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*

This draft: December 2017

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 in the automobile case, inter-industry local externalities (particularly from the carriage and wagon industry) and employee spinouts (particularly due to the high spinout rate in Detroit) play important roles. The presence of other firms in the same industry has a negligible or 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; Location 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, Tom Holmes, Hugo Hopenhayn, Boyan Jovanovic, Bob Lucas, Esteban Rossi-Hansberg and participants at various conferences and seminars for helpful comments; and Christian Hung, Roisin McCord and Erica Paulos 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 do some industries concentrate in certain geographic locations? Ever since Marshall's (1890) seminal work, an intense debate has developed among economists regarding agglomeration externalities. Marshall himself pointed to positive externalities from specialization: regions with specialized production structures tend to be more innovative in their specific industries. 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 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. However, according to Klepper, the mechanism for these self-reinforcing dynamics is quite different from Marshallian externalities; it results from organizational reproduction (that is, 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,

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" (leaving the former term to designate the divestment of a firm's unit).

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 ("competitors," as in the work of Marshall et al); (b) inter-industry spillovers ("complementors," as in the work of Jacobs et al); (c) family network, or spinout, effects ("parents," as in the work of Klepper et al); (d) location fixed effects ("places," as in the work of Ellison and Glaeser).[7]

We first identify six historically important auto production centers from the data: New York City, Chicago, Indianapolis, Detroit, Rochester, and St. Louis. This raises the important question: Why did these six locations rather than other places become prominent auto production centers? The data also shows that the relative importance of Detroit as a production center increased steadily since the beginning of the 20th century. The auto industry eventually became highly concentrated in Detroit. Why?

Our analysis provides answers to the above questions and also refines the existing explanations of industry agglomeration. First, in contrast to Marshall's (1890) hypothesis, we find that the proximate presence of firms from the (narrowly defined) same industry (that is, the presence of 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, regarding the ideas of Jacobs (1969) and others, we find evidence for strong positive inter-industry spillovers. However, as opposed to Jacobs (1969), the spillovers we find operate through the carriage and wagon industry rather than the shipbuilding industry. To the extent that the carriage and wagon industry preceded and was supplanted by the auto industry, the agglomeration effects that Marshall (1890) pointed out to appear valid but need to be interpreted in a broader context.[8]

Third, consistent with Klepper (2007), our findings show that spinouts play an important role on regional agglomeration, which contributed to the increased concentration in Detroit. However, we find that the performance of spinouts is heavily affected by local family members but not by distant ones, which suggests that spinouts may actually benefit from intra-family local spillovers rather than the heredity effect. This finding helps to bridge the gap between Klepper (2007) and Marshall (1890).

Finally, consistent with the work of Ellison and Glaeser (1997, 1999), we find significant location specific effects, particularly in the long run: the search for locally available inputs led founders of auto companies to Detroit; moreover, many of them were previously associated with local carriage and wagon companies, the location of which was in turn largely determined by the availability of local inputs. (To the best of our knowledge, we are the first to make this distinction, both conceptually and quantitatively.)

When discussing the evolution of the U.S. auto industry, a reference to the industry's shakeout is inevitable. The industry started in the 1890s and the number of car producers reached more than 200 around 1910. However, the number of firms declined since the mid-1910s (in spite of the considerable expansion of industry output). Only 40 firms survived by the mid-1920s, and merely 8 made it into 1940s.[9] While the shakeout in terms of firm numbers is a phenomenon of unquestionable interest, it is largely independent of the central issue we study: industry agglomeration around Detroit and

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

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.

8. The positive spillovers from carriage and wagon (the precedent industry) to auto (the new industry) is natural given that early automobiles were constructed in a very similar fashion to carriages, sharing identical parts and mechanics. This is consistent with Klepper and Simons (2000), who find that the U.S. television receiver industry was dominated by firms who had experiences producing radios prior to TVs.

9. As shown by Gort and Klepper (1982), this shakeout pattern is commonly observed in many industries, not just automobile.

other clusters. As Figure 1 shows, well before the industry shakeout took place, agglomeration around the six production centers — with Detroit being the largest one — had already emerged. Accordingly, our analysis focuses on the industry's pre-shakeout stage, which also avoids concerns of survivor bias in explaining industry agglomeration (that is, the idea that a few Detroit-based firms survived the shakeout, so that Detroit became effectively the dominant production center).

Our quantitative exercise begins with a series of reduced-form regressions explaining firm entry and survival. Based on these findings, we then construct a dynamic structural model and calibrate its stationary equilibrium to data from *before* the shakeout. Having calibrated the model, we conduct a series of counterfactual simulations where a selected feature of the model (e.g., spinouts) is shut off. We measure that feature's contribution to agglomeration by the corresponding drop in model goodness of fit (both in absolute terms and as a fraction of the model's total explained variation). Finally, by considering the endogeneity of carriage and wagon (C&W) firm location, we distinguish between short-term and long-term effects.

As mentioned earlier, we find the net contribution of competitors to be negative (that is, agglomeration takes place not because but in spite of the effect of local competitors). 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 the C&W industry — complementors and spinouts come out ahead in the "horse race." In the long-run — that is, considering the endogeneity of C&W location decisions — location fixed effects, particularly due to local input resources, come out ahead (again, together with spinouts).

Specifically, taking C&W locations as given we estimate that inter-industry spillovers contribute 45.89% to explain total agglomeration; local input resources, 23.03%; and spinouts, 59.72%; all leading to a sum total greater than 100%, to compensate for the dispersion effect of local competitors. Taking explicit account of C&W location decisions, we estimate that inter-industry spillovers contribute 9.78%; local input resources, 52.88%; and spinouts, 79.71%. In other words, much of the effect of related industries is transferred to local input resources once we consider endogenous C&W location decisions.

Finally, we note that almost the entirety of the effect of spinouts stems from allowing for a Detroit-specific spinout rate. This is consistent with the work of Marx, Strumsky and Fleming (2009), who stress the role of Michigan Statute 445.761 (of 1905) which prohibits non-compete agreements (and thus facilitates spinouts). This suggests that it's not spinouts per se that account for agglomeration in the auto industry but rather the difference in spinout rates across regions, differences which in turn stem from location-specific effects (regulations in this case). In this sense, our findings can be interpreted as implying that location-specific effects — including local inputs and location-specific spinout rates — accounted for the lion's share of the auto industry agglomeration, both in the short run and in the long run.

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.[10] Our agglomeration accounting exercise complements theirs: we do not focus on the different channels proposed by Marshall;[11] by contrast, we pay close attention to the distinction between intra and inter-industry spillovers. Our findings suggest that the theories of Marshall (1890), Jacobs (1969), and Klepper (2007) are all relevant, but they need to be interpreted in the correct context.

10. 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.'"

11. 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."

Our paper is also related to recent work by Bloom et al (2013). They find that firm performance is affected by two countervailing agglomeration effects: a positive effect from knowledge spillovers; and a negative business-stealing effect from product market rivalry. Our analysis is consistent with this distinction. Specifically, we estimate that, when defining agglomeration effects narrowly, the competition effect dominates; whereas, when defining agglomeration effects broadly, the knowledge spillover effect dominates.

The rest of the paper is structured as follows. In Section 2, we briefly describe the background 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 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 industry output continued to expand, but 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 in the 1940s. This is a common life-cycle pattern observed in many industries, termed as "industry shakeout" in the literature.

In the meantime, the auto industry showed substantial changes in geographic concentration. Based on the data, we identified six historically important auto production centers: St. Louis, Chicago, Indianapolis, Detroit, Rochester, and New York City.[12] Figure 1 presents the evolution of U.S. auto production measured by each location's share of the total number of firms from 1895 to 1942. As can be seen, New York City and Chicago were the most important centers in the late 1890s. Soon after, Detroit and other centers caught up. In 1905, 25% of all active firms were located in Detroit, accounting for more than 50% of total industry output. Meanwhile, 15% of the firms were located in New York City, 10% in Chicago, 8% in Indianapolis, 7% in Rochester, 2% in St. Louis, and the remaining 32% scattered across the country. In the years that followed, Detroit continued to gain an increasing share, both in terms of the number of firms and in terms of 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 generating 12–17 spinouts.[13] As an example, Figure 2 displays the GM/Buick family tree. As can be seen, it's a family with three generations: for example, a former GM employee founded Chevrolet, from which in turn Gardner and Monroe spun out.

Figure 3 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

12. 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).

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

Figure 1
Geographical distribution of U.S. auto producers

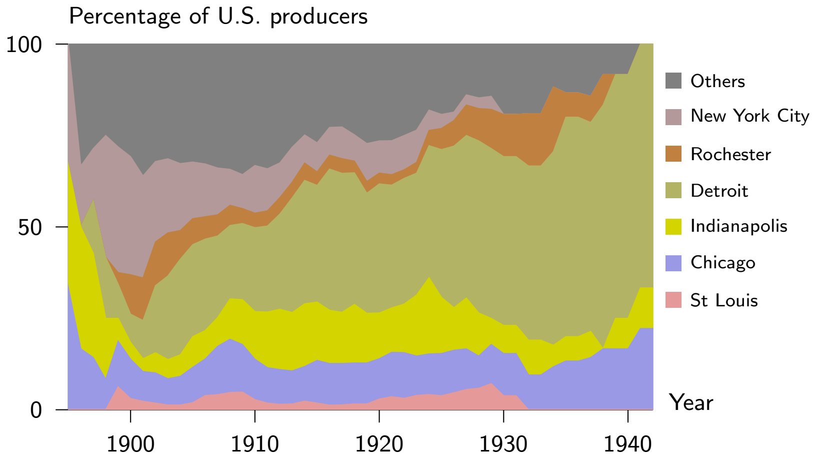
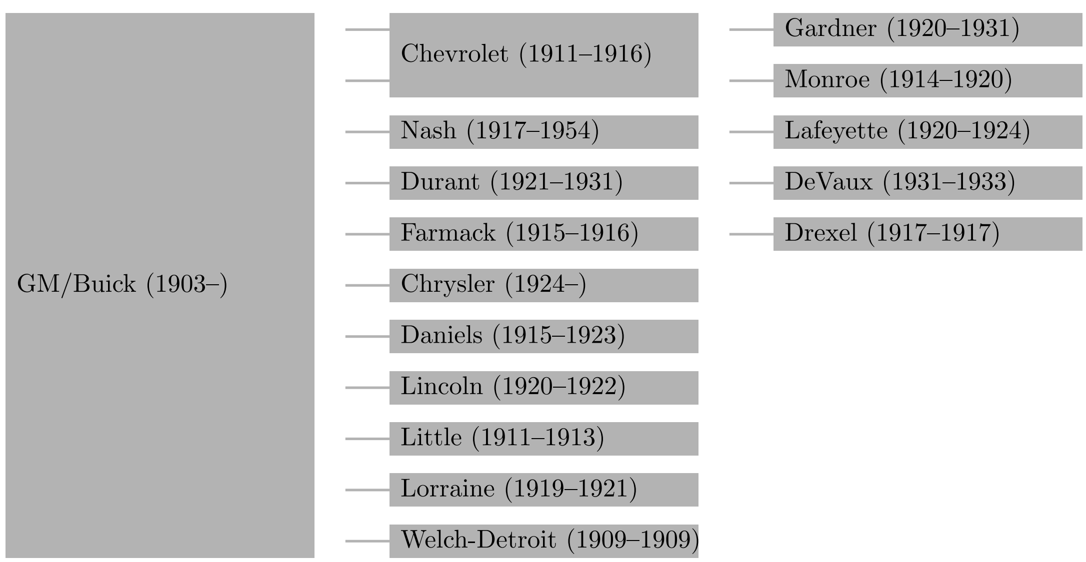


Figure 2
GM/Buick's family tree



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% of the spinouts in the top three families stayed in Detroit.

Although our analysis focuses on the auto industry, there are related industries which play an important role in explaining entry and exit patterns by auto firms. Prominent among these is the carriage and wagon (C&W) industry. The left panel of Figure 4 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 activity level in the C&W industry (measured by C&W employment level in 1905).[14] 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

14. Given that only state-level data are available for 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.

Figure 3
Family size distribution of U.S. auto producers (1900–1925)

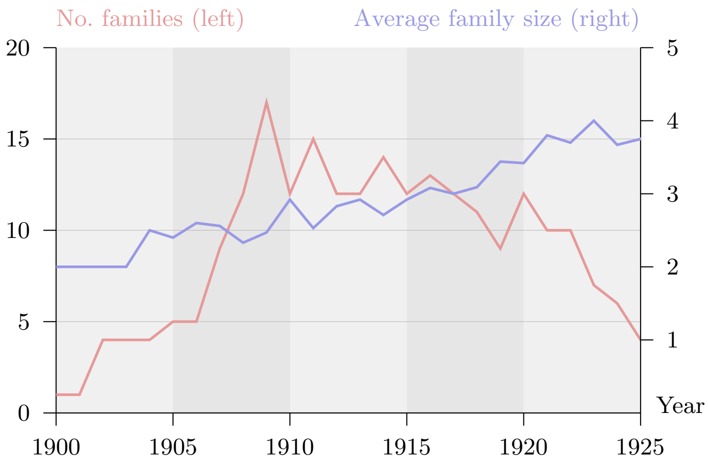
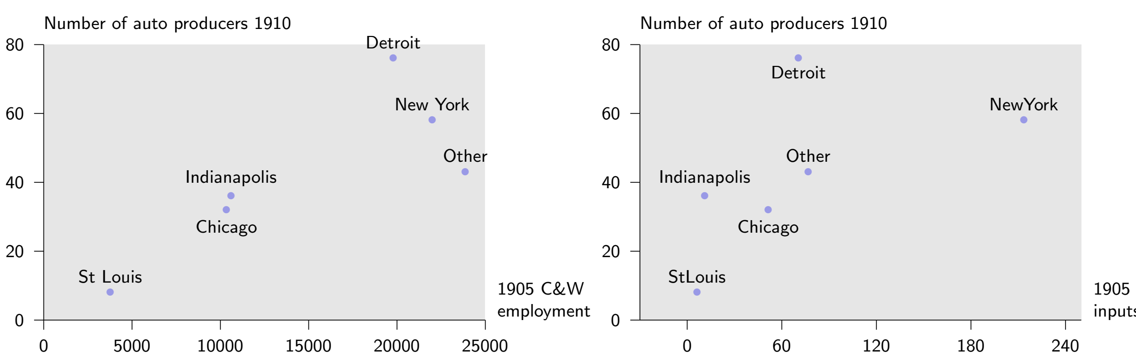


Figure 4
Location-level variables



we explore this relation in greater detail, and the results suggest that location patterns of auto firms (the new industry) were indeed influenced by location patterns of C&W firms (the precedent 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.

Aside from this channel, however, it is likely that other factors could also have contributed to the co-agglomeration of C&W and auto firms. For example, at the time both industries relied heavily on common inputs, such as iron and lumber.[15] The right panel of Figure 4 plots the concentration of the auto industry against the index of local auto-related input resources.[16] There is also a clear

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

positive correlation. In the following analysis, we will try to tease out various effects, and show that the C&W industry and local inputs both affected the auto industry agglomeration but through different channels.

It is worth noting that the geographic concentration pattern of the U.S. auto industry had formed well before the shakeout started around the mid-1910s. Our following analysis, therefore, will largely focus on the pre-shakeout stage of industry development. According to the classic theory of Jovanovic and MacDonald (1994), new industries evolve over two stages: A new product leads to the creation of a new industry, which in a first stage reaches a stationary equilibrium. In a second stage, a major innovation (e.g., the introduction of assembly-line production into the auto industry during the mid-1910s) leads to a survival race where firms unable to adopt the innovation are shaken out of the industry.[17] While many forces that we identified with our reduced-form entry and exit regressions are relevant for both pre- and post-shakeout periods, our structural calibration and agglomeration accounting exercises will focus on the pre-shakeout period, leaving the study of industry shakeout aside. As mentioned earlier, by doing so we also avoid concerns of survivor bias in explaining industry agglomeration.

To the extent that shakeout and geographic agglomeration are both commonly seen in many industries' life cycles, the approach that we develop in this paper could potentially apply to other industries besides automobile. Our analysis points to the importance of distinguishing pre- and post-shakeout stages while conducting the agglomeration accounting exercise, so that the contribution of each agglomeration factor can be quantified in a clearly-defined industry development context.

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 prospects, most likely through the human capital channel. We thus expect entry (respectively, exit) rates to be increasing (respectively, decreasing) in the employment size of related industries.

As shown by Klepper (2007) and others, spinouts are an important factor in the development of a new industry, except for the very beginning stage (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 by firm type.

Finally, as Ellison and Glaeser (1999) pointed out, location fixed-effects could also be important. Accordingly, we include regional population, per capita income, local input resources, and location dummies in our analysis. The first two factors, population and per capita income, may reflect the quantity and quality of local labor supply. The abundance of local input resources is measured by a location's total production values of physical inputs needed for auto manufacturing, including iron and steel, brass and copper, lumber and timber, and rubber. Location dummies capture the remaining location advantages, such as transportation.

■ **Data.** Our data comes from various sources. First, Smith (1970) provides a list of every make

17. Klepper and Simons (2005) and Klepper (2002) provide an alternative theory of shakeout, in which they emphasize incumbents' learning and innovations. However, to the extent that our analysis focuses on the pre-shakeout stage, the distinction between alternative shakeout theories is less of an issue. See also Cabral (2012).

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

of passenger cars produced commercially in the United States from 1895 through 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.[18]

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 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 48 U.S. continental states in the early 1900s. These data come from Easterlin (1960) and 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. In addition, from various issues of *Census of the U.S. Manufactures*, we also collect data of auto industry employment and auto output value for each of the 48 U.S. states from 1899-1935.

■ **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.

Tables 1 through 3 provide summary statistics of the variables used in the regressions. The sample range is set up to the time of U.S. entry into WWII (1895–1942), which includes 771 firms. 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.

18. The entry and exit are based on the first and last year of commercial production.

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

- 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.
- 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.

Table 3
Region level summary stats

Variable	Obs	Mean	St Dev	Min	Max
Population 1900	43	1763.2	2716.7	43	16390
Per capital income 1900	43	123.2	55.6	54	281
C&W employment 1905	43	1852.9	4127.3	0	19245
Shipbuilding employment 1905	43	1237.8	3478.8	0	20854
Iron and steel 1905	43	50.6	178	0.04	1100
Brass and copper 1905	43	2.3	11.1	0	72.2
Lumber and timber 1905	43	13.5	15.1	0	53.1
Rubber 1905	43	1.5	4.8	0	24.2
Local input resources 1905	43	9.989	34.411	0.010	213.557
Auto employment (1899-1935)	1591	5445	29003	0	303668
Auto output value (1899-1935)	1591	43.7	247	0	2960

- 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 1900. Regional population (thousands) in 1900.
- Per capita income 1900. Regional per capita income in 1900.
- C&W employment 1905. Number of workers in the C&W industry in a given region in 1905.
- Shipbuilding employment 1905. Number of workers in the shipbuilding industry in a given region in 1905.
- Iron and steel 1905. Regional production of iron and steel (million dollars) in 1905.
- Brass and copper 1905. Regional production of brass and copper (million dollars) in 1905.
- Lumber and timber 1905. Regional production of lumber and timber (million dollars) in 1905.
- Rubber 1905. Regional production of rubber (million dollars) in 1905.
- Local input resources (million dollars) 1905. 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).
- Auto employment (1899-1935). Annual number of workers in the auto industry in a given region, 1899-1935.
- Auto output value (1899-1935). Annual value of auto output (million dollars) in a given region, 1899-1935.

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
Center size	-0.008 (0.007)		-0.006 (0.007)	
Log auto employment	0.113 (0.082)		0.096 (0.083)	
Log population 1900	0.726 (0.442)	0.853** (0.432)	0.725* (0.434)	0.837** (0.424)
Log per capita income 1900	1.666*** (0.457)	1.886*** (0.420)	1.732*** (0.452)	1.925*** (0.414)
Log C&W employment 1905	0.556* (0.310)	0.552* (0.310)	0.558* (0.305)	0.557* (0.306)
Log shipbuilding employment 1905	-0.067 (0.090)	-0.071 (0.090)	-0.080 (0.087)	-0.084 (0.087)
Log local input resources 1905	0.329 (0.224)	0.442** (0.197)	0.379* (0.217)	0.483** (0.188)
Chicago	-0.996** (0.467)	-1.129** (0.466)	-1.032** (0.453)	-1.157*** (0.446)
Indianapolis	0.008 (0.616)	0.195 (0.604)	0.028 (0.602)	0.189 (0.589)
St Louis	-0.020 (0.433)	0.138 (0.423)	0.017 (0.415)	0.140 (0.406)
Detroit	-0.388 (0.502)	-0.624 (0.492)	-0.472 (0.491)	-0.668 (0.474)
New York	-1.610*** (0.661)	-2.124*** (0.577)	-1.735*** (0.652)	-2.183*** (0.551)
Year	0.033 (0.032)	0.046*** (0.017)	dummies	dummies
Constant	-84.099 (63.999)	-112.469*** (32.219)	-21.808*** (4.399)	-24.940*** (3.198)
Observations	473	473	473	473
Number of center	43	43	43	43
Log Likelihood	-273.4	-274.5	-262.5	-263.2

Notes: Center size one-year lagged; Log auto employment is calculated by taking the logarithm of three-year lagged average of auto employment in each location.
Standard errors in parentheses. Star levels: 10, 5 and 1%.

■ **Marshall vs others: entry by non-spinout firms.** Table 4 presents our first set of regressions. 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.[19] All regressions are based on a negative binomial model; the different specifications refer to different sets of independent variables.

Specifications 1 and 3 test an implication of Marshall's hypothesis, namely that auto companies are attracted to locations where the auto industry already has an important presence. We do so with two independent variables: the number of auto companies and the number of workers employed in the auto industry. To avoid simultaneity issues, we lag the variable Center Size by one year; and we calculate the variable Log Auto Employment by taking the logarithm of three-year lagged average of auto employment in a location.

As can be seen, the coefficients corresponding to the local auto industry are statistically insignif-

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

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%.

icant — and, in the case of the number of local competitors, the sign is opposite of what Marshall would predict.[20] In other words, the regression results do not support the notion of Marshallian economies, at least the narrow notion of Marshallian economies.

Inspired by the work of Jacobs (1969) and Jackson (1988), we also consider the possibility of a broader notion of agglomeration economies. Jacobs (1969) argues that the local presence of the shipbuilding industry played an important role in the emergence of Detroit as auto production center. We test inter-industry agglomeration economies by measuring location-specific employment in two related industries: carriage & wagon (C&W); and shipbuilding. The regression results show that the former has a statistically significant effect, but not the latter.

Other location-level variables that appear statistically significant include: population, per-capita income and input availability. They show positive relations with the entry of non-spinout firms.

Considering that narrow Marshallian economies seem to play no role in auto firm entry decisions, our specifications 2 and 4 excludes the variables measuring local auto industry size (number of firms and employment). Consistent with our expectation, the other variables' regression coefficients are fairly close in value and in statistical significance.

Note that specifications 3 and 4 repeat specifications 1 and 2 with one difference: instead of measuring a time trend we include year-specific dummy variables. While there is increase in fit (as measure by log likelihood), the difference is not very large. More important, the various point estimates seem remarkably robust to changes in specification.

The local presence of the C&W industry, which we measure by total employment in each region, plays an important role in our agglomeration analysis. One possibility is that the size of the C&W industry picks up the effect of local inputs, which are, to a great extent, shared with the auto industry. In fact, early automobiles looked very much like horse carriages with an engine in lieu of a horse. In this respect, our prior is confirmed by the *Census of the U.S. Manufactures* and Leontief's input-output table, which show the same top inputs for carriages and wagons as for cars: iron and steel; lumber and timber; and rubber.

Table 5, where we regress C&W employment to the same set of input availability variables, also

20. We also ran alternative regressions in which the variable Log Auto Employment is dropped. The coefficients of Center Size remain negative.

confirms the commonality of inputs.[21] To the extent that the importance of inputs is not identical for the C&W and auto industries, Table 4, together with Table 5, allows 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. We return to this in the next section, where we calibrate a structural model of 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 four logit regressions, which differ in the set of independent variables considered. The level of observation is firm-year and the sample range is 1900–1935, which results in a total of 3,000 to 3,500 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. The variable Log Auto Employment is defined as before.

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 capability to some extent, in which case the results suggest that higher capability 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 1 and 3 suggest that the greater a firm's family size, the more likely the firm will give birth to a new spinout. In specifications 2 and 4 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 1 and 2. However, when we include year dummies instead of a year trend in the regressions (specifications 3 and 4), we do obtain a significant coefficient, but it is negative, suggesting that, if there is a Marshallian effect, it is outweighed by the competition effect. The other measure of local auto industry size, Auto Employment, shows a positive correlation with spinout entry in specifications 1 and 2, but the sign turns negative in specifications 3 and 4 when year dummies are used, again contradicting to the positive Marshallian economies.

Controlling for firm age and family characteristics, we find local factors (e.g. population, income, related industries and input resources) have no significant effect on the dependent variable. This suggests that firm age and family are sufficient and more direct predictors of firm quality than those local factors.

In the first two specifications, we use a year trend 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 3 and 4 we use year dummies instead of the year trend. Notwithstanding these changes, our estimate of the relevant coefficients are similar.

Last but not least, the Detroit dummy shows much larger effect than other location dummies,

21. The regression coefficients are unlikely to result from any reverse causality given that the inputs used in C&W production were quite small compared with the size of the input sectors. Nationwide, in 1905, iron and steel used in C&W accounted for 0.45% of total iron and steel production; the corresponding percentage for lumber and timber was 1.78%; and for rubber, 4.17%.

Table 6

Logit models of spinout entry, 1900-1935.

Dependent variable: firm gives birth to spinout at time t

	Spec 1	Spec 2	Spec 3	Spec 4
Firm age	0.083*** (0.028)	0.074*** (0.028)	0.093*** (0.029)	0.082*** (0.030)
Center size	-0.009 (0.015)		-0.037* (0.022)	
Family size	0.160*** (0.052)		0.183*** (0.057)	
Center top		-0.079 (0.092)		-0.164* (0.099)
Family top		0.299*** (0.096)		0.311*** (0.095)
Log auto employment	0.320 (0.239)	0.404* (0.232)	-0.217 (0.292)	-0.104 (0.281)
Log population 1900	0.008 (0.555)	0.021 (0.578)	-0.667 (0.650)	-0.556 (0.540)
Log per capita income 1900	-2.375 (2.067)	-2.487 (2.098)	-1.347 (1.856)	-1.441 (1.841)
Log C&W employment 1905	-0.105 (0.563)	-0.246 (0.549)	0.739 (0.764)	0.539 (0.694)
Log shipbuilding employment 1905	-0.036 (0.193)	-0.037 (0.193)	-0.044 (0.188)	-0.045 (0.184)
Log local input resources 1905	-0.089 (0.535)	-0.073 (0.545)	0.285 (0.460)	0.226 (0.441)
Chicago	0.717 (0.671)	0.717 (0.631)	1.146 (0.816)	0.677 (0.636)
Indianapolis	0.116 (0.717)	0.118 (0.667)	0.594 (0.899)	0.067 (0.683)
Detroit	1.049 (0.778)	1.589 (1.250)	2.713** (1.256)	3.174** (1.485)
Rochester	0.615 (0.737)	0.672 (0.735)	0.788 (0.738)	0.651 (0.681)
New York City	1.210 (0.795)	1.296 (0.816)	1.756** (0.857)	1.485** (0.736)
Year	-0.156*** (0.055)	-0.157*** (0.051)	dummies	dummies
Constant	303.441*** (109.951)	306.573*** (102.463)	-3.161 (10.257)	-1.144 (9.686)
Observations	3,475	3,475	2,979	2,979

Notes: Center size, Family size, Center top and Family top one-year lagged; Log auto employment is calculated by taking the logarithm of three-year lagged average of auto employment in each location.
Robust standard errors clustered by firm. Star levels: 10, 5 and 1%.

especially in specifications 3 and 4.[22] 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 4).

For robustness checks, we also repeat the regressions by extending the sample range to 1895-1942 while dropping the variable Log Auto Employment (the data of auto employment is available up to 1935). All the relevant results remain unchanged.

■ **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 four 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, and the variable Log Auto Employment is defined as before.

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. The results also show that other things being equal, de alio and spinout firms are less likely to exit than de novo firms.

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 (specification 3), while the effect of center top is not significant. Auto Employment shows a positive effect on firm exit, and the effect becomes statistically significant in specifications 3 and 4. Again, the evidence does not seem to match the prediction of Marshall-type agglomeration economies in the strict sense. By contrast, family size and family top both show significant positive impact on firm survival in all specifications.

After we control for firm age and family, most other local factors do not show a significant effect on the dependent variable (except population and C&W employment in specifications 3 and 4). 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 performance.** 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 of a spinout firm.

In our data, a small portion of spinout firms happened to locate away from their parents, largely due to exogenous reasons.[23] Comparing the performance of these spinouts with those locating nearby the parents allows us to address the above question. Table 14 in the Appendix displays four logit regressions where the dependent variable, as in the previous table, is firm exit. 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,

22. St. Louis is omitted in the regressions due to the inexistence of any spinout.
23. We investigate whether there could be endogeneity bias related to spinouts' location choices, for instance, whether certain types of spinouts tend to move away from their parents. In doing so, we collected 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.

Table 7
Logit models of firm exit, 1900-1935. Dependent variable: firm exit at time t

	Spec 1	Spec 2	Spec 3	Spec 4
Firm age	-0.070*** (0.010)	-0.061*** (0.009)	-0.078*** (0.010)	-0.069*** (0.010)
De Alio	-0.457*** (0.108)	-0.452*** (0.106)	-0.479*** (0.108)	-0.469*** (0.105)
Spinout	-0.429*** (0.155)	-0.222 (0.159)	-0.421*** (0.159)	-0.219 (0.161)
Center size	0.009 (0.006)		0.014* (0.007)	
Family size	-0.086** (0.037)		-0.106** (0.042)	
Center top		0.017 (0.033)		0.018 (0.034)
Family top		-0.355*** (0.073)		-0.391*** (0.080)
Log auto employment	0.053 (0.060)	0.066 (0.058)	0.133** (0.065)	0.153** (0.067)
Log population 1900	0.168 (0.167)	0.196 (0.165)	0.283* (0.165)	0.311* (0.163)
Log per capita income 1900	0.354 (0.304)	0.330 (0.307)	0.174 (0.318)	0.112 (0.316)
Log C&W employment 1905	-0.122 (0.148)	-0.161 (0.141)	-0.263* (0.150)	-0.315** (0.145)
Log shipbuilding employment 1905	0.026 (0.045)	0.034 (0.045)	0.011 (0.049)	0.021 (0.049)
Log local input resources 1905	-0.082 (0.109)	-0.090 (0.111)	-0.113 (0.112)	-0.120 (0.114)
St Louis	0.179 (0.315)	0.272 (0.318)	0.040 (0.316)	0.182 (0.306)
Chicago	-0.267 (0.223)	-0.116 (0.194)	-0.374 (0.242)	-0.118 (0.195)
Indianapolis	-0.546** (0.266)	-0.404* (0.228)	-0.669** (0.291)	-0.419* (0.235)
Detroit	-0.657** (0.300)	-0.375 (0.487)	-0.931*** (0.362)	-0.457 (0.505)
Rochester	-0.375 (0.286)	-0.330 (0.275)	-0.443 (0.286)	-0.341 (0.270)
New York City	-0.028 (0.220)	0.097 (0.206)	-0.111 (0.234)	0.126 (0.206)
Year	0.024 (0.015)	0.017 (0.014)	dummies	dummies
Constant	-48.243 (29.541)	-34.741 (27.057)	-0.385 (1.942)	-0.121 (1.936)
Observations	3,573	3,573	3,537	3,537

Notes: Center size, Family size, Center top and Family top one-year lagged; Log auto employment is calculated by taking the logarithm of three-year lagged average of auto employment in each location.
Robust standard errors clustered by firm. Star levels: 10, 5 and 1%.

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 14) 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. The finding suggests that spinouts may actually benefit from intra-family local spillovers rather than the heredity effect, which clarifies the conflicting views between Marshall and Klepper.

■ **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).[24] 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.

Our exit regressions pool spinout and non-spinout firms (with a dummy for spinout firms). One advantage of this approach is that it takes into account the possibility that spinout and non-spinout firms are subject to the same survival factors. As an alternative, we also ran separate exit regressions for spinout and non-spinout firms. The results for the separate regressions, which we include in Tables 15 and 16 in the Appendix, are qualitatively and quantitatively similar. One interesting finding is that Auto Employment shows a negative effect on spinout firm exit, but a positive effect on non-spinout firm exit. This is consistent with our previous findings in the sense that the measure of local auto employment may proxy for local family size of spinout firms, but for non-spinout firms it may instead reflect the presence of competitors.

Finally, for robustness checks, we repeat the exit regressions by extending the sample range to 1895-1942 while dropping the variable Log Auto Employment. The results are very similar.

■ **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 measure of other auto firms in the same location. If anything, spinouts are negatively impacted and exits are positively impacted by competitors (cf Tables 6, 7 and 14).

Taken together, the evidence casts doubt on Marshall (1890) type intra-industry externalities in the strict sense. Rather, Jacobs (1969) type co-agglomeration economies and Klepper (2007) type

24. The data show that in the sample range 1895-1942, there were 108 exits (or 14% of all exits during the period) resulting from acquisition.

spinout entry appear relevant, but they may actually work through the Marshallian externalities in a broader context. 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 quantitative model of the U.S. auto industry

In this section, we develop a simple model of industry dynamics in the spirit of Hopenhayn (1992) which generates theoretical predictions consistent with our reduced-form analysis. We then use the calibrated model to quantify the contribution of each agglomeration factor.[25]

In the model, 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.[26] De novo entrants originate outside of the industry. 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 affected by location-specific characteristics. 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 and chooses the same location as the parent. The model does not restrict the source of the hereditary effect: it could result from interactions of member firms within the family network (e.g., through knowledge linkages or business relations) rather than the family-specific capability that reflects common "genes."[27]

With all of these ingredients, we perform an agglomeration accounting exercise where we consider the four components listed in the paper's title: competitors, complementors, parents and places. By "competitors" we mean Marshallian economies in the strictest sense of the word: the presence of other local firms from the same industry. By "complementors" we mean the broader perspective of inter-industry agglomeration economies: the effect of local related industries. By "parents" we mean the effect of spinouts in creating the self-reinforcing dynamics of geographical location. Finally, by "places" we mean local conditions such as income and the availability of production inputs.

As mentioned in Section 2, one remarkable feature of the U.S. auto industry is the significant shakeout it went through between the mid-1910s and the early 1930s. Many authors have focused on industry shakeouts, both theoretically and empirically, dealing both with the auto industry and other industries. Our focus is in explaining geographical agglomeration rather than shakeout. Accordingly, we consider a stationary model and calibrate it with pre-shakeout data (census years 1909 and 1914).

25. Alternatively, one may consider using a reduced-form analysis for explaining/predicting auto outputs across locations. However, reduced-form analysis has its limitations, especially when it comes to identify the endogenous channels (e.g., intra-industry spillovers or spinouts) that affect the output distribution. Because our paper considers intra-industry spillovers, inter-industry spillovers, local conditions and spinouts as four different forces affecting the auto industry agglomeration, a structural analysis is well suited for disentangling these forces and quantifying their relative importance.

26. 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 8: 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.

27. Our reduced-form analysis strongly suggests that the former is more important than the latter. However, for the purposes of the agglomeration accounting exercise we preform in this section, this distinction is irrelevant.

We note that by the early 1910s no major shakeout had yet occurred, but significant agglomeration had already taken place, including the concentration of production in Detroit.[28]

In the remainder of this section we do three things. First, we develop a dynamic model with the above properties. Second, we calibrate the model based on the coefficients obtained in our reduced-form regressions. Finally, we perform a series of counterfactual simulations where each of the model's features is "shut down." In this way we are able to estimate the contribution of each of the four features (competitors, complementors, parents and places) to explaining the observed industry agglomeration, including its ultimate concentration in Detroit.

■ **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 $m_t(s, j)$, the total mass of firms of capability s at location j time t. Given our assumption regarding capability and location, the evolution of $m_t(s, j)$ 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_t as given. Each firm chooses optimal output $q(s; j, p_t, m_t^j)$ based on its capability and location characteristics. Location characteristics include the number of firms in location j, m_t^j, as well as other factors.[29] Their period profit is denoted by $\pi(s; j, p_t, m_t^j)$. We assume $q(s; j, p_t, m_t^j)$ and $\pi(s; j, p_t, m_t^j)$ are continuous, bounded, and strictly increasing in s and p_t.[30]

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 $\overline{p}$ and mass of firms at own location $\overline{m}_j$, an incumbent's problem can be defined as:

$$V_t(s; j, \overline{p}, \overline{m}_j, \phi^x) = \pi(s; j, p_t, m_t^j) + \max\{VC_t, \phi^x\},$$

where the value of continuation is

$$VC_t(s; j, \overline{p}, \overline{m}_j) = \beta \int V_{t+1}(s; j, \overline{p}, \overline{m}_j, \widetilde{\phi}^x)\, dF(\widetilde{\phi}^x).$$

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 in location j faces a sunk entry cost ϕ_j^e. If the potential entrant pays ϕ_j^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_t^j, the probability that the ex-ante expected value of incumbency is greater than the entry cost ϕ_j^e. It follows that the expected number

28. We calibrate our model to the average regional auto output shares between 1909 and 1914. Because the regional shares were stable during this period, the results do not change if we instead calibrate our model to any specific year.

29. For simplicity, we normalize input price to be unit in every location, so a firm that faces lower input prices due to better access to local inputs or related industries would show up as having higher capability, and a location that enjoy an abundance of local inputs or related industries would show up as having a higher fraction of high-capability entrants. Because we feed the structural model with quantity and quality of entrants from the reduced-form regressions, the effect of lower input prices, while not explicitly modeled, is captured by our model simulation.

30. It is less straightforward to make assumptions about the relationship between profit and the local mass of firms m_t^j. We will assume that the force of agglomeration is not too strong such that a stationary industry equilibrium exists. Detailed magnitude of the intra-industry agglomeration force is discussed in our quantitative exercise.

of de novo entrants at location j is given by

$$n_t^j = \Psi_t^j \, M_j = \Pr\left(\sum_s VC_t(s; j, \overline{p}, \overline{m}_j) \, \mu(s,j) \geq \phi_j^e \right) M_j. \tag{1}$$

The second type of entrants, spinouts, originate within the industry. Each period, an incumbent firm at location j has a probability γ_t^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:[31] A potential spinout will enter if its entry value is higher than its random outside option ϕ^x, i.e.,

$$VC_t(s; j, \overline{p}, \overline{m}_j) \geq \phi^x.$$

We assume that if a potential spinout entrant chooses not to enter in the current period, then the opportunity is foregone forever.[32]

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.

■ **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

$$m_{t+1}(s,j) = m_t(s,j) \, (1 + \gamma_t^j) \, \chi_t^{s,j} + n_t^j \mu(s,j), \tag{2}$$

where

$$\chi_t^{s,j} = F\big(VC_t(s; j, \overline{p}, \overline{m}_j) \big) \tag{3}$$

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 $m_t(s,j)\,(1+\gamma_t^j)$ such firms making entry decisions, $m_t(s,j)$ incumbents and $m_t(s,j)\,\gamma_t^j$ potential spinout entrants. Since their continuation value is the same, their continuation/entry probability, $\chi_t^{s,j}$, 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 entrants, n_t^j, is location specific, and de novo entrants at location j are ex ante identical in terms of their expected capability.

Given each firm's output level, $q(s; j, p_t, m_t^j)$, and given the mass of each firm's type, $m_t(s,j)$, 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:

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

■ **Industry equilibrium.** An industry equilibrium is defined by a sequence of prices $\overline{p}^*$, 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

31. Our assumption that the continuation decision of an incumbent and the entry decision of its spinout are equivalent captures the family network effect found in our reduced-form regressions.

32. This is a simplifying assumption that allows us to maintain a constant spinout rate (per firm and per period). Departing from this assumption would require us to keep track of a cumulative pool of heterogeneous potential spinouts, leading to an unmanageable state space.

- 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 $\overline{p}^*$ and $\overline{m}^*_j$, 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.[33] Although our model introduces some specific features — namely the distinction between de novo and spinout entrants — its basic features are similar to (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 (the proofs are 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 14 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." We consider four possible sources of agglomeration economies: intra-industry effects, inter-industry effects, spinouts, and location specific effects (particularly the local inputs).

■ **Functional forms.** In the model calibration, we consider 43 locations ($j = 1, ..., 43$), corresponding to five production centers (i.e. Chicago, Detroit, New York, Indianapolis and St. Louis) and 38 non-center states;[34] 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 $\pi(s; p, m_j)$ by assuming a decreasing returns production function

$$q(s) = [\exp(c_1\, s)\, m_j^\eta]\;\; l^\alpha, \tag{5}$$

33. 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.

34. We combine New York City and Rochester into one center (New York). We also assume that non-center locations are identical.

where c_1 captures the relative advantage of firms with a higher capability, and l is the quantity of input. (We normalize the input price w_l=1, so a firm that faces a lower value of w_l would simply show up as having a higher capability s.) The elasticity η measures the strength of intra-industry spillover. This implies that a firm, taking industry price p as given, has profit and output given by

$$\pi(s;p,m_j) = \left(\frac{1-\alpha}{\alpha}\right)(\alpha\, p)^{\frac{1}{1-\alpha}} \left[\exp(c_1\, s)\, m_j^{\eta}\right]^{\frac{1}{1-\alpha}}, \tag{6}$$

$$q^*(s;p,m_j) = (\alpha\, p)^{\frac{\alpha}{1-\alpha}} \left[\exp(c_1\, s)\, m_j^{\eta}\right]^{\frac{1}{1-\alpha}}. \tag{7}$$

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

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

where standard errors are reported in parentheses.

■ **Calibration.** 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 M_j and Ψ_t^j. Empirically, we assume the observed quantity of de novo entrants in location j, n_j, is explained by the reduced-form regressions in Section 3 through local factors such as population, income, related-industries employment and input resources. We thus feed into the model the estimated average non-spinout entry rates in each location (as explained by our negative binomial regression reported in Table 4, Spec 2), that is, n_j.[35]

In order to quantify the difference between high- and low-quality entrants, in Table 8 we report the results of a series of logit regressions where the dependent variable is non-spinout top entry, that is, we model the probability of entry by a non-spinout firm that becomes a top firm (according to

35. The regression results reported in Table 4 suggest that intra-industry externalities are quantitatively negligible in explaining entry by non-spinout firms, so we use Spec 2 of Table 4 to simulate non-spinout entrants in our model calibration and counterfactual exercises.

Table 8
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 population 1900	-0.351 (0.295)	-4.197** (1.924)	-4.673* (2.748)
Log per capita income 1900	-1.881 (1.634)	4.175 (6.025)	4.762 (7.592)
Log C&W employment 1905		3.275** (1.655)	3.311* (1.827)
Log shipbuilding employment 1905		0.390 (0.567)	0.271 (0.713)
Log local 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%.

our definition of top firm). 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 and 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 input-relevant resource availability (Spec 3) or not (Spec 2). We use the results from specification 3 to calibrate the average entry probability of high- and low-capability firms in each location, that is, $\mu(s, j)$.

So as to quantify the effect of intra-industry spillovers (Marshallian economies, in the strict sense), Table 9 shows the results of a series of regressions with log average output value per firm (in region j) as a dependent variable. We allow firm output to be affected by the number of local competitors through the elasticity parameter $\eta/(1-\alpha)$ in the output function, as in (7). We estimate $\eta/(1-\alpha) = -0.26$. This negative value is consistent with the negative intra-industry spillovers we found in the reduced-form entry and exit regressions.[36]

We set the discount factor at $\beta = 0.925$. This leaves five key model parameters 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 1909–1914 as an industry equilibrium, where the industry agglomeration pattern was relatively stable. We calibrate 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

36. For robustness checks, we also consider smaller intra-industry spillover effects. For instance, we re-ran the model calibration and counterfactual analysis by assuming $\eta = 0$, in which case the effect of intra-industry spillovers is set to zero by construction. The results are reported in the Appendix (Tables 17-20).

Table 9

Average output value per firm.

Dependent variable: logarithm of average output value per auto firm in region j at time t

Sample period	1899–1910	1899–1919	1899–1935
Log center size	-0.255 (0.154)	-0.258*** (0.086)	-0.274*** (0.080)
Log population 1900	0.480*** (0.136)	0.750*** (0.098)	0.893*** (0.131)
Log per capita income 1900	-0.143 (0.468)	0.641** (0.251)	1.098*** (0.262)
Log C&W employment 1905	0.351** (0.160)	0.326*** (0.095)	0.451*** (0.103)
Log shipbuilding employment 1905	-0.205*** (0.042)	-0.242*** (0.032)	-0.109** (0.042)
Log iron and steel 1905	-0.201 (0.222)	-0.371*** (0.114)	-0.567*** (0.129)
Log lumber and timber 1905	-0.015 (0.077)	0.099** (0.045)	0.050 (0.049)
Log brass and copper 1905	0.059 (0.085)	0.049 (0.036)	0.062* (0.034)
Log rubber 1905	0.155*** (0.025)	0.163*** (0.017)	0.131*** (0.019)
Year dummies	Included	included	Included
Constant	8.168** (3.254)	3.913** (1.815)	3.197* (1.857)
Observations	132	283	430
R-squared	0.563	0.828	0.881

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

determined by the parameters c_1, α and σ, while the spinout rates are mainly driven by γ and γ_D. To best match the data, we set $c_1 = 0.6$, $\alpha = 0.7$, $\sigma = 1.96$, $\gamma_D = 0.08$, and $\gamma = 0.02$; the fit is very good, as shown in the following analysis.

Note that the calibrated value of σ, the average value of a firm's outside option, summarizes the effects of the market demand and production technology 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. Therefore, by adjusting the value of σ, we do not need to calibrate any scaling parameters in the market demand function and the production function.

The calibrated values of γ_D and γ suggest that Detroit had a substantial advantage in generating spinouts. Marx, Strumsky and Fleming (2009) offer a possible explanation for this pattern: Michigan was one of the few states that banned employee non-compete contracts at the time, which in turn encouraged the formation of spinout firms.[37]

■ **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 more likely to found in locations with more workers 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 and workers 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 presents 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 6 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 for the cross-location output distribution

Additional variables of interest include

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

37. According to Marx, Strumsky and Fleming (2009),

> At the turn of the 20th century, the metropolitan area of Detroit, Michigan, in many ways resembled the Silicon Valley of the last few decades. Growth of the nascent auto industry was explosive ... Ten years prior, the Michigan legislature in 1905 had passed statute 445.761 (bearing resemblance to California's prohibition): "All agreements and contracts by which any person agrees not to engage in any avocation or employment are hereby declared to be against public policy and illegal and void." This law governed non-compete enforcement until March 27, 1985, when the Michigan Antitrust Reform Act (MARA) repealed section 445 and with it the prohibition on enforcing non-compete agreements.

Related, Franco and Mitchell (2008) study more recent high-tech industries and argue the overtaking of Route 128 by Silicon Valley can be explained by the difference in noncompete policy between Massachusetts and California.

Table 10

Short-run decomposition. Model fit and six counterfactuals: 1. No intra-industry externalities; 2. Uniform C&W employment levels; 3. Uniform input resources; 4. Uniform other factors (population, income, shipbuilding, location dummies etc); 5. No spinout; 6. No Detroit-specific spinout rate

			Counterfactuals					
	Data	Model	1	2	3	4	5	6
Output share								
Chicago	0.060	0.037	0.028	0.037	0.028	0.038	0.074	0.066
Detroit	0.670	0.681	0.798	0.449	0.575	0.659	0.262	0.239
New York	0.140	0.081	0.067	0.054	0.032	0.192	0.166	0.148
Indianapolis	0.070	0.066	0.044	0.030	0.094	0.020	0.163	0.154
St Louis	0.010	0.011	0.008	0.021	0.021	0.009	0.022	0.019
Others	0.060	0.125	0.055	0.409	0.250	0.081	0.313	0.374
HHI	0.477	0.476	0.644	0.211	0.343	0.474	0.131	0.111
Prediction MSE		0.001	0.004	0.030	0.010	0.001	0.040	0.049
Exit rate								
Non-Detroit	0.150	0.133	0.137	0.132	0.133	0.136	0.121	0.122
Detroit	0.130	0.112	0.107	0.116	0.106	0.108	0.119	0.123
Additional moments								
Average regional spinout rate	0.030	0.044	0.047	0.038	0.038	0.041	0	0.018
Detroit top spinout rate	0.079	0.074	0.074	0.074	0.074	0.074	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 is done by adjusting the feed of n_j and $\mu(s, j)$ from our reduced-form regressions, while keeping other model parameters unchanged.[38] The results are shown in the last 6 columns in Table 10. In Counterfactual 1, we set $\eta = 0$, that is, we assume the presence of local auto companies has no effect on an auto firm's production function. In other words, we assume no intra-industry Marshallian effects. In Counterfactual 2, we set the variable "C&W industry employment" to have the same value in all regions (specifically, the average of production centers) when estimating the number and quality of de novo entrants based on the reduced-form regressions (cf Table 4, Spec 2; Table 8, Spec 3); and then feed those into our structural model to recompute the equilibrium. We thus effectively shut off the type of related industries effects described in Jacobs (1969). In Counterfactuals 3 and 4, 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 5 we force spinout rates to be zero in all regions, whereas in Counterfactual 6 we allow for spinouts but force the Detroit spinout rate to be the same as in other regions, hence shutting off the 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 with 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 also confirms that the cross-location output distribution predicted by our model is very close to the data.

Consider first Counterfactual 1, where we assume no intra-industry effects. The value of HHI in

38. When we shut down a specific model source of agglomeration, the industry price in the counterfactual equilibrium adjusts accordingly. This in turn changes the mass of entrants at each location. Since in our benchmark analysis we do not explicitly separate the role of potential entrants and that of entry cost at each location, we rely on the assumption that the industry price moves the mass of entrants at each location by the same proportion (as in their benchmark level). In this case, since the industry structure $m(s, j)$ is linearly homogeneous in the total mass of entrants, all our counterfactual results of cross-location concentration go through.

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

Factor	Δ HHI	% of HHI	% $\sum \Delta$ HHI
Intra-industry externalities	-0.168	-35.27	-29.09
Related industry (C&W)	0.265	55.64	45.89
Local input resources	0.133	27.92	23.03
Other local factors	0.003	0.55	0.45
Spinouts	0.345	72.41	59.72
Total	0.578	121.25	100.00

the counterfactual equilibrium is .644, higher than the base value of .476. In other words, we estimate (Table 9) that the effect of local competitors is negative, that is, the competition effect outweighs the positive externalities described in Marshall (1890). As a result, shutting off the effect of local competitors actually *increases* the degree of agglomeration. In other words, our model estimates that agglomeration took place not because of but rather in spite of intra-industry effects.

Next, consider Counterfactual 2. As mentioned earlier, this corresponds to forcing the values of C&W employment to be uniform across regions, thus shutting off Jacobs' (1969) related-industry effects. The value of HHI drops from .476 to .211. Moreover, the total output share of the five production centers falls significantly, from 87.5% to 59.1%. 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 location fixed factors, which we do in Counterfactual 3 and 4, also affects the value of HHI but to a less extent, dropping from .476 to .343 (input resources) or from .476 to .474 (other local variables). Overall, this suggests that region fixed effects (particularly input resources) 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 5, where we see a significant drop in HHI, from .476 to .131. Finally, Counterfactual 6 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 5. 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).

There are many ways to compute the relative contribution of each model feature. Based on the results from Table 10, Table 11 shows three different indicators: the change in HHI from omitting the particular model feature; the ratio between this change in HHI and the value of HHI in the baseline model; and the ratio between each contribution and the sum total of all contributions. We exclude Counterfactual 6 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.

According to the last column of Table 11, in the horse race between the various explanations for industry agglomeration, the development of C&W industry in each region and spinouts in Detroit seems to come out ahead, with a relative contribution of 45.89% and 59.72%, respectively. These results are largely consistent with the reduced-form regressions presented in Section 3.

■ **Agglomeration accounting: long-run analysis.** In the 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

Table 12

Long-run decomposition. Model fit and six counterfactuals: 1. No intra-industry externalities; 2. Uniform C&W employment levels; 3. Uniform input resources; 4. Uniform other factors (population, income, shipbuilding, location dummies etc); 5. No spinout; 6. No Detroit-specific spinout rate

			Counterfactuals					
	Data	Model	1	2	3	4	5	6
Output share								
Chicago	0.060	0.037	0.028	0.043	0.021	0.036	0.074	0.066
Detroit	0.670	0.681	0.798	0.635	0.489	0.689	0.262	0.239
New York	0.140	0.081	0.067	0.167	0.017	0.120	0.166	0.148
Indianapolis	0.070	0.066	0.044	0.015	0.049	0.033	0.163	0.154
St Louis	0.010	0.011	0.008	0.010	0.031	0.009	0.022	0.019
Others	0.060	0.125	0.055	0.131	0.393	0.113	0.313	0.374
HHI	0.477	0.476	0.644	0.434	0.248	0.492	0.131	0.111
Prediction MSE		0.001	0.004	0.002	0.027	0.001	0.040	0.049
Exit rate								
Non-Detroit	0.150	0.133	0.137	0.134	0.131	0.135	0.121	0.122
Detroit	0.130	0.112	0.107	0.111	0.102	0.105	0.119	0.123
Additional moments								
Average regional spinout rate	0.030	0.044	0.047	0.042	0.034	0.041	0	0.018
Detroit top spinout rate	0.079	0.074	0.074	0.074	0.074	0.074	0	0.019

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.

In other words, one important reason why many new auto firms were located in Detroit is the fact that many C&W firm were already located 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 2 we now treat the residual terms from the C&W employment regression (cf Table 5) as random C&W shocks which are not explained by the location fixed factors we study. Holding everything else constant, we equalize the random C&W shocks across regions (at the average across all 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 2; Table 8, 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.

Similarly, in Counterfactuals 3 and 4, 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, whereas the contribution of C&W drops considerably.

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 and spinouts now appear as the leading contributors, contributing 52.88% and 79.71%,

Table 13
Long-run agglomeration accounting: relative contribution of various model components

Factor	Δ HHI	% of HHI	% $\sum \Delta$ HHI
Intra-industry externalities	-0.168	-35.27	-38.83
Related industry (C&W)	0.042	8.88	9.78
Local input resources	0.229	48.04	52.88
Other local factors	-0.015	-3.21	-3.54
Spinouts	0.345	72.41	79.71
Total	0.433	90.85	100.00

respectively, to explaining the increase in total concentration. The fact these two percentages add up to more than 100% stems from the negative contribution of intra-industry externalities (as explained before), as well other local effects.

■ **Further discussions of agglomeration accounting.** Our agglomeration accounting exercise shows that in the short run, given that the C&W industry is already in place, its employment level contributes the most (together with spinouts) 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 (again, together with spinouts).

The above agglomeration accounting exercises are based on the assumption that the Marshall-type intra-industry spillover effect is negative (and has a value consistent with our reduced-form regressions). However, depending on the sample range we look at (cf. Table 9), the effect of intra-industry spillovers on firm output can be statistically insignificant, a pattern which is also found in some of our entry and exit regressions. As a robustness check, we re-ran our calibration and counterfactual exercises by assuming $\eta = 0$, that is, by imposing zero intra-industry spillovers. The results, shown in the Appendix (Tables 17-20), are somewhat different with respect to the case when we allow for negative intra-industry spillovers. That said, the overall message of our numerical exercise remains valid: In the short run, C&W spillovers and employee spinouts contribute most to the agglomeration; while in the long run, local inputs and employee spinouts matter most.

Moreover, it is worth noting that the effect of spinouts takes place almost exclusively from allowing the Detroit rate to be different, which is related to the fact that the state of Michigan banned non-compete agreements at the time, thus facilitating spinouts. In this sense, the contribution of spinouts to agglomeration relies on region-specific effects, this time the variation in regulations cross states. Accordingly, one could also interpret our findings, regardless of model specifications, as implying that location-specific effects (local inputs and location-specific spinout rates) accounted for the lion's share of the auto industry agglomeration both in the short run and in the long run.

5. Conclusion

Taking the early U.S. auto industry as an example, we evaluate four competing hypotheses regarding regional industry agglomeration: intra-industry local externalities, inter-industry local externalities, employee spinouts, and location fixed effects. Our findings suggest that in the automobile case, inter-industry local externalities (particularly from the carriage and wagon industry) and employee spinouts (particularly due to the high spinout rate in Detroit) play important roles. 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, it appears to play a much more important role. Finally, the presence of other firms in the same industry has a negligible or even negative effect on agglomeration.

Although we focus on the auto industry, we believe the patterns we find are likely found in other industries as well. For example, the relation we find between C&W and auto producers is similar to that found between radio and TV producers (Klepper and Simons, 2000). Also, the relation between parent and spinout in autos is similar to the auto-tire industry clustered in Akron (Buenstorf and Klepper, 2009) and the IT industry clustered in Silicon Valley (Klepper, 2010). Finally, we also find location fixed effects on industry agglomeration that are consistent with many other manufacturing industries (Ellison and Glaeser, 1997, 1999).

Our findings have an important policy implication: in order to boost regional development, policymakers should evaluate location advantages (in terms of input resources and existing industries) before choosing to promote certain target industries. All too often, one finds countries, and regions within countries, eager to create “the next Silicon Valley” by means of a “cluster strategy” (essentially, a strategy that takes advantage of agglomeration and co-agglomeration economies). However, by failing to understand the dynamics of industry agglomeration, as well as the importance of location fixed effects, many of these initiatives fall very short of expectations.

Our study can be extended in several directions. First, while our paper focuses on the historical evolution of auto industry, the analysis can be applied to other industries. In particular, it would be interesting to explore similarities and differences with respect to more recent high-tech industries (both in terms of agglomeration patterns and in terms of driving forces). Second, due to data limitation, we use entry and exit as proxy measures of firm performance. Provided richer data becomes available, future studies could use more direct measures of firm performance. 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.

References

Arrow K.J. (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.

Bloom, Nicholas, Mark Schankerman and John Van Reenen (2013), "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica* **81(4)**, 1347–1393.

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

Cabral, Luís (2012), "Technology Uncertainty, Sunk Costs, and Industry Shakeout," *Industrial and Corporate Change* **21**, 539–552.

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.

Franco, April, and Matthew Mitchell (2008), "Covenants not to Compete, Labor Mobility, and Industry Dynamics," *Journal of Economics and Management Strategy* **17(3)**, 581–606.

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

Gort, M., and S. Klepper (1982), "Time Paths in the Diffusion of Product Innovations," *Economic Journal* **92**, 630–653.

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.

Jovanovic, Boyan and Glenn M. MacDonald (1994), “The Life Cycle of a Competitive Industry,” *Journal of Political Economy* **102** (2), 322–347.

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

Klepper, Steven (2002), “Firm Survival and the Evolution of Oligopoly,” *RAND Journal of Economics* **33** (1), 37–61.

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.

Klepper, Steven and Kenneth Simons (2000), “Dominance by Birthright: Entry of Prior Radio Producers and Competitive Ramifications in the U.S. Television Receiver Industry,” *Strategic Management Journal* **21(10–11)**, 997–1016.

Klepper, Steven and Kenneth Simons (2005), “Industry Shakeouts and Technological Change,” *International Journal of Industrial Organization* **23(1–2)**, 23–43.

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.

Marx, Matt, Deborah Strumsky, and Lee Fleming (2009), “Mobility, Skills, and the Michigan Non-Compete Experiment,” *Management Science* **55(6)**, 875–889.

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.

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

$$\log\left(\sum_{s,j} q^*(s,p,m_j)\,m(s,j)\right) = \log(Z) - b\,\log(p),$$

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

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

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

$$\log(p^*) = \frac{\log(Z)}{\frac{\alpha}{1-\alpha}+b} - \frac{1}{\frac{\alpha}{1-\alpha}+b}\,\log\left(\sum_{s,j}(\alpha)^{\frac{\alpha}{1-\alpha}}\Big(\exp(c_1\,s)m_j^{\eta}\Big)^{\frac{1}{1-\alpha}}\,m(s,j)\right).$$

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.[39] 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:

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

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

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

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

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

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

39. 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.

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

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

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.

■ **Proofs.** The proofs of Propositions 1–3 follow.

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 $VC(s; j, \overline{p})$ is continuous in s and strictly increasing in s for $\overline{p} > 0$. Thus we know that for each period, $F(VC(s; j, \overline{p}))$ is strictly increasing in s, given the same location j. ■

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. ■

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. ■

■ **Additional regression results (Tables 14, 15, 16).**

Table 14
Spinout location and spinout performance.
Logit models of firm exit, 1900-1935. Dependent variable: firm exit at time t

	Spec 1	Spec 2	Spec 3	Spec 4
Firm age	-0.069*** (0.010)	-0.056*** (0.009)	-0.077*** (0.010)	-0.064*** (0.010)
Dealio	-0.458*** (0.107)	-0.454*** (0.105)	-0.479*** (0.108)	-0.471*** (0.105)
Spinout	-0.432*** (0.156)	-0.240 (0.158)	-0.421*** (0.159)	-0.232 (0.160)
Center size	0.009 (0.006)		0.014* (0.007)	
Local family size	-0.102** (0.042)		-0.113** (0.048)	
Non-local family size	-0.053 (0.074)		-0.092 (0.084)	
Center top		0.020 (0.033)		0.022 (0.034)
Local family top		-0.448*** (0.097)		-0.482*** (0.106)
Non-local family top		-0.083 (0.125)		-0.135 (0.149)
Log auto employment	0.055 (0.060)	0.072 (0.058)	0.133** (0.066)	0.155** (0.067)
Log population 1900	0.165 (0.167)	0.185 (0.165)	0.282* (0.165)	0.299* (0.162)
Log per capita income 1900	0.352 (0.302)	0.324 (0.298)	0.175 (0.317)	0.124 (0.312)
Log C&W employment 1905	-0.123 (0.148)	-0.160 (0.139)	-0.263* (0.150)	-0.309** (0.144)
Log shipbuilding employment 1905	0.025 (0.045)	0.032 (0.045)	0.011 (0.049)	0.021 (0.049)
Log local input resources 1905	-0.082 (0.109)	-0.093 (0.112)	-0.114 (0.112)	-0.125 (0.114)
St Louis	0.161 (0.317)	0.180 (0.322)	0.033 (0.316)	0.090 (0.315)
Chicago	-0.275 (0.223)	-0.128 (0.193)	-0.377 (0.241)	-0.130 (0.195)
Indianapolis	-0.551** (0.267)	-0.415* (0.228)	-0.671** (0.291)	-0.430* (0.234)
Detroit	-0.654** (0.301)	-0.372 (0.484)	-0.929** (0.362)	-0.444 (0.501)
Rochester	-0.373 (0.286)	-0.316 (0.273)	-0.442 (0.286)	-0.327 (0.268)
New York City	-0.029 (0.219)	0.106 (0.202)	-0.112 (0.233)	0.133 (0.202)
Year or year dummies	0.023 (0.015)	0.012 (0.014)	year dummies	year dummies
Constant	-46.127 (29.746)	-25.398 (26.749)	-0.365 (1.943)	-0.041 (1.913)
Observations	3,573	3,573	3,537	3,537

Notes: Center size, Local family size, Non-local family size, Center top, Local family top and Non-local family top one-year lagged; Log auto employment is calculated by taking the logarithm of three-year lagged average of auto employment in each location.

Robust standard errors clustered by firm. Star levels: 10, 5 and 1%.

Table 15

Logit models of spinout firm exit, 1900-1935. Dependent variable: firm exit at time t

	Spec 1	Spec 2	Spec 3	Spec 4
Firm age	-0.073*** (0.025)	-0.071*** (0.024)	-0.075*** (0.027)	-0.074*** (0.027)
Center size	0.038** (0.017)		0.043* (0.025)	
Family size	-0.058 (0.044)		-0.081 (0.053)	
Center top		0.143 (0.107)		0.239** (0.121)
Family top		-0.209*** (0.076)		-0.208*** (0.080)
Log auto employment	-0.301 (0.333)	-0.162 (0.323)	-1.556*** (0.552)	-1.860*** (0.708)
Log population 1900	-2.222 (2.022)	-1.039 (1.854)	-6.529** (2.895)	-7.837** (3.898)
Log per capita income 1900	-2.123 (2.171)	-1.225 (2.025)	-3.310 (3.171)	-5.253 (4.571)
Log C&W employment 1905	0.250 (0.825)	0.113 (0.787)	1.282 (1.003)	1.539 (1.016)
Log shipbuilding employment 1905	0.250 (0.538)	0.238 (0.485)	0.293 (0.608)	0.427 (0.583)
Log local input resources 1905	1.137 (1.885)	0.393 (1.729)	3.979 (2.455)	4.893 (3.164)
St Louis	-0.678 (0.973)	-0.326 (0.785)	-0.639 (1.194)	0.184 (1.138)
Chicago	-1.246* (0.745)	-0.781 (0.690)	-1.722* (0.970)	-1.294 (0.836)
Indianapolis	-1.478*** (0.571)	-0.829* (0.430)	-2.027*** (0.764)	-1.369** (0.572)
Detroit	-2.491*** (0.939)	-2.605 (1.726)	-2.145* (1.189)	-3.674* (1.906)
Rochester	-0.753 (0.494)	-0.901* (0.488)	-0.396 (0.609)	-0.434 (0.569)
New York City	-0.129 (0.583)	0.059 (0.570)	-0.268 (0.759)	-0.113 (0.645)
Year	0.129** (0.064)	0.079 (0.057)	dummies	dummies
Constant	-240.071** (122.213)	-146.257 (109.043)	6.269 (9.670)	10.966 (10.993)
Observations	720	720	658	658

Notes: Center size, Family size, Center top and Family top one-year lagged; Log auto employment is calculated by taking the logarithm of three-year lagged average of auto employment in each location.
Robust standard errors clustered by firm. Star levels: 10, 5 and 1%.

Table 16

Logit models of non-spinout firm exit, 1900-1935. Dependent variable: firm exit at time t

	Spec 1	Spec 2	Spec 3	Spec 4
Firm age	-0.064*** (0.011)	-0.039*** (0.012)	-0.079*** (0.013)	-0.055*** (0.014)
De Alio	-0.447*** (0.108)	-0.441*** (0.105)	-0.470*** (0.108)	-0.459*** (0.105)
Center size	0.006 (0.007)		0.010 (0.008)	
Family size	-0.242 (0.184)		-0.226 (0.185)	
Center top		0.005 (0.037)		0.008 (0.037)
Family top		-1.072*** (0.251)		-1.089*** (0.258)
Log auto employment	0.065 (0.062)	0.076 (0.060)	0.174** (0.070)	0.196*** (0.071)
Log population 1900	0.227 (0.177)	0.212 (0.170)	0.335* (0.174)	0.307* (0.166)
Log per capita income 1900	0.348 (0.312)	0.390 (0.323)	0.018 (0.331)	0.023 (0.336)
Log C&W employment 1905	-0.185 (0.157)	-0.155 (0.142)	-0.375** (0.159)	-0.356** (0.146)
Log shipbuilding employment 1905	0.022 (0.047)	0.037 (0.048)	0.004 (0.052)	0.019 (0.053)
Log local input resources 1905	-0.072 (0.111)	-0.117 (0.115)	-0.093 (0.116)	-0.136 (0.118)
St Louis	0.290 (0.354)	0.231 (0.370)	0.198 (0.345)	0.170 (0.361)
Chicago	-0.147 (0.244)	-0.059 (0.211)	-0.220 (0.265)	-0.046 (0.214)
Indianapolis	-0.447 (0.294)	-0.425* (0.254)	-0.555* (0.324)	-0.464* (0.269)
Detroit	-0.439 (0.342)	-0.121 (0.544)	-0.697* (0.395)	-0.237 (0.542)
Rochester	-0.339 (0.303)	-0.279 (0.288)	-0.374 (0.308)	-0.264 (0.291)
New York City	-0.028 (0.242)	0.149 (0.219)	-0.055 (0.254)	0.214 (0.219)
Year	0.016 (0.016)	0.005 (0.015)	dummies	dummies
Constant	-32.965 (31.472)	-10.883 (29.284)	0.469 (2.091)	0.805 (2.087)
Observations	2,853	2,853	2,832	2,832

Notes: Center size, Family size, Center top and Family top one-year lagged; Log auto employment is calculated by taking the logarithm of three-year lagged average of auto employment in each location.
Robust standard errors clustered by firm. Star levels: 10, 5 and 1%.

■ Model Calibration and Simulations without Intra-industry Spillovers ($\eta = 0$)

Table 17
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	0.060	0.034	0.037	0.027	0.036	0.059	0.058
Detroit	0.670	0.675	0.295	0.489	0.492	0.369	0.371
New York	0.140	0.090	0.055	0.032	0.399	0.166	0.164
Indianapolis	0.070	0.101	0.040	0.186	0.016	0.212	0.215
St Louis	0.010	0.008	0.021	0.018	0.007	0.013	0.013
Others	0.060	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	0.150	0.146	0.145	0.145	0.147	0.142	0.142
Detroit	0.130	0.135	0.142	0.134	0.136	0.136	0.137
Additional moments							
Average regional spinout rate	0.030	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

Table 18
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
Other local factors	0.074	15.55	8.12
Spinouts	0.263	55.25	28.87
Total	0.911	191.39	100.00

Table 19
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 20
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
Other local factors	0.018	3.78	2.35
Spinouts	0.263	55.25	34.29
Total	0.767	161.13	100.00