported in Table 5, Equation FE to examine the impact of the changes in the geographical location of a U.S. state as well as changes in a state's natural advantages. Specifically, a change in geographical location is captured by imposing a change in its market potential while a change in natural advantages is captured by varying the share of farm land and coal prices, respectively. All of these changes are investigated for 1890 since it is the only year in which the coal price interaction has the correct sign and is statistically significant, allowing us to compare the agricultural interaction, the coal price interaction and the market potential interactions.[28]

28 We use this equation to err on the side of generosity in the importance of factor endowments. Even so, it is clear that the responsiveness of industrial location to these variables is much weaker than to market-potential.

Table 13. β-coefficients, estimations year-by-year

	1880	1890	1900	1910	1920
	H–O forces				
Agricultural farm land × Agricultural input use	0.05	0.08	0.08	0.06	0.006
Educated population × White-collar workers	0.03	−0.07	−0.23	−0.01	0.06
Coal abundance × Steam power use	0.11	−0.09	0.02	0.05	0.01
	NEG forces				
Market potential × Intermediate input use	0.07	0.14	0.12	0.25	0.23
Market potential × Industry sale	0.11	0.07	0.16	0.06	0.08
Market potential × Size of establishment	0.26	0.24	0.30	0.35	0.20

Note: The table presents only the β-coefficients of the interaction variables. The full set of the β-coefficients is available from the authors upon request. The β-coefficients are defined as $\beta\text{-}(i)=[s(x_i)/s(y)]^*b(x_i)$, where $b(x_i)$ is the estimates of x_i, $s(x_i)$ is the standard deviation of x_i and $s(y)$ is the standard deviation of y. β-Coefficients are calculated from the OLS regressions in Table 5.
Sources: see the text.

To quantify the importance of proximity to large markets, we undertake two hypothetical experiments: first, we increase the market potential of four states, which have very low market potential by 10%; second, we decrease the market potential of four states that have very high market potential by 10%. We see from Table 14, Columns 1 and 2 that an increase of market potential by 10% generates an increase of the state's share of manufacturing employment in total U.S. manufacturing employment that ranges from 13% to 27%, and that a 10% decrease generates a decrease ranging from 29% to 44%. This means that, for example, lowering the market potential of the state of New York by 10%, which causes a 35.8% drop in the New York's share of manufacturing employment in total U.S. manufacturing employment, decreases that share from 16.3% to 10.45%.

Similarly, we examine the effect of the change in coal prices and the share of farm land on the share of the state's manufacturing employment in total U.S. manufacturing employment by considering 10% increases or decreases. The results in Table 14, Columns 3–6 show that the effects are smaller in comparison with the effects of market potential. For example, an increase of the share of farm land by 10% causes an increase of the state's share of total U.S. manufacturing employment by between 0.13% and 2.34%, and a decrease of coal prices by 10% results in a 3% to 8% increase in that share. As in the previous example, let us consider the state of New York. An increase in the price of coal by 10%, which lowers the state's share of manufacturing employment in total U.S. manufacturing employment by 2.2%, decreases that share from 16.3% to 15.9%.

Finally, we can decompose the contribution of a counterfactual change in market potential on a state's share of manufacturing employment into components from each of forward linkages, backward linkages, and plant size.[29] When this is done, it turns out

29 An exercise of this kind is described in Appendix B.

Table 14. Economic importance of market potential, coal prices and farm land on U.S. states' shares of manufacturing employment in 1890 (%)

U.S. state	Change of market potential (10%)		Change of share of farm land (10%)		Change of coal prices (10%)	
	Increase (1)	Decrease (2)	Increase (3)	Decrease (4)	Decrease (5)	Increase (6)
Nebraska	27.06		2.34		4.49	
Utah	14.46		0.13		3.01	
California	13.45		1.13		7.09	
Washington	13.64		0.51		8.07	
Illinois		−29.74		−6.02		−3.03
New York		−35.79		−5.12		−2.19
Ohio		−31.45		−6.33		−2.55
Pennsylvania		−44.24		−4.57		−2.67

Note: The table reports the predicted effect of a change in geographical and economic characteristics of the U.S. states on their share of manufacturing employment on the total U.S. manufacturing employment. The predicted effects are based on the estimated coefficients for the year 1890 in Tables 5 and 6, equation FE.

that forward linkages have about 1.4 times the impact of plant size, which, in turn, has about 1.9 times the impact of backward linkages.

5. Discussion of the Results

We have used an estimation procedure that allows both H–O and NEG interactions between state and industry characteristics to influence industrial location. The results have a clear pattern that can be observed across a variety of estimating techniques and is robust to concerns about endogeneity and clustering. There is robust support for the proposition that both NEG interactions played a part in industrial location decisions around the turn of the 20th century with rather weaker evidence for H–O interactions. However, even if both factor endowments and market potential influence industrial location, the latter was clearly more important and its impact was felt both through forward-linkage effects and the attraction of market access for sectors where plant size was relatively large. Across all our results, these market potential interactions are the most robust influences on industrial location and they become stronger relative to factor endowment interactions over time.

The overall pattern of our results is consistent with the traditional accounts of industrial location reviewed in the introduction. They would not come as a great surprise to the authors of the report in National Resources Committee (1939) who estimated that, in 1935, 743,000 manufacturing jobs were resource oriented compared with 887,000 that were tied to local consumers and 6,881,000 that were 'footloose'. The strong showing of linkage effects matches the account given by Perloff et al. (1960).

On the other hand, the picture that we paint differs considerably from that sketched by Kim (1999). We believe that Kim's failure to take account of linkage effects is an important omission and has led him to exaggerate the role of factor endowments. Our model takes account of, and finds some evidence for, Rybcynski effects through the factor endowments interactions but makes Kim's claim that these are virtually the whole story seem implausible.[30]

It should also be noted that our emphasis on linkage effects also implies a somewhat different explanation for the existence of the manufacturing belt from the one popularized by Krugman (1991a, 1991b). His account stressed the interaction between market potential and plant size in the context of transport costs reduced by the railroad together with many footloose producers.[31] Our results suggest that while there is robust evidence of this effect throughout our period, it is important also to recognize the role of forward-linkage effects in the early 20th century. This has resonance with the findings of Ellison et al. (2010) that in the late 20th century customer–supplier relationships in the form of input–output linkages are the most important Marshallian reason for coagglomeration.

Finally, it is important to note an important qualification to our findings, namely, that our argument applies to the persistence of the manufacturing belt, which we believe was cemented by linkage effects, not to its emergence. We do not have the data to test hypotheses about the latter. It may well be that in some cases the origins of an industrial cluster can be found in the direct or indirect effects of natural resources but that, once established, the cluster was sustained for market access reasons. In fact, a case in point is automobiles. At the turn of the 20th century, Detroit was already a leading city in making small stationary gasoline engines, marine gasoline engines, wagons and carriages. This was largely due to hardwood forests that provided an excellent material for the production of wagons and carriages and the presence of lakes that stimulated the production of gasoline engines that were used to power boats. Having a large market for gasoline engines, wagons and carriages allowed Detroit to offer good supplier access to the automobile components such as bodies, wheels and internal combustion engines and Detroit emerged as industry's leading part supplier. As a result, car producers found the region very attractive and by the 1920s, Detroit became a leading producer of cars.[32]

6. Conclusion

In this article, we have implemented an empirical strategy suggested by Wolf (2007) to investigate the importance of market access and factor endowments in industrial location decisions in order to discover the reasons for the existence of the

30 We allow for an additional factor endowment, human capital, which was not considered by Kim (1999) but this does not have a significant effect. The work of Goldin and Katz (1998) suggests this is not surprising in our period. The relationship of the factor-endowment interactions to Rybcynski effects is made clear in the derivation of our model set out in the Supplementary Material.

31 Krugman (1991b) proposed a simple model in which manufacturing concentrates in one region out of two when $F > tx(1 - \pi)/2$ where F is fixed costs, t is transport cost, x is sales and π is the share of footloose workers. A similar line of reasoning is used by Meyer (1983, 1989) to explain why the Midwest but not the South joined the manufacturing belt. Clearly, plant sizes did increase (Atack, 1985), transport costs fell (Carter et al., 2006, p. 781) and footloose manufacturing grew in relative importance (Perloff et al., 1960).

32 A detailed analysis of the rise of the Midwest as the centre of the automobile industry is provided in Tsai (1999).

manufacturing belt in the USA at the turn of the 20th century. This allows us to give answers to the questions that we posed at the outset.

As far as factor endowments are concerned, we find that the strongest evidence is for the share of farmland in a state to play a role in manufacturing location at the end of the 19th century while there is quite robust evidence that factor endowment interactions are jointly significant at least up to the start of the 20th century. We find that market potential had a substantial impact on the location of manufacturing in the USA throughout the period 1880–1920 and that it was more important than factor endowments. The influence of market potential worked both through forward-linkage effects and scale effects, with the former becoming more important over time. Market access is found to be the central consideration that underpinned the manufacturing belt in the early 20th century.

Supplementary data

Supplementary data for this article are available at http://joeg.oxfordjournals.org/

References

Arellano, M. (1987) Computing robust standard errors for within-group estimators. *Oxford Bulletin of Economics and Statistics*, 49: 431–434.

Atack, J. (1985) Industrial structure and the emergence of the modern industrial corporation. *Explorations in Economic History*, 22: 29–52.

Baltagi, B. H. (2005) *Econometric Analysis of Panel Data*, 3rd edn. Chichester: John Wiley & Sons.

Bound, J., Jaeger, D. A., Baker, R. M. (1995) Problems with instrumental variable estimation when the correlation between the instruments and the endogenous explanatory variables is weak. *Journal of the American Statistical Association*, 90: 443–450.

Cameron, A. C., Gelbach, J. B., Miller, D. L. (2011) Robust inference with multi-way clustering. *Journal of Business & Economic Statistics*, 29: 238–249.

Carter, S., Gartner, S., Haines, M., Olmstead, A., Sutch, R., Wright, G. (2006) *Historical Statistics of the United States*. Cambridge: Cambridge University Press.

Combes, P-P., Mayer, T., Thisse, J-F (2008) *Economic Geography*. *The Integration of Regions and Nations*. Princeton: Princeton University Press.

Crafts, N., Mulatu, A. (2006) How did the location of industry respond to falling transport costs in Britain before World War I? *Journal of Economic History*, 66: 575–607.

Davis, R. D., Weinstein, D. E. (1999) Economic geography and regional production structure: an empirical investigation. *European Economic Review*, 43: 379–407.

Davis, R. D., Weinstein, D. E. (2003) Market access, economic geography and comparative advantage: an empirical test. *Journal of International Economics*, 59: 1–23.

Easterlin, R. A. (1957) State income estimates. In E. S. Lee, A. R. Miller, C. P. Brainerd, R. A. Easterlin (eds) *Population Redistribution and Economic Growth United States, 1870–1950, vol. I: Methodological Considerations and Reference Tables*. Philadelphia: The American Philosophical Society.

Ellison, G., Glaeser, E. L. (1999) The geographic concentration of industry: does natural advantage explain agglomeration? *American Economic Review Papers and Proceedings*, 89: 311–316.

Ellison, G., Glaeser, E. L., Kerr, W. R. (2010) What causes industry agglomeration? Evidence from co-agglomeration patterns. *American Economic Review*, 100: 1195–1213.

Flandreau, M., Zumer, F. (2004) *The Making of Global Finance, 1880–1913*. Paris: OECD.

Goldin, C. (1998) America's graduation from high school: the evolution and spread of secondary schooling in the twentieth century. *Journal of Economic History*, 58: 345–374.

Goldin, C., Katz, L. F. (1998) The origins of technology-skill complementarity. *Quarterly Journal of Economics*, 113: 693–732.
Green, A. G. (1971) *Regional Aspects of Canada's Economic Growth*. Toronto: University of Toronto Press.
Hahn, J., Hausman, J. (2003) Weak instruments: diagnosis and cures in empirical econometrics. *American Economic Review Papers and Proceedings*, 93: 118–125.
Harris, C. (1954) The market as a factor in the localization of industry in the United States. *Annals of the Association of American Geographers*, 64: 315–348.
Hayashi, F. (2000) *Econometrics*. Princeton: Princeton University Press.
Head, K., Mayer, T. (2004) Market potential and the location of Japanese investment in the European Union. *Review of Economics and Statistics*, 86: 959–972.
Head, K., Mayer, T. (2006) Regional wage and employment responses to market potential in the EU. *Regional Science and Urban Economics*, 36: 573–594.
Hillberry, R., Hummels, D. (2003) International home bias: some explanations. *Review of Economics and Statistics*, 85: 1089–1092.
Instituto Nacional de Estadistica, (1990) *Estadísticas Históricas de México,* Mexico City.
Kim, S. (1995) Expansion of markets and the geographic distribution of economic activities: trends in U. S. regional manufacturing structure, 1860–1987. *Quarterly Journal of Economics*, 110: 881–908.
Kim, S. (1999) Regions, resources, and economic geography: sources of U.S. regional comparative advantage, 1880–1987. *Regional Science and Urban Economics*, 29: 1–32.
Kleibergen, F. (2002) Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, 70: 1781–1803.
Klein, A. (2009) State Personal Income Estimates, United States 1880–1910. *Warwick Economic Research Paper No. 916,* Warwick University, Coventry.
Knaap, T. (2006) Trade, location, and wages in the United States. *Regional Science and Urban Economics*, 36: 595–612.
Krugman, P. (1991a) *Geography and Trade*. Cambridge, MA: MIT Press.
Krugman, P. (1991b) History and industry location: the case of the manufacturing belt. *American Economic Review Papers & Proceedings*, 81: 80–83.
Krugman, P. (1992) A Dynamic Spatial Model. *National Bureau of Economic Research Working Paper No. 4219*, Cambridge, MA.
Krugman, P., Venables, A. J. (1995) Globalization and the inequality of nations. *Quarterly Journal of Economics*, 110: 857–880.
Leontief, W. W. (1941) *The Structure of the American Economy, 1919–1939*. New York: Oxford University Press.
Maddison, A. (2004) World population, GDP and per capita GDP, 1-2003 AD. Available at: http://www.ggdc.net/Maddison/ [Accessed 1 February 2007].
Meyer, D. R. (1983) Emergence of the American manufacturing belt: an interpretation. *Journal of Historical Geography*, 9: 145–174.
Meyer, D. R. (1989) Midwestern industrialization and the American manufacturing belt in the nineteenth century. *Journal of Economic History*, 49: 921–937.
Midelfart-Knarvik, K. H., Overman, H. G., Venables, A. J. (2000) Comparative Advantage and Economic Geography: Estimating the Determinants of Industrial Location in the EU. *CEPR Discussion Paper No. 2618,* London.
Mitchell, B. R. (2003a) *International Historical Statistics, Europe 1750–2000*, 5th edn. Basingstoke: Palgrave Macmillan.
Mitchell, B. R. (2003b) *International Historical Statistics, Americas 1750–2000*, 5th edn. Basingstoke: Palgrave Macmillan.
Niemi, A. W. (1974) *State and Regional Patterns in American Manufacturing, 1860–1900*. London: Greenwood Press.
Nunn, N. (2007) Relationship-specificity, incomplete contracts and the pattern of trade. *Quarterly Journal of Economics*, 122: 569–600.
O'Brien, A. P. (1988) Factory size, economies of scale, and the great merger wave of 1898–1902. *Journal of Economic History*, 48: 639–649.
Officer, L. H. (2008) *Dollar-Pound Exchange Rate From 1791, Measuring Worth.* Available online at: http://www.measuringworth.org/exchangepound/ [Accessed 13 December 2008].

Pepper, J. P. (2002) Robust inferences from random clustered samples: an application using data from the panel study of income dynamics. *Economic Letters*, 75: 341–345.
Perloff, H. S., Dunn, E. S., Lampard, E. E., Muth, R. F. (1960) *Regions, Resources, and Economic Growth*. Lincoln: University of Nebraska Press.
Redding, S., Venables, A. J. (2004) Economic geography and international inequality. *Journal of International Economics*, 62: 53–82.
Romalis, J. (2004) Factor proportions and the structure of commodity trade. *American Economic Review*, 94: 67–97.
Santos Silva, J. M. C, Tenreyro, S. (2006) The log of gravity. *Review of Economics and Statistics*, 88: 641–658.
Shea, J. (1997) Instrument relevance in multivariate linear models: a simple measure. *Review of Economics and Statistics*, 79: 348–352.
Staiger, D., Stock, J. (1997) Instrumental variable regression with weak instruments. *Econometrica*, 65: 557–586.
Stock, J. H., Yogo, M. (2005) Testing for weak instruments in linear IV regression. In D. W. K. Andrews, J. H. Stock (eds) *Identification and Inference for Econometrics Models. Essays in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press.
Tsai, L. B.-G. (1999) *The Spatial Aggregation of Automobile Manufacturing Activities in the American Midwest*. PhD Dissertation, University of California, Berkeley.
Thomas, M. (1984) *An Input-Output Approach to the British Economy, 1890–1914*. DPhil Thesis, University of Oxford.
Ullman, E. L. (1957) *American Commodity Flow*. Seattle: University of Washington Press.
U. S. Historical Climatology Network. (2009) Available online at: http://cdiac.ornl.gov/ftp/ushcn_v2_monthly [Accessed 10 February 2011].
White, H. (1984) *Asymptotic Theory for Econometricians*. San Diego: Academic Press.
Whitney, W. G. (1968) *The Structure of the American Economy in the Late Nineteenth Century*. PhD Dissertation, Harvard University.
Wolf, H. C. (2000) International home bias in trade. *Review of Economics and Statistics*, 82: 555–563.
Wolf, N. (2007) Endowments vs. market potential: What explains the relocation of industry after the Polish reunification in 1918?'' *Explorations in Economic History*, 44: 22–42.
Wooldridge, J. M. (2002) *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.
Wooldridge, J. M. (2003) Cluster-sample methods in applied econometrics. *American Economic Review*, 93: 133–138.

U.S. Government Publications

Department of the Interior, Census Office (1883) *Report on the Manufactures of the United States at the Tenth Census, General Statistics*. Washington, DC: Government Printing Office.
Department of the Interior, Census Office (1883) *Statistics of the Population of the United States at the Tenth Census*. Washington, D.C.: Government Printing Office.
Department of the Interior, Census Office (1895) *Report on Manufacturing Industries in the United States at the Eleventh Census:1890. Part I. Total for States and Industries*. Washington, DC: Government Printing Office.
Department of the Interior, Census Office (1896) *Special Census Report on the Occupation of the Population of the United States at the Eleventh Census: 1890*. Washington, DC.
United States Census Office (1902) *Twelfth Census of the United States, Taken in the Year 1900. Manufactures. Part I. United States by Industries (vol. VII)*. Washington, DC.
United States Census Office (1902) *Twelfth Census of the United States, Taken in the Year 1900. Manufactures. Part II. States and Territories (vol. VIII)*. Washington, DC.
Department of Commerce and Labor, Bureau of the Census (1904) *Occupation at the Twelfth Census*. Washington, D.C.: Government Printing Office.

Department of Commerce, Bureau of Census (1913) *Thirteenth Census of the United States Taken in the Year 1910. Manufactures 1909. General Report and Analysis (vol. VIII)*. Washington, DC: Washington, DC: Government Printing Office.
Department of Commerce, Bureau of Census (1912) *Thirteenth Census of the United States Taken in the Year 1910. Manufactures 1909. Reports by States, with Statistics for Principal Cities (vol. IX)*. Washington, DC: Government Printing Office.
Department of Commerce, Bureau of Census (1923) *Fourteenth Census of the United States Taken in the Year 1920. Manufactures 1919. General Report and Analytical Tables. (vol. VIII)*. Washington, DC: Government Printing Office.
Department of Commerce, Bureau of Census (1923) *Fourteenth Census of the United States Taken in the Year 1920. Manufactures 1919. Reports for States, with Statistics for Principal Cities. (vol. IX)*. Washington, DC: Government Printing Office.
Government Printing Office (1886) *Report on the Statistics of Wages in Manufacturing Industries with Supplementary Reports on the Average Retail Prices of Necessaries of Life and on Trades Societies, and Strikes and Lockouts*. Washington, DC.
Government Printing Office (1892) *Retail Prices and Wages. Report by Mr. Aldrich, from the Committee on Finance, Part 2*. Washington, DC.
Government Printing Office (1912) *Retail Prices 1890 to 1911, Bulletin of the United States Bureau of Labor, no. 105, Part 1*. Washington, DC.
National Resources Committee (1939) *The Structure of the National Economy, Part 1: Basic Characteristics*. Washington, DC: Government Printing Office.
Interstate Commerce Commission, Bureau of Transport Economics and Statistics (1951) *Carload Waybill Analyses, 1949. State-to-State Distribution of Manufactures and Miscellaneous and Forwarder Traffic: Statement No. 5110*. Washington, DC.

Appendix A

A.1 Dependent variable

The Share of Manufacturing Labour Force at the Two-Digit SIC Level in the U.S. State: the data are taken from the U.S. Census of Manufactures 1880–1920. We aggregated them into the two-digit SIC level using Niemi (1974) classification. The censuses provide information on the average number of wage earners, and from 1889 on the average number of employees with a breakdown to wage earners and salaried personnel. We have used the average number of wage earners to make the data comparable over time. 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 and jewellery engraving. To make the data comparable, we have excluded the hand trades in other years as well. The Census of Manufactures reports a special industry category called 'All Other'. This industry category contains <1% 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 that 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 and the regression analysis in the main text is conducted with the exclusion of this industry group.

A.2 Independent variables

A.2.1 Industry characteristics

The Share of White-Collar Workers: this is calculated as the share of salaried personnel in the total persons employed. The data are taken from the U.S. Census of Manufactures 1880–1920. Similarly to the data on the manufacturing employment, we aggregated them up to the two-digit SIC level using Niemi (1974) classification. Salaried personnel include officers, clerks and firm members. There are no data on salaried personnel in 1879 and thus we used 1889 shares. The hand trades are excluded for the same reason as in the case of the dependent variable.

Steam Horse Power per $1000 Gross Output: the data are taken from the U.S. Census of Manufactures 1880–1920 and again we aggregated them into the two-digit SIC level. The steam-horse power data in 1879 are provided only for 22 industries, and therefore we have used 1889 figures. The hand trades are excluded for the same reason as stated above.

Plant size: the figures are taken from O'Brien (1988), Tables 5 and 6. Plant size is calculated as the average number of wage earners per establishment. The hand trades are excluded. O'Brien does not provide plant size in SIC 30, Rubber and Plastic Products, in 1879, and therefore we calculated it from the U.S. Census of Manufactures 1879 using the same set of industries belonging to that SIC as used by O'Brien for other years (the industries include belting and hose rubber, and boots and shoe rubber).

Agricultural input use, intermediate input use, sales to industry and mineral resources use: the figures are calculated from Whitney's (1968) input–output table for 1899, and they are expressed relative to the gross value of output. Whitney's input–output table provides a breakdown of the whole economy into 29 sectors, including agriculture, industries and services. We had to aggregate some of the industries to match the two-digit SIC level. In particular, processed food and grain mill products were aggregated into SIC 20, food and kindred products; petroleum products, and coal products into SIC 29, petroleum and coal products; shipbuilding, transportation, and transport equipment into SIC 37, transport equipment. Whitney's input–output table does not allow calculation of the figures for SIC 20, Tobacco and Tobacco Products; SIC 25, Furniture and Fixtures; SIC 34, Fabricated Metal Products and SIC 38, Instruments and Related Products. Therefore, we have used Leontief's 1919 input–output table for SIC 20, 25 and 34 and Thomas's (1984) input–output table for Great Britain in 1907 for SIC 38. Using the figure from the British input–output table does not pose a problem. These products were unlikely to be produced differently in the U.S. and Great Britain since most of these activities did not use mass production technology.

A.2.2 State characteristics

The Share of Population: From Carter et al. (2006), Table Cc125–137, pp. 3-183–3-184.

The Share of Total Manufacturing Labour Force: From Perloff (1960), Table A-6, p. 632.

The Share of Total Agriculture Labour Force: From Perloff (1960), Table A-2, p. 624.

The Share of Total Mining and Quarrying Labour Force: From Perloff (1960), Table A-3, p. 626.

The Share of Skilled Labour Force: The share of the skilled labour force in 1880–1900 is calculated from the U.S. Population Statistics and the U.S. Occupational Statistics. Skilled labour is considered to be the labour force in professional occupations. The data for 1910 and 1920 are from Goldin (1998) (we have used Goldin's 1928 figures since no data for 1920 exist).

The Share of Farm Land: Calculated from Carter et al. (2006), Table Da159-224, pp. 4-50–4-53, Table Cf8-64, pp. 3-346–3-348.

Market Potential: The methodology and some of the sources are outlined in detail in the text. Here, we provide details of the calculation of the foreign market potential. The nominal GDPs and the exchange rates between the foreign currencies and the $U.S. in 1880–1910 are taken from Flandreau and Zumer (2004) except for Canada, Mexico and the $U.S./GBP exchange rate, which is from Officer (2008). The foreign countries include Argentina, Austria–Hungary, Belgium, Brazil, Canada, Denmark, France, Germany, Greece, Italy, Mexico, Netherlands, Norway, Portugal, Russia, Spain, Sweden, Switzerland and Great Britain. The nominal GDP of Mexico and the exchange rate between pesos and $U.S. come from Instituto Nacional de Estadistica (1990). The Canadian nominal GDP is divided into provinces and the figures come from Green (1971), Table B-1, B-2 and B-3. Green provides data for 1890, 1910 and 1929, respectively. 1900 and 1920 figures had to be calculated using the shares of the provinces' GDP on the total Canadian GDP. Specifically, we have taken the average of 1890 and 1910 shares to obtain 1900 shares and the average of 1910 and 1929 to obtain 1920 shares. Then, we used the total Canadian GDP (Mitchell, 2003b, Table J1) in 1900 and 1920, respectively, to calculate the GDP of provinces in those years. To simplify the calculations, we have considered Prince Edward Island, Nova Scotia and New Brunswick as one province as well as Alberta, Manitoba and Saskatchewan. 1880 values were extrapolated using the Canadian nominal GDP growth rate 1880–1890 calculated from Mitchell (2003b), Table J1. The nominal GDP in 1920 are from Mitchell (2003a, 2003b), Table J1 and the foreign countries include Brazil, Canada, Cuba, Denmark, France, Germany, Italy, Netherlands, Norway, Spain, Sweden and Great Britain. Data on Mexico are for 1921 and are taken from Instituto Nacional de Estadistica (1990). The exchange rates between the $U.S. and foreign currencies are calculated from Carter et al. (2006), Table Ee621–636, pp. 5-567–5-572 and Table Ee637–645, p. 5–572.

Coal prices: there are no satisfactory data on the wholesale prices of coal for every U.S. state in 1880–1920 and thus we have to rely on the retail prices. The prices in 1880 are taken from the 'Report on the Statistics of Wages in Manufacturing Industries with Supplementary Reports on the Average Retail Prices of Necessaries of Life and on Trades Societies, and Strikes and Lockouts' (1886); the prices in 1890 are from 'Retail Prices and Wages. Report by Mr. Aldrich, from the Committee on Finance, Part 2' (1892); the prices in 1910 are from 'Retail Prices 1890 to 1911, Bulletin of the United States Bureau of Labor, No. 105, part 1' (1912). The data for Washington, Arizona, Oklahoma and Wyoming are missing and were proxied them by the coal prices from the nearby states, in particular, by Oregon, New Mexico, Texas and Montana, respectively. The coal prices in 1900 and 1920 were obtained by using the index from the U.S. Millennial Statistics (2006), Table Cc125–137, pp. 3-183–3-184.

Average monthly mean temperature: The temperature data are the 5-year mid-point averages of the monthly mean temperature for every U.S. state taken from U.S.

Historical Climatology Network, file 9641C_YYYYMM_F52.avg.gz. The data for 1880 were proxied with 1890 data.

B.1 Appendix B

Since the overall counterfactual change of market potential comes from three market potential interactions—forward linkages, backward linkages, and plant size—we can split the effect of that counterfactual change among those interactions to see their relative contribution to the resulting change in the share of a state's manufacturing employment on the U.S. total manufacturing employment. Table B1 presents the contribution of each of the market potential interactions to the change of the share of manufacturing employment in total U.S. manufacturing employment in the state of New York and California in 1890, respectively.

Let us consider the state of New York again. We see in Column 1 that the share of New York's manufacturing employment in U.S. total manufacturing employment in 1890 is 16.27%. A counterfactual 10% decline of New York's market potential would result in a 35.8% decrease of that share, as we have seen in Table B1, Column 2. This means that the state of New York loses 5.82 percentage points of that share and the resulting share is 10.45%. Columns 2 and 3 present the contribution of the market potential interactions to the percentage change, and to the loss or gain of the manufacturing shares. We see from Column 2 that, for example, forward linkages decrease the manufacturing share by 17.05% out of the total 35.79%. Column 3 translates these percentage changes into the actual shares of manufacturing employment relative to the U.S. manufacturing employment. Following the previous example, Column 3 tells us that out of 5.82 percentage points of manufacturing employment that are lost because of the counterfactual decrease of market potential, nearly 3 percentage points are due to forward linkages.

Table B1. Contribution of forward linkages, backward linkages, and plant size to the change of the share of a U.S. state manufacturing employment in U.S. total manufacturing employment due to the counterfactual change of market potential in 1890

	Contribution of linkages and plant size to change in manufacturing employment			
			Contribution to percentage change	Contribution to the loss or gain of U.S. state manufacturing employment/ U.S. total manufacturing employment as a result of counterfactual
	(1)		(2)	(3)
State of New York (counterfactual decline of NY's market potential by 10%)				
The share of NY's manufacturing employment in U.S. total manufacturing employment in 1890	16.27	Forward linkages	17.05	2.77
Decline of share of NY's manufacturing employment in U.S. total manufacturing employment as a result of counterfactual (percentage change)	35.79	Backward linkages	6.47	1.05
The share of NY's manufacturing employment in U.S. total manufacturing employment that is lost as a result of counterfactual	5.82	Plant size	12.26	1.99
The share of NY's manufacturing employment in U.S. total manufacturing employment as a result of counterfactual	10.45	Total	35.79	5.82
California (counterfactual increase of California's market potential by 10%)				
The share of California's manufacturing employment in U.S. total manufacturing employment in 1890	2.27	Forward linkages	6.41	0.14
Increase of share of California's manufacturing employment in U.S. total manufacturing employment as a result of counterfactual (percentage change)	13.45	Backward linkages	2.43	0.05
The share of California's manufacturing employment in U.S. total manufacturing employment that is gained as a result of counterfactual	0.30	Plant size	4.61	0.1
The share of California's manufacturing employment in U.S. total manufacturing employment as a result of counterfactual	2.57	Total	13.45	0.30

Sources and *notes*: Column 1: the percentage change of a state's share of manufacturing employment in U.S. total manufacturing employment is from Table 14, Columns 1 and 2; the sources of the manufacturing employment shares in 1890 are in Appendix A.1. Column 2: the figures are calculated using the same procedure as in Table 14. An example for the state of New York shows the calculation of the remaining figures: Column 1: 10.45 = 16.27–5.82; Column 3: 5.82 = 16.27*(0.3579); forward linkages: 2.77=5.82*(17.05/35.79).