Working Paper Series

Competitors, Complementors, Parents and Places: Explaining Regional Agglomeration in the U.S. Auto Industry

WP 13-04R

Luís Cabral
New York University and CEPR

Zhu Wang
Federal Reserve Bank of Richmond

Daniel Yi Xu
Duke University and NBER

THE FEDERAL RESERVE BANK OF RICHMOND
RICHMOND ▪ BALTIMORE ▪ CHARLOTTE

Competitors, Complementors, Parents and Places: Explaining Regional Agglomeration in the U.S. Auto Industry

Luís Cabral
New York University and *CEPR*

Zhu Wang
Federal Reserve Bank of Richmond

Daniel Yi Xu
Duke University and NBER

Working Paper No. 13-04R

This draft: August 2014

Abstract. Taking the early U.S. automobile industry as an example, we evaluate four competing hypotheses on regional industry agglomeration: intra-industry local externalities, inter-industry local externalities, employee spinouts, and location fixed-effects. Our findings suggest that inter-industry spillovers, particularly the development of the carriage and wagon industry, play an important role. Spinouts play a secondary role and work as a special type of intra-industry spillovers. The presence of other firms in the same industry has a negligible (or even negative) effect. Finally, local inputs account for some agglomeration in the short run, but the effects are much more profound in the long run.

JEL classification: L26; L6; R1
Keywords: Local externalities; Spinouts; Resource advantages; Industry agglomeration

---

Cabral: Paganelli-Bull Professor of Economics and International Business, Stern School of Business, New York University; IME and PS-PS (IESE) Research Fellow; and Research Fellow, CEPR (London); luis.cabral@nyu.edu. Wang: Research Department, Federal Reserve Bank of Richmond; zhu.wang@rich.frb.org. Xu: Department of Economics, Duke University; daniel.xu@duke.edu.

We thank Ed Glaeser, Boyan Jovanovic, Bob Lucas, Esteban Rossi-Hansberg and participants at the NBER Summer Institute, the 2012 AEA meetings, the 2012 SED meeting and various seminars for helpful comments; and Christian Hung and Roisin McCord for excellent research assistance. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

1. Introduction

Why does economic activity concentrate in cities and more generally in clusters of innovation and growth? Ever since Marshall's (1890) seminal work, an intense debate has developed among economists regarding agglomeration externalities. Marshall himself pointed to the positive externalities from specialization: regions with specialized production structures tend to be more innovative in that specific industry. In particular, Marshall pointed to the importance of knowledge spillovers: each firm learns from neighboring firms in the same industry.

> Employees from different firms in an industry exchange ideas about new products and new ways to produce goods: the denser the concentration of employees in a common industry in a given location, the greater the opportunity to exchange ideas that lead to key innovations.

Marshall's work was later extended by the work of many authors, including Arrow (1962) and Romer (1986).[1] (Knowledge spillovers are sometimes referred to as MAR spillovers.) We should note that, in addition to knowledge spillovers, Marshall also considered other types of externalities, including input sharing and labor pooling. However, as Ellison, Glaeser and Kerr (2010) argue, all of these "predict that firms will co-locate with other firms in the same industry." Accordingly, we refer to Marshallian externalities as economic effects that lead firms to locate close to other firms of the same industry: intra-industry externalities.[2]

In contrast to Marshallian externalities, other authors, most notably Jacobs (1969), proposed an alternative agglomeration thesis, the idea that knowledge spills across different industries, causing diversified production structures to be more innovative. For example, Jacobs (1969) argues that the growth of Detroit's automobile industry may owe a great deal to the prior growth of Detroit's shipbuilding industry.[3] We refer to these externalities as inter-industry, or related-industry, externalities (as opposed to intra-industry, or Marshallian, externalities).[4]

A different perspective on agglomeration is provided by the work of Klepper (2007), who focuses on the role of employee spinouts. He argues that "the agglomeration of the automobile industry around Detroit, Michigan is explained [by] disagreements [that] lead employees of incumbent firms to found spinoffs in the same industry."[5] The effect of spinouts on agglomeration is related to Marshallian externalities in the sense that the number of firms in a given industry is subject to self-reinforcing dynamics: the more industry i firms are located at location j, the more likely new industry i firms will be located at location j.

---

[1] See also Krugman (1991), Glaeser et al (1992) and Henderson et al (1995).

[2] It may be argued that we are taking a very narrow view of Marshallian externalities by restricting to other firms within the same industry. In fact, we will next consider the possibility of externalities across related industries. Our goal is to present an account of factors affecting industry agglomerations that is as detailed as possible. For this reason, we believe it is helpful to make this distinction.

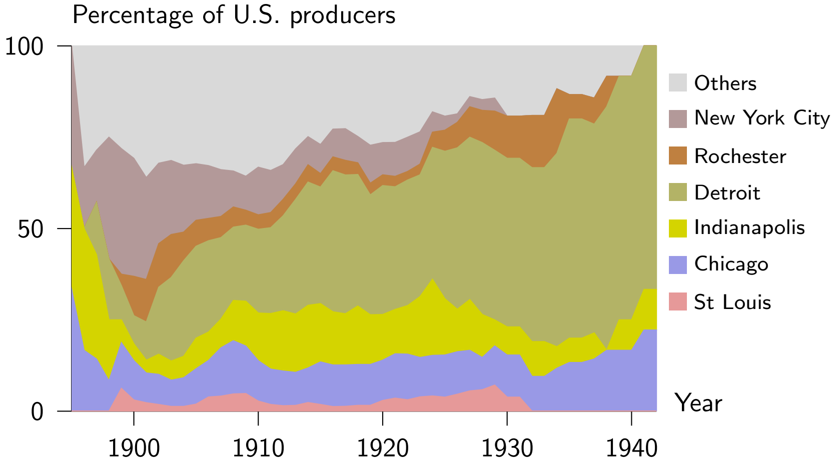
[3] See also Jackson (1988).

[4] These externalities are sometimes referred to as co-agglomeration externalities. Some authors (e.g., Ellison, Glaeser and Kerr, 2010) refer to agglomeration and co-agglomeration externalities as Marshallian externalities. However, we believe the distinction between the two to be relevant. In fact, it corresponds to a central piece of our theoretical and empirical exercise.

[5] Buenstorf and Klepper (2009) paint a similar picture for the tire industry. Note that, while Klepper (2007) uses the term "spinoffs," other authors including ourselves use the term "spinouts."

Figure 1

Geographical distribution of U.S. auto producers



However, according to Klepper, the mechanism for these self-reinforcing dynamics is quite different from Marshallian externalities; rather, it results from organizational reproduction or a "heredity" effect.

Finally, the work of Ellison and Glaeser (1997, 1999) suggests that much of the agglomeration observed in U.S. manufacturing may be due simply to the relative advantage of certain locations, such as the availability of natural resources. For example, the wine industry is located in California, not Kansas, largely due to California's favorable weather.[6]

In this paper, we construct a detailed dataset of the evolution of the U.S. auto industry and run a "horse race" between alternative views of agglomeration: (a) intra-industry spillovers (Marshall et al); (b) related-industry spillovers (Jacobs et al); (c) family network, or spinout, effects (Klepper et al); (d) location fixed effects (Ellison and Glaeser).

Our analysis is motivated by Figure 1, which shows the evolution of U.S. auto production measured by each region's share of the total number of firms. The dataset that we constructed uncovers six historically important auto production centers: New York City, Chicago, Indianapolis, Detroit, Rochester, and St. Louis. This raises an important question: Why did these six locations rather than any other places become prominent auto production centers? The figure also shows that the relative importance of Detroit as a production center increased steadily since the beginning of the century. The auto industry eventually became highly concentrated in Detroit. Why?

Our results provide answers to the above questions and also refine the existing explanations of industry agglomeration. In contrast to Marshall's hypothesis, we find that the proximate presence of firms from the same industry (competitors) has a negligible or even negative effect on firm performance; we interpret this result as implying that the negative competition externality outweighs Marshallian positive spillovers. Second, confirming the ideas of Jacobs (1969) and others, we find evidence for positive spillover effects from related industries located nearby. In particular, the presence of carriage and wagon firms (but not shipbuilding firms) explains why the six auto production centers that we identified

---

[6] Ellison and Glaeser (1997, 1999) also consider the possibility of "spurious" agglomeration due to non-economic motives.

emerged in the first place. Thirdly, consistent with the results in Klepper (2007), we find that spinouts play an important role on regional agglomeration, which led to the increased concentration in Detroit over time. Moreover, the performance of spinouts is heavily affected by local family members but not distant ones, which suggests that spinouts benefit from intra-family local spillovers rather than gene reproduction. This finding bridges the gap between the views of Klepper and Marshall. Finally, consistent with the work of Ellison and Glaeser (1997, 1999), we also find significant location specific effects, particularly due to local input resources. In fact, our long-run analysis suggests that local inputs are quite important, both directly and through the location of related industries.

To put it differently, our estimates suggest that auto companies concentrated in certain regions mainly because of the high density of firms in related industries, especially the carriage and wagon (C&W) industry. Considering that a large number of auto firm founders had prior experience working in the carriage and wagon industry, the C&W channel may have fostered the new auto industry primarily through the human capital channel. Similarly, existing auto firms fostered the entry of spinouts, which then reinforced the agglomeration pattern. Finally, the abundant supply of local inputs, especially iron and lumber, was also important, both directly (as inputs to the auto industry) and indirectly (as inputs to the carriage and wagon industry). In fact, the overall (that is, long-run) importance of local inputs is quite significant; in this sense, the auto industry is not that different from the wine industry.

Our paper's main contribution is to evaluate the relative importance of the four views outlined earlier, that is, the relative effect of competitors, complementors, parents and places. This is possible for three reasons. First, we create a fairly complete dataset of the U.S. auto industry that allows us to evaluate the strength of each effect by means of reduced form estimation.[7] Second, we develop a dynamic structural model that, by means of calibration and counterfactual simulation, allows us to quantify the relative contribution of each theory of industry agglomeration. Third, by identifying — from input-output matrices — the key inputs to co-agglomerating industries, we are able to tease out the direct and indirect effects of local input availability, thus uncovering an important distinction between short-run and long-run analysis.

Quantitatively, we propose the following agglomeration accounting exercise: we first identify factors that contribute to regional agglomeration of the auto industry using reduced-form regressions. Based on those findings, we construct and calibrate our structural model to fit the data (the fit is very good). We then conduct a series of counterfactual simulations where a selected feature of the model (e.g., spinouts) is shut off. We then take the drop in goodness of model fit as that feature's (e.g., spinouts) contribution to the overall fit. Finally, we compute the fraction that each feature represents of the total contributions of "goodness of fit" as a measure of its relative importance.

---

[7] Other authors, most notably Klepper (2007), have also looked at the evolution of the U.S. auto industry. In addition to data on entries, exits and spinouts, we also obtained data on related industries and on the most significant inputs to the auto industry. (To the best of our knowledge, ours is the first paper to work with data of this nature.) Moreover, unlike Klepper (2007), who only examines the evolution of Detroit as a production center, we identify multiple production centers and look at the entire geographic distribution of the U.S. auto industry.

As mentioned earlier, we find the net contribution of competitors to be essentially zero.[8] We also find that the relative contribution of the remaining three channels — complementors, parents and places — depends on the time frame. In the short-run — that is, given location decisions in related industries — the complementors channel seems to win the "horse race." In the long-run — that is, considering the endogeneity of location decisions — location fixed-effects channel win the "horse race." Spinouts also play an important role. The values we obtain are: (a) taking C&W locations as given: inter-industry spillovers, 41%; spinouts, 29%; local inputs, 22%; other local factors, 8%; (b) taking C&W locations as endogenously determined by location conditions: inter-industry spillovers, 13%; spinouts, 34%; local inputs 50%; other local factors 2%.

Ours is not the first attempt at estimating the relative contribution of alternative agglomeration theories. Ellison, Glaeser and Kerr (2010) "exploit patterns of industry coagglomeration to measure the relative importance of different theories of industry agglomeration." In particular, they are interested in teasing out the relative importance of the three Marshallian channels: movement of goods, movement of people and movement of ideas.[9] Our agglomeration accounting exercise complements theirs: we do not focus on the different channels proposed by Marshall;[10] by contrast, we pay close attention to the distinction between intra and inter-industry spillovers. We also consider explicitly the effect of spinouts. Finally, developing a calibrated dynamic structural model allows us to propose a natural agglomeration accounting procedure, as well as to distinguish between short-term and long-term effects.

The rest of the paper is structured as follows. In Section 2, we briefly describe the evolution of the U.S. auto industry. Next, in Section 3 we run a series of reduced-form regressions that test the relative merit of various theories of industry agglomeration. The results motivate Section 4, where we develop and calibrate a dynamic structural model of industry evolution and run a series of counterfactual simulations that allow us to quantify the relative contribution of each agglomeration factor. Section 5 concludes the paper.

2. The U.S. automobile industry

The U.S. auto industry went through tremendous development in its first 75 years, evolving from a small and fragmented infant industry into a gigantic, consolidated triopoly. During this process, the number of firms initially rose and later fell: in its peak years around 1910, there were more than 200 producers, but only 8 survived into the 1940s. Figure 2 documents this evolution, plotting the number of active firms each year, as well as the number of entrants and exiters. As can be seen, two industry "shakeouts" took place, one

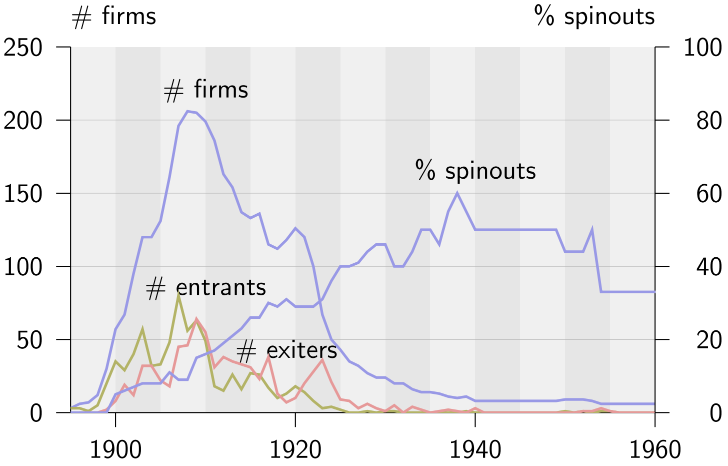
---

[8] Based on the results from reduced-form regressions, we set intra-industry spillovers to zero in our structural model.

[9] Summarizing Marshall's points, Ellison, Glaeser and Kerr (2010) write: "First, he argued that firms will locate near suppliers or customers to save shipping costs. Second, he developed a theory of labor market pooling to explain clustering. Finally, he began the theory of intellectual spillovers by arguing that in agglomerations, 'the mysteries of the trade become no mystery, but are, as it were, in the air.' "

[10] As Ellison, Glaeser and Kerr (2010) rightly point out, "each Marshallian theory predicts that the same thing will happen for similar reasons: plants will locate near other plants in the same industry because there is a benefit to locating near plants that share some characteristic." Their results "support the importance of all three Marshallian theories and the importance of shared natural advantages."

Figure 2
Evolution of the U.S. automobile industry, 1895–1969.



around 1910 and a second one around 1920.

Figure 2 also plots the percentage of active firms that were spinouts, that is, firms that were founded by a former manager or employee of an existing auto firm. This percentage increased from almost zero in 1900 to about 60 percent in 1940.

The auto industry also went through substantial changes in geographic concentration over the years. Figure 3 shows the number of auto firms in each U.S. city over a period of 25 years (from 1900 to 1925). Based on this picture, we identified six historically important auto production centers: St. Louis, Chicago, Indianapolis, Detroit, Rochester, and New York City.[11] As shown in Figure 1, New York City and Chicago were the most important centers in the late 1890s. Soon after, Detroit and other centers caught up. By 1905, 25 percent of all active firms were located in Detroit and they produced more than 50 percent of total industry output. Meanwhile, 16 percent of the firms were located in New York City, 10 percent in Chicago, 8 percent in Indianapolis, 7 percent in Rochester, 2 percent in St. Louis, with the remaining 32 percent scattered across the country. Over time, Detroit gained an increasing share, both in terms of the number of firms and in terms of industry output. By 1920, 35 percent of all active firms were located in Detroit, producing about 70 percent of total industry output.

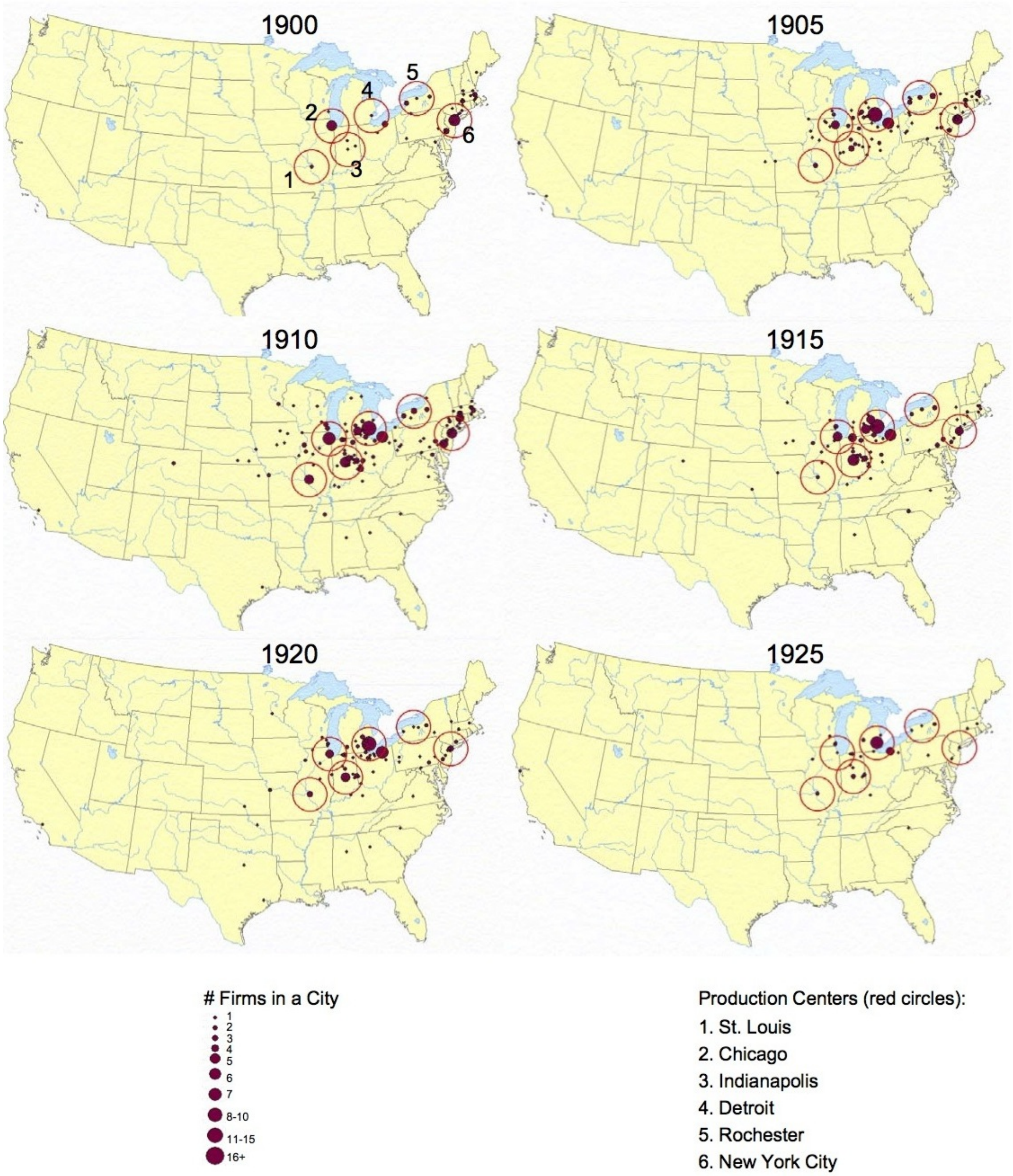
As mentioned earlier, an important fraction of the industry entrants originated in other existing industry firms: a spinout. We will refer to the firm originating the spinout as the "parent" and the spinout firm as a "child." Sometimes, the "child" itself becomes a parent by originating a spinout. Together, spinouts give rise to "families" of auto firms, that is, groups of firms linked together by spinout relationships.

We identified a total of 53 spinout families over the history of the auto industry. The three largest families were GM/Buick, Ford and Oldsmobile, all located in Detroit, each

---

[11] Lacking good data on firm size, we instead use the number of firms as a measure of regional agglomeration. A city is counted as an auto production center city if it had at least five auto producers in 1910 (the peak year of the auto industry in terms of firm numbers). We then define the region within 100 miles of the center city as the production center, named after the center city (we tried different radiuses ranging from 25 miles to 150 miles for the center definition, and the 100 mile radius appears to provide the overall best fit for the data).

Figure 3
Geographic concentration of U.S. auto producers (1900-1925)



generating 12–17 spinouts.[12] As an example, Figure 4 displays the GM/Buick family tree. As can be seen, it's a family with three generations: for example, a former GM employee

---

[12] Klepper (2007) constructed family trees for GM/Buick, Ford, Oldsmobile and Cadillac. The family members that he identified are largely consistent with ours.

Figure 4
GM/Buick's family tree

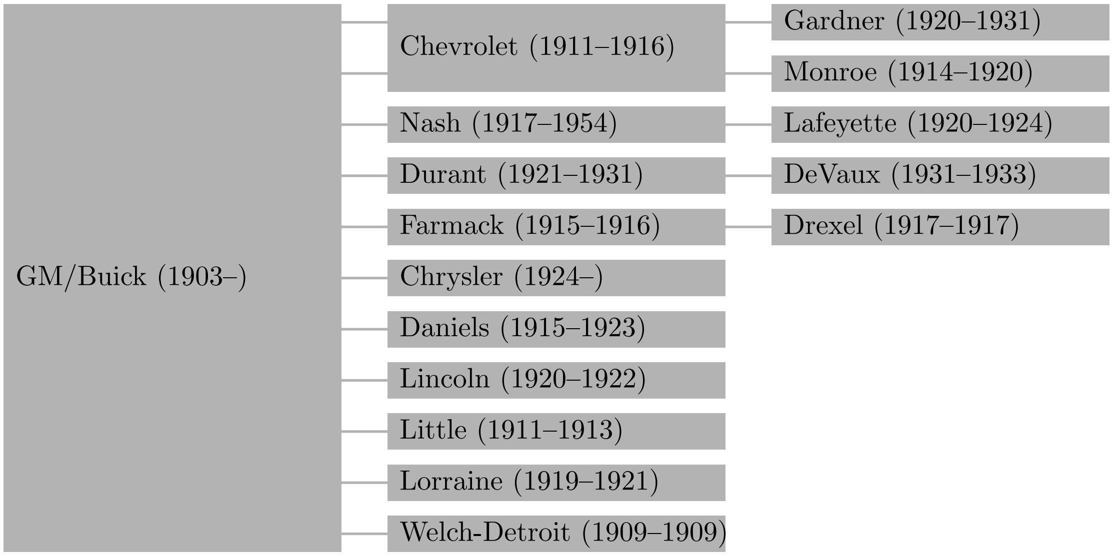
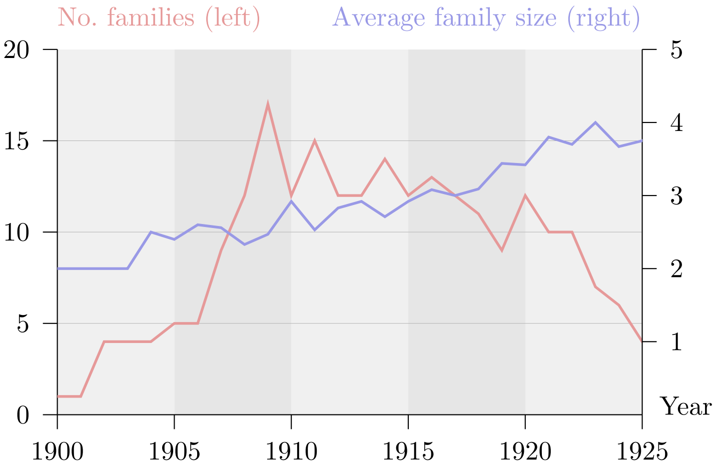


Figure 5
Family size distribution of U.S. auto producers (1900–1925)

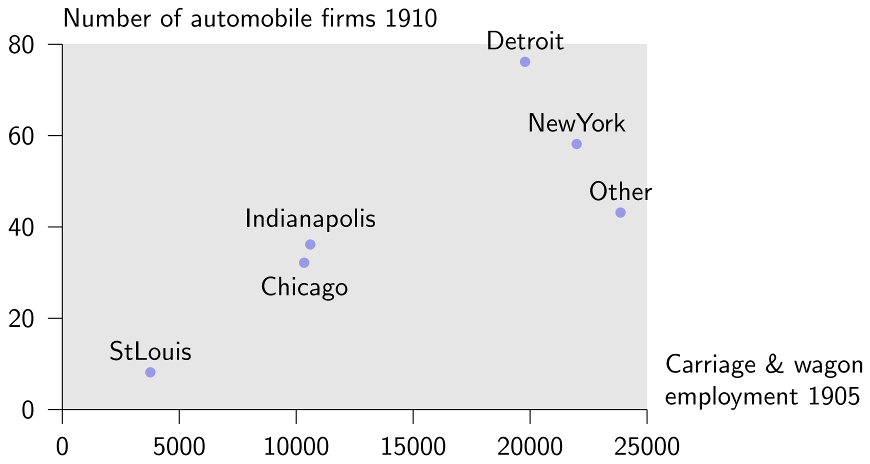


founded Chevrolet, from which in turn Gardner and Monroe spun out.

Figure 5 plots the evolution of the number of spinout families and of family size. Early on, there were very few spinouts. For example, in 1900 only one spinout family existed (which had two members including the parent), out of a total of 57 firms in the industry. In the following two decades, the period of greatest industry turbulence (that is, highest entry and exit rates), the number of families was somewhere between 10 and 15, whereas average family size was somewhere between 3 and 4. By 1920, 41 out of a total of 136 firms belonged to spinout families. Most spinouts located near their parents. For example, 76 percent of the spinouts in the top three families stayed in Detroit.

Although our analysis focuses on the auto industry, there are related industries which

Figure 6
Carriage and wagon industry



play an important role in explaining entry and exit patterns by auto firms. Prominent among these is the carriage and wagon industry (C&W). Figure 6 plots the level of activity in the auto industry (measured by the number of firms in 1910, the peak year of the auto industry in terms of firm numbers) against the C&W industry (measured by C&W employment level in 1905).[13] As can be seen, there is a clear positive correlation between the two. Obviously, at this stage there is little more to be said other than the fact that there is a correlation. Below we explore this relation in greater detail, and the results suggest that location patterns of auto firms (the new industry) may indeed have been influenced by location patterns of C&W firms (the older industry). Similarly, we also examine the influence of the shipbuilding industry, which turns out to be quite insignificant.

Why is the C&W industry important in studying the auto industry? Anecdotal evidence suggests that many auto firms were founded by experienced C&W veterans. One of the prominent figures is William C. Durant, the founder of GM. Before entering the auto industry, he was running the Durant-Dort Carriage Company, based in Flint, Michigan, which was the largest manufacturer of horse-drawn vehicles in the nation at the time. Therefore, it is natural to think that the presence of the C&W industry may have fostered the agglomeration of the new auto industry by providing the human capital that the latter needed.

Besides the above channel, however, it is likely that other factors could also have contributed to the co-agglomeration between C&W and auto. For example, at the time both industries relied heavily on common inputs, such as iron and lumber.[14] Figure 7 plots the concentration of the auto industry against the index of local auto-related input resources.[15] There is also a clear positive correlation. In the following analysis, we will try to tease out

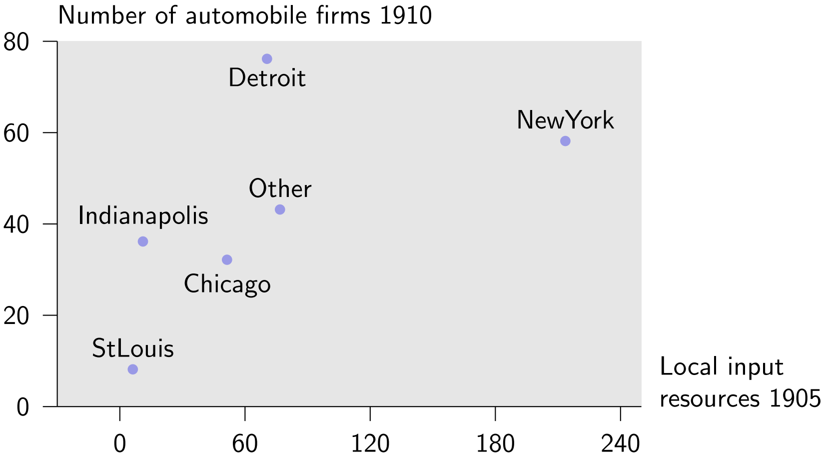
---

[13] Given that only state-level data are available for the C&W industry employment, we combine New York City and Rochester together as New York in the figure. We also add up all other non-center states in the "other" group.

[14] According to *Census of the U.S. Manufactures 1905* and Leontief (1951), iron and steel, lumber and timber, brass and copper, and rubber were top inputs for both industries at the time.

[15] The index of local auto-related input resources is constructed using historical data and the input-output matrix made by Leontief for the U.S. economy in 1919. See Section 3 for more details.

Figure 7
Local input resources



various causal effects, and show that the C&W industry and local inputs both affected the auto industry agglomeration but through different channels.

3. Reduced-form regression analysis

In this section we present a series of reduced-form regressions that provide tests of various theories of industry agglomeration. Marshallian theories imply that a firm's benefit from locating in region i is increasing in the number of other firms located in region i. Everything else constant, we would expect this to be reflected in entry rates and exit rates: entry rates are increasing in the number of firms, whereas exit rates are decreasing in the number of firms in region i.

Regarding co-agglomeration economies, our tests are based on data regarding the importance of related industries, including the carriage and wagon industry and the shipbuilding industry. The theory prediction is that the presence of related industries improves a firm's prospect. We thus expect entry (respectively, exit) rates to be increasing (respectively, decreasing) in the presence of related industries.

As shown by Klepper and others, spinouts are an important factor in the development of a new industry, especially at a second stage of industry evolution (by definition, the first entrant cannot be a spinout). Spinouts per se do not imply agglomeration: if every incumbent firm is equally likely to generate a spinout, then the fraction of industry firms accounted for by region i does not change as a result of spinouts. The question is then whether spinout rates vary systematically from region to region or as a function of region size.

Finally, as Ellison and Glaeser (1999) pointed out, location fixed-effects could also be important. Accordingly, we include regional population, per capita income, various local input resources, and location dummies in our analysis.

■ **Data.** Our data comes from various sources. First, Smith (1970) provides a list of every make of passenger cars produced commercially in the United States from 1895 through

Table 1
Firm level summary stats

Variable	Obs	Mean	St Dev	Min	Max
De novo entrant	771	0.30	0.46	0	1
De alio entrant	771	0.52	0.50	0	1
Spinout entrant	771	0.17	0.38	0	1
Top firm	771	0.06	0.24	0	1
Entry year	771	1908	6	1895	1939

Table 2
Firm-year level summary stats

Variable	Obs	Mean	St Dev	Min	Max
Firm age	4454	6.87	7.24	1	43
Firm exit	4454	0.17	0.38	0	1
Spinout birth	4454	0.02	0.13	0	1
Family size	4454	1.53	1.52	1	10
Family top	4454	.47	1.04	0	5
Local family size	4454	1.37	1.25	1	9
Local family top	4454	.41	.94	0	5
Non-local family size	4454	.16	.70	0	9
Non-local family top	4454	.07	.39	0	5
Center size	4454	35.79	23.43	1	96
Center top	4454	6.03	5.55	0	18

1969. The book lists the firm that manufactured each car make, the firm's location, the years that the car make was produced, and any reorganizations and ownership changes that the firm underwent. Smith's list of car makes is used to derive entry, exit and location of firms.[16]

Second, Kimes (1996) provides comprehensive information for every car make produced in the U.S. from 1890 through 1942. Using Kimes (1996), we are able to collect additional biographical information about the entrepreneurs who founded and ran each individual firm. An entrepreneur is categorized into one of the following three groups: de novo, de alio, or spinout entrants. De alio entrants are firms whose founder had prior experience in related industries before starting an auto firm. Spinouts are firms whose founders previously worked as managers or employees in existing auto firms. Finally, de novo entrants includes all other

---

[16] The entry and exit are based on the first and last year of commercial production.

entrants, those firms whose founders had no experience in either auto or related industries. Kimes' information is also used to derive family linkages between individual firms. In other words, we construct family trees for spinout firms.

The third data source is Bailey (1971), which provides a list of leading car makes from 1896–1970 based on annual sales, specifically, the list of top-15 makes. We use this information to identify top auto producers during this period.

Additionally, we collect information on location specific variables for the 48 U.S. continental states in the early 1900s. These data come from Easterlin (1960) and various issues of the *Census of the U.S. Manufactures* and the *Statistical Abstract of the United States*, which include population and per capita income in 1900, carriage and wagon industry employment, shipbuilding industry employment, local production of iron and steel, brass and copper, lumber and timber, and rubber in 1905. Given that auto was a small infant industry at the time, these location specific variables in the early 1900s can be treated as exogenous to the development of auto industry.

■ **Regression variables and descriptive statistics.** Our dataset includes every U.S. company that ever sold at least one passenger car to the public during the first 75 years of the industry (1895–1969), a total of 775 firms. In our following reduced-form regression analysis, we set our sample range up to the time of U.S. entry into WWII (1895–1942), which includes 771 firms.

Tables 1 through 3 provide summary statistics of the variables used in the regressions. Table 1 includes firm-level variables, as follows:

- De novo entrant. Equals 1 if the firm's founder had no experience in the auto or related industries. 30% of all entrants were de novo entrants.
- De alio entrant. Equals 1 if the firm's founder had previous experience in an auto-related industry, such as carriage and wagon. 52% of all entrants were de alio entrants.
- Spinout entrant. Equals 1 if the firm's founder had previous experience in the auto industry. 17% of all entrants were spinout entrants.
- Top firm. Equals 1 if, at any point during its life, a firm was one of the top car producers, as classified by Bailey (1971). Only 6% of the 771 firms fall into this category.
- Entry year. First year when the firm started commercial production. Varies from 1895 to 1939, with an average of 1908.

We next turn to firm-year level variables. We define location dummies corresponding to St. Louis, Chicago, Indianapolis, Detroit, Rochester, New York City and the others. The summary statistics of firm-year level variables are listed in Table 2.

- Firm age. Difference between current year and entry year. It ranges from 1 to 43 in our sample. The average is 6.87 years, not very different from what is found in other industries.
- Firm exit. Equals 1 if the firm stops commercial production during the current year. The average 0.17 corresponds to a hazard rate somewhat higher than that found in other industries, but one must remember that we are looking at the initial stages of a new industry, where entry and exit rates are typically higher.

Table 3
Region level summary stats

Variable	Obs	Mean	St Dev	Min	Max
Population	43	1763.2	2716.7	43	16390
Per capital income	43	123.2	55.6	54	281
C&W employment	43	1852.9	4127.3	0	19245
Shipbuilding employment	43	1237.8	3478.8	0	20854
Iron and steel	43	50.6	178	0.04	1100
Brass and copper	43	2.3	11.1	0	72.2
Lumber and timber	43	13.5	15.1	0	53.1
Rubber	43	1.5	4.8	0	24.2
Local input resources	43	9.989	34.411	0.010	213.557

- Spinout entry. Equals 1 if a firm generates a spinout entrant in current period, that is, a firm employee founds a new firm in the auto industry. The average spinout birth rate is about 2%.
- Family size. Number of firms belonging to the firm's family (including itself) in the current period. On average, a firm belongs to a family of 1.53 firms; the minimum is 1 and the maximum 10.
- Family top. Number of top firms belonging to the firm's family (including itself) in the current period. On average, there are .47 top firms in a firm's family; the minimum is zero and the maximum 5.
- Local family size. Number of firms belonging to the firm's family (including itself) in the current period that are located in the same region as the firm in question. The average is 1.37, a little lower than 1.53 firms, suggesting that a firm typically locates close to its family.
- Local family top. Number of top firms belonging to the firm's family (including itself) in the current period that are located in the same region as the firm in question.
- Non-local family size. Number of firms belonging to the firm's family (including itself) in the current period that are not located in the same region as the firm in question.
- Non-local family top. Number of top firms belonging to the firm's family (including itself) in the current period that are not located in the same region as the firm in question.
- Center size. Number of firms in a given location and year. It varies from 1 to 96 and has an average of about 36 firms.
- Center top. Number of top firms in a given location and year. It varies from 0 to 18 and has an average of about 6 firms.

Finally, we have the following region-level variables. Given that only state-level data are available, we group the 48 U.S. continental states into 5 production centers (some centers cover multiple states) and 38 non-center regions: St. Louis, Chicago, Indianapolis, Detroit, New York (combining New York City and Rochester), and 38 other states. The summary statistics are listed in Table 3:

- Population. Regional population (thousands) in 1900.
- Per capita income. Regional per capita income in 1900.
- C&W employment. Number of workers in the C&W industry in a given region in 1905.
- Shipbuilding employment. Number of workers in the shipbuilding industry in a given region in 1905.
- Iron and steel. Production of iron and steel (million dollars) in 1905.
- Brass and copper. Production of brass and copper (million dollars) in 1905.
- Lumber and timber. Production of lumber and timber (million dollars) in 1905.
- Rubber. Production of rubber (million dollars) in 1905.
- Local input resources (million dollars). According to the first input-output table made by Leontief for the U.S. economy in 1919, iron and steel, brass and copper, lumber and timber, and rubber were the four most important inputs used in auto production at the time (Leontief, 1951). We then calculate an index of local auto-related input resources, which is the weighted sum of the abundance of the four inputs in a location (measured by the location's production value of each input in 1905) using each input's cost share in auto production as its weight, where cost shares are provided by Leontief's input-output table (1919).

■ **Marshall vs others: entry by non-spinout firms.** Table 4 presents our first set of regressions where the dependent variable is the number of non-spinout entrants in location i in a given year. The sample range is 1900-1910 and we consider 43 different locations: 5 production centers and 38 other states, for a total 473 location-year observations.[17] All regressions are based on a negative binomial model. The different specifications refer to different sets of independent variables.

The first regression directly tests an implication of Marshall's hypothesis, namely that auto companies are attracted to locations with other auto companies. To avoid simultaneity issues, we lag the variable Center Size by one year. As can be seen, the coefficient is statistically not significant — and indeed has the opposite sign of what Marshall would predict. Moreover, several location dummies are statistically and economically significant. In particular, entry rates in Chicago, Indianapolis, New York, and Detroit are higher than the non-center average.

While the first regression suggests that there is agglomeration, it does not explain why. In order to get a better feel for the nature and causes of agglomeration, in the next four regressions we add five different variables: population, income, and input resources, three different measures of "exogenous" location characteristics; and employment in two related industries: carriage and wagon; and shipbuilding.

---

[17] We choose the sample range 1900-1910 to be consistent with our structural calibration in Section 4.

Table 4
Negative binomial models of non-spinout entry, 1900-1910.
Dependent variable: number of non-spinout entrants in region i at time t

	Spec 1	Spec 2	Spec 3	Spec 4	Spec 5
Center size (t-1)	-0.005 (0.006)	0.000 (0.006)	-0.001 (0.006)	-0.003 (0.006)	
Log population 1900		2.160*** (0.244)	1.088** (0.448)	0.868** (0.434)	0.853** (0.432)
Log per capita income 1900		2.856*** (0.386)	2.199*** (0.393)	1.903*** (0.421)	1.886*** (0.420)
Log C&W employment 1905			0.655** (0.312)	0.551* (0.311)	0.552* (0.310)
Log shipbuilding employment 1905			0.050 (0.077)	-0.073 (0.090)	-0.071 (0.090)
Log input resources 1905				0.463** (0.199)	0.442** (0.197)
Chicago	2.414** (1.105)	-0.662* (0.366)	-0.649 (0.412)	-1.146** (0.469)	-1.129** (0.466)
Indianapolis	2.640* (1.452)	1.512*** (0.246)	0.435 (0.602)	0.213 (0.605)	0.195 (0.604)
St Louis	2.303 (1.505)	0.572* (0.331)	0.304 (0.414)	0.125 (0.424)	0.138 (0.423)
Detroit	3.718*** (1.373)	0.311 (0.375)	-0.115 (0.462)	-0.566 (0.504)	-0.624 (0.492)
New York	3.778*** (1.415)	-2.113*** (0.560)	-1.637*** (0.574)	-2.095*** (0.583)	-2.124*** (0.577)
Year	0.070** (0.029)	0.047* (0.028)	0.051* (0.027)	0.058** (0.027)	0.046*** (0.017)
Constant	-133.31** (56.060)	-118.73** (53.209)	-121.16** (52.514)	-136.74*** (52.124)	-112.47*** (32.219)
Observations	473	473	473	473	473
Number of regions	43	43	43	43	43
Log Likelihood	-313.6	-283.2	-276.9	-274.3	-274.5

Notes: Standard errors in parentheses. Star levels: 10, 5 and 1%.

In the second regression we include population and income. Both variables are economically and statistically significant. By contrast, Center Size remains insignificant.

In the third regression, we add employment in two related industries. The results show that employment in the Carriage and Wagon (C&W) industry is significantly correlated with entry into the auto industry. By contrast with C&W and with the conjecture in Jacobs (1969), employment in shipbuilding shows little correlation with entry into the auto

Table 5
A linear model explaining the explanatory variable C&W
Dependent variable: Log Carriage & Wagon Employment 1905

	Spec 1
Log population 1900	0.648*** (0.232)
Log per capita income 1900	-0.750* (0.425)
Log iron and steel 1905	0.341*** (0.124)
Log lumber and Timber 1905	0.083** (0.038)
Log brass and copper 1905	0.027 (0.030)
Log rubber 1905	0.054* (0.031)
Constant	-2.011 (2.570)
Observations	43
R-squared	0.89

Notes: Robust standard errors in parentheses. Star levels: 10, 5 and 1%.

industry. Our complementary evidence that many founders of auto companies had previous experience in the C&W industry suggests human capital as the specific channel for the causality from employment in C&W to entry into the auto industry.

In the fourth regression, we add one more location-specific explanatory variable: the availability of input resources relevant for the auto industry. The regression results suggest that, not surprisingly, the availability of local inputs is an important factor explaining entry in a given region.

Notice that the size of the coefficient on C&W employment is lower in the fourth regression than in the third regression. This suggests that, to the extent that the C&W industry uses a similar input mix to the auto industry, some of the effect of local inputs may be embedded in C&W employment. This is confirmed by the historical data provided by the *Census of the U.S. Manufactures* and Leontief's input-output table, which show the same crucial inputs for carriages and wagons as for cars: iron and steel; brass and copper; lumber and timber; and rubber. Table 5, where we regress C&W employment to the same set of input availability variables, also confirms the commonality of inputs. To the extent that the importance of inputs is not identical for the C&W and auto industries, Regression (4) in Table 4, together with Table 5, allow us to tease out the "direct" and "indirect" effects of local inputs. The direct effect is given by the coefficient in Table 4; the indirect effect is given by the composition of the coefficient on inputs in Table 5 multiplied by the coefficient on C&W employment in Table 4's Regression (4). We return to this in the next section, where we calibrate a structural model of entry into the auto industry.

To conclude our discussion of Table 4, in Regression (5) we exclude the variable Center Size. All coefficients remain constant with respect to Regression (4). Together with the systematically low value of the coefficient on Center Size (from Regression (1) to Regression (5)) we conclude that Marshallian economies, in the strict sense, seem to play no role in the auto industry. By contrast, local conditions, such as income, population and input availability, seem to play an important role; as does the development of related industries, specifically the carriage and wagon industry. With respect to the latter, our results suggest that there is both a "pure" co-agglomeration effect (everything else constant, auto companies benefit from the presence of carriage and wagon firms in a given location); as well as an indirect, long-term effect of local input availability (more local inputs lead to more carriage and wagon firms in a given location, which in turn leads to a higher entry rate into the auto industry).

■ **Are spinouts an agglomeration force?** About 17% of all firm entries in the history of the U.S. auto industry correspond to managers or employees of existing auto firms who leave the company to start their own: a spinout. To the extent that spinout rates vary across regions, it is conceivable that spinouts may act as a force toward agglomeration. We now consider a series of regressions to test this possibility.

Table 6 presents the results of six logit regressions, which differ in the set of independent variables considered. The level of observation is firm-year and the sample range 1895–1942, which results in a total of 3,000 to 4,000 observations approximately (depending on the set of independent variables included). Some of the independent variables — center size, family size, center top and family top — are lagged one year. In this way, we avoid including the new entrants in the measure of existing firms.[18]

Some broad patterns emerge from this set of regressions. First, firm age has a positive and significant coefficient throughout. This suggests that older firms are more likely to give birth to a spinout than younger firms. Note that we do not include direct measures of firm quality on the right-hand side. For this reason, we expect firm age to capture firm ability to some extent, in which case the results suggest that higher ability firms are more likely to give birth to a spinout.

In addition to the "parent's" characteristics (for which age is a proxy), the likelihood of giving birth to a spinout is also a function of the parent's family. Specifications 3 and 5 suggest that the greater a firm's family size, the more likely the firm will give birth to a new spinout. In specifications 4 and 6 we use family top instead of family size. This alternative specification places extra weight on the quality of the parent's family. The coefficient remains significant. It is higher in value, though we should add that both the average and standard deviation of family top is lower than those of family size.

Similarly to Table 4, center size or center top does not seem to have a significant correlation with the dependent variable, as shown in specifications 3 and 4. When they do, as in specifications 5 and 6, the coefficients are rather negative and small.

Meanwhile, after we control for firm age and family, we find local factors (e.g. population, income, related industries and input resources) do not have significant effects on the dependent variable. This suggests that firm age and family are sufficient and more direct predictors of firm quality than those local factors.

---

[18] The regression results are very similar if contemporary values are used.

Table 6
Logit models of spinout entry, 1895-1942.
Dependent variable: firm gives birth to spinout at time t

	Spec 1	Spec 2	Spec 3	Spec 4	Spec 5	Spec 6
Firm age	0.062*** (0.021)	0.064** (0.026)	0.074*** (0.027)	0.060** (0.026)	0.094*** (0.029)	0.083*** (0.029)
Center size		0.015 (0.014)	0.010 (0.014)		-0.043** (0.022)	
Family size			0.170*** (0.050)		0.178*** (0.056)	
Center top				0.042 (0.081)		-0.180** (0.089)
Family top				0.289*** (0.092)		0.310*** (0.095)
Log population 1900			-0.227 (0.544)	-0.217 (0.526)	-0.376 (0.508)	-0.434 (0.501)
Log per capita income 1900			-1.696 (1.691)	-1.782 (1.665)	-1.809 (1.612)	-1.637 (1.634)
Log C&W employment 1905			0.339 (0.430)	0.299 (0.420)	0.358 (0.404)	0.366 (0.405)
Log shipbuilding employment 1905			-0.058 (0.181)	-0.069 (0.179)	-0.033 (0.183)	-0.040 (0.181)
Log input resources 1905			0.062 (0.479)	0.105 (0.468)	0.134 (0.445)	0.158 (0.449)
Chicago	0.608 (0.468)	0.419 (0.564)	0.212 (0.633)	0.325 (0.585)	1.348* (0.791)	0.736 (0.614)
Indianapolis	0.579 (0.457)	0.141 (0.600)	-0.319 (0.734)	-0.109 (0.685)	0.798 (0.878)	0.111 (0.679)
Detroit	1.466*** (0.395)	0.857 (0.745)	0.409 (0.807)	0.216 (1.157)	2.897** (1.251)	3.334** (1.411)
Rochester	0.255 (0.580)	0.180 (0.605)	0.271 (0.658)	0.351 (0.637)	0.936 (0.726)	0.710 (0.666)
New York City	0.738 (0.476)	0.533 (0.586)	0.636 (0.652)	0.671 (0.644)	1.961** (0.839)	1.568** (0.704)
Year or year dummies	-0.057*** (0.021)	-0.065*** (0.023)	-0.086*** (0.025)	-0.084*** (0.025)	year dummies	year dummies
Constant	104.562** (40.770)	118.632*** (43.726)	166.099*** (49.168)	161.193*** (49.948)	1.467 (6.252)	0.947 (6.307)
Observations	4,333	3,585	3,585	3,585	2,979	2,979

Notes: Center Size, Family Size, Center Top and Family Top one-year lagged.
Robust standard errors clustered by firm. Star levels: 10, 5 and 1%.

In the first four specifications, we use year as an explanatory variable. The negative coefficient reflects the fact that entry became more difficult as auto evolved into a mature industry at later stages. Since the evolution is not linear, in specifications 5 and 6 we use year dummies instead of the year trend.

Last but not least, the Detroit dummy shows much larger effect than other location dummies, especially in specifications 5 and 6.[19] Moreover, the coefficient's size is also quite significant in comparison to other determinants of spinout entry. For example, being in Detroit increases the probability of giving birth to a spinout by more than a 5 standard deviation increase in firm age or adding 10 top firms to a firm's family (specification 6).

■ **Survival of the fittest: determinants of exit rates.** As happens in many industries, net entry and exit rates in the auto industry are considerably lower than gross entry and exit rates. Consequently, understanding exit patterns is an important step towards understanding the evolution of industry concentration. Our next set of results pertains precisely to firm exit.

Table 7 displays the results of six logit regressions. In all of them, the dependent variable is firm exit, that is, a dummy variable that takes the value 1 if a firm exits in a given year. Different specifications correspond to different sets of independent variables. Some of the independent variables — center size, family size, center top and family top — are lagged one year.[20]

In all regressions, firm age has a negative coefficient. This is consistent with much of the previous literature on firm exit: older firms are less likely to exit than younger firms. Specifically, a one-standard deviation increase in firm age decreases the odds ratio of exit by about 40% (specification 5).

Other things being equal, de alio and spinout firms are less likely to exit than de novo firms. Specifically, the odds ratios of exit are about 39% lower for de alio firms and 33% lower for spinout firms with respect to de novo firms (specification 5).

We saw earlier that center size does not seem to have a big impact on firm entry (cf Tables 4 and 6). Table 7 suggests that center size may instead have a slightly positive effect on firm exit (specifications 3 and 5), while the effect of center top is not significant (specification 4 and 6). Again, the evidence does not seem to match the prediction of Marshall-type agglomeration economies.

By contrast, family size and family top seem to have a significant impact on survival: an additional "relative" (that is, an additional firm in a firm's family) reduces the odds ratio of exit by about 10% (specification 5), whereas an additional "top relative" reduces the odds ratio of exit by about 33% (specification 6).

After we control for firm age and family, we find none of the local factors (e.g. population, income, related industries, and input resources) has a significant effect on the dependent variable. Again, this suggests that firm age and family are sufficient and more direct predictors of firm quality than those local factors.

Finally, the various center dummies suggest that there are some remaining location specific effects, with firms in Indianapolis and Detroit more likely to survive than firms in other regions.

■ **All in the family: determinants of spinout peformance.** Several of the above regressions suggest that "family matters." Specifically, the size and quality of a family has an important impact on whether a spinout will take place and whether such spinout will survive. We now take a closer look at the mechanism whereby family membership helps the survival

---

[19] St. Louis is omitted in the regressions due to the inexistence of any spinout.

[20] The regression results are very similar if contemporary values are used.

Table 7

Logit models of firm exit, 1895-1942. Dependent variable: firm exit at time t

	Spec 1	Spec 2	Spec 3	Spec 4	Spec 5	Spec 6
Firm age	-0.051*** (0.008)	-0.067*** (0.009)	-0.065*** (0.009)	-0.056*** (0.009)	-0.070*** (0.010)	-0.061*** (0.010)
Dealio			-0.473*** (0.109)	-0.472*** (0.106)	-0.502*** (0.109)	-0.498*** (0.107)
Spinout			-0.413*** (0.157)	-0.200 (0.161)	-0.402** (0.161)	-0.198 (0.162)
Center size		0.013*** (0.005)	0.014*** (0.005)		0.016** (0.007)	
Family size			-0.089** (0.038)		-0.107** (0.043)	
Center top				0.038 (0.028)		0.038 (0.032)
Family top				-0.381*** (0.076)		-0.404*** (0.082)
Log population 1900			0.140 (0.166)	0.173 (0.163)	0.247 (0.165)	0.277* (0.162)
Log per capita income 1900			0.468 (0.293)	0.463 (0.299)	0.474 (0.307)	0.449 (0.307)
Log C&W employment 1905			-0.025 (0.125)	-0.052 (0.119)	-0.063 (0.126)	-0.093 (0.120)
Log ship building employment 1905			0.017 (0.045)	0.025 (0.044)	0.005 (0.048)	0.016 (0.047)
Log input resources 1905			-0.081 (0.108)	-0.085 (0.110)	-0.102 (0.110)	-0.105 (0.112)
St Louis	-0.024 (0.262)	-0.060 (0.315)	0.091 (0.308)	0.208 (0.318)	-0.076 (0.311)	0.072 (0.308)
Chicago	-0.143 (0.147)	-0.426** (0.194)	-0.416** (0.207)	-0.209 (0.185)	-0.494** (0.235)	-0.221 (0.191)
Indianapolis	-0.532*** (0.130)	-0.861*** (0.179)	-0.700*** (0.244)	-0.475** (0.216)	-0.740*** (0.271)	-0.457** (0.223)
Detroit	-0.370*** (0.106)	-1.033*** (0.262)	-0.846*** (0.279)	-0.618 (0.432)	-0.941*** (0.355)	-0.605 (0.487)
Rochester	-0.177 (0.211)	-0.324 (0.231)	-0.405 (0.271)	-0.351 (0.261)	-0.509* (0.284)	-0.416 (0.265)
New York City	0.243** (0.111)	-0.056 (0.164)	-0.146 (0.205)	0.007 (0.191)	-0.235 (0.236)	-0.013 (0.204)
Year or year dummies	0.026*** (0.006)	0.029*** (0.007)	0.034*** (0.009)	0.028*** (0.008)	year dummies	year dummies
Constant	-50.496*** (12.320)	-55.781*** (14.182)	-67.352*** (16.651)	-56.459*** (15.791)	-2.692 (1.665)	-2.789* (1.662)
Observations	4,454	3,683	3,683	3,683	3,602	3,602

Notes: Center Size, Family Size, Center Top and Family Top one-year lagged.
Robust standard errors clustered by firm. Star levels: 10, 5 and 1%.

of a spinout firm. As in many other settings, an interesting question is the split between "nature" and "nurture": do spinouts perform better because they are helped by their parents (nurture) or simply because they have better genes (nature)?

In our data, a small portion of spinout firms happened to locate away from their parents, largely due to exogenous reasons.[21] Comparing the performance of these spinouts with those locating nearby the parents allows us to address the above question. Table 8 displays four logit regressions where the dependent variable, as in the previous table, is firm exit, that is, a dummy variable that takes the value 1 if a firm exits in a given year. Different specifications correspond to different sets of independent variables. Some of the independent variables — center size, family size, center top and family top — again are lagged one year.[22] Differently from the regressions in Table 7, we now split the family size and family top variables by location: local family size now measures the number of relatives in the same location, whereas non-local family size measures the number of relatives located elsewhere (a similar distinction applies to local and non-local family top).

The results are quite striking: whereas the local variables (family size and family top) are statistically significant, the non-local ones are not statistically significant. In terms of coefficient size, local family top shows greater values (in absolute terms) than family top in Table 7. In fact, when we include local family top (specifications 2 and 4 in Table 8) the variable spinout ceases to be statistically significant. This suggests that belonging to a family of high performance firms and being located nearby family relatives is associated with superior spinout performance, whereas if a firm is located far from its family then performance is not statistically different from that of de novo entrants. In other words, the results suggest that, in the case of the auto industry, nurture trumps nature.

■ **Robustness analysis and further notes.** We perform a series of robustness checks on our results regarding firm exit. First, we consider alternative treatments of exit. In our exit regressions, we did not separate exit by acquisition from exit by liquidation. It may be argued that exit by being acquired should not be counted as firm failure (in some cases, it may be quite the opposite). In our sample, we have 762 exits in total, of which 108 (14%) resulting from acquisition. One way to solve the potential problem of confounding the two types of exit is to count exits by acquisition as censored observations of exits. Under the alternative specification, the regression results are similar in signs and values but even stronger in statistical significance.

Second, in our spinout and exit regressions we did not include top firm as an explanatory variable. The reason is that we would like the explanatory variables (e.g. center size, family size, etc) to predict the firm's performance in terms of spinout and exit. Top firm is just another measure of firm performance, which duplicates the dependent variables. Of course, top firm itself could be an imperfect proxy for firm performance (the way we define it as ever being a top firm). Therefore, when we do include it as an explanatory variable, our results still hold in terms of coefficient signs and values but become statistically weaker.

Finally, we ran random-effects logit models and considered different sample ranges (e.g., 1895–1929). The results are very similar.

---

[21] We investigate whether there could be endogeneity bias related to these spinouts' location choices by collecting detailed information on the motive of every spinout from the top three families: GM/Buick, Ford and Oldsmobile. The results show no systematic bias between the motive of a spinout and its subsequent location choice.

[22] The regression results are very similar if contemporary values are used.

Table 8
Logit models of firm exit, 1895-1942. Dependent variable: firm exit at time t

	Spec 1	Spec 2	Spec 3	Spec 4
Firm age	-0.064*** (0.009)	-0.052*** (0.009)	-0.070*** (0.010)	-0.057*** (0.010)
Dealio	-0.474*** (0.109)	-0.475*** (0.105)	-0.503*** (0.109)	-0.500*** (0.106)
Spinout	-0.416*** (0.157)	-0.219 (0.159)	-0.403** (0.161)	-0.213 (0.160)
Center size	0.014*** (0.005)		0.016** (0.007)	
Local family size	-0.102** (0.043)		-0.112** (0.048)	
Non-local family size	-0.062 (0.076)		-0.099 (0.085)	
Center top		0.041 (0.028)		0.041 (0.032)
Local family top		-0.472*** (0.098)		-0.493*** (0.106)
Non-local family top		-0.120 (0.132)		-0.156 (0.151)
Log population 1900	0.137 (0.166)	0.160 (0.163)	0.246 (0.165)	0.263 (0.162)
Log per capita income 1900	0.469 (0.292)	0.469 (0.292)	0.475 (0.307)	0.464 (0.304)
Log C&W employment 1905	-0.023 (0.125)	-0.043 (0.118)	-0.063 (0.126)	-0.083 (0.119)
Log shipbuilding employment 1905	0.016 (0.045)	0.023 (0.044)	0.005 (0.048)	0.016 (0.047)
Log input resources 1905	-0.081 (0.108)	-0.088 (0.110)	-0.102 (0.110)	-0.109 (0.113)
St Louis	0.075 (0.311)	0.121 (0.320)	-0.081 (0.312)	-0.014 (0.314)
Chicago	-0.425** (0.206)	-0.225 (0.185)	-0.496** (0.233)	-0.237 (0.191)
Indianapolis	-0.705*** (0.244)	-0.483** (0.215)	-0.742*** (0.271)	-0.467** (0.222)
Detroit	-0.845*** (0.280)	-0.606 (0.432)	-0.939*** (0.356)	-0.585 (0.483)
Rochester	-0.405 (0.271)	-0.341 (0.258)	-0.509* (0.284)	-0.403 (0.262)
New York City	-0.148 (0.204)	0.013 (0.187)	-0.235 (0.236)	-0.005 (0.200)
Year or year dummies	0.033*** (0.009)	0.025*** (0.008)	year dummies	year dummies
Constant	-66.249*** (16.766)	-50.150*** (15.793)	-2.684 (1.665)	-2.742* (1.648)
Observations	3,683	3,683	3,602	3,602

Notes: Center Size, Family Size, Center Top and Family Top one-year lagged.
Robust standard errors clustered by firm. Star levels: 10, 5 and 1%.

■ **Summary of main empirical results.** We may summarize our empirical findings as follows:

- Entry rates by non-spinout firms in a given location are increasing in: (a) regional population and income; (b) regional employment levels in the C&W industry; (c) local input resources.

- Spinout entrants are more likely to come out of: (a) older firms; (b) larger and better families; (c) Detroit.

- Firm survival rates are higher if: (a) the firm is older; (b) the firm resulted from de alio entry; (c) the firm was spun out of a high-performance parent and remained in the same location as the parent; (d) the firm is located in Detroit or Indianapolis.

- Firm entry and survival rates do *not* depend on center size. If anything, spinouts are negatively impacted and exits are positively impacted by center size or center top (cf Tables 6,7 and 8).

Taken together, the evidence casts doubt on the importance of Marshall (1890) type intra-industry externalities. By contrast, it suggests that Jacobs (1969) type co-agglomeration economies may play an important role, as well as Klepper (2007) type spinout entry. Finally, as suggested by the work of Ellison and Glaeser (1999), the results also unveil some significant location fixed effects, particularly due to local input resources.

The reduced-form analysis is useful for getting a first glance at the sign and size of the various effects. It also provides a useful springboard for our next step: to develop and calibrate a structural model of the U.S. auto industry. Such a model will allow us to perform a series of counterfactual exercises to evaluate the relative weight of each force of industry agglomeration.

4. A structural model of the U.S. auto industry

In this section, we develop a simple model of industry dynamics in the spirit of Hopenhayn (1992). Firms are forward looking, competitive price takers producing a homogeneous product with heterogeneous production capabilities. We consider two types of entrants: *de novo* entrants and *spinout* entrants.[23] De novo entrants originate outside of the industry, that is, their founder was not an industry participant before founding the firm. Spinout entrants, by contrast, are founded by former industry participants, that is, managers or workers previously employed by an existing industry participant.

We assume that the "supply" of de novo entrants is determined by traditional channels such as local population, income and the presence of related industries. Regarding spinout entrants, we assume that each incumbent firm generates *potential* spinouts at a constant per period rate. However, just as with de novo entrants, a potential spinout makes an optimal entry decision. In other words, every actual and potential firm is treated as a rational, forward-looking agent who makes optimal entry and exit decisions.

We also make the assumption that a spinout shares the same capability with its parent firm, that is, firm capabilities are "hereditary." However, the model does not restrict the

---

[23] For simplicity, we conflate de novo and de alio entry into one single category: de novo. We do so for two reasons. First, it keeps our structural model simpler. Second, to the extent that we include C&W and shipbuilding employment for estimating the quality of non-spinout entry in each location (cf Table 9: being a top non-spinout entrant), we effectively allow related industries to influence the quality of non-spinout entrants in our calibration of the structural model.

source of the hereditary effect: the family specific capability may reflect common "genes," or it could result from interactions of member firms within the family network (e.g., through knowledge linkages or business relations).

Finally, consistent with the empirical evidence presented in the previous section, we assume there are no intra-industry effects other than through family effects.

■ **Individual firm's problem.** The model is cast in discrete time and infinite horizon. A continuum of firms produce a homogenous good in a competitive market. Each firm is indexed by its discrete capability $s \in \{0, 1, ..., \bar{s}\}$ and location j. For simplicity, we assume that a firm with capability s starting at location j will retain the same capability and operate at the same location for the rest of its life. The industry structure is thus summarized by $m(s, j)$, the total mass of firms of capability s at location j. Given our assumption regarding capability and location, the evolution of $m(s, j)$ is entirely governed by entry and exit, the main focus of our analysis.

In each period, incumbent firms engage in product market competition by taking the industry price p as given. Each firm chooses optimal output $q(s; j, p)$ based on its capability and location characteristics. Their period profit is denoted by $\pi(s; j, p)$. We assume $q(s; j, p)$ and $\pi(s; j, p)$ are continuous, bounded, and strictly increasing in s and p.

Once an incumbent firm obtains its profit, it decides whether to continue operating or instead to leave the industry and earn an outside options ϕ^x. The value of the outside option is privately known by the firm and i.i.d. according to cdf $F(\phi^x)$. Given its belief of a time-series sequence of industry price $\bar{p}$, an incumbent's problem can be defined as:

$$V(s; j, \bar{p}, \phi^x) = \pi(s; j, p) + \max\{\, VC, \phi^x \}$$

where the value of continuation is

$$VC(s; j, \bar{p}) = \beta \int V(s; j, \bar{p}, \tilde{\phi}^x)\, dF(\tilde{\phi}^x)$$

Potential entrants at each location make their entry decisions at the same time as incumbents. As mentioned earlier, we consider two types of entrants: de novo and spinout. De novo entrants originate outside the industry. We assume the total mass of potential de novo entrants at location j, M_j, is determined by location specific characteristics. Each potential de novo entrant is endowed with a sunk entry cost ϕ^e. If the potential entrant pays ϕ^e then it is given an initial draw of capability s from the distribution $\mu(s, j)$, the discrete density function of capability s at location j. Hence, a potential entrant's probability of entry is given by Ψ_j, the probability that the ex-ante expected value of incumbency is greater than the entry cost ϕ^e. It follows that the expected number of de novo entrants at location j is given by

$$n_j = \Psi_j\, M_j = \Pr\left(\sum_s VC(s; j, \bar{p})\, \mu(s, j) \geq \phi^e \right) M_j \tag{1}$$

The second type of entrants, spinouts, originate within the industry. Each period, an incumbent firm at location j has a probability γ_j of generating a potential spinout. We assume the potential spinout shares the same capability s with its parent and knows its capability when making the entry decision. As a result, the spinout's entry decision is equivalent to its parent's continuing decision: A potential spinout will enter if its entry value is higher than its random outside option ϕ^x, i.e.,

$$VC(s; j, \bar{p}) \geq \phi^x$$

We assume that if a potential spinout entrant chooses not to enter in the current period, then the opportunity is foregone forever.

Note that there are two important differences between the two types of entrants. First, while potential de novo entrants are uncertain about their capability of operating in a new industry, spinout entrants directly inherit their parent's capability draw. This is a sharp assumption we make to highlight the fact that spinout entrants have better knowledge of their own capability given by their industry experience. Second, we assume that de novo entrants need to pay an additional entry cost ϕ^e with respect to spinouts, a difference that corresponds to the extra investment de novo entrants need to make to build up business relations or a customer base in the industry.

■ **Supply and demand.** We next derive the transition of the mass of firms of capability s at location j. This transition depends on the number of exits, spinouts, and de novo entrants at each state (s, j). Specifically, we have

$$m'(s, j) = m(s, j)\,(1 + \gamma_j)\,\chi_{s,j} + n_j\,\mu(s, j) \tag{2}$$

where

$$\chi_{s,j} = F\big(VC(s; j, \bar{p})\big) \tag{3}$$

is the probability of staying in the industry given the cdf function F of the outside option.

The right-hand side of (2) reflects the two sources of entry mentioned earlier. The first term combines the decisions of incumbents and spinout entrants. There are in total $m(s, j)\,(1 + \gamma_j)$ such firms making entry decisions, $m(s, j)$ incumbents and $m(s, j)\,\gamma_j$ potential spinout entrants. Since their continuation value is the same, their continuation/entry probability, $\chi_{s,j}$, is also the same. The second item on the right-hand side is the inflow of de novo entrants. Note that the number of de novo potential entrants, n_j, is location specific, and that moreover de novo entrants at location j are ex ante identical in terms of their expected capability.

Given each firm's output level, $q(s; j, p)$, and given the mass of each firm's type, $m(s, j)$, we determine total supply in location j. Aggregating over locations, we get total supply in the industry. We assume industry demand is given by the inverse demand function $p = D^{-1}(Q)$. Industry price then clears the market in each period so that total supply equals total demand:

$$p = D^{-1}\left(\sum_{s,j} q(s; j, p)\; m(s, j)\right) \tag{4}$$

■ **Industry equilibrium.** An industry equilibrium is defined by a sequence of prices $\bar{p}^*$, a mass of entrants n^*_{jt}, a measure of incumbent firms $m^*(s, j, t)$, and a policy function $\chi^*(s, j, t)$ such that

- n^*_{jt} satisfies the entry condition for de novo entrants each period, that is, n^*_{jt} satisfies (1);
- $m^*(s, j, t+1)$ is defined recursively given $m^*(s, j, t)$, n^*_{jt}, and $\chi^*(s, j, t)$, according to (2).
- $\chi^*(s, j, t)$ solves incumbent firms and potential spinouts' dynamic optimization problem each period, given their belief of $\bar{p}^*$, that is, $\chi^*(s, j, t)$ satisfies (3);
- p^*_t clears product market each period, that is, p^*_t satisfies (4);

In the following analysis, we consider a stationary industry equilibrium.[24] Although our model introduces some specific features — namely the distinction between de novo and spinout entrants — its basic features are similar (and simpler) than the general framework presented in Hopenhayn (1992). With small changes, the equilibrium existence and uniqueness results in Hopenhayn (1992) can therefore be applied in the present context.

■ **Equilibrium properties.** The theoretical model presented above implies a series of equilibrium properties which we now develop formally. (All proofs can be found in the Appendix.)

Proposition 1. *An incumbent (a potential spinout) is more likely to survive (enter) if it belongs to a higher capability family, given the same location and time.*

Proposition 2. *A high-capability family on average has a bigger family size, given the same location and time.*

Proposition 3. *Given positive entry and exit in the stationary equilibrium, spinout firms have lower probability of exit than de novo firms, given the same location and time.*

All of these results are consistent with the empirical evidence presented in Section 3. For example, Table 6 shows that family top has a positive effect on spinout rates, while Table 7 shows that family top has a negative impact on exit rates (cf Proposition 1). Propositions 1 and 2 together imply that family size is positively correlated with the spinout rate but negatively correlated with the exit rate, which are consistent with our findings in Tables 6 and 7. Moreover, Tables 7 and 8 show that spinouts have a lower exit rate than de novo firms (cf Proposition 3).

This correspondence between theory and empirical observation gives us confidence in the model as a good description of the auto industry. We next attempt to calibrate the model's parameters with a view at going beyond qualitative description. Specifically, our goal is to use the calibrated model to estimate the relative contribution of each of the model's features, an exercise we refer to as "agglomeration accounting."

Consistent with these stylized facts, we developed a model with no intra-industry effects. We are thus left with three possible sources of agglomeration economies: inter-industry effects, spinouts, and location specific effects (particularly the local inputs).

---

[24] Naturally, long-run changes in the auto industry can be viewed as comparative statics of the stationary equilibrium. In the Appendix, we show that the model is amenable to rescaling so that any changes to market size or production technology can be modeled equivalently as changing the value of firms' outside option.

■ **Functional forms.** In the model calibration, we assume six production locations ($j = 1, ..., 6$), corresponding to St. Louis, Chicago, New York, Indianapolis, Detroit, and others;[25] and two levels of firm capability ($s = 1, 2$), corresponding to low and high. We exclude any location fixed effects in production functions so as to limit the number of free parameters. Instead, we allow for location-specific effects through differences in entry and spinout rates and types.

We specify the profit function $\pi(s; p)$ by assuming a decreasing returns production function

$$q(s) = \exp(c_1\, s)\, l^{\alpha}$$

where c_1 captures the relative advantage of firms with a higher capability, and l is the quantity of input. This implies that a firm, taking industry price p as given, has profit and output given by

$$\pi(s; p) = \left(\frac{1-\alpha}{\alpha}\right) (\alpha\, p)^{\frac{1}{1-\alpha}} \big(\exp(c_1\, s)\big)^{\frac{1}{1-\alpha}}$$

$$q^*(s; p) = (\alpha\, p)^{\frac{\alpha}{1-\alpha}} \Big(\exp(c_1\, s)\Big)^{\frac{1}{1-\alpha}}$$

Furthermore, we assume that the outside option follows an i.i.d. exponential distribution with parameter σ.

■ **Demand curve.** We estimate an industry demand function using historical annual data of average car price and output from Thomas (1977). The data range is 1900–1929, and we assume a simple log-log per capita demand function:

$$\log\left(\frac{Q_t}{pop_t}\right) = a_t - b\, \log(p_t)$$

In the regression, we control for log U.S. GDP per capita (as a proxy for income) in the demand intercept a_t. Both car price and GDP per capita are in real terms.

To address the issue of potential endogeneity of the price variable, we exploit the model structure. In our theoretical model, industry long-run capability distribution is correlated with price, yet uncorrelated with any transitory demand shock. One proxy for long-run industry capability distribution is the share of spinout firms, which we used as an instrumental variable to estimate the demand slope parameter b. Our IV estimation gives $b = 3.39$ (0.39), with standard error in parentheses. The demand shifter is given by $a_t = 0.04 \times \log(\text{GDP per capita})_t + 17.40$.

The first-stage regression results (adj. $R^2 = 0.86$) are given by:

$$\log(p_t) = \underset{(1.28)}{2.86} + \underset{(0.76)}{1.72} \times \log\left(\frac{GDP_t}{pop_t}\right) - \underset{(0.66)}{5.89} \times (\text{Spinout Share})_t$$

The second-stage regression results (adj. $R^2 = 0.83$) are in turn given by:

$$\log\left(\frac{Q_t}{pop_t}\right) = \underset{(5.55)}{17.40} + \underset{(2.17)}{0.04} \times \log\left(\frac{GDP_t}{pop_t}\right) - \underset{(0.39)}{3.39} \times \log(p_t)$$

---

[25] We combine New York City and Rochester into one center (New York), and the category "others" includes 38 other non-center states.

Table 9
Logit models of top non-spinout entry, 1900-1910.
Dependent variable: being a top non-spinout entrant

	Spec 1	Spec 2	Spec 3
Entry year	-0.293*** (0.097)	-0.326*** (0.110)	-0.325*** (0.109)
Log per capita income 1900	-1.881 (1.634)	4.175 (6.025)	4.762 (7.592)
Log population 1900	-0.351 (0.295)	-4.197** (1.924)	-4.673* (2.748)
Log C&W employment 1905		3.275** (1.655)	3.311* (1.827)
Log ship building employment 1905		0.390 (0.567)	0.271 (0.713)
Log input resources 1905			0.409 (1.501)
Constant	567.034*** (187.758)	600.161*** (209.409)	592.731*** (212.195)
Observations	461	461	461
Log Likelihood	-63.50	-57.49	-57.45

Notes: Standard errors in parentheses. Star levels: 10, 5 and 1%.

where standard errors are reported in parentheses.

■ **Other model primitives.** We observe positive de novo entrants at each location during our sample period. Since all firms compete in a single auto market, the observed entry pattern by location can be rationalized by a free entry equilibrium with location-specific entry costs. We thus feed into the model our estimated average non-spinout entry rates in each location (as explained by our negative binomial regression reported in Table 4, Spec 5). We also use the logit regression reported in Table 9 (Spec 3) to calibrate the average entry probability of high- and low-capability firms in each location, that is, $\mu(s, j)$.[26] Finally, we set the discount factor at $\beta = 0.925$.

■ **Calibrated parameters.** There are five key model parameters left to be calibrated: the parameter reflecting cost heterogeneity across types c_1, the degree of returns to scale α, the average value of the outside option σ, the industry average spinout birth probability γ, and the Detroit specific spinout birth probability γ_D. Our calibration strategy considers 1910-

---

[26] For regressions reported in Table 9, the dependent variable is entry of a non-spinout firm ever being a top seller in the auto industry. The sample range is 1900-1910; it includes 461 non-spinout entrants in 5 production centers and 38 other states. The results suggest that the relative size of the carriage & wagon (C&W) industry, measured by local C&W employment relative to population, significantly raises the chance of being a top non-spinout entrant. Judging from the value of log likelihood, the model fit is equally good whether we include the input resources (Spec 3) or not (Spec 2).

1915 as an industry equilibrium.[27] We pick the parameter values to match the following data moments, as shown in Table 10:

- the distribution of output across the five production centers and other regions
- the firm exit rate at the five production centers and other regions
- the spinout rates at the five production centers and other regions

Intuitively, our model moments of output distribution and firm exit rates across locations are largely determined by the parameters c_1, α and σ, while the spinout rates are mainly driven by γ and γ_D. For this exercise, we set c_1=0.32, α=0.9, σ=0.86, γ_D=0.08, and γ=0.02. Note that the calibrated value of σ, the average value of firms' outside option, captures the levels of the industry demand a and production technology $q(s)$ in the model. In fact, we provide a proof in the Appendix that the industry equilibrium derived from our model is invariant with respect to a rescaling of market size a or production technology $q(s)$, as long as σ is rescaled appropriately.

■ **Agglomeration accounting: short-run analysis.** As mentioned in Section 3, there are two ways of looking at the effect of local conditions and co-agglomeration. Auto entrants are attracted to locations with more firms in the C&W industry. This implies that local factors (population, income, input resources) have an effect on entries into the auto industry in two ways: directly, to the extent that better inputs make for more profitable entry; and indirectly, to the extent that better inputs attract more firms into the C&W industry, which in turn increases the entry rate into the auto industry.

In other words, we may distinguish between a "short-run" and a "long-run" agglomeration accounting exercise. In our short-run decomposition, we take the location decisions in the C&W industry as given; by contrast, in our long-run decomposition, we account for the effect of local conditions on the location's C&W industry size. We expect that our long-run accounting will assign a greater weight to local factors (which we treat as exogenous). The question is, how much greater.

Table 10 shows the results from our short-run decomposition exercise. The second column shows the data, whereas the third column shows the basic model fit. Following that, we have 5 columns which correspond to different counterfactuals. In order to judge the goodness of fit, we present several basic variables of interest related to agglomeration accounting:

- Each region's auto output share
- The HHI index applied to each region's auto output share, that is, $HHI = \sum s_i^2$, where s_i is region i's output share
- The prediction mean squared error

Additional variables of interest include

- Exit rates (Detroit and Non-Detroit)
- *Ex post* spinout rates (Regional average and Detroit top firms)

---

[27] We calibrate our model to the average regional auto output shares between 1910 and 1915. Because the regional shares were stable during this period, the results do not change if we instead calibrate our model to any specific year.

Table 10

Short-run decomposition. Model fit and five counterfactuals: 1. Uniform C&W employment levels; 2. Uniform input resources; 3. Uniform other factors (population, income, shipbuilding, location dummies, etc); 4. No spinout; 5. No Detroit-specific spinout rate

			Counterfactuals				
	Data	Model	1	2	3	4	5
Output share							
Chicago	0.06	0.034	0.037	0.027	0.036	0.059	0.058
Detroit	0.67	0.675	0.295	0.489	0.492	0.369	0.371
New York	0.14	0.090	0.055	0.032	0.399	0.166	0.164
Indianapolis	0.07	0.101	0.040	0.186	0.016	0.212	0.215
St Louis	0.01	0.008	0.021	0.018	0.007	0.013	0.013
Others	0.06	0.092	0.553	0.249	0.051	0.180	0.180
HHI	0.477	0.476	0.101	0.277	0.402	0.213	0.215
Prediction MSE		0.001	0.065	0.016	0.017	0.021	0.021
Exit rate							
Non-Detroit	0.15	0.146	0.145	0.145	0.147	0.142	0.142
Detroit	0.13	0.135	0.142	0.134	0.136	0.136	0.137
Additional moments							
Average regional spinout rate	0.03	0.038	0.033	0.032	0.034	0	0.017
Detroit top spinout rate	0.079	0.073	0.074	0.073	0.073	0	0.019

Finally, in order to account for the effect of each model feature, we run alternative counterfactuals where that feature is shut off. This corresponds to the last 5 columns in Table 10. In Counterfactual 1, we set the variable "Carriage and wagon industry employment" to have the same value in all regions (specifically, the average of production centers) when feeding into the model with estimated number and quality of de novo entrants (cf Table 4, Spec 5; Table 9, Spec 3). We thus effectively shut off the type of related industries effects described in Jacobs (1969). In Counterfactuals 2 and 3, we conduct similar exercises by equalizing input resources or other location fixed factors across all regions, thus effectively shutting off the type of location effects described in Ellison and Glaeser (1999). In Counterfactual 4 we force spinout rates to be zero in all regions, whereas in Counterfactual 5 we allow for spinouts but force the Detroit spinout rate to be the same as in other regions, effectively shutting off the type of spinout effects described in Klepper (2007).

Since our goal is to explain agglomeration, it seems natural to use the HHI index of region output shares as an indicator of fit. Comparing the data column to the model column, we see that the values of HHI are virtually identical, that is, our model does a very good job at explaining the overall level of industry agglomeration. The low level of the prediction mean squared error reinforces this idea.

Consider now Counterfactual 1. As mentioned earlier, this corresponds to forcing the values of C&W employment to be uniform across regions, thus shutting off the Jacobs (1969) related industry effects. The value of HHI drops from .476 to .101, whereas the prediction MSE increases from .001 to .065. Moreover, the total output share of the five production centers falls dramatically, from 91% to 45%. This suggests that the related-industry effect, explained by the variable C&W employment, plays a very important role in explaining the formation of auto production centers.

Shutting off the differences of regional inputs, which we do in Counterfactual 2, also implies a lower value of HHI but to a less extent, dropping from .476 to .277. Omitting other regional effects, as we do in Counterfactual 3, only reduces the value of HHI mildly, from 0.476 to 0.402. For both cases, the increases in prediction MSE are also smaller than Counterfactual 1. Overall, this suggests that region fixed effects account for some agglomeration, but the magnitude is smaller than the C&W.

Turning to the effect of spinouts, we first force all spinout rates to be zero. This corresponds to Counterfactual 4, where we see a significant drop in HHI, from .476 to .213. Meanwhile, the prediction MSE increases from .001 to .021. Finally, Counterfactual 5 qualifies the precise channel through which spinouts operate to explain agglomeration. We allow for positive spinout rates but force the rate to be uniform across all regions, whereas in the base case we allow for a different Detroit spinout rate. As can be seen from the last column in Table 10, the drop in HHI is almost identical to that of Counterfactual (4). In other words, the main contribution of spinout rates to explaining industry agglomeration comes from allowing Detroit to have a different spinout rate, an estimate that seems consistent with the work of Klepper (2007).

Before moving on to agglomeration accounting, it is worth mentioning that, in addition to 1910-1915, we also calibrated the model to 1905 data (an earlier census year). One important difference about 1905 is that at the early stage spinouts were virtually nonexistent. In fact, the drop in explained HHI when shutting off spinouts is nearly zero. We conclude that the effect of spinouts varies greatly along the industry life cycle. This is not entirely surprising: by definition, early entrants could not be spinouts, that is, could not possibly have originated from existing industry firms. As a result, the industry development between 1905 and 1910-1915 is similar to comparing Counterfactual (4) or (5) with the base model, which shows that Detroit's output share rises from 37% to 68% as the result of spinouts. To sum up, out agglomeration accounting exercise is based on our 1910-1915 calibration, and for this reason the estimated contribution of spinouts is likely to be overstated with respect to the overall contribution along the entire industry history.

Based on the results from Table 10, Table 11 presents an estimate of the relative contribution of each of the factors we've considered in the paper. In the second column of Table 11, we note the decrease in HHI that results from omitting a certain feature (as obtained from Table 10). For the first line, intra-industry externalities, the value is obtained by construction, that is, we assume that there are no intra-industry externalities. As mentioned earlier, this follows from the results we obtained in reduced form regressions, where industry variables have very small, or zero, effect on entry and exit rates. For the remaining rows, we compute the change in HHI from Table 10.

We exclude Counterfactual 5 from Table 10 because we think that would amount to double counting. However, it should be understood from Table 10 that when we talk about the effect of spinouts we mean primarily the effect of spinouts in Detroit.

There are many ways to compute the relative contribution of each model feature. Table 11 shows three different indicators: the drop in HHI from omitting the particular model feature; the ratio between this change in HHI and the value of HHI in the data; and the ratio between each contribution and the sum total of all contributions. As can be seen from the second indicator, the sum total of all contributions to HHI is about 191%. This suggests

Table 11
Short-run agglomeration accounting: relative contribution of various model components

Factor	Δ HHI	% of HHI	% $\sum \Delta$ HHI
Intra-industry externalities	0.000	0.00	0.00
Related industry (C&W)	0.375	78.78	41.16
Local input resources	0.199	41.81	21.84
Spinouts	0.263	55.25	28.87
Other local factors	0.074	15.55	8.12
Total	0.911	191.39	100.00

that there are important interactions between the various model features, so that a simple additive contribution accounting is subject to possible overestimation of the contribution of a given feature.

Finally, the last column of Table 11 shows that, in the horse race between the various explanations for industry agglomeration, the development of C&W industry in each region seems to come out ahead, with a relative contribution of 41%. Spinouts contribute a little less than 29%, local inputs contribute about 22%, whereas other local factors account for about 8%. These results are largely consistent with the reduced-form regressions presented in Section 3.

■ **Agglomeration accounting: long-run analysis.** In our above agglomeration accounting exercise, we considered the size of the C&W industry in each location as given. Moreover, we measured the size of the co-agglomeration effect by the relation between C&W industry employment and auto industry entry. This analysis is subject to the caveat that, in the long run, entry into the C&W industry is itself endogenous — just like entry into the auto industry. Specifically, an effect that we are denoting as co-agglomeration may be, in the long run, the result of local (exogenous) conditions. Even more specifically, one important reason why auto entrants were attracted to Detroit was the presence of so many C&W firms in Detroit. But the reason why there were so many C&W firms in Detroit in the first place was the abundance of inputs such as iron and lumber in the Detroit region. In this sense, our previous "short-run" analysis may underestimate the total, ultimate contribution of local conditions, such as the abundance of input resources.

Table 12 corresponds to Table 10 with the difference that we consider the "long-term" contribution of local conditions; that is, in addition to the direct effect of local condition on entry rates, we also consider the indirect effect through C&W employment. Specifically, in Counterfactual 1, we now treat the residual terms from C&W employment regression (cf Table 5) as random C&W shocks which are not explained by the location fixed factors we study. Holding anything else constant, we equalize the random C&W shocks across regions (at the average of production centers) to predict the counterfactual C&W employment size in each region. This allows us to construct the counterfactual number and quality of de novo entrants by region (using Table 4, Spec 5; Table 9, Spec 3) and re-simulate the model. The result hence captures the contribution of random C&W shocks to auto agglomeration but not the C&W effects induced by other location fixed factors considered in our study.

Table 12

Long-run decomposition. Model fit and five counterfactuals: 1. Uniform C&W employment shocks (which are not explained by local conditions); 2. Uniform input resources; 3. Uniform other factors (population, income, shipbuilding, location dummies, etc); 4. No spinout; 5. No Detroit-specific spinout rate

			Counterfactuals				
	Data	Model	1	2	3	4	5
Output share							
Chicago	0.060	0.034	0.044	0.020	0.037	0.059	0.058
Detroit	0.670	0.675	0.524	0.268	0.644	0.369	0.371
New York	0.140	0.090	0.309	0.016	0.202	0.166	0.164
Indianapolis	0.070	0.101	0.012	0.106	0.039	0.212	0.215
St Louis	0.010	0.008	0.007	0.039	0.007	0.013	0.013
Others	0.060	0.092	0.104	0.550	0.072	0.180	0.180
HHI	0.477	0.476	0.373	0.093	0.458	0.213	0.215
Prediction MSE		0.001	0.010	0.070	0.001	0.021	0.021
Exit rate							
Non-Detroit	0.150	0.146	0.147	0.143	0.147	0.142	0.142
Detroit	0.130	0.135	0.137	0.137	0.131	0.136	0.137
Additional moments							
Average regional spinout rate	0.030	0.038	0.036	0.027	0.034	0	0.017
Detroit top spinout rate	0.079	0.073	0.073	0.074	0.073	0	0.019

Table 13

Long-run agglomeration accounting: relative contribution of various model components

Factor	Δ HHI	% of HHI	% $\sum \Delta$ HHI
Intra-industry externalities	0.000	0.00	0.00
Related industry (C&W)	0.103	21.64	13.43
Local input resources	0.383	80.46	49.93
Spinouts	0.263	55.25	34.29
Other local factors	0.018	3.78	2.35
Total	0.767	161.13	100.00

Similarly, in Counterfactuals 2 and 3, we equalize local inputs or other location fixed factors across regions to predict the counterfactual C&W employment size by region (again, using Table 5). The results, together with the counterfactual local inputs or other location fixed factors, are then used to re-simulate the model. Not surprisingly, once we consider indirect effects through C&W employment, the contribution of local inputs to auto agglomeration increases considerably, but the contribution of random C&W shocks becomes small.

Table 13 corresponds to Table 11, again with the difference that we consider the "long-term" contribution of local conditions. Our revised agglomeration accounting leads to different numbers: local inputs now appears as the leading contributor, explaining 50%

of total concentration; next, spinouts contribute 34%; the C&W employment shocks correspond to a mere 13% of the total effect; and other local factors (including population, income, ship building, and location dummies) only add 2%. Finally, as in the short-run decomposition, by construction intra-industry externalities do not contribute to explaining industry agglomeration.

■ **Summary of calibration.** In the short run, given that the C&W industry is already in place, its employment level contributes the most to explaining agglomeration in the auto industry. Although we do not directly measure the channels through which this takes place, we presume that the human capital channel is of paramount importance. In fact, a significant number of the founders of auto companies were previously involved in some venture related to the C&W industry. In the long-run, however, one must understand that employment in the C&W industry is endogenous, and in particular depends on many of the same factors that employment in the auto industry depends. In the long run, the abundance of input resources (that is, local conditions) appears as the main factor. Finally, both the short- and the long-run analysis suggest that spinouts play an important role (again, we presume that through the human capital channel).

5. Conclusion

Taking the early U.S. auto industry as an example, we evaluate four competing hypotheses on regional industry agglomeration: intra-industry local externalities, inter-industry externalities, employee spinouts, and location fixed effects. Our findings suggest that inter-industry spillovers, in particular the development of the carriage and wagon industry, play an important role by fostering non-spinout entrants. Spinouts play a secondary role and work as a special type of intra-industry spillovers. Location fixed effects (particularly local inputs) also explain differences in regional agglomeration. In the short run, this effect appears to be small. In the long run, however, once we account for the effect of local inputs on related industry location, local conditions appear to explain most of the agglomeration effect. Finally, the presence of other firms in the same industry has a negligible (or maybe even negative) effect on agglomeration.

There are some avenues for future research. First, while our paper focuses on the historical evolution of auto industry, the analysis can be applied to other industries, for example, the more recent high-tech industries. It would be interesting to explore any similarities or differences in terms of agglomeration patterns and driving forces. Second, due to data limitation, we use entry and exit as proxy measures of firm performance. In case data becomes available, future studies could use more direct measures of firm performance, such as output, profit, employment, or product variety. Third, our paper points to possible channels through which local factors and externalities contribute to industry agglomeration. It would be interesting to study more precisely the nature and size of local spillovers through those channels. Finally, it would be useful to conduct cross-country comparisons of industry agglomeration in both advanced and developing economies. This will help us better understand the nature of increasing-return technologies and regional spillovers, which are important driving forces for economic growth, development, and international trade.

Appendix: proofs

Proof of Proposition 1: Note that an incumbent's continuing decision is equivalent to a spinout's entry decision. Given that $\pi(s; j, p)$ is strictly increasing in s, continuous, and bounded, standard dynamic programming argument shows that $VC(s; j, \bar{p})$ is continuous in s and strictly increasing in s for $\bar{p} > 0$. Thus we know that for each period, $F(VC(s; j, \bar{p}))$ is strictly increasing in s, given the same location j. $\blacksquare$

Proof of Proposition 2: Note that for each period, all incumbents at location j have the same probability γ_j of having a potential spinout. Also, we show above that an incumbent (a potential spinout) is more likely to survive (enter) if it belongs to a higher capability family. Thus, a higher-capability family has a bigger family size on average, given that $(1 + \gamma_j)\chi_{s,j}$ increases in s. $\blacksquare$

Proof of Proposition 3: The stationary distribution is defined by $m_j^* = m_j^*(1 + \gamma_j)\chi_j^* + n_j\mu_j$, so $m_j^* = \frac{n_j}{1-(1+\gamma_j)\chi_j^*}\mu_j$. The distribution of spinout firms is $m_j^*\chi_j^* = \frac{n_j\chi_j^*}{1-(1+\gamma)\chi_j^*}\mu_j$. Since χ_j^* is strictly increasing in s, the capability distribution of spinout firms strictly dominates that of *de novo* firms, which is μ_j. $\blacksquare$

Appendix: a note on rescaling

This note proves that the industry equilibrium derived from our model is invariant with respect to a rescaling of the model parameters, such as market size a and production technology $q(s)$, as long as firms' average value of outside option σ is rescaled appropriately.

Given the demand function estimated in Section 4,

$$\log\left(\frac{Q}{pop}\right) = a - b\,\log(p)$$

letting $pop \times \exp(a) = Z$, we can rewrite the market equilibrium condition in steady state as

$$\log\left(\sum_s q^*(s,p)\,m(s)\right) = \log(Z) - b\,\log(p)$$

where $q^*(s,p)$ is the firm's profit-maximizing output given the price p. Profit maximization implies

$$q^*(s,p) = (\alpha\,p)^{\frac{\alpha}{1-\alpha}}\Big(\exp(c_1\,s)\Big)^{\frac{1}{1-\alpha}}$$

Thus the market equilibrium price $p^*(m,Z)$ is determined by

$$\log(p^*) = \frac{\log(Z)}{\frac{\alpha}{1-\alpha}+b} - \frac{1}{\frac{\alpha}{1-\alpha}+b}\,\log\left(\sum_s(\alpha)^{\frac{\alpha}{1-\alpha}}\big(\exp(c_1\,s)\big)^{\frac{1}{1-\alpha}}\,m(s)\right)$$

We can then always compute an industry equilibrium which is invariant to the scale of Z as long as the firms' outside options are subject to rescaling.[28] Let the rescaled price

$$\log(\tilde{p}) = \log(p^*) - \frac{\log(Z)}{\frac{\alpha}{1-\alpha}+b}$$

We can similarly rescale profit $\pi(s,p)$, which is again Z dependent:

$$\pi(s,p^*) = \left(\frac{1}{\alpha}-1\right)(\alpha p^*)^{\frac{1}{1-\alpha}}\Big(\exp(c_1\,s)\Big)^{\frac{1}{1-\alpha}}$$

The profit function is given by $\pi(s,p^*) = Z^{\frac{1}{\alpha+b(1-\alpha)}}\,\pi(s,\tilde{p})$.

Finally, recall the firm value in steady state is given by

$$VC(s) = \beta\Big(\pi(s) + F\big(VC(s)\big)\,VC(s) + \Big(1 - F\big(VC(s)\big)\Big)\,E\Big(\phi'|\phi' \geq VC(s)\Big)\Big)$$

Assuming ϕ is exponentially distributed (that is, F is exponential), we have

$$E\Big(\phi'|\phi' \geq VC(s)\Big) = VC(s) + \sigma$$

---

[28] We can also allow production technology $q(s)$ to be rescaled, e.g. changing $q^*(s,p)$ to $\lambda q^*(s,p)$ in the above equation. We can then define a new scaling parameter $\hat{Z} = Z/\lambda$ and all the following proof goes through.

where σ is the mean of the firm's outside option. Then we have

$$\begin{aligned} VC(s) &= \beta\Big(\pi(s) + \Big(1 - F\big(VC(s)\big)\Big)\,\sigma + VC(s)\Big) \\ &= \beta\Big(\pi(s) + \exp\big(-VC(s)/\sigma\big)\,\sigma + VC(s)\Big) \end{aligned}$$

This implies that if the mean outside option also has a scale factor Z such that $\sigma = Z^{\frac{1}{\alpha+b(1-\alpha)}}\,\tilde{\sigma}$, where $\tilde{\sigma}$ is a constant, then we can go from

$$\widetilde{VC} = \Big(Z^{\frac{1}{\alpha+b(1-\alpha)}}\,\tilde{\sigma}\Big)^{-1} VC$$

to

$$\widetilde{VC}(s) = \beta\Big(\tilde{\pi}(s)/\tilde{\sigma} + \exp\Big(-\widetilde{VC}(s)\Big) + \widetilde{VC}(s)\Big)$$

Hence, the firm's rescaled value function, $\widetilde{VC}(s)$, is invariant with respect to Z. Because firm exit rate is simply $\exp(-\widetilde{VC}(s))$, firm entry and exit rates in the steady state are invariant with respect to Z.

We have thus proved that industry equilibrium is invariant with respect to a rescaling of the model parameters governing market size and production technology, as long as firms' average value of outside option is rescaled appropriately.

References

Arrow KJ (1962), "The Economic Implications of Learning by Doing," *Review of Economic Studies* **29**, 155-172.

Bailey, L.S. (1971), *The American Car since 1775*, Automobile Quarterly: New York.

Buenstorf, Guido, and Steven Klepper (2009), "Heritage and Agglomeration: The Akron Tyre Cluster Revisited," *The Economic Journal* **119**, 705–733.

Cabral, Luís, and Zhu Wang (2009), "Spinouts: Theory and Evidence from Early U.S. Automobile Industry," Federal Reserve Bank of Kansas City Working Paper.

Chatterjee, Satyajit, and Esteban Rossi-Hansberg (2012), "Spin-Offs and the Market for Ideas," *International Economic Review* **53** (1), 53-93.

Dumais, G., G. Ellison and E. L. Glaeser (2002), "Geographic Concentration As A Dynamic Process," *Review of Economics and Statistics* **84** (2), 193-204.

Ellison, Glenn, and Edward L. Glaeser (1997), "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy* **105** (5), 889–927.

Ellison, Glenn, and Edward L. Glaeser (1999), "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration," *American Economic Review* **89** (2), 311–316.

Ellison, G. and E.L. Glaeser and W. Kerr (2010), "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," *American Economic Review* **100** (3), 1195-1213.

Easterlin, Richard (1960), "Interregional Differences in Per Capita Income, Population, and Total Income, 1840-1950," in NBER conference volume *Trends in the American Economy in the Nineteenth Century*, UMI.

Franco, April, and Darren Filson (2006), "Spin-Outs: Knowledge Diffusion Through Employee Mobility," *Rand Journal of Economics* **37**, 841–860.

Glaeser, Edward L., Hedi D. Kallal, José A. Sheinkman, and Andrei Schleifer (1992), "Growth in Cities," *Journal of Political Economy* **100**, 1126–1152.

Hanson, Gordon (2001), "Scale Economies and the Geographic Concentration of Industry," *Journal of Economic Geography* **1** (3), 255–276.

Henderson, J.V. (1988), *Urban Development: Theory, Fact, and Illusion*, New York: Oxford University Press.

Henderson, Vernon, Ari Kuncoro, and Matt Turner (1995), "Industrial Development in Cities," *Journal of Political Economy* **103**, 1067–1090.

Hopenhayn, Hugo (1992), "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica* **60** (5), 1127–1150.

Jackson, John (1988), "Michigan," in R. Scott Fosler (Ed), *The New Role of American States*, New York: Oxford University Press.

Jacobs, Jane (1969), *The Economy of Cities*, New York: Vintage Books.

KIMES, BEVERLY R. (1996), *Standard Catalog of American Cars, 1890–1942*, third edition, Iola, WI: Krause Publications Inc.

KLEPPER, STEVEN (2007), "Disagreements, Spinouts, and the Evolution of Detroit as the Capital of the U.S. Automobile Industry," *Management Science* **53** (4), 616–631.

KLEPPER, STEVEN (2010), "The Origin and Growth of Industry Clusters: The Making of Silicon Valley and Detroit," *Journal of Urban Economics* **67**, 15–32.

KRUGMAN, PAUL (1991), *Geography and Trade*, Cambridge: MIT Press.

LEONTIEF, WASSILY W. (1951), *The Structure of American Economy, 1919-1939*, New York: Oxford University Press.

MARSHALL, ALFRED (1890), *Principles of Economics*, Macmillan, London.

ROMER, PAUL M. (1986), "Increasing Returns and Long-run Growth," *Journal of Political Economy* **94**, 1002-1037.

ROSENTAL, STUART AND WILLIAM STRANGE (2004), "Evidence on the Nature and Sources of Agglomeration Economies," *Handbook of Urban and Regional Economics* **4**, 2119–2171.

RUBENSTEIN, JAMES M. (1992), *The Changing US Auto Industry: A Geographical Analysis*, New York: Routledge.

SMITH, PHILIP H. (1970), *Wheels within Wheels: A Short History of American Motor Car Manufacturing*, New York: Funk and Wagnalls.

THOMAS, ROBERT PAUL (1977), *An Analysis of the Pattern of Growth of the Automobile Industry 1895-1929*, New York: Arno Press.