THE UNIVERSITY OF WARWICK

March 2010 No.4
(*Revised October2010*)

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

Dr Alexander Klein (University of Kent) and Professor Nicholas Crafts (University of Warwick)

WORKING PAPER SERIES

Centre for Competitive Advantage in the Global Economy

Department of Economics

CAGE
Competitive Advantage in the Global Economy

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

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

Alexander Klein
(University of Kent)

and

Nicholas Crafts
(University of Warwick)

Abstract

This paper investigates 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. Market potential was generally much higher for states in the manufacturing belt. Natural advantage played a role in industrial location decisions through agricultural inputs which mattered for a 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 Guinnane, Tim Leunig, Chris Meissner, Dennis Novy, Henry Overman, Mark Thomas, Tony Venables, Nikolaus Wolf, and seminar participants at the AEA meeting in Atlanta, Bank of Italy, Copenhagen, Stanford, UC Davis, UC Irvine, Virginia, Warwick, Yale, and York and for their comments.

1. Introduction

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

Beyond this, descriptions of American industrial geography also sought to understand the manufacturing belt. The term 'manufacturing belt' has long been used to describe the remarkable spatial concentration of industry in the 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.

[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.

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

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

There is relatively little modern empirical work on determinants of industrial location in the late-nineteenth and early-twentieth-centuries. The most important papers are Kim (1995) (1999). Kim (1999) estimated the following equation based on the Rybczynski theorem for production of 2-digit manufacturing industries across U.S. states for snapshot years between 1880 and 1987:

$$Y = \alpha_0 + \alpha_1 Labor + \alpha_2 Capital + \alpha 3 Natural\ Resources$$

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

$$Localization = \beta_0 + \beta_1 Scale + \beta_2 Resource$$

He found that U.S. regional specialization in the late 19th and early 20th centuries was positively related to both variables. Thus, the manufacturing belt was based on the rise of large-scale production methods that were intensive in the use of raw materials and energy sources that were relatively immobile.

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

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

We model industrial location taking explicit account of interactions between industrial characteristics and regional characteristics. The approach that we use is grounded in a model of production and trade that takes account both of input-price variations resulting from factor endowments and from costs of intermediate inputs and also of the spatial pattern of demand. We try to explain the existence and persistence of the manufacturing belt around the turn of the twentieth century in an analysis of the shares of employment in 2-digit manufacturing industries across 48 U.S. states using a newly-constructed dataset.[3]

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

1) Which factor endowments affected the location of manufacturing?

2) Did market potential influence the location of manufacturing through linkage effects and/or scale effects?

3) How important were factor endowments and market potential, respectively, as determinants of industrial location?

2. A model and an empirical framework

An Empirical Framework

We wish to examine the relative importance of natural advantages and market potential in explaining the lock in of the manufacturing belt region around the turn of the twentieth century. The literature offers two possibilities: an approach developed

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

by Davis and Weinstein (Davis and Weinstein 1999, 2003) and an approach developed by Midelfart-Knarvik, Overman and Venables (2000, henceforth MK).

The empirical approach developed by Davis and Weinstein (Davis and Weinstein 1999, 2003) uses the home-market effect to empirically separate an increasing returns model of economic geography from a Heckscher-Ohlin model 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 economic geography world of increasing returns, 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 economic geography models of increasing returns 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 those economic geography models is confirmed. Otherwise, other theories are more relevant.

Midelfart-Knarvik, Overman and Venables (2000) 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.[4] Their approach is a synthesis and generalization of two existing approaches in the empirical literature: a literature which estimates the effect of industry characteristics on trade, and a literature which

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

estimates the effect of country characteristics on trade and production.[5] 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. This 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 market potential forces in shaping the location of industries in the EU.

We have decided to use MK methodology to investigate the relative importance of the market potential and natural advantages in explaining the lock-in of the manufacturing belt. The MK approach provides a simple, yet theoretically sound, empirical test which is richer than Davis and Weinstein approach since it enables us separately to estimate the effect of forward and backward linkages on the geographical location of manufacturers.[6] Moreover, it explicitly considers how the characteristics of states interact with those of industries which is a major advantage compared with the approach of Kim (1995) who related the spatial concentration of production only to industrial characteristics, namely, plant size and raw-material intensity.

The MK Model[7]

---

[5] 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.

[6] The MK model has limitations. It abstracts from imperfect competition and does not admit the possibility of multiple equilibria so it is less complex than the Krugman and Venables (1995) model.

[7]The exposition presents a shortened version of the MK model and closely follows Midelfart-Knarvik et al. (2000).

There are *I* countries, *K* industries and *M* primary sectors. The industries are perfectly competitive operating with constant returns to scale technology which uses both primary factors and intermediate goods as factor inputs. Each industry *k* in country *i* produces n_{ik} number of varieties. The number of varieties is assumed to be determined exogenously. The trade costs that industry *k* incurs when shipping goods from country i to *j* are assumed to be of iceberg type denoted as t_{ijk}. The value of each industry's production is determined by the supply of factor inputs, the prices of intermediate goods and the distribution of demand across industries.

The product varieties produced by industry *k* in country *i* are symmetrical in a sense that they face same costs and demand function. The free on board prices equal unit costs, so $p_{ik}=c(z_{ik})$ where z_{ik} is a vector of input prices in industry *k* and country *i*. The vector of input prices consists of the prices of primary factors, $w_{i,}$ and the price of a composite intermediate good, q_i. Iceberg transport costs (t_{ijk}-1) mean that the price of shipping the industry *k*'s product from country *i* to *j* is $c(z_{ik})t_{ijk}$. The value of demand for an industry *k* product which is shipped from country *i* to country *j* is determined by a standard Dixit-Stiglitz maximization exercise which yields the following value of demand and the price index for industry *k* in country *j*:

$$x_{ijk} = \left[c(z_{ik})t_{ijk}\right]^{1-\sigma} E_{jk} \left(P_{ik}\right)^{\sigma-1} \qquad (1)$$

$$P_{ik} = \left[\sum_i n_{ik} \left(c(z_{ik})t_{ik}\right)^{1-\sigma}\right]^{1/(1-\sigma)} \qquad (2)$$

where E_{jk} is the total expenditure on the products of industry *k* in country *j* and σ is the elasticity of substitution. The total value of industry *k* production in country *i* is given by the summation over all product varieties produced by that industry and over all markets *j* where the products are shipped:

$$TV_{ik} = n_{ik} \left[c(z_{ik})t_{ijk}\right]^{1-\sigma} E_{jk} \left(P_{ik}\right)^{\sigma-1} \qquad (3)$$

We take the total value of the production as a numeraire, meaning that $\Sigma_i \Sigma_k TV_{ik} = 1$ where TV_{ik}, with a slight abuse of notation, is now the share of production of industry *k* in country *i*. Also, we define the share of country *i* in total production $s_i = \Sigma_k TV_{ik}$, and the share of industry *k* in total production as $s_k = \Sigma_i TV_{ik}$.

As was mentioned earlier, the number of varieties is assumed to be determined exogenously. Specifically, they are assumed to be set in proportion to the size of an industry and country up to an error term ω_{ik}:

$$n_{ik} = s_i s_k exp(\omega_{ik}) \qquad (4).$$

This assumption departs from the standard monopolistically competitive framework used in the new-economic-geography literature. Having monopolistically competitive industry would imply that the level of output of each variety is determined by the zero profit condition, the number of varieties is endogenously determined by free entry-exit condition, and the cross-country variation in output is due to different number of varieties produced in each country. Assuming away monopolistically competitive industries and setting the number of varieties exogenously allows the derivation of a linear regression equation. This does not mean, however, that the model gives up on a feature that is important in the geography literature - geographical distribution of the demand. Indeed, even though the number of varieties is exogenously fixed, the level of output of each variety can vary across countries according to equation 3.

The estimation equation is based on the double-relative measure of output of each industry in each country. Using equations 3 and 4, we can express that relative measure of output as

$$s_{ik} = TV_{ik} \big/ s_i\, s_k = c\left(z_{ik}\right)^{1-\sigma} \sum_j \left(\tau_{jk}\right)^{1-\sigma} E_{jk} \left(P_{jk}\right)^{\sigma-1} \exp\left(\omega_{ik}\right) \qquad (5).$$

We can see that the resulting expression consists two parts. The first is the cost function $c(z_{ik})$ which captures the input price variation; the second is the summation part which captures the demand effects and can be denoted as the market potential of industry *k* in country *i*:

$$MP\left(m_{ik}\right) = \sum_j \left(t_{ijk}\right)^{1-\sigma} E_{jk} \left(P_{jk}\right)^{\sigma-1} \qquad (6)$$

where m_{ik} denotes industry characteristics that interact with the country's geographical characteristics such as transportation costs or the spatial pattern of demand. The share of output of each industry in each country is now

$$s_{ik} = c\left(z_{ik}\right)^{1-\sigma} MP\left(m_{ik}\right) \exp\left(\omega_{ik}\right) \qquad (7)$$

which means that the cross-country variation of industries' output is determined both by input prices and by market potential.

To derive a linear regression equation, equation 7 is log-linearized around a reference point. The reference point is $c(z^r) = 1$ and $MP(m^r) = 1$. The idea is that there exists an input price vector z^r such that there is no cross-country variation in input costs $c(z^r_k)$ for all *k*, and industry characteristics m^r such that there is no cross-country variation in market potential $MP(m^r_i)$ for all *i*. The resulting equation then captures the variation of industry and input costs characteristics from that reference point:

$$\Delta s_{ik} = \left(1-\sigma\right)\Delta z_i \varepsilon_c + \Delta m_k \varepsilon_{MP} + \omega_{ik} \qquad (8)$$

where Δ denotes a log deviation from the reference point, ε_c is the vector of elasticities of industry *k* costs with respect to input prices z_i, and ε_{mp} is the vector of elasticities of country *i* market potential with respect to industry characteristics m_k.

We can now use the fact that s_{ik} are shares and the deviations from the reference point are both positive and negative which implies that the summation condition $\Sigma_i \Sigma_k TV_{ik} \Delta s_{ik} = 0$ must be satisfied. Using that equation with equation 8 yields

$$\Delta s_{ik} = (1 - \sigma)\left[\varepsilon_c - \sum_i \sum_k TV_{ik}\varepsilon_c\right]\Delta z_i + \left[\varepsilon_{MP} - \sum_i \sum_k TV_{ik}\varepsilon_{MP}\right]\Delta m_k + \omega_{ik} \qquad (9)$$

We see that the terms on the right hand side are the products of industry and country characteristics, both expressed as deviations from the reference points. Specifically, the first bracket gives the product of countries' input prices and costs elasticities with respect to those input prices (which are, in fact, input shares). The second bracket is the product of industries' characteristics and elasticities of countries' market potential with respect to these characteristics. Again, ε_c and ε_{mp} are elasticities; Δz_i and Δm_k are log deviations from the reference point.

We know that the input prices include both primary products and an intermediate good. Midelfart-Knarvik et al. show that it is possible to express the prices of the primary products in country *i* in terms of the factor endowments of country *i* (Midelfart-Knarvik et al. (2000), Appendix 2).[8] Specifically, they show that

$$\Delta z_i = F\,\Delta L_i \qquad (10)$$

where ΔL_i is the variation of the factor endowments from the reference point and F is the matrix of elasticities of factor prices with respect to factor endowments, evaluated at the reference point. As for the price of the intermediate good q_i in a country *i*, Midelfart-Knarvik et al. (2000) show that it can be directly linked to a distance weighted measure of proximity to production in the industry which is a measure of 'supplier access' of country *i*.

Expanding equation 9, remembering that Δ is a log deviation from the reference point and, for the sake of simplicity, denoting country characteristics as *y*, industry characteristics as *x*, and reference points for industry and country χ and γ as respectively, we obtain the following equation:

[8] The derived expression embodies the Rybczynski effect.

$$\ln\left(_{i\,k} \right) = A + y_i\, x_k - \gamma\, x_k - \chi\, y_i \qquad (1)$$

where A is a term containing the sum of products of industry and country reference points. This simplified specification assumes only one country and one industry characteristic. Having *j* industry and country characteristics, that equation then becomes

$$\ln\left(_{i,k} \right) = A + \sum_j \left(y^j{}_i\; x^{j,k} - \gamma^j\; x^{j,k} - \chi^j\; y^{j,k} \right) \qquad (2)$$

where y^j_i is the level of *j*th country characteristic in country *i*; x^{jk} is the industry *k* value of the industry characteristic paired with country characteristic *j*. An econometric application of this equation yields the following regression equation

$$\ln\left(^k_i \right) = c + \sum_j \left(\beta^j y^j{}_i\; x^{j,k} - \beta^j \gamma^j\; x^{j,k} - \beta^j \chi^j\; y^{j,k} \right) + \varepsilon^k_i \qquad (3)$$

This gives a list of independent variables that comprises a constant, country characteristics, industrial characteristics, and interactions between country and industrial characteristics. The estimated coefficients of the country 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. Those estimates, when divided by β^j, provide the estimates of the industry and country reference points. The estimated coefficients of the interaction variables, y^jx^j are estimates of β^j, and *c* is a constant term. The most important estimates from the point of view of what determines the geographical concentration of industries are *the interaction terms* which show the importance of the interplay between industry and country characteristics.

What industry and country characteristics should be considered? With regard to the characteristics related to the endowment of primary products, equation 10 is general enough to include practically any factor-endowment characteristics. As for the characteristics related to the intermediate good, the model allows them to be linked to market potential (the details are in Midelfart-Knarvik et al. (2000), section 4.2) yielding an interaction between the share of intermediate goods in production and market potential which captures 'forward linkages'. However, the number of industry and country characteristics related to market potential directly by equation 6 is limited. In particular, the model allows only two of them: transport intensity, captured by the term $(t_{ijk})^{(1-\sigma)}$ and 'backward linkages' captured by E_{jk}.

The regression equation (13) is a structural equation derived from the model. Its estimation requires that variables are constructed in logs and elasticities as Midelfart-Knarvik et al. (2001) did in their analysis of the location of industries across EU countries. The construction of the relevant variables is, however, data demanding and often infeasible. That might a reason why the literature also views equation 13 as a reduced-form regression equation relating the geographical distribution of industries across countries/regions to industry and country/region characteristics. Various studies have used different state and industry interactions and estimated variants of equation 13. Midelfart-Knarvik et al. (2000) themselves added population and manufacturing employment to equation 13 instead of using a mere constant. In addition, they excluded transport intensity and include a proxy for economies of scale. Wolf (2007), on the other hand, included only the interaction variables and replaced the regional and industry characteristics with region and industry dummies. This approach is acceptable since we are interested only in the interaction terms which capture the effect of industry and regional characteristics on geographical concentration. Another departure from equation 13 is to use employment instead of gross value of production due to the unavailability of production value data (Crafts and Mulatu, 2006; Wolf, 2007). This requires estimation with region-industry dummies to control for the effects that productivity differences might have on the employment-based location quotient (Wolf, 2007).

3. Implementation of the MK empirical framework and data set

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

Regression Equation

In the implementation of the model, we estimate equation (13) using the method of Wolf (2007). We use four state characteristics (share of farm land, share of educated population, coal prices, market potential), six industry characteristics (the share of white-collar workers, steam power use, plant size, agricultural input use, intermediate input use, sales to industry), six interactions and add the state and industry dummies. The estimated equation (14) can be expressed as follows:

$$\begin{aligned}\ln(s_{i,t}^{k}) = {} & \beta_1(\text{FARM LAND x AGRICULTURE INPUT USE})_{i,t} + \\ & + \beta_2(\text{EDUCATED POPULATION x WHITE COLLAR WORKERS})_{i,t} + \\ & + \beta_3(\text{COAL ABUNDANCE x STEAM POWER USE})_{i,t} + \\ & + \beta_4(\text{MARKET POTENTIAL x INTERMEDIATE INPUT USE})_{i,t} + \\ & + \beta_5(\text{MARKET POTENTIAL x SALES TO INDUSTRY})_{i,t} + \\ & + \beta_6(\text{MARKET POTENTIAL x SIZE OF ESTABLISHMENT})_{i,t} + \\ & + \beta_7\text{STATE}_{i,t} + \beta_8\text{INDUSTRY}_{i,t} + \varepsilon_{i,t}^{k} \qquad (14)\end{aligned}$$

The first three of these interactions are predicted by the Heckscher-Ohlin (H-O) theory based on factor endowments. The relative magnitude and statistical significance of β_1 shows the importance of farmland in influencing the location of industry and so on. The last three are predicted by the Krugman and Venables (1995) model 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). The coefficients β_4, β_5 and β_6 show the importance of market potential as a determinant of industrial location. In the original work by Midelfart-Knarvik et al., the authors estimate their version of the equation (13) using OLS, and account for the heteroskedasticity and the country and industry fixed effects. We will 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.[9] The data on the shares of

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

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 labour force in each U.S. state are from Perloff et al. (1960), the share of farm land is calculated from the Historical Statistics of the United States (2006), coal prices are taken from various U.S. government sources, and the data on educated population by states come from the U.S. occupation censuses and Goldin (1998).[10] The share of white-collar workers as well as of steam power use is extracted from the U.S. Censuses of Manufactures 1880-1920. Average plant size is from O'Brien (1988). Forward and backward linkages are evaluated using an input-output table for the U.S. economy in 1899 Whitney (1968).[11]

Panel A in Table 2 reports industrial characteristics obtained from the 1899 input-output table which relate to key aspects highlighted by locational hypotheses based either on market potential (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

[10] We thank Claudia Goldin for providing the data.

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

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 2 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, 2004) with the implication that a gravity equation framework should be used to estimate market potential. However, the resulting methodology requires internal trade flows data which are unavailable for the U.S. for the period 1880-1920. Therefore we use Harris's original approach and calculate the market potential of a U.S. state *i* using the formula $M_i = \sum_j \varphi_{ij}$ GDPj where φ_{ij} is the accessibility of market *j* for goods from the U.S. state *i* defined as $\varphi_{ij} = d_{ij}^{\delta}$ with $\delta = -1$. The market *j* consists of nominal GDP in foreign countries, in other U.S. states, and in the home state *i*. The market accessibility of own U.S. states is calculated as

$$\varphi_{ii} = d_{ii}^{-\delta} = [2/3 \, . \, (area_i/\pi)^{0.5}]^{-\delta} \qquad (15)$$

Nominal GDP of U.S. states in 1880-1910 is taken from Klein (2009) which provides new estimates of 1890 and 1910 nominal GDP for each U.S. state based on the methodology developed by Easterlin (1957), and re-estimates Easterlin's original 1880 and 1900 estimates.[12] Data for 1920 are from Easterlin (1957). The sources of nominal GDP for foreign countries and the corresponding exchange rate are in the Appendix. 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 U.S. states is calculated as the kilometer distance between their capital cities.

[12] 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.

Although there are no US internal trade flows data for the period 1880-1920, we can justify the assumption of δ = -1 in two ways. First, our estimates of market potential are for the railroad era and we believe that by this time physical distances are a reasonable approximation to economic distances inside the United States. Our choice of -1 for δ is consistent with estimates for modern internal U.S. trade (Wolf, 2000; Hillberry and Hummels, 2003, Knaap, 2006). Second, we can analyze US internal railroad commodity trade in 1949. This is the earliest date for which internal trade data exist and it is suitable for our purposes because the manufacturing belt was still intact at that time, and the railroads were still the most important transportation mode.[13] The data come from the Interstate Commerce Commission Carload Waybill Statistics which report commodity flows between the US states at 3-digit level. We estimate the following gravity regression:

$$\ln X_{ij} = EX_i + IM_j + \delta \ln d_{ij} + \beta_j B_{ij} + \varepsilon_{ij} \qquad (16),$$

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

Table 3 displays our estimates of market potential by state for 1880 and 1920. Two points stand out. First, the rank order of market potential is very stable during this period. Second, the 'manufacturing-belt' states tend to have the highest levels of market potential in both years. It should be noted that states with similar GDP inside

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

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 3 shows, market potential of the former was about 5 times that of the latter.

4. Empirical Results

Estimation Issues

In our initial estimations of equation (14) market potential is calculated assuming δ = -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 (13) 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 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 hence its corresponding interactions may be endogenous. This calls for instrumental variable estimation. In our setting, we have to rely on an exogenous geographical determinant such as distance to an eastern seaport, as used in several recent studies (e.g. Redding and Venables 2004, Head and Mayer 2006, Knaap 2006). Specifically, the instrument is the distance to New York City.[14]

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) and perform endogeneity tests on the suspect regressors.

Returning to the time dimension, its presence naturally calls for the use of panel data techniques. However, panel data estimation is done on pooled data, which assumes the same parameters over time and across regions. In our case, pooling the data across time might not be that innocent. Indeed, the period 1880-1920 is known

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

for dramatic changes in the U.S economy, which suggests a cautious approach to pooling the data across time. Consequently, a testing of poolability is carried out to see whether panel data techniques should be used or not (Baltagi, 2005).

As was mentioned in the section on the MK model, the equation (13) can be estimated either with all industry and state controls or with the state and industry dummies, as in Wolf (2007). All of the regressions discussed in the next section were estimated both ways and the differences in the magnitude and the statistical significance of the six interaction variables, which are the main variables of interest, are miniscule. We present the results using Wolf's (2007) specification; the results with the full set of industry and state controls are available from the authors upon request.

The Basic Results

The results of the initial estimation of equation (14) are in Tables 4-6. Table 4 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 market potential 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 4 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-statistic $F(23, 4465)$ is 27.2265 which enables us to reject the null hypothesis that $\beta[j]_t = \beta[j] \ \forall t$ at the 1% significance level. Accordingly, we run separate regressions for 1880, 1890, 1900, 1910, and 1920.

For each of those years, we have estimated equation (14) 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.[15] The results are presented in Table 5.

A general overview of the estimation results suggests that market potential interaction variables matter in each of the years, though some variation exists before 1900. The H-O interactions are less prevalent except for agriculture until 1910. Of the market potential interactions, the plant-size interaction is always statistically significant, usually at the 1% significance level. The backward-linkages interaction is almost always significant, except for 1890 and 1910. Forward linkages are first significant in 1890, after that, they remain significant until 1920. The H-O interactions are very different in terms of significance. The skilled-labor interaction changes signs and is insignificant for most of the time. The coal interaction is significant with a correct sign in 1890 only.[16] The agriculture interaction, on the other hand, is highly statistically significant between 1890 and 1910, before and after which it is insignificant though with the correct sign.[17]

The endogeneity issue regarding market potential and its interactions is addressed by instrumental-variable estimation. As was noted earlier, the instrument is the distance to an eastern seaport – the New York City. Instrumental-variable

---

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

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

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

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 IV/2SLS as well as 2-step GMM, which is more efficient than IV/2SLS. The differences in the magnitude and the statistical significance of the estimated coefficients are very small. Table 5 presents the results of IV/2SLS estimation; the results of 2-step GMM are available from the authors upon request.

For each year, we again estimate equation (14), 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 6 show a very strong correlation between the instruments and the instrumented variables, ranging from 0.74 to 0.78. We have also carried out a formal test of the weak instrument, as suggested by Stock and Yogo (2005). The relevant F-statistics largely exceed the critical values reported by Stock and Yogo (2005) --the F-statistics range from around 47 in 1880 to around 56 in 1920. Finally, the endogeneity test in Table 6 rejects the null hypothesis that the market potential and its interactions are exogenous.

The results in Table 6 show that overall the picture that emerges from Table 5 is preserved. The market potential interaction variables are generally significant and have the correct sign. The plant-size interaction is statistically significant except in 1920 and usually at the 5% significance level, slightly lower than in Table 5. The forward-linkage interaction is significant from 1890, and the significance rises by

1920. The significance of the estimated backward-linkages coefficients remains high throughout the period, except for 1890.

Robustness, Standardized Coefficients and the Economic Significance of the Results

We have also performed additional robustness checks.[18] First, as an alternative way to address endogeneity, we also re-estimated equation (14) 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.

We have also checked the robustness of the H-O interaction variables. Specifically, we have used the share of agricultural labor force (similarly to Crafts and Mulatu, 2006) instead of the share of farm land 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 4-6, with agriculture being the most prevalent among all H-O interaction variables.[20]

---

[18] Results are available from authors on request.

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

[20] The share of agricultural labor force is potentially endogenous too. Therefore, we have also considered agricultural labor force as endogenous and instrumented it with its lagged values. The sign and the statistical significance are the same as when they are treated as exogenous.

Overall, these results show the statistical importance of all the market potential 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 driven by both the agglomeration mechanisms related to market potential as well as natural advantages, though the former seems to prevail. We can support this inference by calculating standardized or so-called beta coefficients of all the interaction variables. The beta coefficients provide a comparison of the relative importance of the interaction variables in determining state shares of manufacturing employment by industry. The results reported in Table 7 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 market potential interactions, scale economies always have a substantial impact but it is noticeable that forward linkages become more important over time and that, by 1920, the contribution of linkages outweighs everything else.

To evaluate the economic significance of the market potential interaction variables and the interaction variables capturing the states' natural advantages, we follow Redding and Venables (2004) and examine their effect on the predicted share of manufacturing employment using counterfactual analysis. We use the estimated coefficients in our preferred specification (Table 5, Equation FE) to examine the impact of the changes in the geographical location of a U.S. state as well as the changes of the U.S. state's natural advantages. Specifically, the change of the U.S. state's geographical location is captured by the change of its market potential; the change of the natural advantages is captured by the changes in the share of the farm land and coal prices respectively. All of these changes are investigated in 1890 since it is the only year in which the coal price interaction has the correct sign and is statistically significant, allowing us to compare the agricultural interaction, the coal price interaction and the market potential interactions.

To quantify the importance of the proximity to large markets, we undertake two hypothetical experiments: first, we increase the market potential of four states which have very low market potential by 50%; second, we decrease the market potential of four states which have very high market potential by 50%. We see from Table 8, columns 1 and 2 that an increase of the market potential by 50% generates an increase of the state's share of the manufacturing employment in total U.S. manufacturing employment that ranges from 34 to 55 percent, and that a 50% decrease generates a decrease ranging from 41 to 45 percent. This means that, for example, lowering the market potential of the state of New York by 50%, which causes 44.8 percent drop in the New York's share of manufacturing employment in total U.S. manufacturing employment, decreases that share from 16.3% to 8.9%.

Similarly, we examine the effect of the change in coal prices and the share of farm land on the share of the state's manufacturing employment in total U.S. manufacturing employment by considering 50% increases or decreases. The results in Table 8, columns 3-6 show that the effects are smaller in comparison with the effects of market potential. For example, an increase of the share of farm land by 50% causes an increase of the state's share of manufacturing employment on the total U.S. manufacturing employment by between 0.04 and 0.16 per cent, and a decrease of coal prices by 50% results in a 0.14 to 0.37 percent increase in that share. As in the previous example, let us consider the state of New York. An increase in the price of coal by 50%, which lowers the state's share of manufacturing employment in total U.S. manufacturing employment by 0.11 percent, decreases that share from 16.3% to 16.2%.

In addition to these counterfactuals, we consider one more hypothetical experiment with the geographical characteristics of the U.S. states. Specifically, we relocate a land-locked U.S. state to the east coast. The relocation is done by changing the state's market potential. As was outlined in the section on the data set, the market potential of a U.S. state comprises the market potential of foreign countries, of the surrounding U.S. states, and that of itself. For example, relocating Nebraska to the east coast means that Nebraska will have higher market potential because it is now closer to foreign countries and is in close proximity to New England and Middle-Atlantic states with very high GDP.[21] The calculations for the relocation of Nebraska in 1890 show that being on the east coast increases Nebraska's share of manufacturing employment in total U.S. employment by 29 percent, which means that it increases from 1.1% to 1.42%.

Since the overall counterfactual change of market potential comes from three market potential interactions – forward linkages, backward linkages, and plant size – we can split the effect of that counterfactual change among those interactions to see their relative contribution to the resulting change in the share of a state's manufacturing employment on the U.S. total manufacturing employment. Table 9 presents the contribution of each of the market potential interactions to the change of the share of manufacturing employment in total U.S. manufacturing employment in the state of New York and California in 1890, respectively.

Let us consider the state of New York again. We see in column 1 that the share of New York's manufacturing employment in U.S. total manufacturing employment in 1890 is 16.27%. A counterfactual 50% decline of New York's market

[21] The counterfactual is what if Nebraska moved to Maryland (a state with very similar GDP per capita).

potential would result in a 44.8% decrease of that share, as we have seen in Table 8, column 2. This means that the state of New York loses 7.29 percentage points of that share and the resulting share is 8.98%. Columns 2 and 3 present the contribution of the market potential interactions to the percentage change, and to the loss or gain of the manufacturing shares. We see from column 2 that, for example, forward linkages decrease the manufacturing share by 24.61% out of the total 44.60%. Column 3 translates these percentage changes into the actual shares of manufacturing employment relative to the U.S. manufacturing employment. Following upon the previous example, column 3 tells us that out of 7.29 percentage points of manufacturing employment that are lost because of the counterfactual decrease of market potential, 4 percentage points are due to forward linkages.

In accounting for changes in a state's share of manufacturing employment as market potential changes, Table 9 reveals that forward linkages have about 2.7 times the impact of plant size which, in turn, has about 2.3 times the impact of backward linkages.

5. Discussion of the Results

The model that we have used for our empirical work maintains that the shares of manufacturing employment in each industry across states depend on input prices and the spatial distribution of demand.[22] Inputs include both primary factors and a composite intermediate good, the prices of which reflect factor endowments and proximity to suppliers, respectively. The model embodies the Rybczynski effect that an increase in the endowment of a factor raises output in the industries that use it

[22] As explained in section 2 above, input prices do not appear in the estimating equation but, using a 'dual' formulation, are represented by factor endowments and market potential

intensively but also allows proximity to suppliers of intermediates to raise output of industries that use intermediates intensively. The spatial distribution of demand has its effect through the attraction of market access that is driven by the geography of GDP and by transport costs.

Our results suggest that input prices and the spatial pattern of demand did indeed matter for industrial location at the turn of the twentieth century. The inputs that matter include intermediates and we find that forward-linkage effects were important. While factor endowments and market potential both influence industrial location, the latter was more important and its impact was felt both through linkage effects and the attraction of market access for sectors where plant size was relatively large.

The overall pattern of our results is consistent with the traditional accounts of industrial location reviewed in the introduction. They would not come as a great surprise to the authors of the report in National Resources Committee (1939) who estimated that, in 1935, 743 thousand manufacturing jobs were resource oriented compared with 887 thousand that were tied to local consumers and 6881 thousand that were 'footloose'. The strong showing of linkage effects matches the account given by Perloff et al. (1960).

On the other hand, the picture that we paint differs quite considerably from that sketched by Kim (1999). We believe that Kim's failure to take account of linkage effects is an important omission and has led him to exaggerate the role of factor endowments. Our model takes account of, and finds some evidence for, Rybcynski effects but makes Kim's claim that these are virtually the whole story seem implausible.[23]

It should also be noted that our emphasis on linkage effects also implies a different explanation for the persistence of the manufacturing belt from the one popularized by Krugman (1991a) (1991b). His account stressed the interaction between market potential and plant size in the context of transport costs reduced by the railroad together with many footloose producers. Our results suggest that this is not the main reason why the manufacturing belt persisted although it could perhaps play a key role in its establishment in the mid nineteenth century.[24]

Finally, it is important to note two caveats to our findings. First, our argument applies to the persistence of the manufacturing belt, which we believe was cemented by linkage effects, not its emergence. We do not have the data to test hypotheses about the latter. It may well be that in some cases the origins of an industrial cluster can be found in the direct or indirect effects of natural resources. In fact a case in point is automobiles. 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

---

[23] We allow for an additional factor endowment, human capital, which was not considered by Kim (1999) but this does not have a significant effect. The work of Goldin and Katz (1998) suggests this is not surprising in our period.

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

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 and 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.[25]

Second, it must be recognized that we have not estimated a fully-fledged NEG model of industrial location. We do not incorporate either monopolistic competition or non-monotonic relationships between transport costs and spatial concentration. So we have certainly not done full justice to the informal ideas in the traditional literature nor the formal treatments of writers like Krugman and Venables (1995). The prominence of linkage effects would surely be expected by these writers but, strictly speaking, we have not investigated their hypotheses.

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. This allows us to give answers to the questions that we posed at the outset.

As far as factor endowments are concerned, we find that the share of farmland in a state is the most important variable and had significant effects around the turn of

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

the twentieth century. Generally speaking, other factor endowments are insignificant. We find that market potential had a substantial impact on the location of manufacturing in the United States throughout the period 1880 to 1920, that it was much more important than factor endowments, and that the influence of market potential worked both through linkage effects and scale effects, especially the former. Our results suggest that market access and linkage effects were the central considerations that locked in the manufacturing belt and accounted 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.

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.

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).

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," *Journal of Economic History* 66 (2006), 575-607.

Davis R.D., and 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).

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

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

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, “Regional Wage and Employment Responses to Market Potential in the EU,” *Regional Science and Urban Economics* 36 (2006), 573-594.

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.

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,” Warwick Economic Research Paper No. 916, 2009.

Knaap, T., “Trade, Location, and Wages in the United States,” *Regional Science and Urban Economics* 36 (2006), 595-612.

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

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

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).

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

Santos Silva, J.M.C, and S. Tenreyro, “The Log of Gravity,” *The Review of Economics and Statistics* 88 (2006), 641-658.

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).

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

Ullman, E. L., *American Commodity Flow* (Seattle: University of Washington Press, 1957)

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.

Wolf, N., "Endowments vs. Market Potential: What Explains the Relocation of Industry after the Polish Reunification in 1918?", *Explorations in Economic History* 44 (2007), pp. 22-42.

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.

The Structure of the National Economy, Part 1: Basic Characteristics, Washington, Government Printing Office, 1939.

Carload Waybill Analyses, 1949, State-to-State Distribution of Manufactures and Miscellaneous and Forwarder Traffic. Interstate Commerce Commission, Bureau of Transport Economics and Statistics, Statement No. 5110, Washington, 1951.

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 they are expressed relative to the gross value of output. Whitney's input-output table provides a breakdown of the whole economy into twenty nine 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.

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: US Passenger Vehicle and Vehicle Parts Trade in 1949

Panel A: US States Exporting Passenger Vehicles to other US States

State	Carloads	%
California	491	30.29
Illinois	42	2.59
Indiana	129	7.96
Michigan	901	55.58
Ohio	57	3.52
Pennsylvania	1	0.06
Total	1621	100.00

Panel B: Imports of Vehicle Parts to Michigan

State	Carloads	%
Illinois	10	1.73
Indiana	54	9.33
Massachusetts	2	0.35
Michigan	204	35.23
Minnesota	1	0.17
Missouri	5	0.86
New Jersey	6	1.04
New York	40	6.91
Ohio	160	27.63
Tennessee	12	2.07
West Virginia	9	1.55
Wisconsin	76	13.13
Total	579	100.00

Sources: see text

Table 2.- 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 3. -Market Potential and the Rank of States Based on Market Potential in 1880 and 1920
Market Potential Estimates Based on δ = -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 4. - Pooled OLS, Panel Data Fixed Effect, 1880-1920
Forward and Backward Linkages based on 1899 Input-Output Table

	I POLS Cluster-Robust SE	II FE Panel Robust SE
	H-O Forces	
Agric. Employment x	0.002***	0.002***
agric. Input use	[0.0003]	[0.0003]
Educated pop. X	0.0008	0.0006
white-collar workers	[0.0006]	[0.0006]
Coal abundance x	0.006	0.009
steam power use	[0.02]	[0.02]
	Market Potential Forces	
Market potential x	0.00003	0.00003
interm. input use	[0.00009]	[0.00009]
Market potential x	0.00028***	0.00027***
industry sale	[0.00006]	[0.00006]
Market potential x	0.00013***	0.00013***
size of establishment	[0.00001]	[0.00001]
State-Industry 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 5.- OLS and Cluster-Specific Fixed Effect Estimations Year by Year
Forward and Backward Linkages based on 1899 Input-Output Table

	1880		1890		1900		1910		1920	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE	OLS	FE
					H-O Forces					
Agric. farm land x	0.0006	0.0006	0.0014**	0.0014**	0.0011**	0.001**	0.0009*	0.0009*	0.00009	0.00009
agric. Input use	[0.0005]	[0.0004]	[0.0006]	[0.0006]	[0.0005]	[0.0005]	[0.0005]	[0.0005]	[0.0004]	[0.0004]
Educated pop. x	0.006	0.0061	-0.001	-0.001	-0.021***	-0.02***	-0.0013	-0.002	0.001	0.001
white-collar workers	[0.004]	[0.0041]	[0.008]	[0.0078]	[0.005]	[0.005]	[0.004]	[0.004]	[0.001]	[0.001]
Coal abundance x	0.16**	0.16**	-0.17**	-0.17**	0.02	0.02	0.06	0.04**	0.01	0.01
steam power use	[0.08]	[0.07]	[0.08]	[0.079]	[0.05]	[0.05]	[0.069]	[0.02]	[0.03]	[0.03]
					Market Potential Forces					
Market potential x	0.0013	0.0013	0.003***	0.003***	0.002***	0.002***	0.0023***	0.0022***	0.0007***	0.0007***
interm. input use	[0.0009]	[0.0009]	[0.0001]	[0.0009]	[0.0007]	[0.0007]	[0.0006]	[0.0006]	[0.0002]	[0.0002]
Market potential x	0.002***	0.002***	0.0008	0.0008	0.0013**	0.0013**	0.0002	0.0002	0.00023**	0.00023**
industry sale	[0.0007]	[0.0007]	[0.0007]	[0.0007]	[0.0006]	[0.0005]	[0.0004]	[0.0004]	[0.00009]	[0.00009]
Market potential x	0.0009***	0.0009***	0.0008***	0.0008***	0.0007***	0.0007***	0.0004***	0.0004***	0.00005***	0.00005***
size of establishment	[0.0002]	[0.0002]	[0.0002]	[0.0002]	[0.0002]	[0.0002]	[0.00008]	[0.00008]	[0.00002]	[0.00002]
Constant	-3.79***	-3.55***	-1.05	-2.29	0.27	1.26	-3.35***	-3.5**	-2.17**	-2.7**
	[0.67]	[0.67]	[1.21]	[1.37]	[1.24]	[1.54]	[1.13]	[1.56]	[0.89]	[1.2]
State-industry dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	912	912	912	912	912	912	912	912	912	912
R-squared	0.75	0.46	0.7	0.45	0.67	0.48	0.64	0.46	0.62	0.36
Breusch-Pagan chi2(1)	304.4***		218.1***		293.3***		236.1***		257.1***	
F-test Joint Significance H-O	2.2*	3.1**	2.96**	3.24**	4.8**	7.7***	1.05	1.92	0.23	0.3

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) for OLS and FE

Table 6.- 2SLS Instrumental Variable Estimation
Forward and Backward Linkages based on 1899 Input-Output Table

	1880	1890	1900	1910	1920
			H-O Forces		
Agric. farm land x	0.0004	0.001*	0.0009*	0.0006	-0.00003
agric. Input use	[0.0005]	[0.0006]	[0.0005]	[0.0005]	[0.0004]
Educated pop. x	0.0056	-0.012	-0.02***	-0.002	0.001
white-collar workers	[0.004]	[0.008]	[0.006]	[0.004]	[0.001]
Coal abundance x	0.15**	-0.2**	0.024	0.07	0.014
steam power use	[0.07]	[0.08]	[0.05]	[0.07]	[0.04]
			Market Potential Forces		
Market potential x	0.0008	0.003**	0.002**	0.0019**	0.0007**
interm. input use	[0.001]	[0.001]	[0.0008]	[0.0009]	[0.0003]
Market potential x	0.0026***	0.001	0.002***	0.0008*	0.0003***
industry sale	[0.0009]	[0.001]	[0.0006]	[0.0005]	[0.0001]
Market potential x	0.0005*	0.0005**	0.0005***	0.0002**	0.00002
size of establishment	[0.0002]	[0.0002]	[0.0002]	[0.00009]	[0.00002]
Constant	-3.42***	-0.57	0.65	-2.75**	-2.09**
	[0.73]	[1.24]	[1.36]	[1.25]	[0.92]
State-industry dummy	Yes	Yes	Yes	Yes	Yes
Observations	912	912	912	912	912
R2	0.74	0.7	0.67	0.64	0.62
Shea Partial R2					
mp1vs2_intermed	0.75	0.76	0.76	0.76	0.76
mp1vs2_sale	0.74	0.76	0.75	0.78	0.76
mp1vs2_plant	0.77	0.77	0.78	0.77	0.76
Endog. C test [chisq (4)]	17.16***	13.32***	9.13*	17.4***	8.12*
Joint Significance					
Heckscher-Ohlin, chi2(3)	6.48*	8.4**	24.3***	3.55	0.89

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 7.- Beta Coefficients, Estimations Year by Year

	1880	1890	1900	1910	1920
			H-O Forces		
Agric. farm land x agric. Input use	0.05	0.09	0.09	0.07	0.008
Educated pop. x white-collar workers	0.035	-0.07	-0.24	-0.02	0.06
Coal abundance x steam power use	0.12	-0.09	0.02	0.05	0.01
			Market Potential Forces		
Market potential x interm. input use	0.06	0.19	0.17	0.33	0.25
Market potential x industry sale	0.13	0.06	0.15	0.04	0.1
Market potential x size of establishment	0.27	0.27	0.36	0.42	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 OLS regressions in Table 4.

Sources: see text

Table 8: Economic importance of market potential, coal prices and farm land in U.S. states' shares of manufacturing employment in 1890 (percentage change).

U.S. State	Change of market potential (50%)		Change of share of farm land (50%)		Change of coal prices (50%)	
	Increase	Decrease	Increase	Decrease	Decrease	Increase
	(1)	(2)	(3)	(4)	(5)	(6)
Nebraska	55.22		0.162		0.21	
Utah	37.95		0.009		0.14	
California	34.43		0.078		0.32	
Washington	33.66		0.036		0.37	
Illinois		-40.85		-0.310		-0.15
New York		-44.82		-0.263		-0.11
Ohio		-42.14		-0.326		-0.13
Pennsylvania		-44.47		-0.233		-0.03

The table reports the predicted effect of a change in geographical and economic characteristics of the U.S. states on their share of manufacturing employment in total U.S. manufacturing employment. The predicted effects are based on the estimated coefficients for the year 1890 in Table 4, equation FE. The relocation of a U.S. state is implemented by changing its market potential.

Table 9: Contribution of forward linkages, backward linkages, and plant size to the change of the share of a U.S. state manufacturing employment in U.S. total manufacturing employment due to the counterfactual change of market potential in 1890.

		Contribution of linkages and plant size to change in manufacturing employment		
			Contribution to percentage change	Contribution to the loss or gain of U.S. state manufacturing employment/ U.S total manufacturing employment as a result of counterfactual
	(1)		(2)	(3)
State of New York (counterfactual decline of NY's market potential by 50%)				
The share of NY's manufacturing employment in U.S total manufacturing employment in 1890	16.27	Forward linkages	24.61	4.00
Decline of share of NY's manufacturing employment in U.S total manufacturing employment as a result of counterfactual (percentage change)	44.80	Backward linkages	6.07	0.99
The share of NY's manufacturing employment in U.S total manufacturing employment that is lost as a result of counterfactual	7.29	Plant size	14.15	2.30
The share of NY's manufacturing employment in U.S total manufacturing employment as a result of counterfactual	8.98	Total	44.80	7.29
California (counterfactual increase of California's market potential by 50%)				
The share of California's manufacturing employment in U.S total manufacturing employment in 1890	2.27	Forward linkages	18.90	0.43
Increase of share of California's manufacturing employment in U.S total manufacturing employment as a result of counterfactual (percentage change)	34.43	Backward linkages	4.66	0.11
The share of California's manufacturing employment in U.S total manufacturing employment that is gained as a result of counterfactual	0.78	Plant size	10.87	0.25
The share of California's manufacturing employment in U.S total manufacturing employment as a result of counterfactual	3.05	Total	34.43	0.78

Sources and notes: Column 1: the percentage change of a state's share of manufacturing employment in U.S. total manufacturing employment is from Table 7, columns 1 and 2; the sources of the manufacturing employment shares in 1890 are in the appendix.
Column 2: the figures are calculated using the same procedure as in Table 7. An example for the state of New York shows the calculation of the remaining figures: column 1: 7.29=16.27*(0.448); 8.98=16.27-7.29; column 3 forward linkages: 4.0=7.29*(24.61/44.8).