Making Sense of the Manufacturing Belt: Determinants of U.S. Industrial Location, 1880-1920

Alexander Klein and Nicholas Crafts

Department of Economics, University of Warwick, Coventry CV4 7AL, United Kingdom

a.klein.1@warwick.ac.uk
n.crafts@warwick.ac.uk

July 7, 2009

Abstract

This paper investigates the ability of the new economic geography to explain the persistence of the manufacturing belt in the United States 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. Natural advantage played a role in industrial location but only through agricultural inputs which were important for a small subset of manufacturing.

Keywords: factor endowments; linkage effects; manufacturing belt; market potential; new economic geography

JEL Classification: N61; N91; R12

Acknowledgments: We are grateful to Gilles Duranton, Tim Leunig, Dennis Novy, Henry Overman, Nikolaus Wolf, and the seminar participants at Stanford, UC Davis, UC Irvine, Virginia, Warwick, and Yale and for their comments.

1. Introduction

The term 'manufacturing belt' has long been used to describe the remarkable spatial concentration of industry in the United States 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.

Krugman (1991) saw the persistence of the manufacturing belt as a classic demonstration of the forces affecting location decisions that are stressed by the New Economic Geography (NEG). In this view, the key characteristic of the manufacturing belt is the market access that it offers to firms rather than an enduring advantage in natural resources. The simplest version of this story is that, when economies of scale became sufficiently large relative to transport costs, firms chose to locate near to demand which in turn locked in greater market access.

A more powerful version of the argument focuses on 'backward' and 'forward' linkages between firms based on the use of manufacturing goods as intermediates in manufacturing production. As in Krugman and Venables (1995), this can generate a process of cumulative causation which creates an industrialized core

[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.

together with a de-industrialized periphery. A large existing market for intermediates makes a region a more attractive place to produce such goods. In turn, better access to these intermediates makes production of final goods cheaper. Once transport costs become low enough to permit realization of the advantages of increasing returns in trade with other regions, a manufacturing-belt phenomenon will be observed. According to Krugman, this occurred with the coming of the railroad and had emerged by the 1860s. Thereafter a path-dependent process was observed in which the initial advantage of the manufacturing belt was locked in by the productivity advantages of proximity which gave rise to external economies. Eventually, when transport costs fell further and become very low, the advantages of proximity to market demand and suppliers would evaporate and manufacturing production would disperse but this did not happen until the later decades of the 20th century.

A similar view of the persistence of the manufacturing belt can certainly be found in the economic-history literature. An account which stresses the importance of market access and which notes the reasons why the Midwest and the North East were both included but the South was not is provided by Meyer (1983) (1989). However, nowhere is there an explicit empirical verification of the NEG hypotheses as to the underpinnings of the manufacturing belt.

Indeed, the most recent quantitative analyses of long-run trends in U.S. industrial location by Kim (1995) (1999) stress the role of natural advantage rather than market access and tend to dismiss the NEG account. Kim (1999) estimated a model of industrial production across states based on the Rybczynski theorem and found that factor endowments were the fundamental explanation for the geographic distribution of U.S. manufacturing from 1880 through 1987. He argued that once factor endowments had been taken into account, there was little left to be explained by NEG forces. Kim (1995) found that U.S. regional specialization in the late 19th and early 20th centuries could be explained by the rise of large-scale production methods that were intensive in the use of raw materials and energy sources that were relatively immobile leading regions to become more specialized, and he argued against the importance of Krugman-type explanations for industrial location.

We have a number of doubts about the strength of Kim's evidence in favor of the natural-advantage hypothesis which we discuss in more detail in a later section of the paper. Here, suffice it to say that neither of his papers tests directly for the role of market access in industrial-location decisions. In this paper, we carry out such a test 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).

We seek to investigate the part that market potential played and to establish whether NEG arguments can explain the existence of the manufacturing belt around the turn of the 20th century. Our framework allows an explicit analysis of the roles of each of scale economies, backward linkages and forward linkages. In essence, we look to see what locked in the manufacturing belt. We do not try to explain the emergence of the manufacturing belt but rather to understand the nature of the path-dependence which sustained it.

In particular, we address the following questions relating to U.S. manufacturing at the 2-digit level.

1) Did market potential matter for the location of manufacturing?

2) Was market potential more important than factor endowments as a determinant of industrial location?

3) Is there evidence to support the hypothesis that market potential influenced the location of manufacturing through linkage effects as well as scale effects?

Answering these questions is essential to establishing whether new economic geography rather than natural advantages can provide a plausible explanation for the longevity of the U.S. manufacturing belt, an issue on which the jury is still out. To answer these questions, we construct a unique data set on 48 U.S. states and 19 two-digit level industries for each of the census years during 1880-1920. This includes, among other data, industry characteristics such as the share of white-collar workers, the use of agricultural products, the use of manufacturing goods as intermediates, sales to other industries, and state characteristics such as two-digit level industrial employment shares and market potential.

2. A model and an empirical framework

A Basic Model[2]

The core of the Krugman and Venables (1995) model rests on the idea that intermediate goods play an important role in a firm's location decision. The firm operates in the standard Dixit-Stiglitz monopolistic competition market and produces differentiated products under increasing returns to scale. The firm's cost function includes labor and the composite manufacturing intermediate good (aggregated by a CES production function) combined with a Cobb-Douglas technology. Economies of

[2] We present only a brief outline. The full model can be found in Fujita, Krugman, and Venables (1999), Baldwin, Forslid, Martin, Ottaviano, and Robert-Nicoud (2003), Combes, Mayer and Thisse (2008).

scale are reflected in the cost of labor that is needed to produce a differentiated product. Specifically, this comprises both a fixed requirement and a marginal requirement. Consumers have preferences over a homogeneous product and the composite differentiated product (aggregated by a CES utility function) which are entered in a Cobb-Douglas utility function. The monopolistic competition feature of the market ensures that no variety is produced by more than one firm and that, in equilibrium, profits are zero due to free entry and exit. Demand for manufacturing goods comes from firms which purchase the intermediate products and from consumers who buy the final products.

A feature of the model, which is a result of the Dixit-Stiglitz modeling strategy, is that the behavior of the firm's competitors is accounted for via the price index of the composite manufactured good. This comes from the fact that the price index is a decreasing function of the number of varieties and since each variety is produced by only one firm, the price index is then a decreasing function of the number of firms. Accordingly, a low (high) price index implies a more (less) competitive product market because there are more (less) firms on the market.

The firm's use of the intermediate products creates input-output linkages between upstream firms, producing intermediate goods, and downstream firms, producing for final consumption. Both backward and forward linkages act as forces for industrial agglomeration. Forward linkages arise as the downstream firms are drawn to the region with good access to the market for intermediate products. Backward linkages arise as the upstream firms are attracted to the region with high demand for their products. Against the agglomeration forces of forward and backward linkages are the dispersion forces of product and labor market competition. More firms in the region imply a lower price index which tends to make the region less profitable, encouraging exit from the region and leading to the geographical dispersion of industry. More firms in the region also imply a higher demand for labor which pushes the money wages up, and makes production more costly, hence making the region less attractive. The balance between those forces depends on the characteristics of industries and transport costs between regions. Transport costs play a crucial role as a potential trigger of agglomeration. The model does not presuppose any exogenous asymmetry between regions since it intends to uncover the reasons for the spatial inequalities on the basis of the market interactions alone.

To see the interplay between industrial characteristics and transport costs, suppose that one region for some reason (for example, natural-resources advantage, proximity to a natural mode of transportation) has a larger manufacturing sector than other regions. The region with a larger manufacturing sector offers access to a large number of manufactured products which lowers the prices that firms and consumers have to pay for them in that region. This lowers the costs of production and consumption, and thus, *ceteris paribus*, attracts firms producing the final products and consumers (forward linkages). This region is also, *ceteris paribus*, a more attractive location for production of intermediate products because of its higher demand for these products (backward linkages), which then fosters the relocation of the firms producing intermediate goods. On the other hand, a larger manufacturing sector pushes down the price index of the manufactured products, which makes the region less profitable and hence less attractive. Furthermore, a larger number of firms increases the local demand for labor, which pushes money wages up. This raises costs of production which also makes the region less attractive.

Suppose that transport costs are very high. This implies that if firms relocate to the region with the larger manufacturing sector, they gain the advantage of forward and backward linkages, but at the same time they face upward pressure on the money wages and downward pressure on the price index which would tend to lower profitability. Since trade is very costly, firms can not serve other regions to compensate for this and thus prefer to stay in their original regions. Now, suppose that transport costs fall. Lower transport costs imply that, when firms relocate to the region with the larger manufacturing sector, this time they will be able to sell their products to other regions as well, increase profits, and thus counterbalance the lower price index and the higher money wages of the region with the larger manufacturing sector.

There is a critical level below which transportation costs have to fall so that firms find it profitable to relocate to the region with larger manufacturing sector when the concentration forces outweigh the dispersion forces. If, however, transport costs continue to fall, the advantage of being close to markets and suppliers gradually evaporates in favor of regions with a smaller manufacturing sector and thus less competitive product and labor markets.[3] This means that there is a critical intermediate level of transportation costs at which some regions will form an industrial core while other regions become a de-industrialized periphery. Of course, in a more detailed model, the relative strength of concentration and dispersion forces would differ across industries and it could well be that falling transport costs induces concentration of some industries and dispersion of others.

[3] The model assumes no interregional, only intersectoral migration. Puga (1999) shows that perfect labor mobility between regions also leads to agglomeration and that the mechanism of concentration and dispersion forces is the same. He also shows that having land in the NEG model does not change its basic qualitative features.

As was mentioned earlier, the Krugman and Venables (1995) model, as is typical in NEG models, assumes no exogenous asymmetry between the regions and thus abstracts from ‘first-nature geography’ such as natural resources. As a result, the distribution of economic activities is not determined uniquely by the model but exhibits multiple equilibria. Which region actually becomes core and which periphery depends on initial conditions such as factor endowments, proximity to transportation modes, or the size of the region. However, once the asymmetry of the regions is introduced, the number of equilibria reduces and the model provides a mechanism that locks in the initial asymmetric distribution of economic activities.[4] In the case of the manufacturing belt, Krugman and Venables (1995) model offers an explanation of what locked in already highly industrialized regions of New England, Middle Atlantic and Midwest into the industrialized core which persisted way until the second half of the twentieth century.

An Empirical Framework

We wish to test the hypothesis that the new economic geography mechanisms rather than natural advantages explain the lock in of the manufacturing belt region around the turn of the twentieth century. We use the methodology of Midelfart-Knarvik, Overman and Venables (2000).[5] The authors developed and econometrically estimated a model of the location of industries across countries, which combines factor endowments with geographical considerations based on the Krugman and Venables (1995) model. Their approach is a synthesis and generalization of two

---

[4] See Baldwin et al. (2003) for an extensive analysis of new economic geography models with exogenously asymmetric regions.

[5] The discussion is based on Midelfart-Knarvik et al. (2000).

existing approaches in the empirical literature: a literature which estimates the effect of industry characteristics on trade, and a literature which estimates the effect of country characteristics on trade and production. It is similar to Ellison and Glaeser (1999) but differs in the sense that the theoretical specification is derived from trade rather than location theory. The core idea of the Midelfart-Knarvik et al. (MK) model is the following. States differ in their factor endowments and face transportation costs on their trade. Industries in those states use both primary products and intermediate goods to produce differentiated goods. In equilibrium, the location of industries is determined both by factor endowments and by geography. Factor endowments matter for obvious reasons. Transport costs mean that the location of demand matters: states at different locations have different market potential which shapes their industrial structure. Intermediate demand and prices vary across locations, which means that forward and backward linkages are present and that industries may find it optimal to locate close to supplier and customer industries. The model generates a regression equation which contains interaction variables between the characteristics of states and the characteristics of industries to determine the industrial structure of states. [6] This

[6] An alternative approach was developed by Davis and Weinstein (Davis and Weinstein 1999, 2003). It uses the home-market effect to empirically separate NEG models from the models of comparative advantage. Their argument is that in a world of comparative advantage, a strong demand for a good will make that good, *ceteris paribus*, an import. However, in an NEG world, a location with a strong demand for a good makes it a preferable place to locate production and thus the location becomes the exporter of that good. This 'home market effect' of demand on trade distinguishes NEG from comparative advantage models. In the empirical analysis, the home-market effect is then captured by a variable which measures the association between changes in demand and changes in output. If an increase in demand leads to more than proportional increase in output, then the mechanism of NEG is confirmed. Otherwise, other theories are more relevant. We use the MK approach because it is richer than the Davis and Weinstein approach. In particular, the MK methodology enables us to estimate the

empirical strategy was used to examine the location of production in the European Union (Midelfart-Knarvik et al. 2000; Midelfart-Knarvik and Overman, 2002), and the studies confirmed the importance of NEG forces in shaping the location of industries in the EU.

The MK approach provides a simple, yet theoretically sound empirical test. However, it has limitations which should be recognized. To make a tractable link between theory and econometrics, the model omits some complexities of NEG models. In particular, NEG models imply a non-monotonic relationship between location and transportation costs, which, as we know, creates a multiplicity of equilibria. As a consequence, there is no unique mapping from characteristics of states and industries to industrial location. The MK approach leaves the estimation of multiple equilibria for future research.

Before we move to the baseline econometric specification, another methodological approach needs to be mentioned. Kim (1995) tries to determine the geographical distribution of industries in the U.S. by estimating the effect of scale economies and the factor endowments on an index of localization of production. His methodology uses plant size to capture economies of scale and the raw material intensity of industries to reflect factor endowments. The methodology used in our paper has a major advantage over Kim's approach not only because it is richer, but also because it uses the characteristics of states along with the characteristics of industries.

Formally, the basic model can be written as follows:

---

effect of market potential and distinguish between forward and backward linkages. Moreover, it makes possible estimation of the effect of various factor endowments on geographical location, similar to Ellison and Glaeser (1999).

$$\ln\left(s_{i,t}^{k}\right)=\alpha\ln\left(pop_{i,t}\right)+\phi\ln\left(man_{i,t}\right)+\sum_{j}\beta^{j}\left(y^{j}{}_{i,t}-\gamma^{j}\right)\left(x_{t}^{j,k}-\chi^{j}\right)+\varepsilon_{i,t}^{k} \quad (1)$$

where $s_{i,t}^{k}$ is the share of industry *k* in state *i* and time *t*, $pop_{i,t}$ is the share of population in state *i* and time *t*, and $man_{i,t}$ is the share of manufacturing employment in state *i* and time *t*; $y^{j}_{i,t}$ is the level of *j*th 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*, 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 α, φ, β^{j}, γ^{j}, and χ^{j} are coefficients to be estimated.

To understand this specification, consider one particular characteristic, say *j* = skilled labor.[7] So x[skilled labor]$_{t}^{k}$ is white-collar worker intensity of industry *k* at time *t*, and y[skilled labor]$_{i,t}$ 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-labor intensity χ[skilled labor] such that its location is independent of state skilled-labor abundance. Second, there exists a level of skilled-labor abundance γ[skilled labor] such that the state's share of any industry is independent of the skilled-labor intensity of the industry. Third, if β[skilled labor] > 0, then industries with skilled labor intensities greater than χ[skilled labor] will be induced to locate in states with skilled-labor abundance greater than γ[skilled labor]. Estimation of the model will produce the following key parameters for each interaction variable: β[j], γ[j], and χ[j] with *j* running over the interactions. If, for example, skilled labor is an important determinant of location patterns, we should see a high value of β[skilled labor].

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

[7] The discussion follows Crafts and Mulatu (2006).

$$\ln\left(s_{i,t}^{k}\right)=c+\alpha\ln\left(pop_{i,t}\right)+\phi\ln\left(man_{i,t}\right)+\sum_{j}(\beta^{j}y^{j}{}_{i,t}\,x^{j,k}{}_{t}-\beta^{j}\gamma^{j}\,x^{j,k}{}_{t}-$$

$$-\beta^{j}\chi^{j}\,y^{j,k}{}_{t})+\varepsilon_{i,t}^{k} \qquad (2)$$

This gives a list of independent variables that comprises scaling terms, state characteristics, industrial characteristics, and interactions between state and industrial characteristics. The coefficients of the two size variables, α and ϕ, are straightforward, and *c* is a constant term. 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, y^jx^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 x white collar workers* provides us with a measure of how important this factor endowment was in influencing the location of industries in the United States.

3. Implementation of the MK empirical framework and data set

In this section, we describe the data used in the paper (a detailed description of the variables is in the appendix) and the implementation of the MK model.

Regression Equation

In the implementation of the model, we consider four state characteristics (in addition to the population and the manufacturing labor force), six industry characteristics, and six interactions. The estimated equation (2) can be expressed as follows:

$$
\begin{aligned}
\ln(s_{i,t}^{k}) &= \mathrm{C} + \alpha\ln(POP_{i,t}) + \phi\ln(MAN_{i,t}) + \\
&+ \beta_1\text{AGRIC EMPL}_{i,t} + \beta_2\text{EDUC POP}_{i,t} + \beta_3\text{COAL ABUNDANCE}_{i,t} + \\
&+ \beta_4\text{MARKET POTENTIAL}_{i,t} + \beta_5\text{WHITE COLLAR WORKERS}_{i,t} + \\
&+ \beta_6\text{STEAM POWER USE}_{i,t} + \beta_7\text{AGRICULTURE INPUT}_{i,t} + \\
&+ \beta_8\text{INTERMEDIATE INPUT USE}_{i,t} + \beta_9\text{SALES TO INDUSTRY}_{i,t} + \\
&+ \beta_{10}\text{SIZE OF ESTABLISHMENT}_{i,t} + \\
&+ \beta_{11}(\text{AGRIC EMPL x AGRICULTURE INPUT USE})_{i,t} + \\
&+ \beta_{12}(\text{EDUC POP x WHITE COLLAR WORKERS})_{i,t} + \\
&+ \beta_{13}(\text{COAL ABUNDANCE x STEAM POWER USE})_{i,t} + \\
&+ \beta_{14}(\text{MARKET POTENTIAL x INTERMEDIATE INPUT USE})_{i,t} + \\
&+ \beta_{15}(\text{MARKET POTENTIAL x SALES TO INDUSTRY})_{i,t} + \\
&+ \beta_{16}(\text{MARKET POTENTIAL x SIZE OF ESTABLISHMENT})_{i,t} + \varepsilon_{i,t}^{k} \qquad (3)
\end{aligned}
$$

The state characteristics are captured by the share of agricultural employment, share of educated population, coal prices, and market potential; industries are characterized by the share of white-collar workers, steam power use, plant size, agricultural input use, intermediate input use, and sales to industry.

The interaction variables are the following: educated population availability and white-collar worker intensity, coal abundance and steam power use, share of agricultural employment and agricultural input use, market potential and intermediate input use, market potential and sales to industry, and market potential and plant size. The first three of these interactions are predicted by the Heckscher-Ohlin (H-O) theory based on factor endowments; the last three are predicted by NEG to be activated when transport costs are in the right "intermediate" range such that the pull of centrality kicks in. 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 final consumer. The second market-potential interaction is based on backward linkages and presumes that industries which sell relatively large fraction of their output to other firms rather than 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).

In the original work by Midelfart-Knarvik et al., the authors estimate their version of the equation (2) using OLS, and account for the heteroskedasticity and the country and industry fixed effects. We also address additional estimation issues including endogeneity and clustered-sample methods.

Data Set

We created a unique data set of the employment shares for 48 U.S. states and 19 two-digit level industries, six industry and four state and characteristics including market potential for each census year during 1880-1920.[8] The data on the share 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 the Historical Statistics of the United States (2006). The data on labor force and agricultural employment in each U.S. state are from Perloff (1960), 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

[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.

[9] We would like to thank Claudia Goldin for providing the data.

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. There are two such tables available for our time-period. One is due to Whitney (1968) who constructed an input-output table for 1899; the other is Leontieff's (1941) seminal work which provides an input-output table for 1919. Both tables can be adjusted to the two-digit industrial level.

Panel A in Table 1 reports industrial characteristics obtained from the 1899 input-output table which relate to key aspects highlighted by locational hypotheses based either on new economic geography (cols. 1 and 2) or on natural advantages (Cols. 3 and 4). 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 does not. Overall, it is noticeable that many sectors have substantial linkages (medians in cols.1 and 2 are both 26 per cent) whereas few sectors rely heavily on inputs of primary products (medians in cols. 3 and 4 are 0.4 per cent and 1.3 per cent, respectively). Panel B in Table 1 shows the distribution of two-digit manufacturing employment between the manufacturing belt states and the states outside the belt. We see that 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 which 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. Panel B 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 which 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, 2002) with the implication that a gravity equation framework should be used to estimate market potential. We use the methodology of Head and Mayer (2004) which is based on Redding and Venables (2004). The market potential of a U.S. state *i* is calculated using the formula $M_i = \sum j\ \varphi_{ij}\ GDPj$ where φ_{ij} is the accessibility of market *j* for goods from the U.S. state *i*. The market *j* consists of nominal GDP in foreign countries, in other U.S. states, and in the home state *i*.

The market accessibility of foreign countries is calculated as follows:

$$\varphi_{ij} = \exp(-\beta_j + \lambda L_{ij}) d_{ij}^{-\delta} \qquad (4),$$

where β is the home-bias coefficient, λ is the language coefficients, δ is the distance coefficients, L_{ij} is an indicator variable with value of 1 if the language between *i* and *j* is the same and 0 otherwise, and d_{ij} is the distance between *i* and *j*. The market accessibility of other U.S. states is calculated as

$$\varphi_{ij} = d_{ij}^{-\delta} \qquad (5),$$

and own U.S. state

$$\varphi_{ii} = d_{ii}^{-\delta} = [2/3 \,.\, (\text{area}_i/\pi)^{0.5}]^{-\delta} \quad (6).$$

In (5) and (6), the subscripts *i* and *j* refer to the U.S. states. The regression equation to estimate β, λ, and δ is

$$\ln X_{ij} = EX_i + IM_j + \delta \ln d_{ij} + \beta_j B_{ij} + \lambda L_{ij} + \varepsilon_{ij} \quad (7),$$

where X_{ij} is the aggregate value of the country's *i* export to country *j*, EX_i and IM_j are exporter and importer fixed effects, B_{ij} is a dummy variable which is 1 if $i \neq j$, β is the home-bias coefficient, λ is the language coefficient, δ is the distance coefficient, L_{ij} is an indicator variable with a value of 1 if the language between *i* and *j* is the same and 0 otherwise, and d_{ij} is the distance between *i* and *j*.

The main issue here is data availability. The estimation of equation (7) to obtain the estimates of β, λ, δ for equation (4) – the accessibility of foreign countries – uses data for trade between the U.S. and foreign countries from Jacks et al. (2008) with total internal trade flows for the U.S., X_{ii}, constructed by subtracting total U.S. exports from U.S. GDP based on Maddison (2007).[10] Unfortunately, no internal trade data for the U.S. states in 1880-1920 exist to estimate equation (7) in order to obtain δ for equations (5) and (6). Here we have two options: either to use -1 for δ as suggested by Harris (1954), which is often used in the literature when the gravity equation similar to (7) is impossible to estimate, or to follow the suggestion of Head and Mayer (2004) and to use the estimate of δ from (7) in (5) and (6). We implement both approaches to see how sensitive the results are to different market potential estimates. The area of U.S. states is taken from the Historical Statistics of the United States (2006), the distance between the U.S. states and the foreign countries is the kilometer distance between the corresponding capitals, and the distance between the

[10] We are grateful to Chris Meissner for providing the trade data.

U.S. states is calculated as the kilometer distance between their capital cities. Our estimates are for the railroad era and we believe that by this time physical distances are a reasonable approximation to economic distances inside the United States. Our choice of -1 for δ is consistent with estimates for modern internal U.S. trade (Wolf, 2000; Hillberry and Hummels, 2003).

To complete the calculation of market potential, nominal GDP estimates are required. Nominal GDP of the 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 GDP of the foreign countries and the corresponding exchange rate are in the Appendix.

Table 2 displays our estimates of market potential by state for 1880 and 1920 based on the $\delta = -1$ variant. Two points stand out. First, the rank order of market

[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 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.

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, Rhode Island and Washington have very similar GDP but, as Table 2 shows, market potential of the former was about 5 times that of the latter.

In addition to the data for 1880-1920, we need market potential estimates for each U.S. state in 1870 to construct the instruments used in the instrumental variable regression presented in the following section. The estimation follows the methodology outlined above and the nominal GDP of each U.S. state in 1870 is calculated from the real GDP estimates of Turner et al. (2006).[12]

4. Empirical Results

Estimation Issues

In our initial estimations of equation (3) market potential is calculated assuming $\delta = -1$, and forward and backward linkages are based on the 1899 input-output table in Whitney (1968); then other variants are presented by way of sensitivity analysis. This section discusses the statistical properties 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 U.S. state suggests that

[12] We are grateful to Robert Tamura for providing us with the dataset to calculate these estimates.

we might face an unobserved cluster effect coming from the U.S. states. In this case, cluster-robust standard errors should be used (White, 1984, Arellano, 1987); failure to do so 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.

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 (Cameron et al., 2005, Wooldridge, 2003, 2006). Second, market potential and the corresponding interactions may be endogenous. This calls for instrumental variable estimation. In our setting, we have to rely on lagged variables since finding an alternative is very difficult. Econometrics research in recent years 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, Cameron et al., 2005). This is exacerbated when the correlation between instruments and instrumented variables is weak, leaving us with IV estimation of low precision (Staiger et al. 1997; Kleibergen, 2002; Hahn et al., 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, p. 104) who says: "Often we must choose between a possibly inconsistent estimator that has relatively small standard errors

(OLS) and a consistent estimator that is so imprecise that nothing interesting can be concluded (2SLS). One approach is to use OLS unless we can reject exogeneity of the explanatory variables." Therefore, we perform endogeneity tests on the suspect regressors.

Coming back 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).

The Basic Results

The results of the initial estimation of equation (3) are in Tables 3-5. Table 3 presents the results for the pooled sample 1880-1920. Column I shows the results of estimation with cluster-robust standard errors since the data are clustered at the state level and heteroskedasticity is present, as confirmed by the Breusch-Pagan test (which rejects the hypothesis of homoskedastic standard errors at the 1% significance level). The estimation results show that out of three H-O interaction variables, only agriculture is statistically significant (at 1%), and has a correct sign; the other two are insignificant. As for the NEG interactions, two of them are highly statistically significant and with the correct sign – backward linkages and plant size – while the forward linkages interaction variable is insignificant, though with the correct sign.

The time dimension potentially allows us to use panel-data estimation. Because of heteroskedasticity, a robust Hausman test (Cameron et al., 2005, p. 718) was used to test between fixed- and random-effects models and the test statistics (see Table 3) favor the fixed-effects model. Column II presents the results of the fixed-effects estimation with panel-robust standard errors. The results confirm the previous findings and provide support for the pooled OLS estimates.

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. Indeed, the forward linkages in Table 3 are not statistically significant despite the fact that many industries have substantial linkages, as discussed in the previous section. Therefore, we carried out a Chow test to determine whether the data should be pooled or not. The calculated F-statistics 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. Accordingly, we run separate regressions for 1880, 1890, 1900, 1910, and 1920.

For each of those years, we have estimated equation (3) with OLS using cluster-robust standard errors and cluster-specific fixed effects. The reason for using cluster-robust standard errors is, as with the earlier regressions, the possibility that there is an unobserved cluster effect which needs to be taken into account. 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.[13] The results are presented in Table 4.

---

[13] 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, 2006). We have also estimated

A general overview of the estimation results suggests that NEG interaction variables are present in each of the years, though some variation exists before 1900. The H-O interactions are less prevalent except for agriculture until 1900. Of the NEG forces, the plant-size interaction is always statistically significant, usually at the 1% significance level. The backward-linkages interaction is almost always significant, except for 1910. Forward linkages are first significant in 1890 (at 10%). After that, they remain significant until 1920 with an increase up to the 1% level. The H-O interactions are very different in terms of significance. Both coal and skilled-labor interactions change signs and are insignificant for most of the time. The agriculture interaction, on the other hand, is highly statistically significant until 1900, after which it becomes insignificant and changes sign in 1920.[14]

The endogeneity issue regarding market potential and its interactions is addressed by instrumental-variable estimation. As was noted earlier, finding an instrument in our setting is difficult and we have to rely on lagged variables. 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, pp 233-234). Instrumental variable estimation is carried out using 2-step GMM, which is more efficient than IV/2SLS. The results are presented in Table 5.[15]

---

the cluster-specific random effect model, and the results remain qualitatively unchanged; they are available from the authors upon request.

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

[15] The share of manufacturing labor force and the share of agricultural labor force are potentially endogenous too. The manufacturing labor force is mobile in the standard NEG model which suggests

For each year, we again estimate equation (3), and we use cluster-robust standard errors. First, we check the correlation between our instruments and instrumented market potential and the corresponding interactions. Shea's partial R^2 in Table 3 show a very strong correlation between the instruments and instrumented variables, ranging from 0.88 to 0.99. We have also carried out a formal test of the weak instrument suggested by Stock and Yogo (2005). The relevant F-statistics largely exceed the critical values reported by Stock and Yogo (2005) in their Tables 1-4 (the F-statistics range from around 500 in 1880 to around 8000 in 1920).

The endogeneity test (Table 5) rejects the null hypothesis that the market potential and its interaction are exogenous in all cases except for 1890.[16] Nevertheless, we estimate the model for that year anyway as a sensitivity check. The results in Table 5 show that overall the picture that emerges from Table 4 is preserved. The NEG interaction variables are almost always significant and have the correct sign, with the plant-size interaction having the strongest significance, followed by the backward-linkages and then the forward-linkages interaction. The forward-linkage interaction is significant from 1890, and the significance rises from 1900. The significance of the estimated backward-linkages coefficients remains high throughout the period, except for 1910.

---

that it might be endogenous in our regression equation. The agricultural labor force is considered immobile in the original Krugman (1991) model but is treated as mobile across sectors in Krugman and Venables (1995) as well as in Puga (1999). Therefore, we have also considered both manufacturing and agricultural labor force as endogenous and instrumented them with their lagged values. The sign and the statistical significance are the same as when they are treated as exogenous.

[16] The endogeneity test in 1920 does not reject the null hypothesis of exogenous market potential interaction at the conventional significance levels; however, the null hypothesis can be rejected at the 11% significance level.

Finally, 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 omitting own GDP. The results that were obtained (available on request) 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.

Robustness, Standardized Coefficients and Counterfactuals

We have also performed additional robustness checks with respect to our data. As was mentioned in the section on data, market potential was calculated twice: first, with $\delta = -1$ and, secondly, with the estimated coefficients of equation (6). The estimations in Table 3-5 use the former market potential data. Therefore, Tables 6 and 7 show the results of the estimation techniques used in Tables 4 and 5 with the market potential figures calculated using the gravity equation estimates.[17] We see that the results are similar, with the agriculture interaction being the only important H-O force and NEG forces being important throughout the whole period.[18]

---

[17] We have also performed a similar sensitivity analysis with the estimation techniques used in Table 3, and the results are qualitatively unchanged. We do not report the results here, they are available from the authors upon request.

[18] In addition to market potential calculated using the gravity equation estimates, we have calculated market potential using both the gravity equation estimates of δ and $\delta = -1$. The former was used to calculate the market access of the foreign countries, the latter for the market access of the home U.S. state as well as other U.S. states. The regression results are qualitatively unchanged and are available from the authors upon request.

The crucial NEG interaction variables are forward and backward linkages estimated using Whitney's 1899 input-output table. However, as was mentioned earlier, 1880-1920 was a period of dramatic changes in the U.S. economy and therefore using the same input-output table might raise the issue of the accuracy of the estimated forward/backward linkages. Unfortunately, for the period before 1899, there are no input-output tables. For 1920, we can perform a sensitivity analysis using the well-known input-output table for 1919 constructed by Leontief (1941). Table 8 presents the results of OLS estimation with cluster-robust standard errors, cluster-specific fixed effect estimations, and 2-step GMM for that year; estimation was done with both variants of the market potential figures. We see that when we use the market potential figures calculated using the gravity regression estimates, the forward linkages lose statistical significance in the OLS cases, though the sign remains positive; however, the cluster-specific estimates of that linkage are significant, as in Table 4. The qualitative results for all the remaining variables are unchanged.

In unreported results, we have also performed two additional robustness checks. First, we have checked the robustness of the H-O interaction variables. Specifically, we have used the share of farm land (similarly to Ellison and Glaeser, 1999) instead of the share of agricultural labor force in the agricultural-interaction variable, and the share of coal inputs in gross product instead of the ratio of horse power to gross output in the coal-interaction variable.[19] In both cases, the qualitative results are similar to the results in Tables 3-8, with agriculture being the most prevalent among all H-O interaction variables. Second, we have re-estimated all the regressions in Tables 3-8, as well as the unreported regressions, with industry-specific

[19] The share of farm land is calculated from the Historical Statistics of the United States (2006); the share of coal in gross product comes from Whitney (1968) and Leontief (1941).

dummy variables to control for unobserved industry-specific effects and the results are virtually unchanged.

Overall, these results show the importance of all the NEG and some of the H-O forces, consistently throughout the whole period 1880-1920 irrespective of the estimation technique. This suggests that industrial location was indeed substantially driven by the agglomeration mechanisms related to market potential. We can support this inference by calculating beta coefficients for the relative importance of the interaction variables in determining state shares of manufacturing employment by industry. The results reported in Table 9 show that throughout 1880 to 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 NEG 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.

It is also possible to illustrate the importance of market potential for industrial location by making some counterfactual calculations based on the estimated coefficients in our preferred specification (Table 4, Equation FE). We consider cases of states in 1900 with similar GDP and GDP per person but different market potential. Pairs of states with these characteristics include California and New Jersey, Nebraska and Maryland, Utah and Delaware, and Washington and Rhode Island. Moving the peripheral state to the location of their manufacturing-belt counterpart would raise the predicted shares of the state in overall manufacturing employment by 30%, 22%, 48%, and 56%, respectively.

5. Discussion of the Results

In the preceding section we have argued that econometric analysis provides robust support for an interpretation of the manufacturing belt around the turn of the 20th century that relies quite heavily on new economic geography. In this section we further investigate the plausibility of this claim.

If the trigger for agglomeration is that economies of scale rise relative to transport costs, then this is surely the generally-accepted story of American manufacturing by the second half of the 19th century. While the costs of transporting goods was perhaps 8 times higher in 1890 than in 2000 (Glaeser and Kohlhase, 2004), it was still much lower than in the early 19th century; average rail freight rates fell from 6.2 cents per ton-mile in 1833 to 0.73 cents per ton-mile in 1900 (Carter et al., 2006, p. 781).

With regard to economies of scale, estimates of the cost-dual of a Leontief production function by Cain and Paterson (1986) showed that scale economies were prevalent at the two-digit industry level between 1850 and 1919 with the exceptions only of Food (SIC 20) and Leather (SIC 31). Atack (1985) found that plant size generally rose considerably between 1870 and 1900 as the potential of new technologies introduced from the mid-19th century was realized and that in industries singled out by Chandler (1977) as the pioneers in mass production and mass distribution (for example, iron and steel, flour milling) the average scale of operation in 1900 was much larger than efficient scale in 1870.

Traditional accounts of the determinants of industrial location in the early 20th century do not stress the role of H-O factors. Instead, they emphasize the importance of manufactured intermediates to manufacturing production as a key factor promoting regional concentration of manufacturing and tend at the same time to downplay the role of natural resources (Harris, 1954; Perloff et al., 1960, pp. 394-5). Moreover, there was little correlation between the spatial distribution of employment in coal-mining and in manufacturing; only in the high market-potential state of Pennsylvania did they really coincide. Human capital of the workforce is little discussed by these authors but the work of Goldin and Katz (1998) suggests that it is not surprising that the educated population-white collar workers interaction is insignificant. They convincingly argue that in this 'factory-production' phase of manufacturing, physical capital was a substitute for skill and technological advance was downgrading the role of skilled labor.

Various industry studies also lend a support to our findings with the automobile industry being an excellent example. By the 1920s, the automobile industry was heavily concentrated in southeast Michigan and the Detroit area became the leading car manufacturing region. The literature recognizes that linkages were among the most important reasons for that development (e.g. Klier and Rubenstein (2008), Brinkley (2003), Flink (1990), Nevins (1954)). The automobile was a combination of components that were being produced by manufacturing firms which originally produced closely related products such as gasoline engines, carriage bodies and wheels. At the turn of the twentieth 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 which 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. Once gasoline engines defeated steam and electricity engines as the main source of power, Detroit emerged as industry's leading part supplier. As a result, the car producers found the region very attractive and by the 1920s, Detroit became a leading producer of cars.[20]

The importance of the suppliers is well documented by the Ford Motor Co. When the company opened its business in 1903, ninety five percent of the costs of a car came from the costs of buying parts, with the chassis (engines, transmissions, and axles) supplied by the Detroit machine shop of John F. and Horace E. Dodge, and bodies supplied by C.R. Wilson Carriage Co. (Klier and Rubenstein (2008), pp. 37-38).

While our results are consistent with this traditional literature, they clearly differ from the findings of Kim (1995) (1999) who stressed the importance of natural advantage. Kim's approach is indirect since in neither paper was market potential considered as a variable. Kim (1995) related plant size and raw-materials intensity of industries to an index of localization of production but took no account of either regional characteristics or other industrial characteristics. Kim (1999) analyzes the industrial location in the U.S. between 1880 and 1987 by relating measures of factor endowments to levels of production assuming that the former were exogenous. He uses the goodness of fit of the regressions to measure the effect of increasing returns on the U.S. economic geography and concludes that the factor endowments rather than increasing returns explain a large amount of geographical variation in U.S. manufacturing. A more appropriate specification would have allowed for capital and labor mobility and would have sought to explain shares in U.S. production, i.e., the geographic distribution of activities. Moreover, the high adjusted-R^2 that was obtained

[20] A detailed analysis of the rise of Midwest as the centre of the automobile industry is provided, for example, in Tsai (1999).

for what are in effect crude estimates of a production function has little to say about the scope for new economic geography forces to influence location. The MK estimation framework, developed after Kim (1995) (1999) and used in this study, seems preferable since it provides a direct test of NEG and H-O forces.

6. Conclusion

In this paper we have implemented a version of a model originally developed by Midelfart-Knarvik et al. (2000) to investigate the importance of market access and factor endowments in industrial location decisions in order to discover the reasons for the persistence of the manufacturing belt in the United States at the turn of the 20th century. Our results show that, in each case, the answer to the three questions that we posed in the introduction is 'yes'.

We find that market potential did matter for the location of manufacturing in the United States throughout the period 1880 to 1920, that it was more important than factor endowments, and that the influence of market potential worked both through linkage effects and scale effects, more so in later years. We believe that this is the first empirical validation of the claims made by Krugman (1991) that the manufacturing belt was a classic demonstration of the explanatory power of the new economic geography. Our results suggest that market access was the central consideration that locked in the manufacturing belt and accounts for the path dependence in the location of American manufacturing in the late 19th and early 20th centuries.

References

Arellano, M., "Computing Robust Standard Errors for Within-Group Estimators," *Oxford Bulletin of Economics and Statistics* 49 (1987), 431-434.

Atack, J., "Industrial Structure and the Emergence of the Modern Industrial Corporation", *Explorations in Economic History* 22 (1985), 29-52.

Baldwin, R., R. Forslid, P. Martin, G. Ottaviano, and F. Robert-Nicoud, *Economic Geography and Public Policy* (Princeton: Princeton University Press, 2003).

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

Bound, J., D. A. Jaeger, and R. M. Baker, "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 (1995), 443-450.

Brinkley, D., *Wheels for the World. Henry Ford, His Company, and a Century of Progress* (Penguin, 2003).

Cain, L. P. and Paterson, D. G., "Biased Technical Change, Scale, and Factor Substitution in American Industry, 1850-1919", *Journal of Economic History* 46 (1986), 153-164.

Cameron, A.C., and P.K. Trivedi, *Microeconometrics: Methods and Application* (Cambridge: Cambridge University Press, 2005).

Carter, S., Gartner, S., Haines, M., Olmstead, A., Sutch, R., and Wright, G., *Historical Statistics of the United States*, vol. 4 (Cambridge: Cambridge University Press, 2006).

Chandler, A., *The Visible Hand* (Harvard: Harvard University Press, 1977).

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

Crafts, N., and A. Mulatu, "How Did the Location of Industry Respond to Falling Transport Costs in Britain Before World War I," *The Journal of Economic History* 66 (2006), 575-607.

Davis R.D., D.E Weinstein, "Economic Geography and Regional Production Structure: An Empirical Investigation," *European Economic Review* 43 (1999), 379-407.

Davis R.D., and D.E. Weinstein, "Market Access, Economic Geography and Comparative Advantage: An Empirical Test," *Journal of International Economics* 59 (2003), 1-23.

Easterlin, R.A., "State Income Estimates", in: *Population Redistribution and Economic Growth United States, 1870-1950, Volume I: Methodological Considerations and Reference Tables*, by Everett S. Lee Ann Rather Miller, Carol P. Brainerd, Richard A. Easterlin, (Philadelphia: The American Philosophical Society, 1957).

Ellison, G., and E.L. Glaeser, "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" *American Economic Review Papers & Proceedings* 89 (1999), 311-316.

Estadísticas Históricas de México, (Instituto Nacional de Estadística, 1990).

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

Flink, J., *The Automobile Age* (MIT Press, 1990).

Fujita, M., P. Krugman, and A. Venables, *The Spatial Economy* (Cambridge, MA: MIT Press, 1999).

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

Goldin, C., "America's Graduation from High School: The Evolution and Spread of Secondary Schooling in the Twentieth Century", *The Journal of Economic History* 58 (1998), 345-374.

Goldin, C. and Katz, L. F., "The Origins of Technology-Skill Complementarity", *Quarterly Journal of Economics* 113 (1998), 693-732.

Green, A.G., *Regional Aspects of Canada's Economic Growth* (Toronto and Buffalo: University of Toronto Press , 1971).

Hahn, J., and J. Hausman, " Weak Instruments: Diagnosis and Cures in Empirical Econometrics," *American Economic Review Papers and Proceedings* 93 (2003), 118-125.

Harris, C., "The Market as a Factor in the Localization of Industry in the United States," *Annals of the Association of American Geographers* 64 (1954), 315-348.

Hayashi, F., *Econometrics*, (Princeton: Princeton University Press, 2000).

Head, K., and T. Mayer, "Market Potential and the Location of Japanese Investment in the European Union," *Review of Economics and Statistics* 86 (2004), 959-972.

Hillberry, R. and Hummels, D., "Intranational Home Bias: Some Explanations", Review of Economics and Statistics 85 (2003), 1089-1092.

Jacks, D., Meissner, C., and D. Novy, "Trade Costs 1870-2000," *American Economic Review Papers & Proceedings* 98 (2008), 529-534.

Kim, S., “Expansion of Markets and the Geographic Distribution of Economic Activities: Trends in U. S. Regional Manufacturing Structure, 1860-1987”, *Quarterly Journal of Economics* 110 (1995), 881-908.

Kim, S., “Regions, Resources, and Economic Geography: Sources of U.S. Regional Comparative Advantage, 1880-1987”, *Regional Science and Urban Economics* 29 (1999), 1-32.

Klein, A., “State Personal Income Estimates, United States 1880-1910,” *University of Warwick mimeo*, 2009.

Klier, T. and J. Rubestein, *Who Really Made Your Car. Restructuring and Geographic Change in the Auto Industry* (Kalamazoo, Michigan: W.E. Upjohn Institute for Employment Research, 2008).

Krugman, P., *Geography and Trade* (Cambridge, MA.: MIT Press, 1991).

Krugman, P., “History and Industry Location: The Case of the Manufacturing Belt,” *American Economic Review Papers & Proceedings* 81 (1991), 80-83.

Krugman, P., “A Dynamic Spatial Model,” National Bureau of Economic Research working paper no 4219 (1992).

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

Leontief, W.W., *The Structure of the American Economy, 1919-1939* (New York: Oxford University Press, 1941).

Maddison, A., “World Population, GDP and Per Capita GDP, 1-2003 AD”, *Groningen Growth and Development Centre*, 2004. URL: http://www.ggdc.net/Maddison/

Meyer, D. R., “Emergence of the American Manufacturing Belt: an Interpretation”, *Journal of Historical Geography* 9 (1983), 145-174.

Meyer, D. R., “Midwestern Industrialization and the American Manufacturing Belt in the Nineteenth Century”, *Journal of Economic History* 49 (1989), 921-937.

Midelfart-Knarvik, K.H., H.G. Overman, S. Redding, and A.J. Venables, “The Location of European Industry,” report prepared for the Directorate Generate for Economic and Financial Affairs, European Commission, Economic Papers No. 142, (Brussels: European Commission, April 2000).

Midelfart-Knarvik, K.H., H.G. Overman, and A.J. Venables, “Comparative Advantage and Economic Geography: Estimating the Determinants of Industrial Location in the EU,” CEPR Discussion Paper No. 2618 (2000).

Midelfart-Knarvik, K.H., H.G. Overman, “Delocation and European Integration. Is Structural Spending Justified?” *Economic Policy* (2002), 323-359.

Mitchell, B.R., *International Historical Statistics, Europe 1750-2000, Fifth Edition* (Basingstoke: Palgrave Macmillan, 2003).

Mitchell, B.R., *International Historical Statistics, Americas 1750-2000, Fifth Edition* (Basingstoke: Palgrave Macmillan, 2003).

Nevins, A., Ford. *The Times, The Man, The Company* (New York: Charles Scribner’s Sons, 1954).

Niemi, A.W., *State and Regional Patterns in American Manufacturing 1860-1900* (London: Greenwood Press, 1974).

O'Brien, A.P., ”Factory Size, Economies of Scale, and Great Merger Wave of 1898-1902,” *The Journal of Economic History* 48 (1988), 639-649.

Officer, L.H., *Dollar-Pound Exchange Rate From 1791*, MeasuringWorth, 2008. URL: http://www.measuringworth.org/exchangepound/

Pepper, J.P., "Robust Inferences from Random Clustered Samples: An Application Using Data from the Panel Study of Income Dynamics," *Economic Letters* 75 (2002), 341-345.

Perloff, H.S., E.S. Dunn, E.E. Lampard, and R.F. Muth, *Regions, Resources, and Economic Growth* (Lincoln: University of Nebraska Press, 1960).

Puga, D., "The Rise and Fall of Regional Inequalities", *European Economic Review* 43 (1999), 303-334.

Redding, S., and A. Venables, "Economic Geography and International Inequality,"*Journal of International Economics* 62:1 (2004), 53-82.

Shea, J., "Instrument Relevance in Multivariate Linear Models: A Simple Measure," *Review of Economics and Statistics* 79 (1997), 348-352.

Staiger, D., and J. Stock, "Instrumental Variable Regression with Weak Instruments," *Econometrica* 65 (1997), 557-586.

Stock, J.H., and M. Yogo, "Testing for Weak Instruments in Linear IV Regression", in Andrews, D.W.K., and J. H. Stock, *Identification and Inference for Econometrics Models. Essays in Honor of Thomas Rothenberg* (Cambridge: Cambridge University Press, 2005).

Tsai, L. B.-G., *The Spatial Aggregation of Automobile Manufacturing Activities in the American Midwest*, PhD Dissertation (Berkeley, CA: University of California, Berkeley, 1999).

Turner, C., R. Tamura, S.E. Mulholland, and S. Baier, "Education and Income of the States of the United States: 1840-2000," *Journal of Economic Growth* 12 (2006), 101-158.

Thomas, M., *An Input-Output Approach to the British Economy, 1890-1914*, D. Phil. Thesis (University of Oxford, 1984).

White, H., “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity,” *Econometrica* 48 (1980), 817-838.

White, H., *Asymptotic Theory for Econometricians* (San Diego: Academic Press, 1984).

Whitney, W.G., *The Structure of the American Economy in the Late Nineteenth Century*, PhD Dissertation (Cambridge, MA: Harvard University, 1968).

Wolf, H. C., “Intranational Home Bias in Trade”, *Review of Economics and Statistics* 82 (2000), 555-563.

Wooldridge, J.M., *Econometric Analysis of Cross Section and Panel Data* (Cambridge, MA: MIT Press, 2002).

Wooldridge, J.M., “Cluster-Sample Methods in Applied Econometrics,“ *American Economic Review* 93 (2003), 133-138.

Wooldridge, J.M., “Cluster-Sample Methods in Applied Econometrics: An Extended Analysis,” *Michigan State University mimeo*, 2006.

U.S. Government Publications

Report on the Manufactures of the United States at the Tenth Census, General Statistics. Department of the Interior, Census Office. Washington, Government Printing Office 1883.

Statistics of the Population of the United States at the Tenth Census. Department of the Interior, Census Office. Washington, Government Printing Office 1883.

Report on Manufacturing Industries in the United States at the Eleventh Census:1890. Part I. Total for States and Industries. Department of the Interior, Census Office. Washington, D.C.: Government Printing Office. 1895.

Special Census Report on the Occupation of the Population of the United States at the Eleventh Census: 1890. Washington, D.C.: Government Printing Office. 1896.

Twelfth Census of the United States, Taken in the Year 1900. Manufactures. Part I. United States by Industries. (Volume VII). United States Census Office. Washington, 1902.

Twelfth Census of the United States, Taken in the Year 1900. Manufactures. Part II. States and Territories. (Volume VIII). United States Census Office. Washington, 1902.

Occupation at the Twelve Census. Department of Commerce and Labor, Bureau of the Census. Washington, Government Printing Office, 1904.

Thirteenth Census of the United States Taken in the Year 1910. Manufactures 1909. General Report and Analysis. (Volume VIII). Department of Commerce, Bureau of Census. Washington, Government Printing Office, 1913.

Thirteenth Census of the United States Taken in the Year 1910. Manufactures 1909. Reports by States, with Statistics for Principal Cities. (Volume IX). Department of Commerce, Bureau of Census. Washington, Government Printing Office, 1912.

Fourteenth Census of the United States Taken in the Year 1920. Manufactures 1919. General Report and Analytical Tables. (Volume VIII). Department of Commerce, Bureau of Census. Washington, Government Printing Office, 1923.

Fourteenth Census of the United States Taken in the Year 1920. Manufactures 1919. Reports for States, with Statistics for Principal Cities. (Volume IX). Department of Commerce, Bureau of Census. Washington, Government Printing Office, 1923.

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, Government Printing Office, 1886.

Retail Prices and Wages. Report by Mr. Aldrich, from the Committee on Finance, Part 2, Washington, Government Printing Office, 1892.

Retail Prices 1890 to 1911, Bulletin of the United States Bureau of Labor, no. 105, Part 1, Washington, Government Printing Office, 1912.

Appendix

Dependent Variable

The share of manufacturing labor 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, jewelry 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 less than one percent of the state's total manufacturing employment and includes the industries with a small number of firms to prevent the identification of those firms. As a result, this category contains a heterogeneous set of industries which makes it difficult to assign it to any of the SIC categories. We have decided to perform the analysis with this industry category assigned to SIC 39, miscellaneous, as well as without that industry. The results are virtually unchanged and the regression analysis in the main text is conducted with the exclusion of this industry group.

Independent variables

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), Table 4. 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, Mineral Resources Use: The figures are calculated from Whitney's (1968) input-output table for 1899 and from Leontief's (1941) input-output table for 1919, 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, 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. Leontief's input-output table breaks down the economy into 41 sectors including agriculture and industries. Again, we had to aggregate some of the industries to match the two-digit SIC level. SIC 20, Food and Kindred Products, includes flour and grist products, canning and preserving, bread and bakery products, sugar and glucose and starch, liquor and beverages, slaughtering and meat packing, butter and cheese and etc, other food industries; SIC 23, Apparel and Related Products, includes clothing, and other textile products; SIC 26, Paper and Allied Products, contains paper and wood pulp, and other paper products; SIC 29, Petroleum and Coal Products, consists of refined petroleum, coke, and manufactured gas; SIC 31, Leather and Leather Products, includes leather tanning, leather shoes, and other leather products; SIC 33, Primary Metal Products, includes blast furnaces, steel works and rolling mills, and smelting and refining. SIC 38, Instruments and Related Products, is again taken from Thomas (1984).

State characteristics

The share of population: from U.S. Millennial Statistics (2006), Table Cc125-137, pp. 3-183-3-184

The share of total manufacturing labor force: from Perloff (1960), Table A-6, p. 632.

The share of total agriculture labor force: from Perloff (1960), Table A-2, p. 624.

The share of total mining and quarrying labor force: from Perloff (1960), Table A-3, p. 626.

The share of skilled labor force: The share of the skilled labor force in 1880-1900 is calculated from the U.S. Population Statistics and the U.S. Occupational Statistics. Skilled labor is considered to be the labor 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 U.S. Millennial Statistics (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 $US in 1880-1910 are taken from Flandreau and Zumer (2004) except for Canada, Mexico, and the $US/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 $US come from Estadicas Historicas de Mexico (1990). The Canadian nominal GDP is divided into provinces and the figures come from Green (1971), Table B-1, B-2, 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, 2003, 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 (2003), Table J1. The nominal GDP in 1920 are from Mitchell (2003), 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 Estadicas Historicas de Mexico (1990). The exchange rates between the $US and foreign currencies are calculated from U.S. Millennial Statistics (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.

Table 1.- Industry Characteristics in 1899 and Manufacturing Employment in 1880 and 1920.

Panel A.- Industry Characteristics, 1899

	SIC	Intermediate Input Use	Sales to Industry	Agricultural Input Use	Mineral Resources Use
Food and kindred product	20	18.2	11.7	23.6	1.3
Tobacco and tobacco product	21	1.7	0	18.9	0.1
Textile mill product	22	24.6	57.8	19.9	0.7
Apparel and related products	23	46.2	9.0	1.7	0.2
Lumber and wood products	24	38.9	54.2	7.1	0.1
Furniture and fixtures	25	43.2	5.9	0.0	0.5
Paper and allied products	26	38.5	63.0	6.7	2.4
Printing and publishing	27	23.9	14.3	0.0	0.9
Chemicals and allied products	28	37.3	42.8	11.2	4.3
Petroleum and coal products	29	23.4	33.1	0.0	10.7
Rubber and plastic products	30	22.4	30.3	0.0	1.2
Leather and leather products	31	51.1	37.4	8.2	0.2
Stone, clay, and glass products	32	21.0	23.5	0.0	10.3
Primary metal products	33	47.8	58.4	0.0	4.6
Fabricated metal products	34	10.4	25.6	0.0	0.7
Machinery	35, 36	32.3	22.6	0.0	10.4
Transportation equipment	37	25.9	35.7	0.4	2.1
Instruments and related products	38	51.6	15	0.0	0.02
Miscellaneous manufacturing	39	26.8	15.7	1.3	10.2

Panel B.- Manufacturing Employment (%) 1880, 1920

		1880		1920	
		MB	Outside MB	MB	Outside MB
Food and kindred product	20	75.25	24.75	61.05	38.95
Tobacco and tobacco product	21	78.97	21.03	71.27	28.73
Textile mill product	22	94.63	5.37	75.79	24.21
Apparel and related products	23	93.73	6.27	88.97	11.03
Lumber and wood products	24	77.00	23.00	40.69	59.31
Furniture and fixtures	25	87.58	12.42	81.62	18.38
Paper and allied products	26	95.76	4.24	92.61	7.39
Printing and publishing	27	83.15	16.85	74.08	25.92
Chemicals and allied products	28	69.25	30.75	72.48	27.52
Petroleum and coal products	29	91.31	8.69	54.25	45.75
Rubber and plastic products	30	99.97	0.03	98.35	1.65
Leather and leather products	31	84.88	15.12	88.87	11.13
Stone, clay, and glass products	32	81.09	18.91	80.72	19.28
Primary metal products	33	90.22	9.78	92.31	7.69
Fabricated metal products	34	89.68	10.32	88.22	11.78
Machinery	35, 36	89.35	10.65	93.00	7.00
Transportation equipment	37	86.16	13.84	73.03	26.97
Instruments and related products	38	94.36	5.64	95.07	4.93
Miscellaneous manufacturing	39	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

Notes: The figures in Panel A are for the manufacturing sector and are expressed as the percentages of the gross output. The figures in Panel B are the percentages of the U.S. total in the corresponding category. MB stands for the Manufacturing Belt. Sources: Panel A: Whitney (1968), SIC 21, 25, and 34 are from Leontief (1941), SIC 38 is from Thomas (1984). Panel B: U.S. Census of Manufactures 1880, 1920, Perloff (1960), U.S. Millennial Statistics (2006).

Table 2. -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 $US

	1880		1920			1880		1920	
	Market Potential	Rank	Market Potential	Rank		Market Potential	Rank	Market Potential	Rank
Rhode Island	32.13	1	209.97	2	Alabama	12.62	28	81.59	29
Connecticut	31.88	2	212.41	1	Nebraska	12.61	29	83.56	28
Massachusetts	30.21	3	195.34	4	Arkansas	12.37	30	82.31	29
New Jersey	28.51	4	197.30	3	Mississippi	11.94	31	77.13	31
New York	28.32	5	188.45	5	Florida	10.99	32	70.54	33
New Hampshire	26.75	6	170.47	8	Louisiana	10.91	33	69.97	34
Pennsylvania	26.06	7	172.66	7	Oklahoma	10.23	34	72.58	32
Delaware	25.47	8	174.78	6	South Dakota	9.69	35	63.87	35
Maryland	25.41	9	167.74	9	North Dakota	9.24	36	59.09	37
Vermont	23.15	10	145.70	10	Wyoming	8.91	37	58.42	38
Ohio	21.33	11	142.00	11	Colorado	8.71	38	57.28	39
Indiana	20.07	12	131.91	12	Texas	8.69	39	59.54	36
West Virginia	18.98	13	127.26	14	Nevada	8.09	40	55.84	40
Illinois	18.97	14	129.24	13	New Mexico	7.84	41	50.76	41
Kentucky	18.86	15	123.05	16	Utah	7.37	42	47.32	44
Virginia	18.84	16	123.17	15	Montana	7.34	43	46.30	45
Maine	18.63	17	112.23	18	California	7.23	44	47.53	43
Michigan	18.22	18	121.63	17	Idaho	7.00	45	45.38	46
Wisconsin	16.13	19	107.03	19	Washington	6.74	46	47.70	42
Missouri	15.88	20	106.90	20	Oregon	6.71	47	44.44	47
North Carolina	15.70	21	102.30	21	Arizona	6.66	48	42.02	48
Tennessee	15.65	22	102.11	22					
Iowa	15.18	23	98.73	23					
South Carolina	13.90	24	89.66	24					
Georgia	13.81	25	89.53	25					
Kansas	13.13	26	87.99	26					
Minnesota	12.89	27	84.09	27					

Source: see text

Table 3. - Pooled OLS, Panel Data Fixed Effect, 1880-1920
Forward and Backward Linkages based on 1899 Input-Output Table
Market Potential Estimates Based on $\delta = -1$

	I POLS Cluster-Robust SE	II FE Panel Robust SE
	H-O Forces	
Agric. Employment x agric. Input use	0.002***	0.002***
	[0.0003]	[0.0003]
Educated pop. X white-collar workers	0.0008	0.0006
	[0.0006]	[0.0006]
Coal abundance x steam power use	0.006	0.009
	[0.02]	[0.02]
	NEG Forces	
Market potential x interm. input use	0.00003	0.00003
	[0.00009]	[0.00009]
Market potential x industry sale	0.00028***	0.00027***
	[0.00006]	[0.00006]
Market potential x size of establishment	0.00013***	0.00013***
	[0.00001]	[0.00001]
	Industry and State Controls	
ln (Population)	1.6***	0.87***
	[0.14]	[0.068]
ln (Manuf. Empl)	5.67***	1.48*
	[0.58]	[0.77]
% Agricultural Empl	0.06***	-0.028
	[0.015]	[0.026]
% Educated Population	0.007	-0.003
	[0.019]	[0.017]
Agricultural Input	-0.09***	-0.09***
	[0.017]	[0.017]
% White Collar Workers	0.05**	0.05**
	[0.02]	[0.02]
Coal Abundance	0.069**	-0.008
	[0.03]	[0.018]
Market Potential	-0.007*	-0.03***
	[0.004]	[0.005]
Steam Power Use	1.25***	1.23***
	[0.27]	[0.26]
Intermediate Input Use	-0.03***	-0.03***
	[0.01]	[0.01]
Sales to Industry	-0.02***	-0.02***
	[0.006]	[0.005]
Size of establishment	-0.02***	-0.02***
	[0.001]	[0.001]
Time Dummies	Yes	Yes
No. observations	4560	4560
Adj. R-squared	0.64	0.53
Breusch-Pagan heteroskedasticity test: chi-square(2) = 1129.7***		
robust Hausman test: chi-square (11)=298.757***		

Sources: see text; Notes: * significant at 10%; ** significant at 5%; *** significant at 1%
Note: POLS - Pooled OLS, FE - Fixed Effect, clustered standard errors at the U.S. state level

Table 4.- OLS and Cluster-Specific Fixed Effect Estimations Year by Year
Forward and Backward Linkages based on 1899 Input-Output Table
Market Potential Estimates Based on δ = -1

	1880		1890		1900	
	OLS	FE	OLS	FE	OLS	FE
			H-O Forces			
Agric. Employment x	0.003***	0.003***	0.002**	0.0017**	0.0015***	0.0015***
agric. Input use	[0.0004]	[0.00040]	[0.0007]	[0.0007]	[0.0005]	[0.0005]
Educated pop. X	0.006	0.006	-0.008	-0.008	-0.018***	-0.018***
white-collar workers	[0.004]	[0.004]	[0.0078]	[0.007]	[0.005]	[0.005]
Coal abundance x	0.16**	0.16**	-0.16*	-0.16**	0.03	0.03
steam power use	[0.07]	[0.07]	[0.08]	[0.08]	[0.05]	[0.05]
			NEG forces			
Market potential x	0.0001	0.0001	0.002*	0.0017*	0.001*	0.001*
interm. input use	[0.001]	[0.0009]	[0.0009]	[0.0009]	[0.0006]	[0.0006]
Market potential x	0.002***	0.002***	0.001*	0.001*	0.002***	0.002***
industry sale	[0.0007]	[0.0007]	[0.0007]	[0.0007]	[0.0005]	[0.0005]
Market potential x	0.0009***	0.0008***	0.0007***	0.0007***	0.0006***	0.0006***
size of establishment	[0.0002]	[0.0002]	[0.0002]	[0.0002]	[0.0002]	[0.0002]
			Industry and State Controls			
ln (Population)	1.09***	-	3.04***	-	2.13***	-
	[0.10]	-	[0.31]	-	[0.19]	-
ln (Manuf. Empl)	4.32***	-	4.25***	-	4.40***	-
	[0.61]	-	[0.78]	-	[0.75]	-
% Agricultural Empl	0.07***	-	-0.00	-	0.03	-
	[0.02]	-	[0.03]	-	[0.03]	-
% Educated Population	-0.02	-	-0.19	-	0.28*	-
	[0.05]	-	[0.17]	-	[0.15]	-
Agricultural Input	-0.06**	-0.06**	-0.09**	-0.09**	-0.09***	-0.09***
	[0.02]	[0.02]	[0.04]	[0.04]	[0.03]	[0.03]
% White Collar Workers	-0.07**	-0.07**	0.17**	0.17**	0.23***	0.23***
	[0.03]	[0.03]	[0.07]	[0.07]	[0.06]	[0.06]
Coal Abundance	0.01	-	0.21**	-	-0.01	-
	[0.08]	-	[0.08]	-	[0.06]	-
Market Potential	0.01	-	-0.13***	-	-0.13***	-
	[0.05]	-	[0.04]	-	[0.03]	-
Steam Power Use	1.25	1.25	2.44***	2.44***	0.40	0.40
	[0.84]	[0.84]	[0.78]	[0.78]	[0.45]	[0.45]
Intermediate Input Use	0.03	0.03	-0.08***	-0.08***	-0.05**	-0.05**
	[0.02]	[0.02]	[0.03]	[0.03]	[0.02]	[0.02]
Sales to Industry	-0.09***	-0.09***	-0.03	-0.03	-0.05***	-0.05***
	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]
Size of establishment	-0.03***	-0.03***	-0.04***	-0.04***	-0.04***	-0.04***
	[0.003]	[0.003]	[0.00]	[0.004]	[0.004]	[0.004]
Constant	1.57	-3.33***	7.64**	-2.17***	5.45**	-1.58***
	[1.77]	[0.45]	[2.88]	[0.34]	[2.69]	[0.39]
Observations	912	912	912	912	912	912
R-squared	0.66	0.35	0.61	0.35	0.59	0.39
Breusch-Pagan chi2(1)	212.2***		268.9***		297.3***	
F-test Joint Significance H-O	16.20***	16.31***	3.62**	3.64**	9.77***	9.84***

Sources: see text; 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)

Table 4. - Continued

	1910		1920	
	OLS	FE	OLS	FE
	H-O Forces			
Agric. Employment x	0.0004	0.0004	-0.0003	-0.0003
agric. Input use	[0.0008]	[0.0008]	[0.0008]	[0.0008]
Educated pop. X	-0.001	-0.001	0.001	0.001
white-collar workers	[0.004]	[0.004]	[0.001]	[0.001]
Coal abundance x	0.06	0.06	0.01	0.01
steam power use	[0.06]	[0.06]	[0.03]	[0.03]
	NEG forces			
Market potential x	0.002***	0.002***	0.0007***	0.0007***
interm. input use	[0.0006]	[0.0006]	[0.0002]	[0.0002]
Market potential x	0.0003	0.0003	0.0002**	0.0002**
industry sale	[0.0004]	[0.0004]	[0.00009]	[0.00009]
Market potential x	0.0004***	0.0004***	0.00005***	0.00005***
size of establishment	[0.00008]	[0.00008]	[0.00002]	[0.00002]
	Industry and State Controls			
ln (Population)	2.04***	-	2.45***	-
	[0.16]	-	[0.14]	-
ln (Manuf. Empl)	4.11***	-	4.96***	-
	[1.05]	-	[0.78]	-
% Agricultural Empl	0.06*	-	0.04**	-
	[0.03]	-	[0.02]	-
% Educated Population	0.24*	-	0.00	-
	[0.13]	-	[0.03]	-
Agricultural Input	-0.06*	-0.06*	-0.03	-0.03
	[0.03]	[0.03]	[0.03]	[0.03]
% White Collar Workers	0.03	0.03	0.02	0.02
	[0.05]	[0.05]	[0.04]	[0.04]
Coal Abundance	0.07	-	0.03**	-
	[0.06]	-	[0.02]	-
Market Potential	-0.07***	-	-0.03***	-
	[0.02]	-	[0.01]	-
Steam Power Use	0.38	0.38	2.53**	2.53**
	[0.65]	[0.65]	[0.97]	[0.97]
Intermediate Input Use	-0.18***	-0.18***	-0.12***	-0.12***
	[0.04]	[0.04]	[0.03]	[0.03]
Sales to Industry	0.00	0.00	-0.04***	-0.04***
	[0.02]	[0.02]	[0.01]	[0.01]
Size of establishment	-0.04***	-0.04***	-0.01***	-0.01***
	[0.00]	[0.003]	[0.00]	[0.002]
Constant	3.42	0.07	6.19***	-1.10**
	[2.45]	[0.46]	[1.91]	[0.45]
Observations	912	912	912	912
R-squared	0.55	0.36	0.53	0.26
Breusch-Pagan chi2(1)	216.8***		290.5***	
F-test Joint Significance H-O	0.71	0.46	0.74	0.42

Sources: see text; 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)

Table 5.- Two-Step GMM Instrumental Variable Estimation
Forward and Backward Linkages based on 1899 Input-Output Table
Market Potential Estimates Based on δ = -1

	1880	1890	1900	1910	1920
			H-O Forces		
Agric. Employment x	0.003***	0.002**	0.0016***	0.0003	-0.0004
agric. Input use	[0.0004]	[0.0007]	[0.0005]	[0.0008]	[0.0008]
Educated pop. X	0.006	-0.008	-0.018***	-0.0013	0.001
white-collar workers	[0.004]	[0.007]	[0.005]	[0.004]	[0.001]
Coal abundance x	0.17**	-0.16**	0.03	0.06	0.01
steam power use	[0.07]	[0.08]	[0.05]	[0.06]	[0.03]
			NEG Forces		
Market potential x	0.0001	0.002*	0.001*	0.002***	0.0007***
interm. input use	[0.0009]	[0.0009]	[0.0006]	[0.0006]	[0.0002]
Market potential x	0.002***	0.001**	0.002***	0.0003	0.0002**
industry sale	[0.0007]	[0.0006]	[0.0006]	[0.0004]	[0.00009]
Market potential x	0.001***	0.0007***	0.0005***	0.0004***	0.00005***
size of establishment	[0.0002]	[0.0002]	[0.0002]	[0.00008]	[0.00002]
			Industry and State Controls		
ln (Population)	1.09***	3.04***	2.13***	2.04***	2.45***
	[0.10]	[0.31]	[0.19]	[0.16]	[0.14]
ln (Manuf. Empl)	4.32***	4.25***	4.40***	4.11***	4.96***
	[0.60]	[0.76]	[0.74]	[1.03]	[0.76]
% Agricultural Empl	0.07***	-0.005	0.03	0.06*	0.04**
	[0.02]	[0.03]	[0.03]	[0.03]	[0.02]
% Educated Population	-0.03	-0.19	0.29*	0.24*	0.004
	[0.05]	[0.17]	[0.15]	[0.13]	[0.02]
Agricultural Input	-0.06***	-0.09**	-0.09***	-0.06**	-0.03
	[0.02]	[0.04]	[0.03]	[0.03]	[0.03]
% White Collar Workers	-0.07**	0.17**	0.23***	0.03	0.02
	[0.03]	[0.07]	[0.06]	[0.05]	[0.04]
Coal Abundance	0.01	0.21***	-0.01	0.07	0.03**
	[0.07]	[0.08]	[0.06]	[0.06]	[0.02]
Market Potential	0.02	-0.13***	-0.13***	-0.06***	-0.03***
	[0.05]	[0.04]	[0.03]	[0.02]	[0.01]
Steam Power Use	1.27	2.44***	0.38	0.39	2.48**
	[0.81]	[0.77]	[0.44]	[0.64]	[0.97]
Intermediate Input Use	0.03*	-0.07***	-0.05**	-0.17***	-0.12***
	[0.02]	[0.03]	[0.02]	[0.04]	[0.03]
Sales to Industry	-0.08***	-0.03*	-0.06***	0.01	-0.05***
	[0.02]	[0.02]	[0.02]	[0.02]	[0.01]
Size of establishment	-0.04***	-0.04***	-0.04***	-0.04***	-0.01***
	[0.004]	[0.004]	[0.004]	[0.003]	[0.002]
Constant	1.42	7.62***	5.40**	3.24	6.40***
	[1.72]	[2.81]	[2.64]	[2.40]	[1.87]

Table 5. - Continued

Observations	912	912	912	912	912
R2	0.74	0.7	0.67	0.64	0.62
Shea Partial R2					
mp1vs2	0.98	0.99	0.99	0.99	0.99
mp1vs2_intermed	0.99	0.99	0.99	0.99	0.99
mp1vs2_sale	0.99	0.99	0.99	0.99	0.99
mp1vs2_plant	0.88	0.93	0.94	0.97	0.92
Endog. C test [chisq (4)]	22.3***	5.2	15.7**	13.5**	7.4
Joint Significance	50.4***	11.7***	25.4***	1.4	1.4
Heckscher-Ohlin, chi2(3)					

Sources: see text

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

Table 6.- OLS and Cluster-Specific Fixed Effect Estimations Year by Year
Forward and Backward Linkages based on 1899 Input-Output Table
Market Potential Estimates Based on Gravity Equation

	1880		1890		1900	
	OLS	FE	OLS	FE	OLS	FE
			H-O Forces			
Agric. Employment x	0.003***	0.003***	0.002***	0.002***	0.0016***	0.0016***
agric. Input use	[0.0004]	[0.0004]	[0.0007]	[0.0007]	[0.0005]	[0.0005]
Educated pop. X	0.005	0.005	-0.01	-0.01	-0.02***	-0.02***
white-collar workers	[0.004]	[0.004]	[0.008]	[0.008]	[0.005]	[0.005]
Coal abundance x	0.2**	0.2**	-0.16*	-0.16*	0.035	0.03
steam power use	[0.08]	[0.08]	[0.08]	[0.08]	[0.05]	[0.05]
			NEG forces			
Market potential x	0.0001	0.0001	0.0002**	0.0002**	0.0001**	0.0001**
interm. input use	[0.0002]	[0.0002]	[0.00008]	[0.00008]	[0.00006]	[0.00006]
Market potential x	0.0005***	0.0005***	0.0001*	0.0001*	0.0002***	0.0002***
industry sale	[0.0002]	[0.0002]	[0.00006]	[0.00006]	[0.00005]	[0.00005]
Market potential x	0.0002***	0.0002***	0.00005**	0.00005**	0.00005***	0.00005***
size of establishment	[0.00005]	[0.00005]	[0.00002]	[0.00002]	[0.00002]	[0.00002]
			Industry and State Controls			
ln (Population)	1.08***	-	3.05***	-	2.12***	-
	[0.10]	-	[0.33]	-	[0.20]	-
ln (Manuf. Empl)	4.09***	-	4.32***	-	4.31***	-
	[0.64]	-	[0.84]	-	[0.81]	-
% Agricultural Empl	0.06***	-	-0.01	-	0.03	-
	[0.02]	-	[0.03]	-	[0.03]	-
% Educated Population	-0.02	-	-0.17	-	0.34**	-
	[0.05]	-	[0.17]	-	[0.15]	-
Agricultural Input	-0.06***	-0.06***	-0.10***	-0.10***	-0.10***	-0.10***
	[0.02]	[0.02]	[0.04]	[0.04]	[0.03]	[0.03]
% White Collar Workers	-0.07**	-0.07**	0.20***	0.20***	0.26***	0.26***
	[0.03]	[0.03]	[0.07]	[0.07]	[0.06]	[0.06]
Coal Abundance	0.03	-	0.20**	-	-0.01	-
	[0.08]	-	[0.08]	-	[0.06]	-
Market Potential	0.00	-	-0.01***	-	-0.01***	-
	[0.01]	-	[0.00]	-	[0.00]	-
Steam Power Use			2.45***	2.45***	0.34	0.34
	[0.86]	[0.86]	[0.77]	[0.77]	[0.45]	[0.45]
Intermediate Input Use	0.02	0.02	-0.13**	-0.13**	-0.09**	-0.09**
	[0.02]	[0.02]	[0.05]	[0.05]	[0.04]	[0.04]
Sales to Industry	-0.10***	-0.10***	-0.06	-0.06	-0.09***	-0.09***
	[0.02]	[0.02]	[0.04]	[0.04]	[0.03]	[0.03]
Size of establishment	-0.04***	-0.04***	-0.05***	-0.05***	-0.05***	-0.05***
			[0.01]	[0.01]	[0.01]	[0.01]
Constant		-3.33***	10.77***	-2.17***	7.74**	-1.58***
	[1.92]	[0.44]	[3.48]	[0.34]	[3.24]	[0.41]
Observations	912	912	912	912	912	912
R-squared	0.66	0.34	0.61	0.34	0.59	0.38
Breusch-Pagan chi2(1)	212.2***		268.9***		297.3***	
F-test Joint Significance H-O	16.7***	16.8***	4.5***	4.5***	15.01***	15.1***

Sources: see text; 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)

Table 6. - Continued

	1910		1920	
	OLS	FE	OLS	FE
		H-O Forces		
Agric. Employment x	0.0008	0.0008	-0.00005	-0.00005
agric. Input use	[0.0008]	[0.0008]	[0.0008]	[0.0008]
Educated pop. X	-0.003	-0.003	0.0009	0.0009
white-collar workers	[0.004]	[0.004]	[0.001]	[0.001]
Coal abundance x	0.05	0.05	0.005	0.005
steam power use	[0.06]	[0.06]	[0.04]	[0.04]
		NEG forces		
Market potential x	0.0002***	0.0002***	0.00008***	0.00008***
interm. input use	[0.00008]	[0.00008]	[0.00003]	[0.00003]
Market potential x	0.00005	0.00005	0.00003**	0.00003**
industry sale	[0.00004]	[0.00004]	[0.00001]	[0.00001]
Market potential x	0.00003***	0.00003***	0.000001**	0.000001**
size of establishment	[0.00001]	[0.00001]	[0.0000001]	[0.0000001]
		Industry and State Controls		
ln (Population)	2.03***	-	2.43***	-
	[0.18]	-	[0.14]	-
ln (Manuf. Empl)	4.10***	-	4.87***	-
	[1.09]	-	[0.80]	-
% Agricultural Empl	0.05	-	0.04*	-
	[0.03]	-	[0.02]	-
% Educated Population	0.25*	-	0.01	-
	[0.13]	-	[0.02]	-
Agricultural Input	-0.07**	-0.07**	-0.04	-0.04
	[0.03]	[0.03]	[0.03]	[0.03]
% White Collar Workers	0.06	0.06	0.03	0.03
	[0.05]	[0.05]	[0.04]	[0.04]
Coal Abundance	0.07	-	0.04**	-
	[0.06]	-	[0.02]	-
Market Potential	-0.01***	-	-0.00***	-
	[0.00]	-	[0.00]	-
Steam Power Use	0.42	0.42	2.64**	2.64**
	[0.66]	[0.66]	[1.01]	[1.01]
Intermediate Input Use	-0.26***	-0.26***	-0.19***	-0.19***
	[0.07]	[0.07]	[0.06]	[0.06]
Sales to Industry	-0.02	-0.02	-0.07***	-0.07***
	[0.03]	[0.03]	[0.02]	[0.02]
Size of establishment	-0.04***	-0.04***	-0.02***	-0.02***
	[0.01]	[0.01]		
Constant	6.61**	0.07	9.06***	-1.10**
	[2.88]	[0.48]	[2.54]	[0.44]
Observations	912	912	912	912
R-squared	0.55	0.35	0.53	0.26
Breusch-Pagan chi2(1)	216.8***		290.5***	
F-test Joint Significance H-O	1.0	1.0	0.24	0.24

Sources: see text; 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)

Table 7.- Two-Step GMM Instrumental Variable Estimation, Sensitivity Analysis
Forward and Backward Linkages based on 1899 Input-Output Table
Market Potential Estimates Based on Gravity Equation

	1880	1890	1900	1910	1920
			H-O Forces		
Agric. Employment x	0.003***	0.002***	0.002***	0.0007	-0.00005
agric. Input use	[0.0004]	[0.0007]	[0.0004]	[0.0008]	[0.0008]
Educated pop. X	0.006	-0.01	-0.02***	-0.003	0.0009
white-collar workers	[0.004]	[0.008]	[0.005]	[0.004]	[0.001]
Coal abundance x	0.18**	-0.15*	0.03	0.05	0.005
steam power use	[0.07]	[0.08]	[0.05]	[0.06]	[0.04]
			NEG Forces		
Market potential x	0.0001	0.0001**	0.0001**	0.0002***	0.00008***
interm. input use	[0.0002]	[0.00008]	[0.00006]	[0.00008]	[0.00003]
Market potential x	0.0005***	0.0001**	0.0002***	0.00005	0.00003**
industry sale	[0.0002]	[0.00006]	[0.00005]	[0.00004]	[0.00001]
Market potential x	0.0002***	0.00005**	0.00004*	0.00004***	0.000005**
size of establishment	[0.00005]	[0.00002]	[0.00002]	[0.00001]	[0.000002]
			Industry and State Controls		
ln (Population)	1.08***	3.05***	2.12***	2.03***	2.43***
	[0.10]	[0.32]	[0.20]	[0.17]	[0.14]
ln (Manuf. Empl)	4.09***	4.32***	4.31***	4.10***	4.87***
	[0.62]	[0.82]	[0.80]	[1.07]	[0.78]
% Agricultural Empl	0.06***	-0.01	0.03	0.05	0.04**
	[0.02]	[0.03]	[0.03]	[0.03]	[0.02]
% Educated Population	-0.02	-0.17	0.35**	0.25*	0.01
	[0.05]	[0.17]	[0.14]	[0.13]	[0.02]
Agricultural Input	-0.06***	-0.10***	-0.10***	-0.07**	-0.04
	[0.02]	[0.04]	[0.03]	[0.03]	[0.03]
% White Collar Workers	-0.07**	0.20***	0.27***	0.06	0.03
	[0.03]	[0.07]	[0.05]	[0.05]	[0.04]
Coal Abundance	0.02	0.20**	-0.01	0.07	0.04**
	[0.08]	[0.08]	[0.06]	[0.06]	[0.02]
Market Potential	0.001	-0.01***	-0.01***	-0.008***	-0.003***
	[0.001]	[0.003]	[0.003]	[0.002]	[0.0009]
Steam Power Use	1.13	2.35***	0.33	0.43	2.60***
	[0.83]	[0.78]	[0.44]	[0.65]	[1.00]
Intermediate Input Use	0.02	-0.12**	-0.09**	-0.25***	-0.20***
	[0.02]	[0.05]	[0.04]	[0.07]	[0.06]
Sales to Industry	-0.10***	-0.07*	-0.11***	-0.01	-0.07***
	[0.02]	[0.04]	[0.03]	[0.03]	[0.02]
Size of establishment	-0.04***	-0.05***	-0.04***	-0.05***	-0.02***
	[0.005]	[0.01]	[0.009]	[0.007]	[0.003]
Constant	1.16	10.98***	7.53**	6.40**	9.43***
	[1.87]	[3.37]	[3.18]	[2.87]	[2.49]

Table 7. - Continued

Observations	912	912	912	912	912
R2	0.74	0.7	0.67	0.64	0.62
Shea Partial R2					
mp	0.98	0.97	0.99	0.99	0.99
mp_intermed	0.99	0.98	0.98	0.99	0.99
mp_sale	0.98	0.97	0.98	0.99	0.98
mp_plant	0.83	0.88	0.81	0.92	0.81
Endog. C test [chisq (4)]	19.1***	3.8	12.7**	6.9	12.1**
chi2(3) Joint Significance Heckscher-Ohlin	52.6***	13.8***	59.3***	3.02	0.7

Sources: see text

Notes: regression with cluster-robust se, Cluster at the U.S. state level,

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8.- OLS, Cluster-Specific Fixed Effect, and Two-Step GMM Instrumental Variable Estimations in 1920, Sensitivity Analysis

Forward and Backward Linkages based on 1919 Input-Output Table

	1920a			1920b		
	OLS	FE	GMM	OLS	FE	GMM
				H-O Forces		
Agric. Employment x	0.0008***	0.0002	-0.0006	0.0009***	0.0005	-0.0001
agric. Input use	[0.0003]	[0.0007]	[0.0009]	[0.0003]	[0.0007]	[0.0009]
Educated pop. X	0.001	0.001	0.0008	0.001	0.001	0.0004
white-collar workers	[0.001]	[0.001]	[0.0011]	[0.001]	[0.001]	[0.001]
Coal abundance x	-0.003	-0.004	-0.0003	-0.01	-0.01	-0.003
steam power use	[0.03]	[0.03]	[0.033]	[0.03]	[0.03]	[0.03]
				NEG Forces		
Market potential x	0.0001**	0.0002**	0.0008***	0.00001	0.00001*	0.00008***
interm. input use	[0.00007]	[0.00007]	[0.0002]	[0.000008]	[0.000001]	[0.00003]
Market potential x	0.0002**	0.0003***	0.0004***	0.00002**	0.00002***	0.00005***
industry sale	[0.00009]	[0.00009]	[0.0001]	[0.00001]	[0.00001]	[0.00001]
Market potential x	0.00005***	0.00005***	0.00005***	0.000001**	0.000001**	0.000003
size of establishment	[0.00002]	[0.00002]	[0.00002]	[0.0000001]	[0.0000001]	[0.000002]
				Industrial and State Controls		
ln (Population)	2.45***	-	2.45***	2.43***	-	2.4***
	[0.14]	-	[0.14]	[0.14]	-	[0.14]
ln (Manuf. Empl)	4.96***	-	4.96***	4.87***	-	4.87***
	[0.78]	-	[0.76]	[0.80]	-	[0.78]
% Agricultural Empl	0.04**	-	0.04***	0.03*	-	0.04**
	[0.02]	-	[0.02]	[0.02]	-	[0.02]
% Educated Population	0.00	-	0.008	0.00	-	0.02
	[0.03]	-	[0.02]	[0.03]	-	[0.02]
Agricultural Input	-0.07***	-0.04	0.007	-0.07***	-0.05**	0.01
	[0.02]	[0.02]	[0.03]	[0.02]	[0.02]	[0.04]
% White Collar Workers	0.03	0.04	0.08**	0.03	0.04	0.16***
	[0.04]	[0.04]	0.04	[0.04]	[0.04]	[0.04]
Coal Abundance	0.04**	-	0.04**	0.04**	-	0.04**
	[0.02]	-	[0.02]	[0.02]	-	[0.02]
Market Potential	-0.01*	-	-0.03***	-0.00	-	-0.003***
	[0.01]	-	[0.007]	[0.00]	-	[0.0008]
Steam Power Use	2.94***	3.06***	4.14***	2.96***	3.07***	5.4***
	[0.94]	[0.94]	[1.07]	[0.93]	[0.93]	[1.4]
Intermediate Input Use	-0.06***	-0.06***	-0.12***	-0.06***	-0.06***	-0.18***
	[0.02]	[0.02]	[0.03]	[0.02]	[0.02]	[0.05]
Sales to Industry	-0.04***	-0.04***	-0.05***	-0.05***	-0.05***	-0.09***
	[0.01]	[0.01]	[0.01]	[0.02]	[0.02]	[0.02]
Size of establishment	-0.01***	-0.01***	-0.01***	-0.02***	-0.02***	-0.01***
	[0.002]	[0.002]	[0.002]	[0.003]	[0.003]	[0.004]
Constant	4.04**	-1.46***	4.5	4.12*	-1.46***	5.5
	[1.80]	[0.42]	[1.7]	[2.17]	[0.43]	[2.04]
Observations	912	912	912	912	912	912
R-squared	0.53	0.26	0.56	0.53	0.26	0.6
Breusch-Pagan chi2(1)	281.8***			276.2***		
F-test Joint Significance H-O	3.1**	0.38		3.65**	0.49	

Sources: see text; Notes: * significant at 10%; ** significant at 5%; *** significant at 1%,

OLS - cluster-robust se, FE - cluster-specific fixed effect with cluster-robust se at the U.S. state level
GMM – two step GMM with clustered-robust se; Degrees of Freedom in F-test are (3, 47) for OLS and FE, a refers to market potential estimates based on $\delta = -1$, b refers to market potential estimates based on gravity equation

Table 9.- Beta Coefficients, Estimations Year by Year

	1880	1890	1900	1910	1920
	Market Potential Estimates Based on δ = -1				
	H-O Forces				
Agric. Employment x agric. Input use	0.18	0.10	0.10	0.02	-0.02
Educated pop. X white-collar workers	0.03	-0.05	-0.21	-0.02	0.06
Coal abundance x steam power use	0.12	-0.08	0.03	0.05	0.01
	NEG Forces				
Market potential x interm. input use	0.007	0.11	0.10	0.30	0.26
Market potential x industry sale	0.15	0.09	0.19	0.06	0.10
Market potential x size of establishment	0.25	0.24	0.33	0.40	0.23

Note: The table presents only the beta coefficients of the interaction variables. The full set of the beta coefficients is available from the authors upon request. The beta coefficients are defined as beta(i)=[s(xi)/s(y)]*b(xi) where b(xi) is the estimates of xi, s(xi) is the standard deviation of xi and s(y) is the standard deviation of y. Beta coefficients are calculated from the FE regressions in Table 4.

Sources: see text