The Determinants of Geographic Concentration of Industry: A Quantitative Analysis*

Luis Cabral, Zhu Wang and Daniel Yi Xu†

October, 2011

Abstract

Taking the early U.S. automobile industry as an example, we evaluate two competing hypotheses on geographic concentration of industry: local externalities *versus* employee spinoffs. Our findings suggest that both forces contribute to industry agglomeration through their specific channels, and the spinoff effect can be viewed as a special form of local externalities. Calibrating our model to the quantitative pattern of industry evolution reveals that traditional local externalities are main driving forces of agglomeration. Particularly, the local economy and related industries play an important role by fostering new entrants. Spinoffs play a secondary role and contribute to an increased concentration at later stages of the industry life cycle.

JEL classification: J6; L0; R1

Keywords: Local externalities; Employee spinoffs; Industry agglomeration

*We thank Ed Glaeser, Boyan Jovanovic, Bob Lucas and participants at the 2010 NBER Summer Institute Macro, Trade and Productivity Meeting for helpful comments, and Christian Hung and Elizabeth Antonious 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.

†**Cabral**: Department of Economics, Stern School of Business, New York University. Email: lcabral@stern.nyu.edu. **Wang:** Research Department, Federal Reserve Bank of Richmond. Email: zhu.wang@rich.frb.org. **Xu**: Department of Economics, Duke University. Email: daniel.xu@duke.edu.

1 Introduction

Geographic concentration is a striking feature of many industries. Detroit and Silicon Valley are two most prominent examples. This concentration is too great to be explained by exogenous spatial differences in natural advantage (Ellison and Glaeser, 1999), and an abundance of theories has been proposed.

Many authors consider that agglomeration results from local externalities: the presence of other firms in the same location increases each given firm's productivity (Marshall 1920, Krugman 1991, Rosental and Strange 2004). The nature of such externalities, and the channel through which they take effect, may vary: A firm may benefit from the overall size of the local economy due to better access to input suppliers or final consumers; at the industry level, location proximity allows for labor pooling and knowledge spillovers among industry participants; and finally, a firm may benefit from the influence of related industries in the same location. Although these externalities all predict conglomeration, it remains at issue to assess their validity and relative importance.

In contrast, some recent studies argue that agglomeration is primarily driven by organizational reproduction and heredity (Klepper 2007, 2010, Buenstorf and Klepper 2009). It was pointed out that a large number of industry entrants result from spinoffs – former managers or employees who leave an existing firm to start a new firm in the same industry. As a result, agglomeration may form just because spinoffs are generally high performers and they locate near parents. However, some key questions remain: What's the quantitative importance of spinoffs versus traditional local externalities in accounting for agglomeration? Is spinoff's heredity related to traditional local externalities?

In this paper, we try to address these issues and investigate the competing views. First, we construct a model of industry dynamics that incorporates two types of entrants: *De Novo* entrants and *Spinoffs*. The former refers to new entrants from outside the industry and the latter refers to employee spinoffs from within the industry. Both entrants are subject to various local externalities, and spinoffs are also affected by other firms sharing the same family origin.

Second, we put together a unique dataset of the early U.S. automobile industry. Guided by our theory, we test the impact of local externalities and spinoffs on firm turnover and how that affects the evolving pattern of geographic concentration over the industry life cycle. We find that traditional local externalities and spinoffs work through different channels. The local economy and related industries have positive influences on the quantity and quality of *De Novo* entrants. Spinoffs' performance is heavily affected by local family members but not distant ones, which suggests that spinoffs may indeed enjoy a special type of local externalities rather than gene reproduction because the latter would not depend on local proximity. Controlling for spinoff family effects, we find that the local industry size has a negative effect on firm performance, suggesting the presence of competition-induced congestion, rather than a positive externality.

Finally, we calibrate our model to the quantitative pattern of industry evolution. Our findings suggest that traditional local externalities are main driving forces of agglomeration. Particularly, the local economy and related industries play a very important role by fostering new entrants. In comparison, spinoffs play a secondary role and contribute to an increased concentration at later stages of the industry life cycle.

Our paper is related to the growing literature on spinoffs but focuses on a different perspective. Many recent studies in the literature try to identify and explain various motivations of spinoffs.[1] In contrast, this paper aims to quantify the impact of spinoffs on industry agglomeration. Therefore, we abstract from different motivations of spinoffs but rather emphasize the fact that being a member of a spinoff family creates a special social network for sharing information and resources, and this may have important implications on industry agglomeration.

The paper is organized as follows. In section 2, we provide a model of industry

[1] For example, Klepper and Thompson (2005) point out that spinoffs likely result from "disagreement" between employees and employers. Chatterjee and Rossi-Hansberg (2007) argue that due to adverse selection, employees who have better information about the value of their ideas may choose to start a new firm. Franco and Filson (2006) stress the fact that employees acquire know-how while working for a firm and eventually capitalize on that know-how by starting their own firm. Cabral and Wang (2009) develop a "passive learning" model of firm entry by spinoffs: Firm employees leave their employer and create a new firm either when they learn they are good entrepreneurs or they learn their employer's prospects are bad.

dynamics that incorporates both *De Novo* and spinoff entrants, and derive testable hypotheses. In section 3, we test the spinoff "family effect" and other local externalities using a unique dataset of the early U.S. automobile industry. In section 4, we calibrate our model to match quantitative patterns of the U.S. automobile industry, and disentangle the impact of different driving forces of agglomeration. Section 5 concludes.

2 The Model

In this section, we provide a simple model of industry dynamics in the spirit of Hopenhayn (1992). Firms are heterogeneous in their production capabilities and compete in a competitive market. They are forward-looking and make optimal entry and exit decisions. There are two types of entrants in our model. The first type is new entrants from outside the industry, so-called *De Novo* entrants. The potential number of *De Novo* entrants are affected by traditional channels of local externalities, such as local population, income, as well as the presence of related industries. The second type is experienced entrants from within the industry, so-called *spinoffs*. We assume an incumbent firm has a constant rate generating a *potential* spinoff each period, however, these potential spinoffs need to make their optimal entry decision. We also make the theoretical abstraction that a spinoff shares the same capability with its parent firm. This family-specific capability may result from knowledge linkage or business relation within the family network. To some extent, we may regard it as a special form of local externalities, which we name it *family network effect*.

2.1 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 starting at location j will operate at the same location for the rest of its life. The industry structure is summarized by $m(s, j)$, the total mass of firms of capability s at location j.

At each period, all incumbent firms engage in product market competition by taking industry price p as given. Each firm decides on the optimal quantity of output $q(s; j, p)$ based on its capability and location characteristics. Their period profit is denoted by $\pi(s; j, p)$. Both functions $q(s; j, p)$ and $\pi(s; j, p)$ have the standard properties, e.g., strictly increasing in s, decreasing in p, continuous and bounded. Once an incumbent firm obtains its profit, it decides whether to continue operating in this industry or to leave by taking some outside options ϕ^x . The distribution of these outside options $F(\phi^x)$ is i.i.d. and privately known by individual firms. Given its belief of a sequence of industry price $\bar{p}$, an incumbent's problem can be defined as:

$$V(s; j, \bar{p}, \phi^x) = \pi(s; j, p) + max\{VC, \phi^x\}, \tag{1}$$

where the value of continuation is

$$VC(s; j, \bar{p}) = \beta \int V(s; j, \bar{p}, \phi^{x\prime}) dF(\phi^{x\prime}). \tag{2}$$

Potential entrants at each location make their entry decisions at the same time that incumbents are making their exit decisions. The first type, "*De Novo* entrants", is from outside the industry. The mass of potential *De Novo* entrants M_j at location j is determined by location specific characteristics. A potential *De Novo* entrant pays a random sunk cost ϕ^e, and has the initial draw of capability s from distribution $\mu(s, j)$. Hence, a potential entrant's probability of entry is

$$\Psi_j = \Pr(\int_s VC(s; j, \bar{p}) d\mu(s, j) \geq \phi^e). \tag{3}$$

As a result, the expected number of *De Novo* entrants at location j is

$$n_j = M_j \Psi_j. \tag{4}$$

The second type of entrants, "spinoffs," is from within the industry. Each period, an incumbent firm at location j has a probability γ_j of generating a potential spinoff. We assume the potential spinoff shares the same capability s with its parent, and knows its capability while making the entry decision. A potential spinoff will enter if its value of entry is higher than its random outside option ϕ^x, i.e.

$$VC(s; j, \bar{p}) \geq \phi^x. \tag{5}$$

If it chooses not to enter in the current period, its opportunity is foregone once for all.

There are several important differences between these two types of entrants. First, while potential *De Novo* entrants are uncertain about their capability of operating in a new industry, spinoff entrants directly inherit their parent's capability draw. This is a sharp assumption we make to highlight the fact that spinoff entrants have better knowledge of their own capability given their experience of working in this industry. Second, we assume that *De Novo* entrants need to pay an additional entry cost ϕ^e compared with spinoffs, which corresponds to the extra investment *De Novo* entrants need to make to build up business relation and customer base in the industry.

2.2 Transition of Industry Structure and Price

Let's first define the transition of the mass of firms of capability s at location j, with current industry price p. It depends on the number of exits, spinoffs, and *De Novo* entries at each state (s, j). Explicitly, we have

$$m'(s, j) = m(s, j)(1 + \gamma_j)\chi_{s,j} + n_j \mu(s, j), \tag{6}$$

where $\chi_{s,j} = F(VC(s; j, \bar{p}))$ is the probability of staying in the industry given the cdf function F of the outside option. The first item on the right hand side combines the decisions of incumbents and spinoff entrants. There are in total $m(s, j)(1 + \gamma_j)$ firms making their decisions on whether to take the outside option. Given that their value of continuation is the same, their probability of continuation $\chi_{s,j}$ is also the same. The second item on the right hand side is the inflow of *De Novo* entrants. The potential *De Novo* firms are ex ante identical in terms of their expected capability, however, the size of potential entrants pool is location specific. This explains the heterogeneity in n_j. The distribution of capability $\mu(s, j)$ is allowed to vary across locations in our theoretical setup, but in the empirical analysis we will restrict it to be the same across locations to avoid over-parametrization of our model.

Next, we define the industry output market. Total industry demand is given by the inverse demand function $p = D^{-1}(Q)$. Industry price clears the market each period such that

$$p = D^{-1}\{\sum_{s,j} q(s; j, p)m(s, j)\}. \tag{7}$$

2.3 Industry Equilibrium

An industry equilibrium is defined by a sequence of prices $\bar{p}^*$, a mass of entrants n_{jt}^* at each location in each period, a measure of incumbent firms $m^*(s, j, t)$, and policy function χ^* such that

- χ^* solves incumbent firms and potential spinoffs' dynamic optimization problem each period, given their belief of $\bar{p}^*$;
- n_{jt}^* satisfies entry condition for *De Novo* entrants each period;
- p^* clears product market each period;
- $m^*(s, j, t + 1)$ is defined recursively given $m^*(s, j, t)$, n_{jt}^*, and χ^*.

To aid our empirical analysis, we further restrict the notion of our equilibrium to be stationary. Following immediately Hopenhayn (1992), there exists a stationary equilibrium defined as an time-invariant output price p^*, a mass of entrants n_j^* at each location, a measure of incumbents $m^*(s, j)$, and policy function χ^* so that the equilibrium condition holds. This stationary equilibrium allows us to further derive several testable implications of our model and solve it quantitatively. Specifically we can take a certain stage of the industry life cycle and fit our model quantitatively as long as firms are assumed to have stationary belief about the industry price.

2.4 Model Implications

In this section we introduce several testable implications of our model. We specifically focus on the entry and survival patterns of spinoff entrants and contrast those with *De Novo* entrants. Note that the first two propositions are derived directly as a result of individual firm's optimization condition, so they do not depend on the stationary equilibrium concept.

Proposition 1 *Conditional on the same location, a potential spinoff firm is more likely to enter than another one if it belongs to a higher capability family.*

Proof. Given that $\pi(s; j, p)$ is strictly increasing in s, continuous, and bounded, standard dynamic programming argument shows that $VC(s; j, \bar{p})$ is continuous in s and strictly increasing in s for $\bar{p} > 0$. Thus we know that for each period, $F(VC(s; j, \bar{p}))$ is strictly increasing in s, given the same production location j. ■

Proposition 2 *A high-capability family has higher probability of generating spinoffs and thus on average has a bigger family size, conditional on the same location and time.*

Proof. Note that all incumbents at location j have the same probability γ_j of having a potential spinoff, while the spinoff's probability of entering $\chi^*_{s,j,t}$ is increasing in s. More specifically, the average size of a capability s family at location j and time t is $(1 + \gamma_j)\chi^*_{s,j,t}$. As shown above, higher capability incumbents have higher probability of generating spinoffs and they themselves are less likely to exit. Thus higher-capability family has bigger family size on average. ■

Proposition 3 *Given positive entry and exit in the stationary equilibrium, spinoff firms have lower probability to exit than De Novo entrant, given the same location j.*

Proof. The stationary distribution is defined by $m^* = m^*(1+\gamma_j)\chi^* + n_j\mu$, so $m^* = \frac{n_j}{1-(1+\gamma_j)\chi^*}\mu$. The distribution of spinoff firms is $m^*\chi^* = \frac{n_j\chi^*}{1-(1+\gamma)\chi^*}\mu$. Since χ^* is strictly increasing in s, the capability distribution of spinoff firms strictly dominates that of *De Novo* firms, which is μ. ■

3 Empirical Analysis

In this section, we test our theoretical hypotheses using a unique dataset of the U.S. automobile industry. The 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 with their characteristics.

3.1 Data Sources

The data comes from several sources. First, Smith (1970) provides a list of every make of automobile 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 was used to derive entry, exit and location of firms.[2]

Second, Kimes (1996) provides comprehensive information for every automobile make produced in the U.S. from 1890 through 1942. Using Kimes (1996), we were able to collect additional biographical information about the entrepreneurs who founded and ran each individual firm. An entrepreneur was categorized into one of the following three groups: *De Alio* entrants, Spinoff entrants and *De Novo* entrants.[3] The first group includes entrepreneurs who had prior experience in related industries before starting an automobile firm. The second group includes entrepreneurs who had worked as employees

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

[3] Note that the definition of *De Novo* entrants in our empirical study is slightly different from that in our theory. Particularly, we separate a subgroup and denote them as *De Alio* entrants.

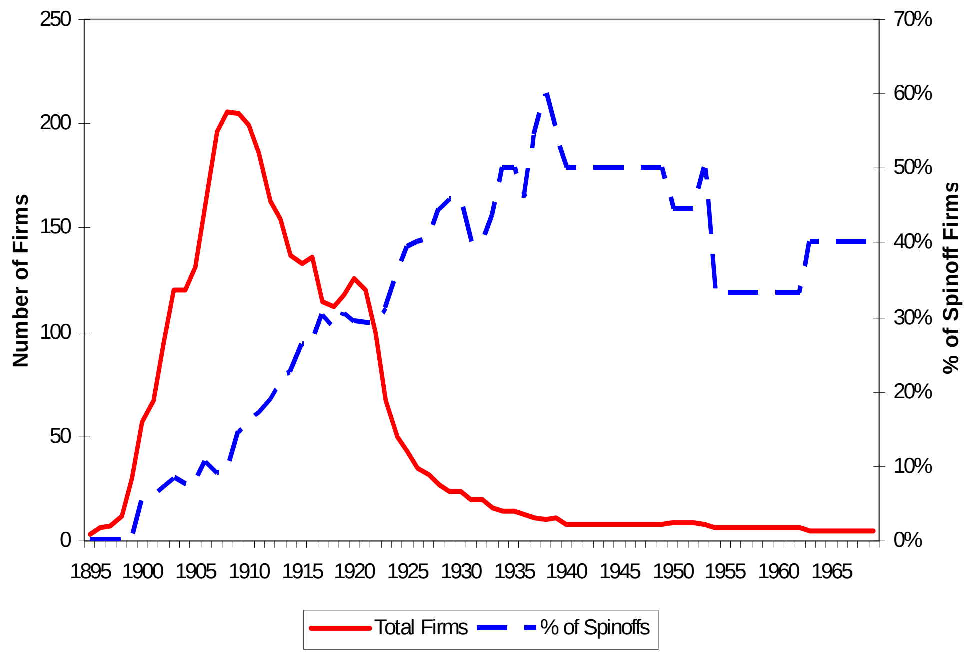


Figure 1: The Evolution of US Automobile Industry: 1895-1969

in existing automobile firms before starting their own. The last group includes those with no identifiable background information. Kime's information was used to derive family linkages between individual firms. In other words, we constructed family trees for spinoff firms.

Third, Bailey (1971) provides a list of leading automobile makes from 1896-1970 based on top-15 annual sales. We used this information to identify top automobile producers during these periods.

3.2 Industry Overview

As shown in Figure 1, the automobile industry went through a tremendous development in its first 75 years, evolving from a small infant industry to a gigantic, concentrated, mature industry. During the process, the number of firms initially rose and later fell. In its peak years around 1910, the industry had more than 200 producers, but only eight firms survived into 1940s. Meanwhile, the percent of spinoff firms increased from almost zero in 1900 to 60 percent in 1940. After that, the ratio of spinoffs started to decline as some top parent firms, which were not spinoffs themselves, eventually outlived their spinoffs.

The industry also went through substantial changes of geographic concentration pattern over those years. Using the number of firms as the criterion, we identified six historically important automobile production centers, namely St. Louis, Chicago, Indianapolis, Detroit, Rochester and New York City.[4] As shown in Figures A1 and A2 in the Appendix, the industry initially started in New York City and Chicago in the late 1890s. Soon after, Detroit and other centers caught up. In 1905, 25 percent firms located in Detroit and they produced more than 50 percent industry output. Meanwhile, 16 percent firms located in New York City, 10 percent in Chicago, 8 percent in Indianapolis, 7 percent in Rochester, 2 percent in St. Louis, and the remaining 32 percent firms scattered in other places across the country. Over time, Detroit gained an increasing share of firms as well as industry output. In 1920, 35 percent firms located in Detroit and they produced more than 70 percent of industry output.

The number of firms also varied substantially across spinoff families over time. Using the family trees that we constructed, we identified a total of 53 spinoff families that ever existed in the industry. The three largest families were GM, Ford and Oldsmobile, all located in Detroit, each generating 12-17 spinoffs (Table A1 in the Appendix describes the family tree of GM as an example). As expected, most spinoffs located near their parents, for example, 76 percent of spinoffs in the top three families stayed in Detroit. Figure A3 in the Appendix describes the variation of the family size distribution over time. As it shows, few spinoffs were around in the early years of the industry. For example, only 12 firms, out of a total of 155 firms in 1905, belonged to multi-member families. Over time, spinoffs became increasingly important in the industry. In 1920, there were 41 firms, out of a total of 136 firms, belonging to multi-member families.

[4] A city is counted as an automobile production center city if it had at least five automobile producers in year 1910. We then define the region within 100 miles of the center city as the production center, named after the center city.

3.3 Regression Analysis

To summarize, our automobile industry dataset includes the following information: (1) the entry year of each firm, (2) the exit year of each firm, (3) the type of each exit, (4) the background of entrepreneurs: *De Novo, De Alio* and Spinoffs, (5) the capability of firms in terms of whether they were ever top sellers in the industry, (6) the location of each firm, (7) the six automobile production centers, and (8) the family linkage of each spinoff firm.

Using the above information, we create the following dummy variables used in our regressions (indexed by firm and year). Whenever needed, additional explanation is given in parentheses.

◦ Firm Death (A firm exits in the current period).

◦ Spinoff Birth (A firm generates a spinoff entrant in the current period).

◦ Production Center (Seven dummies corresponding to St. Louis, Chicago, Indianapolis, Detroit, Rochester, New York City and the other places).

◦ *De Novo* (The firm was founded by an unexperienced entrant).

◦ *De Alio* (The firm was founded by an experienced entrant from related industries).

◦ Spinoff (The firm was founded by a spinoff in the auto industry).

In addition, we create the following variables:

◦ Center Size (The number of auto firms in the center where the firm locates).

◦ Center Top (The number of top auto firms in the center where the firm locates).

◦ Family Size (The number of auto firms in the family to which the firm belongs).

◦ Family Top (The number of top auto firms in the family to which the firm belongs).

◦ Local Family Size (Family Size within where the firm is located).

◦ Local Family Top (Family Top within where the firm is located).

◦ Non-Local Family Size (Family Size outside where the firm is located).

◦ Non-Local Family Top (Family Top outside where the firm is located).

◦ Firm Age.

◦ Year.

Tables A2 and A3 in the Appendix provide descriptive statistics of the regression variables, both at the firm level and the firm-year level. Table A2 shows that the average firm death rate was about 17 percent per year and the average age of a firm was about 7 years. These are not unusual for a growing industry. Meanwhile, the firm spinoff (birth) rate is about 2 percent per year. At the firm-year level, Table A2 also shows that 21 percent of firms were *De Novo* entrants, 59 percent were *De Alio* entrants, 20 percent were Spinoffs, and 20 percent firms ever made the top firm list. On average, a location had 36 firms (the range was from 1 to 96), among which 6 were top ones (the range was from 0 to 18). An average firm belonged to a family of 1.5 members (the range was from 1 to 10), among which 0.5 were top ones (the range was from 0 to 6). Table A3 shows that, at the firm level, 30 percent of firms were *De Novo* entrants, 52 percent were *De Alio* entrants and 17 percent were Spinoffs. Among all firms, 6 percent of them ever made the top producer list.

In the following analysis, we ran regressions using firm-year observations with "Firm Death" (for incumbents) or "Firm Birth" (for spinoff or non-spinoff entrants) being the dependent variable. According to our theory, these exercises can be viewed as estimating firms' policy functions of entry and exit. The data range we use is from 1895-1942, including 771 firms and 4454 firm-year observations.[5]

3.3.1 Firm Death Analysis

Assuming firms' outside options ϕ^x follows a logistic distribution, Eq (1) in our theory suggests a logit regression model for firm exit analysis. The basic logit model is equivalent to a discrete-time duration model under the assumption that the baseline hazard is constant over time. However, by including firm age and year effects, we allow the hazard rate to vary over time. The inclusion of year effects also addresses the potential nonstationarity of firm numbers during the period we run our regressions.

Table 1 reports the logit regression results with "Firm Death" (A firm exits in the

[5]Using the information provided in Kimes (1996), we collect biographical information about the entrepreneurs up to 1942, before the U.S. entered the WWII.

Table 1: Logit Models for Firm Death

Firm Death	(1)	(2)	(3)	(4)	(5)	(6)
De Alio	-0.460***	-0.461***	-0.457***	-0.456***	-0.488***	-0.482***
	(0.092)	(0.092)	(0.092)	(0.090)	(0.095)	(0.092)
Spinoff	-0.612***	-0.614***	-0.452***	-0.235*	-0.470***	-0.260*
	(0.126)	(0.124)	(0.140)	(0.138)	(0.143)	(0.140)
St. Louis	0.013	0.303	0.341	0.287	0.009	0.164
	(0.246)	(0.272)	(0.269)	(0.274)	(0.302)	(0.282)
Chicago	-0.160	0.047	0.057	-0.047	-0.154	-0.089
	(0.143)	(0.165)	(0.164)	(0.149)	(0.189)	(0.154)
Indianapolis	-0.450***	-0.268*	-0.281*	-0.414***	-0.502***	-0.489***
	(0.132)	(0.149)	(0.148)	(0.143)	(0.166)	(0.154)
Detroit	-0.292***	-0.247**	-0.199*	-0.765***	-0.252**	-0.555*
	(0.105)	(0.105)	(0.106)	(0.264)	(0.109)	(0.307)
Rochester	-0.182	0.094	0.107	-0.026	-0.169	-0.100
	(0.227)	(0.256)	(0.256)	(0.238)	(0.271)	(0.238)
New York	0.243**	0.461***	0.470***	0.318***	0.289*	0.317***
	(0.116)	(0.143)	(0.144)	(0.117)	(0.167)	(0.122)
Center Size		0.006**	0.006***		0.0005	
		(0.002)	(0.002)		(0.004)	
Family Size			-0.085**		-0.091**	
			(0.036)		(0.041)	
Center Top				0.059***		0.039
				(0.022)		(0.027)
Family Top				-0.370***		-0.373***
				(0.072)		(0.075)
Firm Age	-0.049***	-0.048***	-0.048***	-0.037***	-0.049***	-0.040***
	(0.007)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)
Year	0.030***	0.036***	0.037***	0.032***		
	(0.007)	(0.007)	(0.007)	(0.007)		
Constant	-58.359***	-69.145***	-71.007***	-61.868***	-1.527***	-1.682***
	(13.123)	(13.411)	(13.437)	(12.626)	(0.406)	(0.391)
Year Dummies					Y	Y
Observations	4454	4454	4454	4454	4360	4360

Note: Robust standard errors (clustered by firm) are reported in parentheses under coefficient values. One, two and three * indicate statistical significance at the 10, 5 and 1% levels, respectively.

current period) being the dependent variable. The findings support the implications of our theory. First, the variable "Family Size" shows a negative coefficient as predicted by Propositions 1 and 2, and the coefficient is always statistically significant (Specifications 3 and 5). The magnitude of the coefficient is also economically significant: The corresponding odds ratio implies that the relative firm death rate may drop by 8.5 percent as the number of firms in its family increases by one. Given the fact that firms are of different size in reality, we also try "Family Top" as an alternative measure (Specifications 4 and 6). Again, it has a negative coefficient as predicted by Propositions 1 and 2. The coefficient is also statistically significant, and the odds ratio implies that the relative death rate of a firm may drop by 31 percent as the number of top firms in its family increases by one.

Second, the variable "Spinoff" has a negative coefficient as predicted by Proposition 3, and the coefficient is statistically significant for all specifications. The effect of "Spinoff" appears weaker when we include "Family Size" or "Family Top" in the regression. This is because the "Family" variables separate out a part of the difference between Spinoff and *De Novo* firms that are due to the family effect.

Third, the variable "Center Size" has a positive coefficient, and it is statistically significant (Specifications 2 and 3). This suggests that there is no local positive externality, but rather local congestions. We also try "Center Top" as an alternative measure (Specification 4). As expected, it has a larger positive coefficient than that of "Center Size," which implies that top firms in a location cause bigger congestions. The effects of "Center Size" and "Center Top" become weaker and no longer statistically significant when we include year dummies instead of a year trend in the regression (Specifications 5 and 6).

Fourth, among all six location dummies, three are statistically significant (Detroit, Indianapolis and New York). Particularly, the two dummies "Detroit" and "Indianapolis" have negative coefficients, which suggest some location-specific advantage. In contrast, "New York" has a positive coefficient, suggesting some location-specific disadvantage.

Finally, the variables "*De Alio*" and "Firm Age" both have negative coefficients and statistically significant. This is consistent with the explanation that firm age and experience indicate its capability. Meanwhile, we use the "Year" trend to capture the changing threshold of surviving in the industry. As expected, the coefficient is positive and statistically significant.

For robustness checks, we also ran random-effects logit models or used different sample ranges (e.g., 1895-1929). The results are all similar.

3.3.2 Firm Birth Analysis

Entry of Spinoff Firms Assuming firms' outside options ϕ^x follows a logistic distribution, Eq (1) in our theory also suggests a logit regression model for spinoff entry analysis. Table 2 presents the logit regression results with "Spinoff Birth" (A firm generates a spinoff entrant in the current period) being the dependent variable. The findings support the implications of our theory.

First, the variable "Family Size" has a positive coefficient as predicted by Propositions 1 and 2. The coefficient is always statistically significant (Specifications 3 and 5). The magnitude of the coefficient is also economically significant: The corresponding odds ratio implies that the relative birth rate of firm may increase by 36 percent as the number of firms in a family increases by one. We also try "Family Top" as an alternative measure (Specifications 4 and 6). As expected, the results become even stronger.

Second, the coefficient of "Center Size" or "Center Top" shows a positive sign but is never statistically significant (Specifications 2-4). Moreover, when we introduce year dummies, the coefficient turns negative, which again suggests local congestions (Specifications 5-6).

Third, most center dummies show positive signs, which suggest some advantages of encouraging spinoff entries compared with non-center locations. However, many of those coefficients are not statistically significant across different model specifications.

Finally, the variable "Firm Age" has a positive coefficient and is always statistically significant. This is consistent with the explanation that firm age indicates its capability. Meanwhile, we use the "Year" trend to capture the changing threshold of entering the

Table 2: Logit Models for Spinoff Birth

Spinoff Birth	(1)	(2)	(3)	(4)	(5)	(6)
Chicago	0.608	1.308**	1.180*	0.797	-0.018	0.256
	(0.468)	(0.655)	(0.636)	(0.510)	(0.596)	(0.493)
Indianapolis	0.579	1.027	0.926	0.618	-0.300	-0.064
	(0.457)	(0.627)	(0.609)	(0.502)	(0.601)	(0.532)
Detroit	1.466***	1.743***	1.447***	0.797	1.110***	3.063***
	(0.395)	(0.463)	(0.446)	(0.784)	(0.396)	(0.956)
Rochester	0.255	1.067	0.911	0.489	-0.505	-0.112
	(0.580)	(0.800)	(0.772)	(0.605)	(0.744)	(0.575)
New York	0.738	1.425**	1.250**	0.800	0.242	0.743
	(0.476)	(0.662)	(0.632)	(0.489)	(0.571)	(0.491)
Center Size		0.015	0.012		-0.020	
		(0.011)	(0.010)		(0.015)	
Family Size			0.168***		0.174***	
			(0.049)		(0.055)	
Center Top				0.042		-0.183**
				(0.078)		(0.087)
Family Top				0.291***		0.315***
				(0.090)		(0.095)
Firm Age	0.062***	0.065***	0.072***	0.056**	0.090***	0.079***
	(0.021)	(0.025)	(0.025)	(0.024)	(0.028)	(0.027)
Year	-0.057***	-0.064***	-0.080***	-0.078***		
	(0.021)	(0.024)	(0.025)	(0.024)		
Constant	104.562***	117.157***	146.479***	143.176***	-3.838***	-4.472***
	(40.770)	(45.353)	(46.764)	(44.699)	(1.163)	(1.034)
Year Dummies					Y	Y
Observations	4333	3585	3585	3585	2979	2979

Note: St. Louis is omitted in the regressions due to no spinoff. Robust standard errors (clustered by firm) are reported in parentheses under coefficient values. One, two and three * indicate statistical significance at the 10, 5 and 1% levels, respectively.

industry. As excepted, the coefficient is negative and statistically significant.

For robustness checks, we also ran random-effects logit models or used different sample ranges (e.g., 1895-1929). The results are all similar.

Entry of Non-Spinoff Firms Eqs (3) and (4) in our theory describe the entry process of non-spinoff firms. Based on that, we estimate a discrete version by using the random-effects negative binomial model, in which the expected number of non-spinoff entrants at location j is determined by

$$n_j = \sum_e \binom{e}{M_j} \Psi_j^e (1 - \Psi_j)^{M_j - e}.$$

Table 3A reports the regression results with the dependent variable being the number of "*De Novo* entrants", "*De Alio* entrants" or both.

The sample range is 1895-1915 for 35 locations, which includes 5 production centers and 30 other states.[6] The explanatory variables include log carriage & wagon industry employment in 1904, log population in 1900, log per capita income in 1900 and the year trend.[7]

The results show that carriage & wagon employment and per capita income significantly raise the *De Alio* and *De Novo* entry in the automobile industry. This suggests that the absolute size of related industries and the development of local economy have positive impact on the quantity of non-spinoff entrants. Moreover, after we control for the carriage & wagon employment, the local population no longer affects the quantity of non-spinoff entrants.

In addition, we also estimate a logit model to check whether local economic conditions affect the quality distribution of non-spinoff entrants. Table 4A reports the results with the dependent variable being "Top Non-Spinoff Entry" (Entry of a non-spinoff firm which was ever a top seller in the auto industry).

[6] Note that we combine the two centers New York City and Rochester together because only state level data is available for most explanatory variables used in the regressions.

[7] Data sources: *Census of the U.S. Manufactures, 1904*; *Statistical Abstract of the United States, 1900-1910*.

Table 3. Negative Binomial Model: Non-Spinoff Entry
(1895 - 1915)

3A. Estimation: Annual Entry

Variable	Non-Spinoff Spec. 1	Non-Spinoff Spec. 2	DeNovo	DeAlio
Constant	-73.630***	-78.619***	-90.565***	-67.270***
	(20.053)	(19.769)	(27.864)	(21.854)
Year	0.031***	0.034***	0.038***	0.028**
	(0.011)	(0.010)	(0.015)	(0.012)
Log population 1900	0.976***	0.192	-0.119	0.121
	(0.201)	(0.240)	(0.255)	(0.339)
Log per capita income 1900	1.078***	1.150***	1.583***	1.226***
	(0.399)	(0.284)	(0.406)	(0.296)
Log C&W employment 1904		0.722***	1.071***	0.828***
		(0.176)	(0.184)	(0.239)
Log likelihood	-493.383	-485.294	-272.800	-393.696
Number of Observations	735	735	735	735

Notes: Standard errors are in parentheses. One, two and three * indicate statistical significance at the 10, 5, and 1% levels, respectively.

3B. Prediction: Annual Entry

Location	C&W Employment 1904	Per Capita Income 1904	Non-Spinoff Entry Actual	Non-Spinoff Entry Model (Spec. 2)
New York	21991	287	7.492	6.909
Detroit	19786	208	7.143	3.711
Chicago	10342	236	3.571	2.712
Indianapolis	10601	182	2.381	1.688
St. Louis	3765	188	1.000	0.864
Non-Detroit Centers (average)	11675	223	3.595	3.043
Non-center Locations (average)	794	182	0.171	0.205

The sample range is 1895-1915 for 560 non-spinoff entrants in 35 locations. The results show that the relative size of related industries, measured by local carriage & wagon employment over population ratio, significantly raises the entry of top non-spinoff firms. The firm entry year is shown to have a negative effect, and per capita income and Detroit dummy are not statistically significant.

Our findings shed new light on why Detroit became a top automobile production center in the early 1900s. The predominant view among automotive historians is that Detroit became the center through the accident that the most successful innovators and entrepreneurs happened to live there around the time. However, our findings suggest that a big part of the story can be explained by economic conditions.

As shown in Table 3B, New York and Detroit were the two largest carriage and wagon production centers in the country, each having about 20,000 carriage and wagon employees in 1904. In comparison, Chicago and Indianapolis had about 10,000 employees, St. Louis had 3,800 and other non-center states each had about 800 on average. Together with the relatively high per capita income, this explains why New York and Detroit had the largest number of non-spinoff entry of auto firms at the time. Figure A4 in the Appendix provides further evidence showing the striking correlation between geographic concentration of carriage & wagon production and automobile production across the country in 1909.

On the other hand, Table 4B shows that Indianapolis and Detroit ranked as the top two centers in terms of the relative size of carriage and wagon industry in early 1900s, measured by the local carriage & wagon employment over population ratio (0.42% in Indianapolis and 0.3% in Detroit). Compared with New York (0.13%), Chicago (0.15%), St Louis (0.12%) and other locations (0.09% on average), this advantage allowed new auto entrants in Indianapolis and Detroit to benefit from complementary industries while avoiding resource competition from unrelated industries, which explains why Indianapolis and Detroit had the highest chance of top non-spinoff entry.

Overall, our findings suggest that the formation of automobile production centers, including Detroit, can be largely explained by economic conditions, particularly the

Table 4. Logit Model: Top Non-Spinoff Entry
(1895 - 1915)

4A. Estimation: Top Entry

Variable	Spec. 1	Spec. 2	Spec. 3
Constant	530.291***	525.903***	533.575***
	(114.088)	(120.587)	(121.230)
Entry Year	-0.279***	-0.281***	-0.286***
	(0.060)	(0.063)	(0.064)
Per capita income 1900	-0.010**	0.013	0.022
	(0.005)	(0.012)	(0.016)
C&W employment / population ratio 1904		1.142**	1.385**
		(0.474)	(0.640)
Detroit			0.744
			(0.546)
Log likelihood	-92.104	-86.946	-85.959
Number of Observations	560	560	560

Notes: Standard errors are in parentheses. One, two and three * indicate statistical significance at the 10, 5, and 1% levels, respectively.

4B. Prediction: Top Entry

Location [# Entry]	C&W Employment / Population Ratio 1904 ($^0/_{00}$)	Probability of Top Entry: Actual	Probability of Top Entry: Model (Spec. 2)	Probability of Top Entry: Model (Spec. 3)
Indianapolis [50]	4.210	0.100	0.130	0.099
Detroit [150]	3.005	0.073	0.055	0.073
New York [156]	1.342	0.032	0.043	0.039
Chicago [75]	1.499	0.013	0.018	0.011
St. Louis [21]	1.211	0.000	0.007	0.002
Non-Detroit Centers (average) [302]	1.847	0.036	0.049	0.040
Non-center Locations (average) [108]	0.890	0.037	0.028	0.028

auto-related industries and local economy. Of course, this is not meant to be an exclusive explanation. Some other factors, including local entrepreneurial talent and early adoption of gasoline engine, could have played some additional role in making Detroit the leading production center in the country (Rubenstein 1992).

3.3.3 Family Network Effect vs. Family Gene Effect

The above regression results suggest that the "local externalities" and the spinoff "family effect" both contribute to the geographic concentration of U.S. automobile industry. One further question is where the "family effect" comes from. Naturally, there are two competing hypotheses. One may be referred to as the "family network effect," which means family members share some sort of local externalities within the family network, e.g., through business relation and knowledge linkage. The other may be referred to as the "family gene effect," which means spinoff firms inherit some predetermined capability from their parents.

In order to distinguish the "family network effect" from the "family gene effect," we consider a group of spinoff firms that happened to locate away from their parents. Note that in the top three spinoff families that located in Detroit (GM, Oldsmobile and Ford), about 24 percent spinoffs moved away from Detroit.[8] Presumably, the "family network effect" would diminish with the distance, so a firm's performance is influenced more by local family members than non-local family members. In contrast, the "family gene effect" would not depend on local proximity. Therefore, we use this as an identification strategy to tell apart the "family network effect" from the "family gene effect."

We ran similar regressions as we did in the previous firm exit analysis, using Firm Death (A firm exits in the current period) as the dependent variable. However, instead of using "Family Size" or "Family Top" as an explanatory variable, we split the "Family Size" into two variables – "Local Family Size" and "Non-Local Family Size," and similarly we split the "Family Top" into "Local Family Top" and "Non-Local Family Top."

[8] For each spinoff from the top three families, we used historical information to identify the motive of the spinoff. The results show no systematic relation between the motive of a spinoff and its choice on whether or not to locate close to the parent firm.

Table 5. Logit Models for Firm Death

Firm Death	(1)	(2)	(3)	(4)
De Alio	-0.457***	-0.458***	-0.488***	-0.483***
	(0.092)	(0.090)	(0.095)	(0.092)
Spinoff	-0.456***	-0.254*	-0.472***	-0.272**
	(0.140)	(0.137)	(0.143)	(0.139)
St. Louis	0.332	0.215	-0.001	0.091
	(0.271)	(0.276)	(0.303)	(0.286)
Chicago	0.055	-0.051	-0.157	-0.095
	(0.164)	(0.149)	(0.189)	(0.155)
Indianapolis	-0.275*	-0.399***	-0.499***	-0.479***
	(0.148)	(0.145)	(0.166)	(0.155)
Detroit	-0.186*	-0.725***	-0.241**	-0.510*
	(0.109)	(0.264)	(0.112)	(0.307)
Rochester	0.109	-0.020	-0.170	-0.095
	(0.256)	(0.237)	(0.271)	(0.237)
New York	0.470***	0.319***	0.288*	0.315***
	(0.143)	(0.114)	(0.166)	(0.120)
Center Size	0.006***		0.001	
	(0.002)		(0.004)	
Local Family Size	-0.101**		-0.106**	
	(0.042)		(0.047)	
Non-Local Family Size	-0.054		-0.064	
	(0.069)		(0.074)	
Center Top		0.062***		0.041
		(0.022)		(0.027)
Local Family Top		-0.484***		-0.485***
		(0.094)		(0.100)
Non-Local Family Top		-0.086		-0.105
		(0.115)		(0.120)
Firm Age	-0.047***	-0.032***	-0.048***	-0.035***
	(0.009)	(0.008)	(0.009)	(0.009)
Year	0.036***	0.028***		
	(0.007)	(0.007)		
Constant	-69.863***	-55.234***	-1.513***	-1.682***
	(13.453)	(12.624)	(0.406)	(0.391)
Year Dummies			Y	Y
Observations	4454	4454	4360	4360

Note: Robust standard errors (clustered by firm) are reported in parentheses under coefficient values. One, two and three * indicate statistical significance at the 10, 5 and 1% levels, respectively.

The regression results are shown in Table 5. We find that “Local Family Size” or “Local Family Top” has significantly negative effects on a firm’s death rate, but the effect from “Non-Local Family Size” or “Non-Local Family Top” is much less and not statistically significant. These results suggest that the “family effect” is likely due to the “family network effect” rather than the “family gene effect.”

4 Model Calibration

The findings of our reduce-form regressions point to the specific channels through which industry agglomeration takes place. In order to quantify the contribution of each of the channels to the agglomeration, we now take a step further to calibrate our theoretical model to the data.

4.1 Functional Forms

In this section, we calibrate our theoretical model to match quantitative patterns of the U.S. automobile industry evolution. We assume five production locations $j = 1, ..., 5$ (corresponding to St. Louis, Chicago, New York, Indianapolis and Detroit) and two levels of firm capability $s = 1, 2$ (corresponding to low and high).[9] We exclude any location fixed effects in production functions to restrain the number of free parameters, and also to focus on more interesting (and more meaningful) explanatory factors.

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

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

where c_1 captures the relative advantage of firms with a higher capability, and l is the quantity of input. This implies that a price-taking firm will have profit and output as:

$$\pi(s; p) = (\frac{1-\alpha}{\alpha})(\alpha p)^{\frac{\alpha}{1-\alpha}}[exp(c_1 s)]^{\frac{\alpha}{1-\alpha}},$$

[9]Note that we combine the two centers New York City and Rochester due to the data limitation.

$$q(s;p) = (\alpha p)^{\frac{\alpha}{1-\alpha}}[exp(c_1 s)]^{\frac{1}{1-\alpha}}.$$

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

4.2 Model Preliminaries

To calibrate our model, we first estimate an industry demand function using historical annual data of automobile prices and output from Thomas (1977).

$$p_t = D^{-1}[\sum_{s,j} q(s;p_t)m_t(s,j)] = \frac{a}{b} - \frac{1}{b}[\sum_{s,j} q(s;p_t)m_t(s,j)].$$

Second, we take the estimated non-spinoff entry rate in each location from the result of our negative binomial regression in Section 3.3.2.

Third, we set the discount factor $\beta = 0.95$, and the rate of return of the production function $\alpha = 0.9$.

4.3 Calibration

Our calibration takes two steps and uses 1904 and 1919 as two benchmark years.[10]

First, we consider 1904 as an industry steady state with no spinoff. This is based on the fact that spinoffs were insignificant at the time. Accordingly, we set $\gamma_j = 0$ and pick the set of parameters (c_1, σ, μ) to match the following data moments in 1904:[11]

- the distribution of output across the five production centers,
- the firm exit rate at the five production centers.

We set $c_1 = 0.33$, $\sigma = 0.85$, $\mu(s = 1,\text{non-Detroit}) = 0.96$, $\mu(s = 1,\text{Detroit}) = 0.82$. The model calibration matches the data moments well, which show that more than half

[10] One reason for choosing 1904 and 1919 as benchmark years is because industry census data are available for these two years. Also, few spinoffs existed in 1904 (less than 8 percent of firms), but the ratio rose to more than 30 percent in 1919.

[11] Data sources: *Census of the U.S. Manufactures, 1904, 1919.*

of the total industry output concentrated in Detroit, followed by New York, Chicago, Indianapolis and St. Louis. Detroit also had a lower firm exit rate than other production centers. Our exercise suggests that the early industry concentration pattern in Detroit as well as in the other centers was mainly driven by non-spinoff entrants. As we found before, the quantity and quality of those entrants were mainly influenced by the local economy and auto-related industries.

Table 6. Model Fit

	1904		1919	
	Data	Model	Data	Model
Share of output				
St. Louis	0.01	0.07	0.03	0.04
Chicago	0.08	0.11	0.08	0.06
New York	0.31	0.15	0.11	0.08
Indianapolis	0.06	0.10	0.05	0.06
Detroit	0.54	0.56	0.73	0.76
Exit rate				
Non-Detroit	0.18	0.20	0.17	0.22
Detroit	0.11	0.14	0.16	0.15
Spinoff rate				
Center average			0.03	0.03
Detroit			0.07	0.07

In the second step, we conduct a counterfactual experiment: We take the above parameter values as given but set $\gamma_{\text{detroit}} = 0.08$ and $\gamma_{\text{non-detroit}} = 0.02$ to match the Detroit and industry average spinoff rates in 1910s (0.07 vs. 0.03). Considering 1919 as an industry steady state with spinoffs, our model calibration again matches very well the data moments, including output shares and firm exit rates across production centers in 1919. Compared with 1904, the result suggests that spinoffs accounted for an additional 25 percent of total industry output added to Detroit as well as the changing concentration pattern in each of the other production centers.

The calibration exercise allows us to quantify the impact of different driving forces on agglomeration in different stages of industry life cycle. The findings reveals that at the early stage of an industry, agglomeration is mainly driven by non-spinoffs, whose entries are largely influenced by the local economy and related industries. Spinoffs and family network effect become important only at later stages and they contribute to an increased concentration as the industry matures.

5 Conclusion

Taking the early U.S. automobile industry as an example, we evaluate two competing hypotheses on geographic concentration of industry: local externalities *versus* employee spinoffs. Our findings suggest that both forces contribute to industry agglomeration through their specific channels. The local economy and related industries have positive influences on the quantity and quality of non-spinoff entrants. Spinoffs' performance is heavily affected by local family members but not distant ones, which suggests that spinoffs may indeed enjoy a special type of local externalities rather than gene reproduction because the latter would not depend on local proximity. Controlling for spinoff family effects, we find that the local industry size has a negative effect on firm performance, suggesting the presence of competition-induced congestion, rather than a positive externality.

Calibrating our model to the quantitative pattern of industry evolution, we find that traditional local externalities are main driving forces of agglomeration. The local economy and related industries play an especially important role by fostering new entrants. In comparison, spinoffs play a secondary role and contribute to an increased concentration at later stages of the industry life cycle.

There are some avenues for further research. First, our paper focuses on the evolution of an individual industry. It would be useful to extend the analysis to multiple industries. For example, we may compare our findings with cross-industry studies using census data. Another possible extension is to compare the early automobile industry with recent high-tech industries, such as the agglomeration of IT firms in Silicon Valley.

Second, due to data limitation, we use entry and exit as proxy measures of firm performance. If data allows, future studies may consider using more direct measures, such as output, profit, employment or product variety.

Third, our paper points to the channels through which local externalities and spinoffs contribute to the industry agglomeration. It would be interesting to identify and measure more precisely the nature and size of local spillovers through those channels.

Finally, it would be useful to conduct cross-country comparison of industry agglomeration, in both advanced economies and developing economies. Understanding the causes and consequences of industry agglomeration would shed light on the nature of increasing-return technologies and spillovers which are important driving forces of growth, development and international trade.

Figure A1. Geographic Concentration of US Auto Producers (1900-1925)

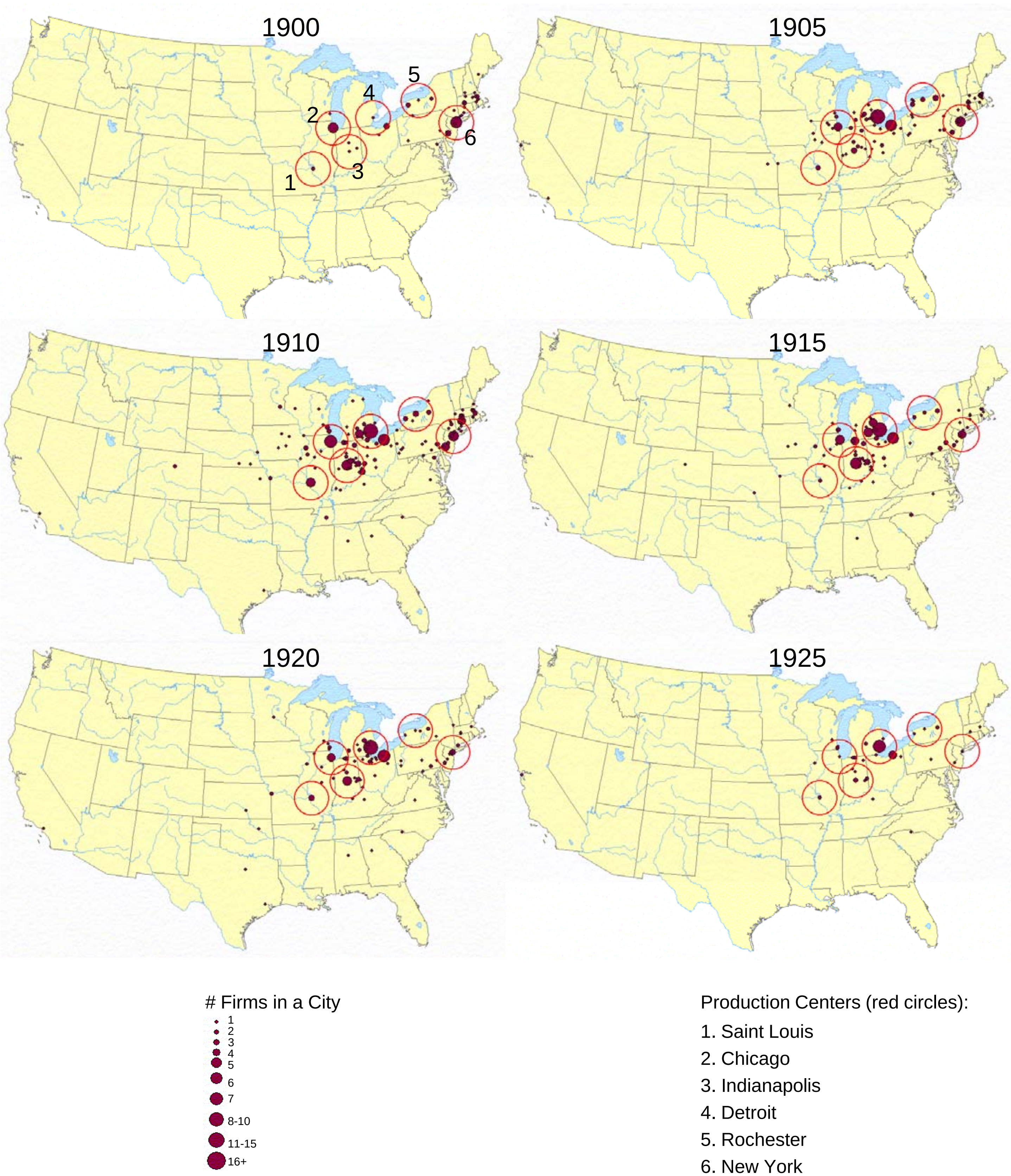


Figure A2. Center Shares of US Auto Producers (1900-1925)

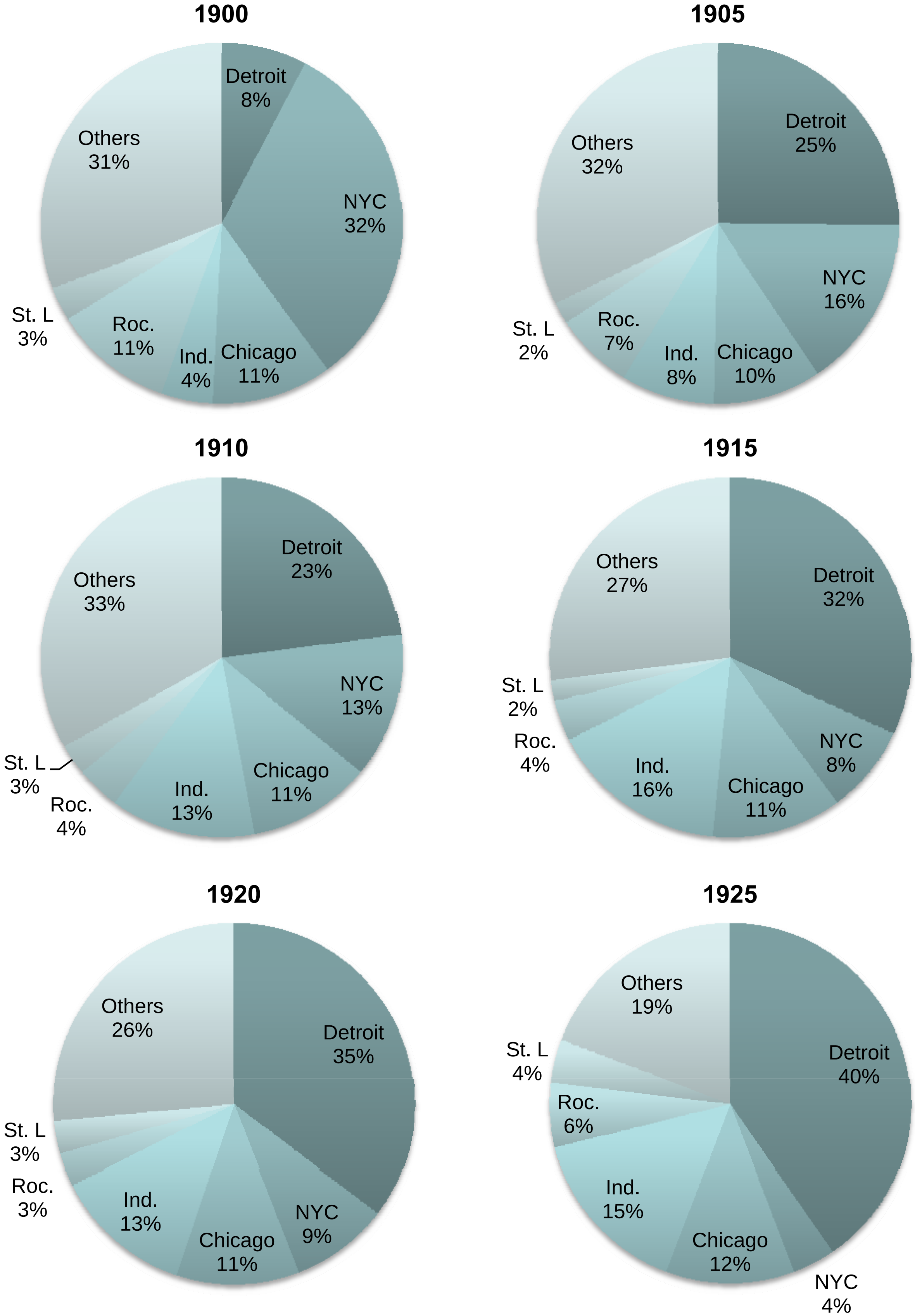


Figure A3. Family Size Distribution of US Auto Producers (1900-1925)

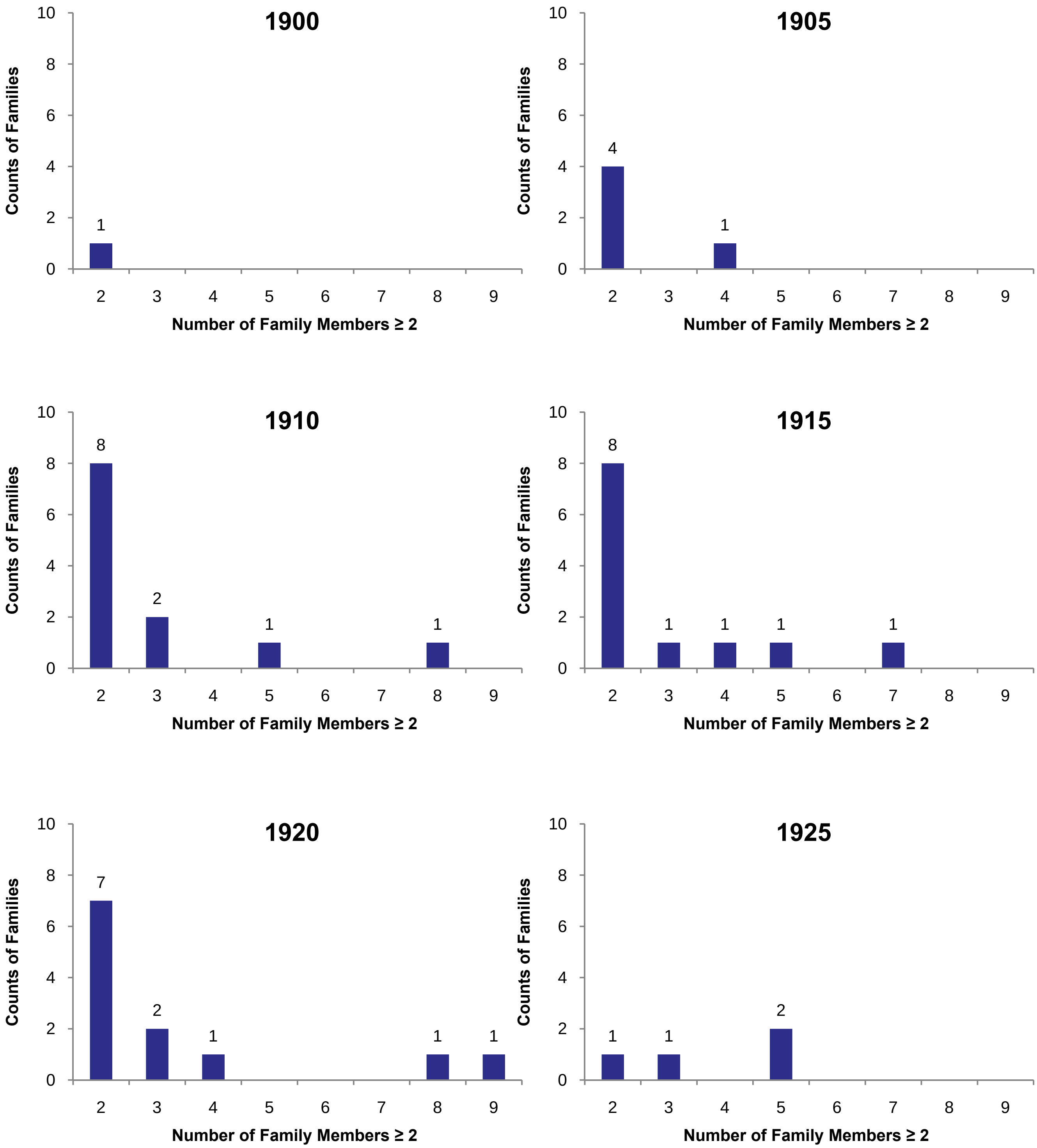
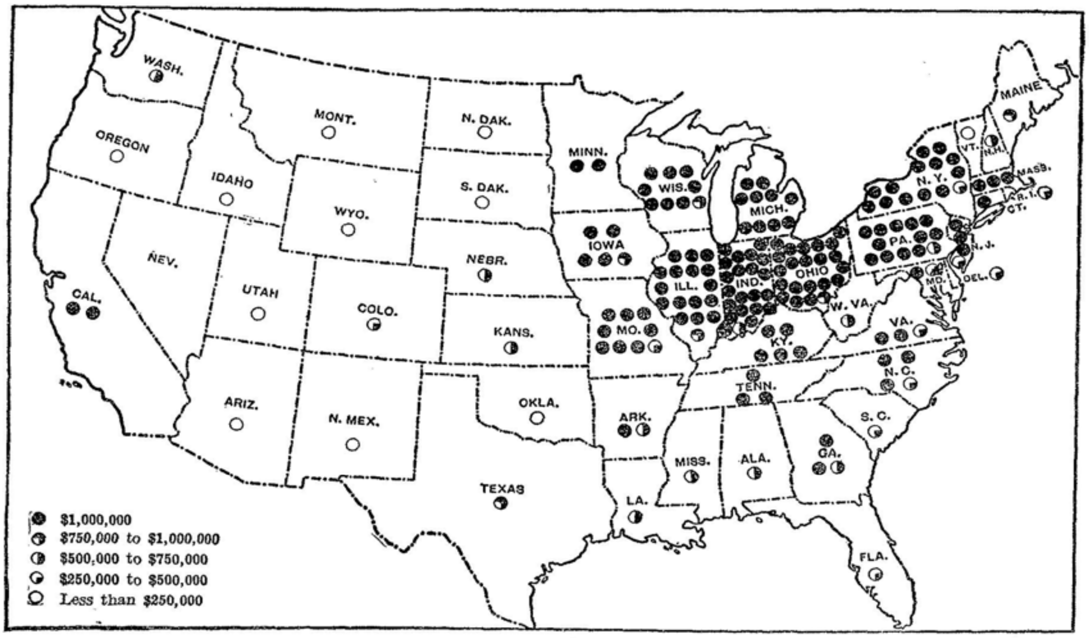
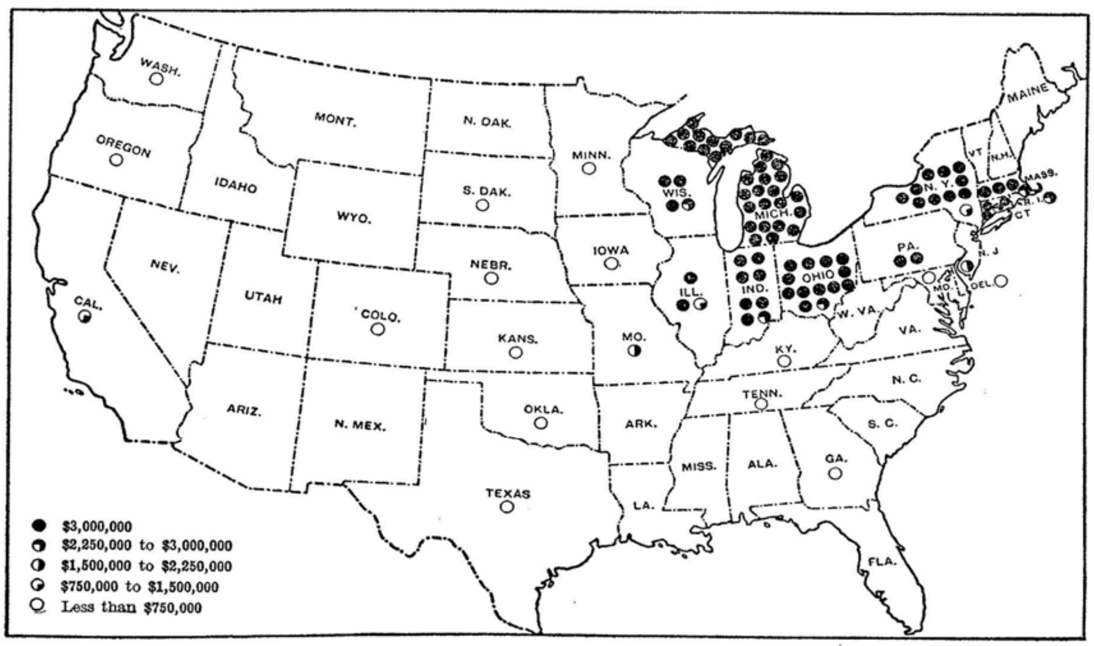


Figure A4. Geographic Concentration of Production: Carriage and Wagon *vs.* Automobile



The Carriage and Wagon Industry – Value of Products, By States: 1909



The Automobile Industry – Value of Products, By States: 1909

Source: *The 13th Census of the U.S. Manufactures*, U.S. Census Bureau, 1910.

Table A1. Spinoff Tree of General Motors

General Motors/Buick (1903 -)	Chevrolet (1911 - 1916)	Gardner (1920 - 1931)
		Monroe (1914 - 1920)
	Nash (1917 - 1954)	Lafeyette (1920 - 1924)
	Durant (1921 - 1931)	DeVaux (1931 - 1933)
	Farmack (1915 - 1916)	Drexel (1917 - 1917)
	Chrysler (1924 -)	
	Daniels (1915 - 1923)	
	Lincoln (1920 - 1922)	
	Little (1911 - 1913)	
	Lorraine (1919 - 1921)	
	Welch-Detroit (1909 - 1909)	

Notes: Buick was formed in 1903 before GM was created in 1908 as a holding company for it. Chevrolet was formed by ex-GM president William C. Durant and Louis Chevrolet following Durant's first departure from General Motors. Nash was formed by Charles Nash after a stint as president of General Motors. Durant was formed by William C. Durant following his second departure from General Motors. Farmack was founded by Albert J. Farmer, who had previously designed motors for General Motors. Chrysler was founded by Walter Chrysler, who had had previous experience as president of the Buick Division of General Motors. Daniels was founded by George E. Daniels, a former vice president of General Motors. Lincoln was founded by Henry Martin Leland after he left General Motors in a dispute with William C. Durant. Little was formed by William C. Durant following his first departure from General Motors. He was concurrently involved with Chevrolet, but this was a separate business entity. Lorraine was formed by David Dunbar Buick, who had previously founded Buick. Welch-Detroit was the project of A.B.C. Hardy, who was previously with General Motors. Gardner was founded by Russell E. Gardner after being involved with Chevrolet. Monroe was co-founded by R.F. Monroe and William C. Durant, then of Chevrolet. As with Little, Monroe was a separate business entity from Chevrolet. Lafeyette was founded by Charles Nash. He was then also running the Nash firm as a separate venture. DeVaux was founded by Norman de Vaux, who was involved with the West Coast branch of the Durant company. Drexel was formed by Albert J. Farmer from his financially troubled Farmack venture.

Table A2. Data Summary Statistics: Firm×Year Level

(Sample Range: 1895 – 1942)

Variable	Obs	Mean	Std. Dev.	Min	Max
De Novo	4454	0.21	0.41	0	1
De Alio	4454	0.59	0.49	0	1
Spinoff	4454	0.20	0.40	0	1
Top Firm	4454	0.20	0.40	0	1
Firm Death	4454	0.17	0.38	0	1
Spinoff Birth	4454	0.02	0.13	0	1
Center Size	4454	35.79	23.43	1	96
Family Size	4454	1.53	1.52	1	10
Center Top	4454	6.03	5.55	0	18
Family Top	4454	0.47	1.04	0	5
Local Family Size	4454	1.37	1.25	1	9
Local Family Top	4454	0.41	0.94	0	5
Non-Local Family Size	4454	0.16	0.70	0	9
Non-Local Family Top	4454	0.07	0.39	0	5
Firm Age	4454	6.87	7.24	1	43
Year	4454	1913	8	1895	1942

Table A3. Data Summary Statistics: Firm Level

(Sample Range: 1895 – 1942)

Variable	Obs	Mean	Std. Dev.	Min	Max
De Novo	771	0.30	0.46	0	1
De Alio	771	0.52	0.50	0	1
Spinoff	771	0.17	0.38	0	1
Top Firm	771	0.06	0.24	0	1
Entry Year	771	1908	6	1895	1939

References

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

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

[3] Cabral, Luis and Zhu Wang (2009). "Spinoffs: Theory and Evidence from Early U.S. Automobile industry," Federal Reserve Bank of Kansas City Working Paper.

[4] Chatterjee, Satyajit, and Esteban Rossi-Hansberg (2007), "Spin-Offs and the Market for Ideas," Federal Reserve Bank of Philadelphia Working Paper.

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

[6] Ellison, G. and E. L. Glaeser, (1999). "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?," *American Economic Review*, 89(2): 311-316,

[7] Ellison, G. and E.L. Glaeser. (1997). "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy,* 105: 889-927.

[8] Ellison, G. and E.L. Glaeser and W. Kerr (2007). "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." Harvard Institute of Economic Research, Discussion Paper No. 2133.

[9] Franco, April, and Darren Filson (2006). "Spin-Outs: Knowledge Diffusion through Employee Mobility," *Rand Journal of Economics* 37, 841-860;

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

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

[12] Hopenhayn, Hugo (1992). "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 60(5): 1127-50.

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

[14] Klepper, Steven (2010). "The Origin and Growth of Industry Clusters: The Making of Silicon Valley and Detroit," Journal of Urban Economics, 67: 15-32.

[15] Klepper, Steven (2007). "Disagreements, Spinoffs, and the Evolution of Detroit as the Capital of the U.S. Automobile Industry," *Management Science*, 53 (4): 616-631.

[16] Klepper, Steven and Peter Thompson (2005). "Spinoff Entry in High-tech Industries: Motives and Consequences," Working Papers 0503, Florida International University, Department of Economics.

[17] Krugman, Paul (1991). *Geography and Trade*. Cambridge, MA: MIT Press.

[18] Marshall, A. (1920). *Principles of Economics.* New York: Macmillan.

[19] Rosental, Stuart and William Strange (2004). "Evidence on the Nature and Sources of Agglomeration Economies." *Handbook of Urban and Regional Economics*, Vol 4, 2119-2171.

[20] Rubenstein, James M. (1992). *The Changing US Auto Industry: A Geographical Analysis,* New York: Routledge.

[21] Smith, Philip H. (1970). *Wheels within Wheels: A Short History of American Motor Car Manufacturing*, New York: Funk and Wagnalls.

[22] Thomas. Robert Paul (1977). *An Analysis of the Pattern of Growth of the Automobile Industry 1895-1929*. New York: Arno Press.