

*Department of Economics
Seminar Series*

*Francine Lafontaine
University of Michigan*

***“Too Far Away? Examining How
Distance to Headquarters Hurts Business
Establishment Performance”***

*Friday,
December 4, 2009
3:30 p.m.
210 Middlebush*

Too Far Away? Examining How Distance to Headquarters Hurts Business Establishment Performance[†]

Arturs Kalnins
School of Hotel Administration
Cornell University
atk23@cornell.edu

Francine Lafontaine
Ross School of Business
University of Michigan
laf@umich.edu

[†] We would like to thank seminar participants at Baruch College, Carnegie Mellon University, Cornell University, Lehigh University and Tilburg University, and at the University of Maryland and the University of Montreal for their comments. We are also grateful to Robert Picard for his expert assistance with the data. The usual caveat applies.

ABSTRACT

In the population of over 1.6 million sales-tax collecting businesses founded in Texas from 1990 to 2003, we show that distance to owner headquarters, and even distance to previous owner HQ if applicable, are both associated with shorter establishment longevity. For the lodging industry, where we have revenue data, increases in distance to headquarters due to HQ-moving owners or acquisitions are associated with reductions in revenues per room. We use a variety of analyses to explore why distance to headquarters has this detrimental effect on performance, and conclude that owners with distant headquarters face higher agency and information asymmetry costs that hinder the performance of their businesses.

1- Introduction

How distance affects trade patterns, and the survival and performance of firms and plants, domestically and globally, has attracted much attention in economics. For example, recent analyses in international trade that take firm heterogeneity into account have shown how the profitability of export, and the success of exporting firms, vary with distance from country of origin to destination (see e.g. Helpman, Melitz, and Rubinstein, 2008). In their survey of literature, Anderson and van Wincoop (2004) suggest that beyond contributing to increasing transportation and search costs, distance is associated with informational frictions, such as increased costs of monitoring the behavior of subsidiaries and trading partners and disciplining misbehavior. In domestic settings, industrial organization economists historically have emphasized the role of distance in affecting firm profits, through models of spatially differentiated markets and agglomeration economies dating back at least to Hotelling (1929) and Marshall (1920) respectively. Most recently, authors have conducted single-firm and single-industry studies using new detailed data on buyer and seller geographic location to examine how distance affects costs and returns. For example, Holmes (2008) considers how economies of density arising from locating stores near each other can explain the success of Walmart, while Coval and Moskowitz (2001) and Chen et al. (2009) examine the effect of distance on the performance of professional investment firms, and Hortacsu et al. (2009) show that, even on the internet, distance between buyer and seller remains an impediment to trade.

In this paper, we examine the effect of distance for firm performance at a very micro level, and in a different way. Specifically, we assess how the geographic distance from a business establishment to its owner's headquarters location affects the performance of the very many small retail and personal services businesses in the economy. We do this using data on all sales-tax collecting businesses, i.e. retail and services that serve the end consumer, founded

between 1990 and 2003 inclusively, in the state of Texas, more than 1.6 million of them.¹

We focus on the retail and small-scale service sectors in part because, relative to their size and growing importance in the economy, these sectors receive little attention from economists.² Yet the performance of firms in these industries affects the daily lives of a very large number of workers in the US economy, workers whose jobs are, by definition, local—they cannot be exported.³ In that sense, factors that affect these jobs, and the welfare of the many business owners who earn their living while creating them, should be of particular interest to economists and policy makers alike. Further, the short survival spans for many of these businesses (see, e.g., Table 2 below) suggest that substantial welfare gains—for owners, workers, and customers alike—could result from a better understanding of the causes of poor performance.

We examine the effect of distance to owner headquarters for the businesses in these sectors in particular because of the crucial role of the small business owner in ensuring the success of his or her firm. The owner of a small business is expected, among other things, to provide managerial expertise, keep up with local demand and supply factors, identify opportunities, participate in local operations when needed, and, of course, recruit, train, and monitor employees. If distance from a business establishment to its owner's HQ increases the

¹ In retail and service industries, the “headquarters” can be an owner's home, a dedicated office, a business establishment, or a post office box. We come back to this further below.

² There are some notable exceptions, including work on markdowns (Pashigian, 1988; Pashigian and Bowen, 1991) and on shopping malls, (Pashigian and Gould, 1998; Gould, Pashigian and Prendergast 2005), on supermarkets (e.g. Chevalier, 1995a, 1995b), on gasoline stations (Shepard, 1993, Slade, 1997) and on the franchise system of distribution (Pashigian, 1961, Brickley and Dark, 1987, Lafontaine 1992). More recently, there has been renewed interest in the study of retailing, largely as a result of the success of large retailers like Wal-Mart (Basker, 2005, Holmes, 2008, Jia, 2008) and other chains (Foster, et al., 2006). Still, the mere fact that one can enumerate studies of this type indicates how unusual it has been for economists to examine questions relating to this important segment of the economy, and the dire need for much more work in this area.

³ According to the Bureau of Labor Statistics, in 2004, 15 of the 110 million employees in the U.S. worked in manufacturing, while 89 million workers were involved in private service industries, including 15 million in retailing and another 13 million in the leisure and hospitality sector (see Table B-3: Employees on nonfarm payrolls by major industry sector and selected industry detail). Regarding the growing importance to the economy, Oi (2006) finds that the share of employment involved in services rose from 51% in 1950 to 76% in 1990.

cost of these activities, as theory suggests it should, then it should also have a detrimental effect on the firm's performance.

Many trade articles support the idea that distance to HQ matters for the performance of small businesses with detailed anecdotes (e.g., Chain Drug Review, 1994). Even Ray Kroc, the founder of McDonald's, in his autobiography, writes: "One thing that I liked about that [my] house was that it (sic) perched on a hill looking down on a McDonald's store on the main thoroughfare. I could pick up a pair of binoculars and watch business in that store from my living room window. It drove the manager crazy when I told him about it. But he sure had one hell of a hard-working crew!" (p. 141). The fact that this type of close monitoring could still affect the effort of employees in a business as standardized as a McDonald's restaurant speaks volumes about the importance of such supervision in retail and small scale service firms, for example.

Of course, if increased distance to HQ leads to poor performance, owners, especially those with just one or a few businesses, would be expected to counter these effects by choosing both their business and headquarters locations in a way that minimizes these effects or offers compensating advantages. Consequently, the distance-to-HQ effect observed in the data may be biased downward. As we detail further below, we rely on duration and instrumental variable regressions for various subsets of our data, as well as matched-pairs analyses, to ascertain the effect of distance while mitigating potential biases due to the endogeneity of distance to HQ, as well as biases associated with selection and various forms of unobserved heterogeneity.

Across all our methodologies, we find pervasive evidence of a strong significant negative relationship between distance to HQ and establishment survival. For the lodging industry, where we have a sixteen-year quarterly panel of revenue data, we show that increases in distance to owner HQ resulting from HQ moves, with owner identity remaining the same, or from acquisitions by new owners, are associated with reductions in revenues per room, the standard

top line measure of performance in the industry. We view our results as establishing a causal link between distance to HQ and performance, and we believe we make a compelling case that this effect arises from the increased costs that distant owners face in monitoring employees and managing information asymmetries generally.

The paper is organized as follows. In the next section, we briefly present the main theoretical arguments relating geographic distance from HQ to the performance of establishments. In Section 3, we describe our data, namely the population of sales-tax collecting businesses, and the revenue panel data for the lodging industry. In Section 4, we present our results from duration and instrumental variables regressions, from matched pair comparisons, and from panel regressions with fixed-effects. Concluding remarks are found in Section 5.

2. Why Should Distance to Headquarters Affect Small Business Performance

Owners of small businesses play a central role in ensuring the success of their businesses. Not only are they often involved in day-to-day operations, but they are also responsible for keeping track of supply and demand factors in the local market, and monitoring and disciplining employees to obtain the desired level of effort.⁴ Distance between an owner's headquarters and his or her businesses, we argue, makes it more difficult or costly for owners to put forth effort towards these activities, and hence can affect business performance, as implied by the statement of the McDonald's founder mentioned above. But if distance to HQ has a negative effect on firm outcomes, one might question why owners would not choose both establishment and HQ locations to minimize this distance. There are factors that enter into these decisions that may make it difficult for owners to locate their HQ and the business near each other, however. For one thing, conventional wisdom dictates that in retail and small-scale service firms, owners must

⁴ For example, Enz (2004) surveyed 484 restaurant owners and managers and found that hiring, retaining, and monitoring employees was their greatest concern by far, and took up most of their time.

locate their businesses near customers.⁵ While the location of the establishment is constrained by such demand considerations, the location of the HQ may be constrained by either household characteristics (where the owner lives), or by the business owner's potential involvement in other activities or businesses. As described further in our methodology section, we make use of fixed-effects and IV methods, and exploit our very large data set to generate subsets of our data where owner HQ is predetermined, for example, to address the issue of HQ distance endogeneity in our analyses. In the remainder of this section, however, we focus on the different mechanisms through which the effect of HQ distance may operate. We highlight throughout how different methodologies, and analyses for subsets of owners or businesses, can be used to identify the importance of these different mechanisms.

2.1 Agency Theory, Commuting, and Employee Monitoring

The most developed literature on the implications of headquarters distance for business performance relates to the idea that business owner effort is needed to monitor and motivate employee effort (e.g., Alchian and Demsetz, 1972). Krueger (1991), among others, discusses how franchising, by allocating ownership rights to a local business person, may reduce employee moral hazard. Consistent with this argument, Brickley and Dark (1987) find that chains are more likely to franchise those establishments far from the chain headquarters, and Lafontaine (1992) shows that chains with more geographically dispersed establishments tend to franchise more. Moreover, Kalnins and Lafontaine (2004) find that franchisors are more likely to allocate new establishments to existing franchisees with a nearby headquarters. When franchising is not used, this literature implies, the managers and employees at a business distant from the owner's HQ

⁵ That consumers will only travel some distance to obtain the goods sold in most retail and small-scale service stores is well understood by firms in these industries. For example, Salvaneschi (1996: 76) describes how McDonald's believed that "A restaurant's core part of its RTZ (retail trade zone) was under three minutes of travel time (by car)." The common adage in retailing, "location, location, location," also underscores this same point.

may exert lower effort than those of a nearby owner who monitors the behavior of employees more regularly. This, in turn, would be expected to lead to lower productivity in businesses further away from owner headquarters.

For sole proprietors who have no employees to monitor, the cost of a long commute to their business might entice them to close a business. In a data set containing a potentially large number of such sole proprietors, evidence that more distant businesses exit more rapidly might just reflect such decisions rather than being a sign of agency issues. To distinguish agency costs from pure commuting costs, we look for evidence that distance to HQ affects performance for businesses owned by multi-establishment owners—where such owners must hire employees to staff and manage at least some of their businesses—as an indication that monitoring costs are an important mechanism through which the distance-to-HQ effect operates.

2.2 Knowledge of Local Market Supply and Demand Conditions

Owners with nearby HQs are expected to enjoy informational advantages beyond those related to employee monitoring. Hayek (1945, p. 524), for example, suggests that the “man on the spot” has the best knowledge of local demand and supply conditions and can therefore make the best decisions.⁶ Barringer and Greening (1998) present a detailed set of case studies that support Hayek’s conjecture. In one, an owner of four women’s clothing stores observed that “it never occurred to me that one of the reasons that I was successful in the original location was that I had a good feel for what the market wanted in terms of fashion.” (p. 481). She explained that customer preferences varied widely among her sites and that she had initially missed this difference. In the end, she shut down all but the original establishment. Barringer and Greening (1998) also document the case of a drugstore owner who located a new establishment 26 miles

⁶ Communities, in turn, may possess the best knowledge of owners with local HQs, possibly yielding favorable outcomes for businesses near their HQs. Granovetter (2005) reviews the literature and Kalnins (2006) provides examples.

from his home location, selecting a site where he believed a major thoroughfare was planned. Unbeknownst to him, however, the thoroughfare plans had been scrapped. The new location provided insufficient revenues to support the store, and the drugstore soon closed down.

While information benefits from a nearby HQ are difficult to distinguish empirically from benefits associated with better employee monitoring, they do differ in one important regard. Knowledge of the tastes, doings and customs of a local area will remain with an owner for some time after she has left the area, while the benefits of a local presence for monitoring should dissipate immediately after the owner leaves an area. Thus we examine the effect of distance not only to an owner's current HQ location, but also to that of a previous HQ location, if applicable.

2.3 Market Power

Proximity to HQ might also lead to superior performance because of an increase in market power arising from the presence of several establishments of the same owner near her headquarters, and a reduced density of the same in further away locations. As retail and service establishments are geographically separated, each establishment possesses some degree of market power regardless of HQ location, because of transportation costs faced by consumers. In a Hotelling-style address model, the owner of two adjacent establishments will enjoy even more market power, and be able to set higher prices, than would two owners each controlling only one of the same two establishments (e.g. Levy and Reitzes, 1992). For owners with multiple neighboring establishments near the HQ, then, market power could lead to better performance and longer survival times. We address this issue below by considering how distance to HQ affects the survival of owners' very first establishments, since market power cannot be invoked to explain a negative correlation between distance to HQ and performance for this set.

2.4 Possible Endogeneity of Distance to Headquarters

The arguments above are all causal in nature. However, location decisions are endogenous. Here we explore the different ways in which such endogeneity might bias results empirically, and what we do to assuage concerns over this issue.

First, as mentioned above, business owners cognizant of a possible negative effect of distance to HQ on their businesses' performance might select establishment and HQ locations to minimize this effect. This would bias any observed negative effect of distance on survival toward zero. We address this issue using an instrumental variable approach, described further in our methodology section.

Second, owners might choose to open distant locations only in those cases where they forecast higher expected value of revenues and profits to compensate for a greater likelihood of exit, where the latter again may be due to information asymmetries. If such a high risk/high payoff strategy were common, we might find a negative relationship between distance and survival but no correlation, or even a positive correlation, between distance and revenue. Not only do we address this issue again using IV estimation, but our lodging data allows us to ascertain the validity of the revenue portion of this argument directly.

Third, as mentioned by Caves (1971) in the context of the foreign direct investment literature, high-quality owners may be systematically more likely to venture far from home than low-quality owners because they know they can succeed where others cannot. This selection effect may help explain the success of geographically dispersed retail chains (see e.g. Foster et al., 2006), where many establishments can be quite far from the chains' HQs. We address this concern empirically by analyzing within-owner effects.

Fourth, if entrepreneurs who find a particularly lucrative location for their business also choose to designate that location as their HQ, while owners of less profitable businesses are less

systematic in their choice of HQ location, we could observe a negative distance-to-HQ effect that would have no causal interpretation. As we discuss further below, we address this possibility empirically by considering establishments with pre-existing headquarters.

Fifth, a negative “within owner” relationship between HQ distance and performance might occur if entrepreneurs tend to commence operations at the best of many alternative locations, and also locate their HQ there or near that location. Businesses opened later and farther from HQ would be at less lucrative sites. Like the case of market power, we address this issue below by considering how distance to HQ affects the survival of owners’ very first establishments, thereby excluding any effect that could arise from subsequent and possibly less lucrative establishments. In addition, we hold founding dates or order of entry constant in some of our analyses, thereby controlling for the potential impact of the sequencing of establishments.

2.5 Possible Endogenous Establishment Closure Criteria:

Separate from endogeneity issues arising from location decisions, a negative distance to headquarter effect for firm performance might also result from endogenous closure decisions. In particular, it has been argued that individuals may prefer, for non-pecuniary reasons, to locate and maintain in operation those businesses that are close to a personal residence, where the residence also may function as the HQ (Katona and Morgan, 1952; Krumme, 1969). And indeed, Watts and Kirkham (1999) and Landier et al. (2009) find evidence that plants near headquarters are less likely to be shut down, even if they under-perform other plants. They argue that this arises due to a managerial perquisite problem: managers obtain a social benefit from treating their local communities better, and thus do so at the expense of shareholders. In other words, the closure decision may be endogenous to the distance to the firm's HQ.

Since the vast majority of the establishments in our data are owned and operated by the person who makes the decision to remain in operation or not, there is far less scope for

managerial perquisite taking to play a role in our context. However, we can also directly ascertain whether distance from HQ is associated with true performance differences rather than simply longer survival by combining survival and revenue analyses. If owners or managers prefer nearby businesses to remain open for non-pecuniary reasons, we should observe a negative relationship between distance and survival, but no relationship, or even a positive one, between distance and revenue.

3. Data and Basic Data Patterns

3.1 The Texas Sales Tax and Hotel Revenue Tax Data Sets

Our data set of retail and service businesses in Texas is uniquely comprehensive. We can identify every entity that collected sales taxes in the state—and hence every establishment that sold goods or services to end consumers—at any point in time from 1990 through October 2006. We constructed our database by combining information from five downloads of the Texas Sales and Use Tax Permit Holder Information File, in 1995, 1998, 2001, 2004, and 2006.⁷ After grouping records relating to the same business establishment, and eliminating those lacking valid SIC codes (primarily due to the switch to NAICS in the 2006 data), those with out-of-state headquarters, those founded before 1990, and some with missing data, we are left with records for 1,683,653 separate businesses for our analyses. Appendix A provides more detail.

For each establishment in the data, we know the identity and address of the owner and the name and address of the business itself. The owner's stated address, which we refer to as its headquarters or HQ, can be one of three types. First, for 702,272 establishments, or 41.71% of them, the HQ address is that of an establishment – the one under consideration or a previously opened one – from which an owner administers its business(es). Second, for 650,224 (38.62%) establishments, the owner HQ address is a street address unrelated to any establishment. A street

⁷ To our knowledge, Texas and California are the only states that make such data public. However, the California database lacks out-of-business date information.

address HQ likely represents a residential home address or a dedicated office address, but we cannot distinguish these. Third, for 331,157 (19.67%) establishments, the owner address is a post office box. Of course, no administrative activity can take place at the latter. Still, an owner's choice of a PO Box distant from its business as its address suggests that the owner has reasons to be at this faraway location, either because of where he lives, works, or operates other businesses. More generally, we cannot ascertain how much administrative activity actually takes place at the HQ addresses provided by owners. However, if the addresses do not function as owners' true administrative bases, then analyses based on them should yield null results.

For each business, we also know the date at which it began operations under its owner, which we refer to as the founding date, and we know when the business goes out of business or is acquired if applicable. The fact that we have the "doing business as" (DBA) name of the establishment allows us to identify business acquisitions separately from closures. In particular, we can identify when an establishment continues to operate under the same DBA name despite acquisition by a new owner, which occurs for 70,596, or 4.2% of the 1.68 million establishments in our data.

The business name also allows us to identify franchised businesses. However, for franchisee-owned businesses, franchisor HQ location does not play a role in our analyses. If an owner owns a La Quinta hotel as a franchisee, for example, we measure the distance to the franchisee's HQ location, not the distance to La Quinta's corporate HQ in Dallas. Only if the La Quinta chain itself owns and operates the property, as it does many in Texas, will the relevant distance be that to the Dallas HQ.

We also know each owner's legal ownership form. Proprietorships are the form of choice for owners of 1,182,974 establishments in our data, or 70.26% of them. Texan corporations own another 303,867 (18.05%) of establishments, while 150,094 (8.91%) are operated by general

partnerships, and 25,838 (1.53%) by limited partnerships. Finally, a few establishments, or 20,880 (1.24%) of them, are owned by foreign (non-Texan) corporations with HQs in Texas.⁸ These owners are public companies based in, and operated from, Texas but registered in another state, often for tax purposes. Others are held by out-of-state private equity firms. Businesses in these categories include Blockbuster Video, Radio Shack, and 7-11 convenience stores.

Finally, we know the four-digit SIC code of the business, based on the 1987 definition in the Economic Census. The latter is only available through early 2004, however, so we do not include establishments founded in 2004 or after in any of our analyses below.⁹ However, we still rely on the post-2004 information in our data to ascertain the longevity of businesses through October 2006.

For one industry, lodging, the Texas Comptroller's Office makes available, in a separate database, revenue and size (number of rooms) data in addition to all the information described above, with the exception of legal ownership form. This, in turn, allows us to analyze revenue effects using an unbalanced panel of 138,652 quarterly observations from the 3625 hotels of in-state owners in operation in Texas between 1990 and 2005. Many hotels are excluded from the larger sales tax data set because they do not collect sales taxes for room rentals.

For all the businesses in our data, we geocoded latitude and longitude coordinates for both the business and owner HQ address, more than three million addresses total. Geocoding allowed us to calculate the distance in miles from each of the businesses to its owner HQ.¹⁰

⁸ Recall that we limit our data to establishments with Texan HQs. Note that all our results below remain the same if we exclude from the data the establishments of non-Texan corporations.

⁹ In the data downloaded in 2006, the source had switched from SIC to NAICS. Because there is not a one-to-one correspondence between the two classification systems, we cannot assign SIC codes to individual businesses started after our 2004 data download (in July).

¹⁰ The precision of the geocoding varies by address and address type. The 3.4 million full street addresses were geocoded in batch mode using ESRI's ArcGIS v.9.2 Geocoding tool. We were able to get exact matches for about 53% of these addresses, followed by another 20% with geocoding scores of 80-99, another 11% with scores of 50-79, and 3% with a score of 30 to 49. The vast majority of other addresses, including all those that were not full street addresses (e.g. PO Box addresses), were geocoded at the zip code centroid.

Table 1 shows the distribution of these distances for the establishments in the sales tax data set. Two patterns emerge from the data here. First, a substantial number of owners open even their first few establishments some distance away from their HQ: 228,335 (20% of) very first establishments opened by new owners between 1990 and 2003 report a HQ address eight or more miles away from the establishment, and 10% are more than 17.6 miles away. The mileage distance for the 80th percentile jumps to 10.9 miles (18.7 miles) for owners' second (third through fifth) establishments.

TABLE 1: DISTANCE (IN MILES) TO HQ, BUSINESSES FOUNDED 1990 TO 2003

Percentile	All estabs	Owners' 1 st estabs	Owners' 2 nd estabs	Owners' 3 rd - 5 th	Owners' 6 th - 10 th	Owners' 11 th +
10	0.0	0.0	0.0	0.0	0.2	6.1
20	0.0	0.0	0.0	0.0	2.3	13.0
30	0.0	0.0	0.0	0.3	5.2	22.5
40	0.0	0.0	0.4	1.7	9.0	46.5
50	1.2	0.6	1.7	3.4	14.4	111.3
60	2.9	2.0	3.3	5.9	23.7	165.3
70	5.5	4.1	5.9	10.4	46.4	204.4
80	10.7	8.0	10.9	18.7	126.1	239.1
90	26.1	17.6	24.9	59.9	223.8	294.1
100	760.4	760.4	741.5	755.3	757.2	746.0
Total	1,683,653	1,141,675	331,365	154,084	21,936	34,593

Second, distance to HQ is smaller for an owner's first establishments. Owners are more likely to use their very first establishment than subsequent ones as the HQ location, as indicated by the 40%+ establishments (or more precisely, 42.5%) that have a 0.0 distance from the owner HQ. This proportion decreases monotonically as owners open additional establishments.¹¹ Similarly, the median distance of an establishment from its owner's HQ increases monotonically, from 0.6 miles for an owner's first establishment, to 14.4 miles for owners' 6th-10th, and 111.3

¹¹ The 0.0 distance to HQ for establishments opened after an owner's first arises from (1) new establishments opened at the same address as the first establishment (e.g. a mall), and operated simultaneously with the first, (2) new establishments opened nearby, in cases where geocoding lacks perfect accuracy, or (3) owners' moves to a new HQ.

miles for the 11th+. The latter increase is quite large because the final column includes the many establishments of the large wholly-owned chains with Texan HQs.

Table 2 shows the survival duration for the cohort of businesses founded in 1990, the year for which we have the longest histories. The percentiles shown in the table are out of the 90% or so of establishments in each column that did not survive through to October 2006. Within this group, more than 20% of owners' first through 10th businesses opened survive one year or less. The difference in median survival times between the owners with 11+ businesses previously opened and the owners' first establishments are not so large, with a 3.5 year median survival for the former group compared with a 2.5 year median for the latter. In that sense, the data in Table 2 confirm the large amount of churn among retail and service businesses, not only for owners' first establishments but even for experienced owners with more than ten establishments.

TABLE 2: SURVIVAL (IN YEARS) FOR BUSINESSES FOUNDED IN 1990.

Percentile of Exits	All estabs	Owners' 1 st estabs	Owners' 2 nd estabs	Owners' 3 rd - 5 th	Owners' 6 th - 10 th	Owners' 11 th +
10	0.49	0.49	0.48	0.41	0.51	0.85
20	0.88	0.88	0.83	0.75	1.00	1.39
30	1.26	1.25	1.25	1.14	1.50	1.84
40	1.88	1.88	1.75	1.67	2.19	2.97
50	2.55	2.50	2.50	2.33	3.08	3.55
60	3.36	3.32	3.33	3.24	4.17	5.08
70	4.42	4.32	4.42	4.51	5.41	7.07
80	6.17	6.00	6.25	6.37	7.45	9.09
90	9.46	9.41	9.33	9.09	9.74	11.38
100	16.76	16.76	16.63	16.66	16.51	16.55
Exited before Oct. 2006	99,176 (92.6%)	71,169 (93.1%)	18820 (92.1%)	4418 (92.1%)	1401 (89.2%)	3368 (89.0%)
Operating Oct. 2006	7877 (7.4%)	5305 (6.9%)	1610 (7.9%)	379 (7.9%)	169 (10.8%)	414 (11.0%)
Total	107,053	76,474	20,430	4797	1570	3782

4. Methods and Results

In this section, we describe how we use our very large data set to show how important and pervasive the negative distance-to-HQ effect is, and identify the sources of this effect. As mentioned previously, we establish the latter using numerous sub-populations and matched pair analyses. To that end, we first summarize, in Table 3, the implications of the theoretical arguments discussed in Section 2 for the sub-populations of interest, in terms of both establishment survival and revenues (for our lodging data). In other words, the table entries indicate the different types of analyses that we rely on to differentiate among the theoretical arguments and interpret the relationship between distance to headquarters and firm performance in our data. We then explain, in the remainder of this section, the different methodologies and specifications we rely on, and present results from each set of analyses as well.

TABLE 3: A SUMMARY OF IMPLICATIONS

THEORY/ARGUMENT:	Predicted Effect of Distance to Headquarters			
	Business Survival		Revenues	
	Effect	Notes	Effect	Notes
1- Agency theory/employee monitoring	Negative	Applies to owner's current HQ only; cannot explain effect for sole proprietors with no employees	Negative	
2- Owner commuting costs	Negative	Applies to owner's current HQ only; becomes an agency/monitoring argument for those owners that are employers	None	Not relevant, all hotels have employees
3- Information Asymmetry; Knowledge of local tastes and opportunities	Negative	Applies to owner's current and past (if applicable) HQ location	Negative	
4- Market power arising from greater density near HQ	Negative	Cannot explain effect for owners with only one establishment	Depends	Cannot explain effect for owners with only one hotel
Possible Endogeneity of Distance To Headquarters:				
5- Business location or HQ chosen to counteract the negative distance effect	Biased towards zero	Addressed via IV analyses	Biased towards zero	
6- Distant locations are high risk/high payoff	Negative	Addressed via IV analyses	Positive	Cannot explain effect in HQ-mover analysis
7- Selection: only high quality owners dare locate establishments far away	Positive	Cannot explain within-owner effect	Positive	
8- Locate HQ in or near first establishment, developed first because judged best. Sequentially open others with less potential.	Negative	Cannot explain effect when analyzing only owners' 1 st establishment, nor when founding date, or order of entry held constant	Negative	Cannot explain within-hotel effects
9- Locate HQ in or near what is expected to be highest performing establishment	Negative	Cannot explain effect for new establishments with pre-existing HQ	Negative	Cannot explain within-hotel effects
Possible Endogenous Closure Criterion:				
10- Favoritism towards local community	Negative	Owner must have a choice of establishment to close; see multi-business owner analyses	Positive	Near HQ, low performing establishments stay open

4.1 Duration Analyses

We analyze the survival of the 1,683,653 establishments in our data using a Weibull regression model. In this model, the hazard, or instantaneous transition from origin (active business) to destination state (business exit) given that the establishment has survived to time t can be written as $h(t) = h_o(t) g(X)$, where $h_o(t) = pt^{p-1}$ and p is the shape parameter for the Weibull distribution. A popular choice for the non-negative function of the covariates, $g(X)$, is to set it equal to $e^{X\beta}$, where X is the vector of independent variables, and β a vector of coefficients.

Establishment duration is measured by subtracting the date an owner commences operations at an establishment from the exit date (if closed permanently), the acquisition date (for the 4.2% of establishments that are acquired), or from October 2006, the date of our last download, if the establishment is still open for business at that time (right censored observations). We conduct extensive robustness tests, described in Section 4.7, regarding the treatment of acquisitions. We find no evidence that treating these as exits affects our results.

We include the log of the distance in miles between the establishment and its owner's HQ, calculated using the geocoded addresses, and the (log of) number of establishments previously opened by the owner, in all our regressions. In addition, to control for various dimensions of heterogeneity that could affect exit rates, all regressions include separate intercepts for the five forms of organization (proprietorships, limited and general partnerships, and Texan and non-Texan corporations), fourteen year-of-founding dummy variables to control for macroeconomic conditions at the time of business formation, and seven geocoding quality dummy variables for the establishment, and another seven for the owner HQ, addresses.¹²

Finally, in columns 3-6, we add dummy variables for each four-digit SIC code present in the

¹² The seven geocode quality dummy variables represent exact matches, scores of 80-99, scores of 50-79, scores of 30 to 49, zip code centroid, a few addresses that were coded at the city centroid level, and an "other" category for the very few that were coded using Yahoo Maps API (using Geocoding Tool v3.1.xls) or Google.

data, while in columns 5-6 we further include dummy variables for the zip code where the establishment is located.

Since the Weibull model exhibits the “proportional hazard rate” property, changes in regressors shift the baseline hazard, $h_o(t)$, and the exponentiated coefficients capture the effect of a one unit increase in a particular variable on the exit hazard ratio. Specifically, if the exponentiated coefficient b is greater than one, the difference $(b-1)*100$ indicates the percentage by which a one unit increase in the explanatory variable of interest would increase the exit rate of the establishment. Similarly, if $b < 1$, the difference $(1 - b)*100$ represents the percentage reduction in the exit rate that would result from a one unit increase in the variable of interest. Because exponentiated coefficients are easily interpreted, we present our results in Table 4 using these. As is standard in survival analyses, however, due to the skewed distribution of these coefficients, we present standard errors for the original coefficients, and assess the levels of significance – indicated by stars in the table – based on original coefficients and standard errors.

TABLE 4: DURATION REGRESSIONS; DESTINATION STATE IS ESTABLISHMENT EXIT.

Log Distance to HQ (in miles)	1.09** (.001)	1.11** (.001)	1.09** (.001)	1.10** (.001)	1.09** (.001)	1.10** (.001)
Log Number of Previous Estabs		.86** (.001)		.86** (.001)		.86** (.001)
Dummy variables:						
SIC Code Dummies			556	556	556	556
Estab. Zip Code Dum.					1442	1442
Chi Square	149377.0	161058.2	179728.7	191155.1	204339.7	215878.9
Number of Establishments = 1,683,653; Average Duration = 1241 days						

Notes: Coefficients are exponentiated. All regressions also include 7 dummy variables for HQ geocode quality, 7 for establishment geocode quality, 5 for organization type, and 14 for year of founding.

** Significance at the 1% level. Standard errors (based on non-exponentiated coefficients) in parentheses.

Results in Table 4 clearly show the importance of distance to HQ in relation to establishment exit. Every additional unit of the (log of) distance from the establishment to the HQ results in a 9% increase in exit rate. This effect is quite consistent across our specifications,

going up only slightly when we add the number of establishments previously opened by the owner among the regressors, and remaining basically the same when we include the sets of 556 SIC code and 1442 zip code dummy variables in the regressions. The consistency of the results across columns in Table 4 strongly suggests that unobserved industry or geographic heterogeneity do not confound the negative effect of a distant headquarters on survival.

Based on our discussion of theory and endogeneity issues, one might be concerned that some of the distance-to-HQ effect in Table 4 may be driven by the endogeneity of the HQ location decision. As mentioned in Section 2 and Table 3 (see row 9), the most promising establishment plausibly might be the most desirable place for an owner to locate his or her HQ. To address this concern, we examine the effect of distance to HQ for establishments of owners whose first and only HQ address pre-dates the establishment under observation by one, three, and five years respectively. Owners who move their HQ are excluded from these analyses. We analyze HQ movers separately, for a different reason, in a later section.

Results, in Table 5, show that there is a strong distance effect even among the sub-population where the owners' HQ locations can be considered exogenous or pre-determined. Even for the set of establishments where the HQ has been in use for five years before the establishments' founding date, the distance to headquarters strongly affects survival. Moreover, the effect of distance is strikingly similar across all three columns. While the sub-populations analyzed in each column have some overlap because the data in the second (third) column is a subset of that analyzed in the first (second), the consistency is impressive. We conclude that the distance-to-HQ effect is not due to locating HQs in what are the best establishments generally. Further, we find it noteworthy that the distance-to-HQ effects for this group are twice as large, at 22 to 23% per unit increase in the (log of) distance, as those for the population in Table 4. We revisit this issue further below.

TABLE 5: DURATION REGRESSIONS; ESTABLISHMENTS WITH PRE-EXISTING HQS

	HQ pre-dates establishments under observation by:		
	1 year	3 years	5 years
Log Distance to HQ (in miles)	1.22** (.002)	1.23** (.003)	1.23** (.004)
Log Number of Previous Establishments	.905** (.005)	.938** (.006)	.941** (.007)
SIC Code Dummies	551	549	545
Establishment Zip Code Dummies	1441	1437	1435
Number of Establishments (N)	208,031	127,994	80,595

Notes: Coefficients are exponentiated. All regressions also include 7 dummy variables for HQ geocode quality, 7 for establishment geocode quality, 5 for organization type, and 14 for year of founding.

** Significance at the 1% level. Standard errors (based on non-exponentiated coefficients) in parentheses.

The significant coefficient of the “Log Number of Previous Establishments” variable in all regressions in Tables 4 and 5 also provides evidence of a tendency for the exit rates of owners' later establishments to be smaller than those of their first few establishments. The combination of this result with the information from Table 1, where an establishment opened by an owner after many others had a tendency to be farther from the owner's HQ location, suggests that the detrimental effect of distance to HQ might dissipate after an owner has opened many establishments. We consider how these effects interact by presenting duration results for sub-populations defined on the basis of owners’ previously opened number of establishments in Table 6. In this table, we show results from specifications equivalent to the most flexible regressions in Tables 4 and 5, namely those that include the SIC and zip code dummy variables.¹³ The last column of Table 6 gives the number of establishments in each subset.

¹³ Results are equivalent for the specifications that do not include all these dummy variables.

TABLE 6: DURATION REGRESSIONS, BY NUMBER OF ESTABLISHMENTS OPENED PREVIOUSLY

Establishment is owner's:	Log distance to HQ		Total Establishments
	Coefficient	Std Error	
First	1.09**	0.003	1,141,675
2nd	1.13**	0.001	331,365
3 rd	1.13**	0.003	101,383
4 th	1.13**	0.004	36,669
5th	1.13**	0.007	16,032
6 th -7 th	1.1**	0.008	13,674
8 th -10 th	1.11**	0.011	8262
11 th -15 th	1.1**	0.014	6734
16 th -25 th	1.07**	0.014	6702
26 th -50 th	1.04*	0.015	7234
51 st +	1.00	0.01	13,923

Notes: Coefficients are exponentiated. All regressions include the (log of) number of establishments previously opened, 7 dummy variables for HQ address geocode quality, 7 for establishment address geocode quality, 5 for organization type, and 14 for year of founding, as well as dummy variables for every SIC code and zip code. ** Significance at the 1% level. * 5% level. Standard errors and significance levels are for original coefficients.

The results in Table 6 are important because they allow us to refute three of the mechanisms listed in Table 3 (rows 2, 4, and 8) as potential drivers of the negative relationship between survival and distance to HQ. First, the fact that we find strong results when analyzing owners' very first establishments goes counter to the market power explanation. An owner who owns a single establishment cannot benefit from extra market power just because the establishment is in the vicinity of the HQ. Second, the results for first establishments also refute the idea that the distance effect is driven by owners choosing the best location for their first establishment, and opening an HQ near or in that establishment, and then slowly expanding to other, presumed less desirable, locations. If this type of sequential expansion from most to least desirable were driving our results, we would not expect a distance effect among first establishments. Third, our results for establishments founded well after the first allow us to reject the argument that a lengthy commute for a sole proprietor with no employees, who therefore faces no agency issues, is a main driver of the distance-to-HQ result in our data.

Owners of 5 or 10 establishments must rely on employees and salaried managers in at least some of their establishments. In a world without agency or information asymmetry problems, these multi-unit owners would rarely need to commute to the establishments manned by their employees. As mentioned earlier, for such owners, commuting costs can only be invoked as a reason to close far away establishments because they increase the cost of monitoring employee behavior and otherwise keeping track of the local market.

Several additional aspects of the results in Table 6 are noteworthy. The distance between the establishment and HQ has a strikingly uniform effect on exit rates for the 485,449 establishments (28.8% of the 1,683,653 establishments analyzed in total) that are an owner's 2nd through 5th. This suggests a very consistent effect of HQ distance across four establishment sub-populations with no overlap. In fact, the distance-to-HQ effect only begins to substantially diminish after an owner has opened 15 establishments, while only after an owner has opened 50 establishments does the effect disappear completely.

Owners with 50+ establishments previously opened fall into three categories: (1) mall chain stores with a Texas HQ address, such as Zale's Jewelers, Gordon's Jewelers, JC Penney and Radio Shack, (2) the company owned outlets of franchisors based in Texas, such as Blockbuster Video, 7-11 food stores, and Whataburger, and (3) large Texas-based franchisees of national chains. The lack of a distance effect for these firms suggests that they rely on other mechanisms, such as relative performance evaluation across stores (Carmichael, 1983) and/or regional monitoring locations (Brickley and Dark, 1987), that allow them to achieve good outcomes in their many stores despite distance. In addition, for the mall stores, the mall developer attracts patrons by choosing the mix of stores and engages in some monitoring activities as well (see, e.g., Pashigian and Gould, 1998; Gould et al, 2005). Thus there is likely less remaining scope for locally-based owners to affect outcomes in these chains.

Finally, the one perhaps surprising result here is the lower effect for owners' very first establishments compared to their 2nd through 5th establishments opened. This difference is consistent with the results of Table 5, however, which are to a large degree made up of owners' 2nd, 3rd, and 4th establishments (85.4%, 80.7% and 75.8% for the 1-, 3-, and 5-year pre-existing HQ groups, respectively). There too we observe a greater hazard ratio than in Table 4 for the HQ distance variable. This higher distance sensitivity for an owner's few establishments beyond the first, however, is quite consistent with studies that suggest expanding beyond a single establishment is a very stressful endeavor for most small business owners (e.g., Barringer and Greening, 1998). Distance from these subsequent establishments to the HQ likely exacerbates this stress.

In sum, our duration analyses have shown a clear link between HQ distance and survival. They have also allowed us to rule out the possibility that endogenous HQ location choices, market power, and commuting alone are the underlying factor generating the distance-to-HQ effect in our data. Still, unobserved owner and business location characteristics and associated endogeneity and selection issues (see Table 3, rows 5-7) might be driving these. We rely on two distinct approaches to address these potential sources of bias. First, we estimate binary probit regressions where we instrument for distance from HQ or endogenous establishment location decisions. Second, we follow the spirit of the large literatures on human twins (e.g., Ashenfelter and Krueger, 1994) and patents (e.g., Jaffe et al., 1993), and the small literature on plant closings by economic geographers (e.g., Watts and Kirkham, 1999) and present a series of matched pairs analyses of twins and siblings. Analyses of relative survival within matched pairs of twins and siblings have the advantage, relative to duration or binary choice regression models, that the owners are held constant. This is simply infeasible in duration analyses due given our large populations of establishments and owners.

4.2 IV probit regressions

Table 7 shows results from a binary probit model of survival where the dependent variables are whether an establishment has survived for at least two, three or five years respectively. The main advantage of the probit model is that we can rely on instrumental variables to address the issue of HQ distance endogeneity. Per Table 3, row 5, for example, we might expect owners to be more likely to go far from HQ if a particular business location is very desirable. Such correlation between distance and the desirability of a particular location would bias downward the estimate of the negative effect of distance in our survival analyses.

We use a zip-code-level commuting statistic from the 1990 Population Census, evaluated for the zip code where the HQ is located, as our main instrument for distance to HQ. Specifically, we use the percentage of the working population, 92% of which are wage-earning employees, whose commute to their workplace from home is fifteen minutes or less.¹⁴ We focus on firms founded in 1991 or later in Table 7 to ensure that the 1990 census data are exogenous.¹⁵ Our instrument captures the tendency of individuals to not travel far to go to work when their starting point is in the area of the owner's HQ. We expect owners to have a greater tendency to locate their establishments farther from their HQ in locales where it is the norm to commute far away, and conversely to keep distance to HQ smaller in those areas where people do not commute far. In other words, shorter commute times for workers living in a particular area should reflect the fact that some areas are more self-contained regarding places of work and their supporting businesses, while longer commute times reflect greater dependence on, and ties to,

¹⁴ This statistic is based on Question 25 of the 1990 Census. The question reads: "How many minutes did it usually take this person to get from home to work LAST WEEK?" The Census file 3B provides 13 zip code-based counts of commuters that fall within specified time increments, zero through five minutes, five through ten minutes, etc. (p50_1 through p50_13). To get the proportion that commutes less than 15 minutes, we added p50_1 through p50_3, and divided by the total number of commuters in the zip code. This statistic was used because of the particularly strong correlation with the log of distance to an owner's establishment from the HQ.

¹⁵ We do this to ensure that the 8% of respondents that are business owners, some of whom may own businesses in our data, cannot affect the reported commuting times.

surrounding neighborhoods or cities. This, in turn, leads to the high correlation between HQ distance and our instrument that is necessary to make our instrument valid, and that we observe empirically. At the same time, we do not expect commuting times for the population where the HQ is located to be systematically related to the survival of establishments, which may or may not be near that location. And indeed, empirically, we find no direct effect of our commuting variable on establishment survival.¹⁶

TABLE 7: PROBIT AND IV PROBIT RESULTS.

	Survival duration:					
	2 years		3 years		5 year	
	Probit	IV Probit	Probit	IV Probit	Probit	IV Probit
Log Distance (in miles) to HQ	-.060**	-.055**	-.070**	-.078**	-.087**	-.103**
	[-.024]	[-.022]	[-.027]	[-.030]	[-.027]	[-.032]
	(.001)	(.009)	(.001)	(.010)	(.001)	(.012)
Log Previous Estabs Opened	.120**	.117**	.133**	.137**	.143**	.152**
	[.048]	[.047]	[.051]	[.051]	[.044]	[.047]
	(.002)	(.005)	(.002)	(.006)	(.002)	(.007)
SIC Code Dummy Vars	552	552	552	552	551	551
Estab. Zip Code Dummy Vars	1442	1442	1442	1442	1441	1441
Total Establishments (N)	1395336	1395336	1287824	1287824	1089399	1089399
Proportion that survives	.506	.506	.381	.381	.239	.239
Chi Square	116029	103819	114899	101636	94261	81088
Wald test of Exogeneity		.320		.620		2.02

Notes: Marginal effects (dy/dx), shown in square brackets, are calculated at mean values, where y is the proportion of establishments that survive the specified number of years.

All regressions also include 7 dummy variables for HQ geocode quality, 7 for establishment geocode quality, 5 for organization type, and 14 for year of founding.

Standard errors are in parentheses. **: p < 0.01

The first, third, and fifth columns in Table 7 show results from binary probit regressions that treat distance to HQ as exogenous. Note that since the dependent variables in Table 7 are survival of some minimum duration, a negative coefficient denotes a reduction in survival rate.

¹⁶ Note that while one might expect that when the population in the vicinity of the HQ tends to commute far, the establishment will be able draw customers from far away, which could increase survival. However, this would also mean that the establishment faces more competition from further away businesses, which would have the reverse effect. In the end, we do not expect any particular effect, and find none in our data. In other words, our instrument satisfies the technical criteria for a good instrument, as described in Wooldridge (2002).

Thus results in this table are consistent with the duration results presented earlier, indicating again that distance to HQ has a negative effect on survival. To give an idea of magnitudes, we show marginal effects, calculated at the means, in square brackets. We find that for an establishment with the mean characteristics, a one-unit increase in the (log of) mileage distance to HQ decreases the likelihood that the establishment survives at least two years from 0.506 to 0.482 (= 0.506 – 0.024), a reduction of about 6%. For three-year survival, the likelihood is decreased from 0.381 to 0.354, a reduction of about 7%. Finally, the five-year survival rate decreases from 0.239 to 0.212, or 8.7%. Note that these percentages are similar in magnitude to those in the duration regressions of Table 4, where a one-unit increase in distance to HQ led to a 9% reduction in duration.

We show instrumental variable probit results in the second, fourth, and sixth columns of Table 7. These are all very similar to the corresponding probit results, suggesting that endogeneity issues are not important here. The effect of HQ distance remains highly statistically significant, though the standard errors are much higher, as is typical of instrumental variable estimation generally. The Wald test of exogeneity is insignificant for all three survival durations. In other words, the null hypothesis that the (log of) mileage distance between an establishment and its owner's HQ can be treated as exogenous cannot be rejected.¹⁷

4.3 Matched Pairs of Twins and Siblings

While the probit results above suggest that HQ-distance endogeneity is not a major concern in our data, unobserved owner heterogeneity may still bias our results in important ways. For example, better owners may be better able to establish businesses far from HQ (see row 7 in Table 3), leading again to a positive correlation between HQ distance and survival that could bias

¹⁷ With a single endogenous variable, the test is simply a Wald test that the correlation parameter ρ between the error term in the structural equation (the survival equation) and that in the reduced-form equation for the potentially endogenous variable is equal to zero.

our negative distance results towards zero. In this section we address the issue of owner heterogeneity head on by focusing on twin and consecutive non-twin sibling pairs of establishments. The simple outcome that we examine is which establishment of the pair survives longer under the current owner. Theory, and our findings so far, suggest that for each pair of establishments, the one closer to the HQ should, more often than not, out-survive the one that is more distant from the HQ. We designate as twins those pairs of establishments opened by the same owner in the same industry (four-digit SIC code), and founded, or acquired, on the exact same day. Consecutive non-twin siblings are pairs of establishments opened by the same owner in the same industry, and founded at different times, but with no other establishment opened in the intervening period. Note that by holding owner constant, we control for all owner characteristics, such as their quality and reservation wage, but also for HQ type and HQ location, which, as explained further below, are held fixed at the owner level in all our analyses.

Our measure of distance in these analyses is different from that used in the regression analyses above, however. So far, we have measured distance between the geocoded addresses of an owner's HQ and establishments using a simple mileage distance. In our matched pair analyses we employ the minimum number of zip codes that one would have to traverse to get from the HQ zip code to the establishment's zip code, a measure formally known as the "order of contiguity" among economic geographers. This measure has the benefit of being discrete, thereby allowing us to compare relative survival rates within pairs of establishments that have distance differences of "one zip", "two zips," etc. Compared to using arbitrary distance cut-offs, this measure also better captures the fact that the same distance in miles may not have the same impact in say rural and urban areas. Note that all our regression-based analyses in Tables 4 through 7 yield equivalent results to those shown above if we use the zip code "order of

contiguity” measure of distance in place of mileage distance. Further, the conclusions we draw from the pairs analyses below also hold when we do measure distances in miles.

For each owner, we sample randomly one pair of twins that exhibits within-pair variation in distance to HQ for the analyses in Table 8, and one pair of sequential non-twin siblings for those in Table 9.¹⁸ Some owners, however, move HQs over time or change their legal ownership form (e.g. they might incorporate). While this does not affect our twin analyses, for siblings we only include those establishments of owners whose ownership form and HQ address remain identical for both, thereby ensuring that HQ type and location, and organizational form, indeed are all held constant in our within-owner analyses.

We analyze twins separately from the more general category of siblings because holding the founding date constant eliminates potential heterogeneity in owner characteristics that may result from the passage of time. The common founding date also eliminates potential sources of bias associated with the sequencing of establishment openings as discussed in Section 2 and Table 3. Still, sibling results are valuable because they represent a much larger, and yet still “within owner,” sample. Confirming the need to analyze twins separately from siblings, we found that twin establishments survive longer, with a median of 913 days compared to 658 days for non-twin siblings. This is consistent with the idea that owners of twins must possess more resources and business acumen to open two establishments at once.

Analyses of returns to schooling with human twins have been criticized for neglecting the role of within-twin variation and its impact on the choice of schooling (Bound and Solon, 1999).

We believe this concern is not relevant here because, HQ-establishment distance, the choice

¹⁸ Sampling as we do reduces the likelihood that the idiosyncrasies of one or a few large owners opening many establishments at once would affect results disproportionately. If we used all pairs, an owner opening say 10 establishments on the same day would account for $10 \times 9 = 90$ observations, whereas an owner opening two establishments would only account for one. We also exclude potential twins from the sample of siblings so there are no common establishments in the two sets of analyses. We analyzed many independently drawn samples of both twins and siblings, with little variation in results from those presented in Tables 4 and 5.

analogous to that of schooling, is determined at the establishments' founding date. A hypothetical parallel with returns to schooling would be a case where the twins' schools are chosen at birth, and thus not based on any characteristics they exhibit later. Nonetheless, we recognize that our analyses of twins and siblings do not completely address potentially endogenous HQ location. Complementing our analyses in Table 5, for some of our twin and sibling analyses, we select sub-samples of establishments whose owners' HQs were founded well before the establishments themselves.

TABLE 8: RELATIVE SURVIVAL WITHIN PAIRS OF TWINS

	% of cases where twin closer to HQ survives longer	Duration of survival in days						KS
		25 th percentile		50 th percentile		75 th percentile		
		closer	farther	Closer	farther	closer	farther	d-val
1. All twin pairs	60.8% of 6806**	425	330	1022	760	2012	1613	.10**
<i>Separated by number of establishments opened previously by the owner</i>								
2. Twins are very 1 st estabs	62.5% of 4312**	455	332	1065	769	2081	1636	.11**
3. One previous estab.	59.9% of 1146**	364	302	882	645	1856	1461	.11**
4. 2 - 4 previous estabs	58.2% of 865**	365	298	854	698	1826	1430	.10**
5. 5+ previous estabs	51.9% of 483	545	442	1234	1098	2375	2101	.07
<i>Sub-samples with previously existing HQs</i>								
6. Same HQ exists > 1 yr	58.1% of 1366**	390	334	1025	741	2160	1704	.10**
7. Same HQ exists > 3 yrs	57.3% of 915**	457	364	1168	910	2372	1885	.10**
8. Same HQ exists > 5 yrs	58.6% of 592**	548	396	1551	1185	2556	2102	.12**

** : $p < 0.01$ using a binomial distribution test with $H_0 = 50\%$.

In the first column of Tables 8 and 9, we report the proportion of pairs for which the twin and non-twin sibling closer to the owner HQ out-survives the more distant one. Next, we report quartiles of establishment duration, in days, for the closer and farther establishment.¹⁹ In the last column, we report the Kolmogorov-Smirnov test statistic, which tests the null hypothesis that the durations of the closer and farther establishments are generated by the same underlying distribution. A positive and significant statistic indicates that the distribution of duration for the closer-to-HQ establishments has a thicker tail (i.e. more long-duration outcomes) than that of

¹⁹ Note that the survival days of establishments still in operation in October 2006 are truncated. Thus we are slightly biasing our measure of days of survival downwards. However, when we examine establishments founded in the early 1990s, very few of which survive until 2006, we find days of survival very similar to those we report.

farther away establishments.²⁰

TABLE 9: RELATIVE SURVIVAL WITHIN PAIRS OF NON-TWIN SIBLINGS

	% of cases where sibling closer to HQ survives longer	Duration of survival in days						KS d-val
		25 th percentile		50 th percentile		75 th percentile		
		closer	farther	closer	farther	closer	farther	
1. All non-twin sibling pairs	59.3% of 112628**	320	257	729	548	1528	1125	.11**
1a. First sibling closer	57.5% of 34409**	274	244	716	516	1645	1085	.11**
1b. Second sibling closer	60.1% of 78219**	353	260	730	576	1461	1161	.13**
<i>Separated by number of establishments opened previously by the owner</i>								
2. Siblings are 1 st estabs	59.9% of 82157**	294	230	700	502	1460	1040	.12**
3. One previous estab.	57.9% of 20693**	364	333	821	708	1794	1368	.09**
4. 2-4 previous estabs	58.1% of 8141**	364	345	913	729	1826	1429	.10**
5. 5+ previous estabs	52.9% of 1637*	509	456	1308	1163	2618	2452	.06*
<i>Sub-samples with previously existing HQs</i>								
6. Same HQ exists > 1 yr	60.2% of 11714**	364	273	850	576	1913	1185	.15**
7. Same HQ exists > 3 yrs	60.9% of 6910**	394	305	1035	665	2220	1385	.16**
8. Same HQ exists > 5 yrs	61.1% of 3857**	457	347	1238	730	2537	1641	.17**

** : p < 0.01 using a binomial distribution test with H₀ = 50%; * : p < 0.05.

Results in Tables 8 and 9 again show a strong survival benefit associated with being closer to the owner's HQ. The first row of each table presents the most general results. For 60.8% of the 6806 pairs of twins, and for 59.3% of the 112,628 pairs of non-twin siblings, the closer establishment remains in business longer. In other words, the closer establishment remains in business longer about 60% of the time, whereas the farther away establishment does so only in the remaining 40% of the cases. The median duration of the closer twin is a substantial 262 days longer than that of the farther twin (1022 vs. 760), a 34.4% difference. For non-twin siblings, the closer establishment survives 181 days longer (729 vs. 548), a very similar 33% difference.

Rows 1a and b in Table 9 address an issue that does not arise for the twin establishments: whether the order of entry for the closer and farther establishment affects relative survival. We find a significant survival advantage for the closer-to-HQ establishment regardless of whether it

²⁰ The test statistic is the maximum vertical deviation between the two cumulative distributions. It is positive if the cumulative distribution of durations for farther establishments is above that of closer establishments at the largest deviation point.

is opened earlier or later. Complementing the duration result for owners' first establishments from Table 6, this result alleviates the concern that survival advantages might in fact be driven by the sequence of entry (Table 3, row 8) rather than distance per se.

The remainder of Tables 8 and 9 displays the relative likelihood of longer survival for the closer establishment for sub-samples constructed to reproduce analyses in Tables 5 and 6.

Rows 2-5 separate the twin and sibling pairs by the number of establishments previously opened by their owner. For the 4312 twin and 82,157 sibling pairs that are the owner's first two establishments, the establishment closer to HQ survives longer 62.5% and 59.9% of the time, respectively. When the owner has opened five or more establishments before opening the twin or sibling pair, these numbers are far lower, at 51.9% and 52.9%. Consistent with our regression results from Table 6, this decrease in the importance of HQ distance with increases in owner experience certainly helps explain the existence and success of those large chains in the retail and small scale service sectors that do not rely on franchising.

Finally, rows 6-8 present results equivalent to those of Table 5, for establishments whose owners have operated previous establishments from the same HQ for at least one, three, and five year(s) before opening the twins or the earlier of the two consecutive siblings. Consistent with results in Table 5, even in the subsets of pairs with well-established and thus exogenous HQ locations, the advantage of being closer to HQ is still strong, at 58.1%, 57.3%, and 58.6% for the one-, three-, and five-year time frames respectively for the twins. For the siblings these numbers are even greater, at 60.2%, 60.9%, and 61.1% respectively.

4.4 Distinguishing the Effects of Monitoring and Information Asymmetries

We believe our results so far have shown conclusively that the geographic separation of headquarters and operations reduces the longevity of small business establishments. We have

addressed potential endogeneity and selection issues that could affect these results, and ruled out many of the potential explanations for the observed negative distance-to-HQ effect. As noted above, in our view, distance to HQ has a negative effect on performance because it increases the cost to the business owner of carrying out activities that affect performance, namely employee monitoring and information gathering in the local market. In this section, we explore the relative importance of these two sources of distance effects. More specifically, we want to assess the effects of distance to current HQ, which determines current monitoring costs and informational asymmetries, separately from those of historical information asymmetries. We do this by showing, in Table 10, duration regression results for establishment exit for all establishments where an owner has moved HQ locations—and shut down all establishments associated with the previous HQ location *before* opening the establishments under observation. For these establishments, then, there are two relevant distances: that to the current HQ and that to the previous HQ.²¹ The former variable should still capture both the ability to gain currently available local information and to monitor employees. The second HQ distance variable, the distance to the owner's previous HQ, however, should capture only residual effects of historical local knowledge from the owner's time at that location. In other words, this location is unlikely to provide ongoing employee monitoring or information gathering benefits for the owner.

The analysis of HQ-moving owners is carried out separately for owners with different numbers of previously opened establishments. The first column of Table 10 thus shows results of a regression of the duration of owners' 2nd establishments, where the current HQ address is in a different zip code than the previous HQ. The second column presents results for owners' 3rd establishments. In this case, the HQ move might have taken place either after the first or the

²¹ Cases where establishments listed as affiliated with a previous HQ remain open at the time new establishments are opened from a new HQ certainly are present in the data. In these cases, we cannot determine whether the owner simultaneously operates from both HQ locations. For this reason, we analyze only those cases where the owner associates no establishment with the previous HQ by the time the new businesses are established.

second establishment was opened. In the former case, the first establishment has been shut down while in the latter case, both have stopped operations. The columns for the 4th, 5th, and 6th establishments apply the same logic.

TABLE 10: DURATION REGRESSIONS FOR HQ-MOVING OWNERS

	Owners' 2 nd estabs	Owners' 3 rd estabs	Owners' 4 th estabs	Owners' 5 th estabs	Owners' 6 th estabs
Log Mileage Distance to HQ	1.11** (.003)	1.19** (.005)	1.19** (.008)	1.18** (.015)	1.29** (.031)
Log Distance to Previous HQ (all establishments affiliated with this HQ are shut down)	1.07** (.004)	1.03** (.005)	1.03** (.007)	1.00 (.013)	1.01 (.024)
SIC Code Dummies	542	504	419	335	256
Establishment Zip Code Dummies	1406	1381	1264	1068	828
Total Establishments (N)	63739	36189	14466	5710	2517
Chi Square	15656.5	6863.9	3856.8	2610.1	2163.1

Notes: Coefficients are exponentiated. All regressions also include 7 dummy variables for HQ geocode quality, 7 for establishment geocode quality, 5 for organization type, and 14 for year of founding.

** Significance at the 1% level. Standard errors (based on non-exponentiated coefficients) in parentheses.

Not surprisingly, we find that the distance to the current HQ is the more important of the two distance measures in affecting duration. This is as expected since managerial benefits of a nearby HQ result from both the ability to gain local information and monitor employees. Importantly, however, distance to the previous HQ location of the owner remains a significant predictor of survival for the 2nd through 4th establishments opened. Because the owner cannot conduct monitoring from the shut-down HQ location, we conclude that it is information garnered while at the previous HQ location, which is more valuable when near the establishment, that makes distance to previous HQ matter in these regressions. The lack of effect for this variable for the owners' 5th and 6th establishments is not surprising in this regard. Because most of the HQ moves take place after the first or second establishment, a significant amount of time has passed and the historical local information has likely dissipated by the time the owner opens a 5th or 6th

business. Note that the median (and modal) second establishment from Column 1 of Table 10 is opened one day after the previous establishment and associated HQ are both shut down. The median fifth establishment is opened 1,081 days (almost 3 years) after the HQ move takes place.

4.5 HQ Moves, Acquisitions, and Revenues in the Lodging Industry

In this section, we present regression results using lodging industry data, the one industry for which the Texas Comptroller makes establishment-level revenue data public. As mentioned in Section 2 and Table 3, an examination of revenues complements our survival analyses in two ways. First, the HQ-moving hotel owners provide another setting where owner can be held constant while analyzing distance effects. In fact, unlike in our twin analyses above, we can control not only for owners, but also for establishment characteristics via fixed effects. Specifically, we include fixed effects for every hotel/owner combination in our HQ-mover analyses, and for every hotel in our acquisitions analyses. Thus, the identification of the distance-to-HQ effect is achieved strictly from changes due to owners moving their headquarters or from changes in ownership. Second, our examination of revenue effects for hotels allows us to shed further light on the cause of the longer survival periods for establishments close to their owners' HQs. If we observe revenue gains when more proximate owners acquire hotels, then we can conclude that performance has improved as a result of the new ownership. Otherwise, we might conclude that non-pecuniary or perquisite-taking motives, or even the reliance by owners on a high-risk high reward strategy, could explain the longer survival times of establishments that are closer to their owner's HQ (see Table 3, rows 6 and 10).

We rely on data from the 3625 hotel properties of in-state owners in operation in Texas between 1990 and 2005. All our conclusions remain the same if hotels of out-of-state owners are included in the analyses, but we exclude them to be consistent with our other data. The

dependent variable is taxable room revenues per calendar quarter per available room (RevPAR), the standard top-line performance metric used in the industry.²² Our independent variable of interest is once again the distance from the establishment (the hotel) to the owner's HQ. Between 1990 and 2005, 3548 owner HQ moves and acquisitions took place among the 3625 properties.²³

In addition to hotel or hotel/owner fixed effects, we include 60 calendar quarter fixed effects to control for macroeconomic fluctuations and potential seasonality, as well as fixed effects for all 51 branded chains (e.g., Holiday Inn, Econolodge) in the data. A total of 1848 hotels operated under a brand at some point during our data period, most of these under the ownership of franchisees. Brand fixed effects can coexist with hotel fixed effects because many hotels switch brand affiliation, or go from affiliated to independent operations or vice versa.

Each row in Table 11 presents the coefficients for the HQ distance variable from two separate regressions, followed by the number of fixed effects and observations for the regression in question. In the very last column, we show median quarterly RevPAR for the hotels satisfying the criterion for each row. The first set of regression results on each row refers to the sample of hotels where the owner has moved its headquarters while continuing to operate the hotel. In each regression we include fixed effects for the 400 hotel/owner combinations, or the appropriate subsamples thereof, along with the (log of) size of each owner in terms of number of establishments, and all the fixed effects mentioned above. Because fixed effects are included for each hotel/owner combination in the HQ-mover regressions, changes in HQ distance due to acquisitions have no effect on the distance coefficient. In other words, when a hotel is acquired

²² We verified the consistency of our RevPAR data for chain hotels with an independent source. The median taxable quarterly RevPAR for chain hotels shown in row 7 of Table 11 is \$2473 while the average is \$2917. Enz and Canina (2002) relied on the private STR database and reported an average quarterly RevPAR between \$2700 and \$3000 in the 1990s for chain hotels in the South Central U.S. region.

²³ Acquisitions are far more prevalent among hotels than among sales-tax collecting businesses. We believe this is the result of the dedicated nature of hotel assets. Once built, hotels have few alternative uses. Further, ownership transfers are easier to identify in these data.

by a new owner, the relevant observations are assigned a new fixed effect, leaving no variation to be explained due to acquisition-related HQ changes.

In the second set of three columns, we report regression results for acquisitions, which rely on only periods of ownership where the owners never moved HQs, for the 3433 hotels that have at least one non-moving owner. Including the HQ-moving owners yields equivalent results, but we exclude them so that the acquisition results are net of potential HQ mover effects. Again, we show only the coefficient for the HQ distance variable, but the regressions also include the (log of) owner size (in number of establishments), the (log of) number of times the hotel has gone through an ownership change since 1990, and fixed effects for each hotel, as well as brand, and calendar quarter fixed effects.

TABLE 11: QUARTERLY REVENUE PER AVAILABLE ROOM REGRESSIONS: 1990-2005 [†]

	HQ Mover Subsample (Owner remains constant)			Acquisitions by New Owners			Median Rev- PAR per Quarter
	Distance to HQ Coef. (std. error)	# of hotel /owner dummy vars	# of Obs.	Distance to HQ Coef. (std. error)	# of Hotels ^{††}	# of Obs.	
1. All in-state owned hotels	-29.65** (5.75)	400	17,816	-24.21** (2.35)	3433	120,836	\$1452
<i>Separated by size of owner (in establishments)</i>							
2. Hotels of only single-estab. owner(s)	-31.46** (7.48)	132	4569	-22.62** (3.80)	1799	61,446	\$1134
3. Hotels with some multi-estab. owner(s)	-8.10 (8.34)	268	13,247	-30.60** (3.11)	1634	59,390	\$1755
<i>Separated by size of hotel ^{†††}</i>							
4. Small hotels (< 58 rooms)	-28.95** (10.53)	105	3946	-26.51** (2.58)	1888	66,954	\$915
5. Large hotels (≥ 58 rooms)	-12.07+ (7.21)	299	13870	-15.02** (3.97)	1739	53,882	\$2053

Standard errors in parentheses; **: $p < 0.01$, +: $p < 0.1$.

[†]: All regressions include 60 calendar quarter fixed effects, and 51 branded chain fixed effects.

^{††}: We exclude periods of ownership by HQ-moving owners so that the HQ-mover and Acquisition sub-populations have no overlap of data.

^{†††}: The sums of small and large hotels (rows 4 & 5) are greater than the respective totals in row 1 because some properties expand. In these instances, the same property is included for part of its lifespan in both rows.

Results in the first row of Table 11 show that the loss in quarterly RevPAR associated with increased distance to HQ due to a HQ move is \$29.65 for each unit increase in the (log of)

mileage distance to HQ. For changes in distance to owner HQ that result from an acquisition, this loss is \$24.21.

In rows 2-3, we present results separately for hotels owned by only single-hotel owners from 1990 through 2005, and then for hotels with owners that owned more than one at some point in time (HQ moves) or that were acquired by at least one multiple-hotel owner between 1990-2005. This distinction is important because changes in hotel ownership from a single- to a multiple-hotel owner could increase ownership concentration and thus market power locally. The distance to HQ variable has a statistically significant coefficient of -\$31.46 for HQ-moving single-hotel owners, and -\$22.62 for acquisitions where both the acquiror and acquiree are single-hotel owners. As these are in line in terms of magnitude with the population effects, we conclude again that market power is not the source of the HQ-distance effects found here.

In rows 4-5, we present results separately for small and large properties, defined as below and above or at the median size of 58 rooms, respectively. We find that for small hotels, the losses in quarterly RevPAR associated with an HQ move or a change in ownership that increases the (log of) distance to HQ by one unit are \$28.95 and \$26.51, respectively. These effects are statistically significantly smaller (\$12.07 and \$15.02) for large hotels, despite their substantially higher median RevPAR. This suggests that the owners of larger properties are more effective at managing across distance, presumably because they can afford more highly trained and incentivized staff (see e.g. Krueger, 1991).

In sum, our revenue results confirm that revenues per room go down with increases in the distance to HQ. Of course, these estimates may be somewhat biased towards zero because HQ-movers and acquirors would be unlikely to move their HQs to a distant location if that would severely hamper their operations. Thus we find it noteworthy that we observe a negative HQ-distance effect at all. Moreover, because we do, we can conclude that perquisite taking by

managers as in Landier et al (2009), and other non-pecuniary motives that may lead managers to keep establishments open longer than warranted by performance alone, either are not at play in our context, or are outweighed by the real performance effect induced by more hands-on management by local owners. We can also conclude that distant owners are not systematically operating establishments in a high-risk, high-reward fashion.

4.7 Robustness Tests

We have conducted many robustness tests for all regressions and pair comparisons made in Tables 4 through 11. All tables that use mileage distances to analyze survival in the text (Tables 4-7, 10) were reproduced using the “zip code distances” that we rely on for the twins and siblings analyses. Likewise, the twins and siblings tables (Tables 8 and 9) were reproduced using mileage distances. All conclusions remained the same.²⁴ We conclude that our results are not driven by idiosyncrasies associated with using any particular measurement of distance.

We also reproduced results from Table 4 and the first rows from Tables 8 and 9 for a variety of sub-populations. First, and most importantly, we handled the 4.2% of all establishments that are acquired in several different ways. In our main analyses above, we treated establishment exit and acquisition as equivalent events, consistent with the conventional wisdom that, due to market inefficiencies and “lemons” problems, small business owners often receive less for their businesses than cash flows would warrant (see e.g., Fraser, 1999). However, our results continue to hold if we use the final exit date to measure survival spans for all establishments, including those that are acquired. Further, our results hold if we simply exclude from our analyses all establishments that are acquired, or, for the duration regressions, if we treat acquisitions as a censoring event rather than as an exit. Our results even hold for the set of

²⁴ Results were somewhat weaker in only one case, namely for the HQ-moving owners of Table 11, when we used the zip code contiguity measure of distance.

acquired establishments analyzed in isolation. In other words, we find that distance to HQ significantly accelerates the time to acquisition as well as that to final business exit.

Second, we separated our data by year of founding, and estimated regressions separately for each year. We found that distance to HQ was an important factor throughout, but, as expected, found no evident trend in this effect or in the survival times of firms in our data.

Third, we separated the observations into rural and urban areas. We found that the negative distance-to-HQ effect was important in both sets.

Fourth, our results also hold separately and consistently for retail (SICs of 5200-5999), services (SICs of 7000-7999), and an “other” category.

Fifth, the results also hold separately for all five major ownership forms, but are smaller though still significant for general partnerships (8.91% of establishments) and the non-Texan corporations (1.24%). We believe that for partnerships, the effect may be smaller because there may be two or more de facto HQs for these, one for each partner, making the effect of the listed HQ less important. As for the non-Texan corporations, these are large public firms with many establishments, so the reduced effect is consistent with that shown in Table 6.

Sixth, statistically significant results hold separately for the three HQ address types (establishment address, PO Box, or separate street address). While the effect of distance to HQ was smaller in magnitude for the PO Box sub-population, it is noteworthy because it suggests that, for the 19.74% of establishments that list a Box as the HQ address, the address still represents a meaningful proxy for the location of business owner activity.

Finally, given we mentioned McDonald's founder Ray Kroc's quote about monitoring, it is interesting to note that statistically significant negative effects of distance to HQ on establishment survival obtain when we analyze separately the duration of the 1175 McDonald's restaurants with Texan-based franchisees founded during our data period. There is not a

sufficient number of McDonald's restaurants in the data for us to include zip code intercepts in these regressions, but our results hold even when we include separate intercepts for each franchisee, to sweep away any owner-specific effects. Even for a business as standardized as McDonald's, then, we find that the distance from an establishment to its franchisee's HQ is detrimental to the long-term viability of that establishment.

5. Summary and Conclusion

In this paper, we have investigated the geographical separation of headquarters and operations in the retail and service industries. We have established the existence of a robust, pervasive and economically important negative relationship between a business establishment's distance from its owner's HQ and (1) its survival duration under that owner and (2) its revenues under that owner. We find it noteworthy that our survival and revenue results are strong throughout despite owners' incentives to select sites for their establishments and for their HQs that reduce the problems of distance or that at least have compensating advantages. In reality, the results of our instrumental variable probit regressions suggest that distance to HQ can indeed be treated as exogenous in our data.

We view our results as supportive of Hayek's contention – and McDonald's founder Ray Kroc's view - that the “man on the spot” has clear advantages in monitoring and using local information when operating retail and service establishments. While we cannot fully distinguish the effects of monitoring and other information asymmetries, our finding that distance to owners' previous HQs has an effect on the survival of establishments, even after all establishments affiliated with that previous HQ have been shut down, suggest a clear role for local information. Nonetheless, the much larger effect of distance to current headquarters implies that both monitoring costs and information problems are more acute for owners with distant HQs.

On the one hand, our conclusion that distance harms performance seems intuitive and consistent with existing academic work and practitioner observation. On the other hand, the conclusion may appear inconsistent with the great success and growth of the many large retail chains in the global economy (see notably Basker, Klimek and Van, 2008). We did find evidence that proximity to HQ offers advantages even for large owners of up to 50 establishments, but dissipates above that number. This finding suggests that owners of larger numbers of establishments have found mechanisms – including perhaps local monitoring headquarters whose presence was not necessarily captured in our data - allowing them to operate efficiently outside the immediate vicinity of their HQ location. Other owners, in particular those with a potentially lucrative business concept but with less ability to manage across distance, have the option of franchising their concept to others whose HQs will then be more local to the individual establishments (e.g., Caves and Murphy, 1976; Brickley and Dark, 1987). We view our results as consistent with the growth in chains in part because so many of the chains whose success is noted in the literature, such as, for example, McDonald’s, Subway, and Holiday Inn, to name a few, are indeed franchised.

References:

- Alchian, A.A. and H. Demsetz, (1972), "Production, Information Costs, and Economic Organization," *American Economic Review*, 62: 777-795.
- Anderson, James E., and Eric van Wincoop. 2004. "Trade Costs." *Journal of Economic Literature*, 42(3): 691–751.
- Ashenfelter, O. and A. Krueger. 1994 Estimates of the Economic Return to Schooling from a New Sample of Twins. *American Economic Review*, 84: 1157-1173.
- Barringer, B. and Greening, D. 1998. Small Business Growth through Geographic Expansion: A Comparative Case Study. *Journal of Business Venturing*, 13: 467-492.
- Basker, Emek. 2005. Job Creation or Destruction? Labor-Market Effects of Wal-Mart Expansion, *Review of Economics and Statistics* 87:174-183.
- Basker, E., S. Klimek, and Van, P.H. 2008. Supersize It: The Growth of Retail Chains and the Rise of the "Big Box" Retail Format. Mimeo, University of Missouri.
- Bound, John and Gary Solon, 1999, Double Trouble: Pitfalls in Twins-based Estimates of the Returns to Schooling. *Economics of Education Review*, 18:169-182.
- Brickley, J., and Dark, F. 1987. The choice of organizational form: The case of franchising. *Journal of Financial Economics*. 18: 401-420.
- Carmichael, L.H. 1983. The Agent-Agents Problem: Payment by Relative Output, *Journal of Labor Economics*, 1983, 1: 50-65
- Caves, R. E. 1971. International Corporations: the Industrial Economics of Foreign Investment, *Economica*, 38, 1-27.
- Caves, R. E. 1998. Industrial organization and new findings on the turnover and mobility of firms, *Journal of Economic Literature*, 36: 1947-1982.
- Caves, R. E., and W. F. Murphy II. 1976, Franchising: Firms, Markets , and Intangible Assets. *Southern Economic Journal* 42, 572-586.
- Chain Drug Review. 1994. Taylor adapting to expansion of third-party Rx. *Chain Drug Review*, 16(9): 143.
- Chen, H., P. Gompers, A. Kovner, and J. Lerner. 2009. Buy Local? The Geography of Successful and Unsuccessful Venture Capital Expansion, mimeo, Harvard University.
- Chevalier, Judith A. 1995a. Capital Structure and Product-Market Competition: Empirical Evidence from the Supermarket Industry. *The American Economic Review*, 85: 415-435.
- Chevalier, Judith A. 1995b. Do LBO Supermarkets Charge More? An Empirical Analysis of the Effects of LBOs on Supermarket Pricing. *The Journal of Finance*, 50: 1095-1112.
- Coval, J. and Moskowitz, T. 2001. The Geography of Investment: Informed Trading and Asset Prices, *Journal of Political Economy*, 109: 811-841.
- Dunne, T., M. Roberts and L. Samuelson, 1989. The growth and exit of U.S. manufacturing

- plants, *Quarterly Journal of Economics* 104: 671-698.
- Enz, Cathy A. (2004). "Issues of Concern for Restaurant Owners and Managers," *Cornell Hotel and Restaurant Administration Quarterly*. 45 (4) 315 – 332.
- Enz, C. and L. Canina. 2002. The best of times, the worst of times: Differences in hotel performance following 9/11. *Cornell Hotel and Restaurant Administration Quarterly*, 43: 41-52.
- Foster, L., Haltiwanger, J., and Krizan, C. 2006. Market selection, reallocation, and restructuring in the U.S. retail trade sector in the 1990s. *Review of Economics and Statistics*, 88: 748-758.
- Fraser, J. A. 1999. Business for sale update. *Inc.* 21(7): 60-69.
- Gould, Eric D. B. Peter Pashigian, Canice J. Prendergast, 2005. Contracts, Externalities, and Incentives in Shopping Malls. *Review of Economics and Statistics*, 87: 411-422
- Granovetter, Mark. 2005. The Impact of Social Structure on Economic Outcomes, *Journal of Economic Perspectives*, 19 (1): 33-50.
- Hayek, F. 1945. The Use of Knowledge in Society. *American Economic Review*, 35: 519-530.
- Helpman, E., M. Melitz, and Y. Rubinstein. 2008. Estimating Trade Flows: Trading Partners and Trading Volumes, *Quarterly Journal of Economics*, 123: 441-487.
- Holmes, T. 2008. Diffusion of Wal-Mart and Economies of Density, mimeo, U. of Minnesota.
- Hortaçsu, Ali, F. Asís Martínez-Jerez, and Jason Douglas. 2009. The Geography of Trade in Online Transactions: Evidence from eBay and MercadoLibre, *American Economic Journal: Microeconomics*, 1: 53-74.
- Jaffe, A., M. Trajtenberg, and R. Henderson. 1993. Geographic localization and knowledge spillovers as evidenced by patent citations, *Quarterly Journal of Economics*, 108: 577-598.
- Jia, Panle. 2008. What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Industry, *Econometrica*, forthcoming.
- Kalnins, A., and F. Lafontaine. 2004. Multi-Unit Ownership in Franchising: Evidence from the Fast-Food Industry in Texas. *RAND Journal of Economics*, 35, 747-761.
- Kalnins, A. 2006. Markets: The U.S. Lodging Industry, *Journal of Economic Perspectives*, 20(4): 203-218.
- Katona, G. and J. Morgan. 1952. The quantitative study of factors determining business decisions, *Quarterly Journal of Economics* 66: 67-90
- Kroc, Ray. 1977. Grinding it Out: The Making of McDonald's, Chicago: Contemporary Books.
- Krueger, Alan B. 1991. Ownership, Agency and Wages: An Examination of the Fast Food Industry. *Quarterly Journal of Economics*, 106: 75-101.
- Krumme, G. 1969. Toward a geography of enterprise, *Economic Geography*, 45: 30-40

- Lafontaine, F. 1992. Agency theory and franchising: Some empirical results. *RAND Journal of Economics*, 23: 263-283.
- Landier, A., V. Nair and J. Wulf. 2009. Trade-offs in Staying Close: Corporate Decision Making and Geographic Dispersion. *The Review of Financial Studies*, 22: 1119-1148.
- Levy, David T. and James Reitzes (1992), "Anticompetitive Effects of Mergers in Markets with Localized Competition," *Journal of Law, Economics and Organization*, 8: 427-440.
- Oi, W. 2006. Measuring Productivity in the Retail Trade and Services. Mimeo, University of Rochester.
- Pashigian, Peter B. 1961. *The Distribution of Automobiles: An Economic Analysis of the Franchising System*. Englewoods Cliffs, NJ: Prentice-Hall.
- Pashigian, Peter B. 1988. Demand Uncertainty and Sales: A Study of Fashion and Markdown Pricing. *The American Economic Review*, 78: 936-953
- Pashigian, Peter B. and Brian Bowen. 1991. Why Are Products Sold on Sale?: Explanations of Pricing Regularities. *The Quarterly Journal of Economics*, 106: 1015-1038.
- Pashigian, Peter B. and Eric D. Gould. 1998. Internalizing Externalities: The Pricing of Space in Shopping Malls. *The Journal of Law and Economics*, 41: 115-142.
- Shepard, Andrea. 1993. Contractual Form, Retail Price, and Asset Characteristics. *Rand Journal of Economics* 24: 58-77.
- Slade, Margaret E. 1998. Strategic Motives for Vertical Separation: Evidence from Retail Gasoline. *Journal of Law, Economics, & Organization* 14: 84-113.
- Watts, H. and J. Kirkham, 1999. "Plant Closures by Multi-locational Firms: A Comparative Perspective," *Regional Studies*, 33: 413-424.

Appendix A: Data Appendix

Our dataset was constructed from publicly available Sales Tax datasets obtained directly from the Texas Comptroller of Public Accounts/Open Government Section. In Texas, the sales tax is imposed on all retail sales, leases and rentals of most goods, as well as taxable services. Moreover, the tax code specifies that “Each seller must apply to the comptroller and obtain a tax permit for each place of business.” See:

http://www.window.state.tx.us/taxinfo/sales/faq_permit.html#permit1

for more details on who is required to get a permit. As best as we can ascertain, the requirements as to who must get a permit have not changed substantively during the period of our data.

After each request, we received two datasets. The active dataset contains a list of all active sales tax permits at the time the file was prepared. The inactive dataset contains all sales tax permits for businesses that are no longer in business. We obtained the data on the following dates:

	Active		Inactive	
	Date obtained	Number of records	Date Obtained	Number of records
1995	12/12/1995	678,939	12/12/1995	666,156
1998	06/02/1998	675,449	08/28/1998	444,243
2001	12/17/2001	676,858	01/08/2002	477,472
2004	07/12/2004	734,105	07/16/2004	366,478
2006	10/26/2006	705,272	10/26/2006	511,601
Total		3,470,623		2,465,950

Note: The number of records in these tables excludes the 19,950 records attached to places of business whose addresses were outside Texas as these are not relevant for our purposes. They include, of course, places of business that belong to owners whose address is outside Texas.

The State of Texas purges its inactive dataset of older tax permits. In 1995, it kept permits for 5 years after the out-of-business date. After that, permits were purged within 4 years and then 3 years after the out-of-business date, which is why these files are smaller in 1998 onward.

For each record, the data contains the following information:

- Owner identification code
- Establishment (place of business) identification code
- Owner name
- Owner address (street number, street, city, state, and zip code)
- Owner organization type (corporation, individual owner, partnership, etc)
- Establishment name
- Establishment address (street number, street, city, state, and zip code)
- Establishment SIC code until the 2004 file (NAICS in the 2006 data file)
- Establishment “first sale” date (year, month, day) and “end of business” date (year, month, day)

By tracking owners and their establishments across the data sets, we end up with a list of all places of business, i.e., business establishments, that have been in operation from 1990 to 2006.

Specifically, for each of the 5,936,573 records in the data, we rely first on the owner and establishment identifier codes, which we found to be quite reliable overall. We assigned new ids, however, to some businesses that seemed to have been assigned the same number though they looked different (different name or address or owner name, and so on). We later regrouped these if they satisfied matching criteria described below.

Once we separated those few records we thought may not belong together, we grouped all records that belonged to the same business establishment. This allowed us to identify a total of 3,326,945 potentially different businesses in the data.

We then examined records again to identify those that belonged to the same business firm but that for some reason were not grouped by the Comptroller's identification numbers. One reason we knew this occurred in the data is that we were told that some owner identification numbers owned by individuals were changed between 2001 and 2004 to maintain confidentiality.

We addressed this issue by first standardizing business establishment and owner names and addresses for more accurate matching. The standardization involved simplifications and corrections, such as correcting city names, removing punctuation, standardizing abbreviations (e.g., systematically using hwy rather than highway, st rather than street), removing spaces, replacing middle names when spelled out by middle initials as the latter were used often, and so on. We then created new owner identification numbers that grouped together owners whose owner name and organization type matched, and who were found to own a given establishment at a given address, based on corrected establishment name and address. Once owners were better identified, we then grouped establishment records under the same owner and establishment identification number if we found that the establishment name and address were the same. We then took the earliest first-sale date ever associated with the business establishment, and the latest end-of-business date ever associated with the same. As a result, we removed 381,224 redundant records that belong to the same business establishment, leaving 2,945,721 potential business establishments.

We then removed from the data 197,617 establishments that appeared at some point in the active files, but never appeared again in a subsequent active or any inactive file. We view these as establishments that were planned but never really existed. We also removed 49,313 records where the end date listed was before the establishment's first-sale date. We believe that these establishments, too, were planned but never existed. These subtractions led to a set of 2,698,791 different business establishments in existence at some point in the state of Texas between 1990 and 2006.

Because of the importance of SIC code and zip code for our analyses, we then removed records that lacked verifiable information about these. This included 322,294 records that did not have valid SIC codes, largely due to the 2004 switch to NAICS. We then removed 174,391 establishments where the establishment or HQ did not have a valid Texas zip code. A vast majority of these were establishments with out-of-state owners. We then removed 76,732 that were of organizational forms other than the large five (proprietorships, Texan corporations, foreign corporations, general partnerships, and limited partnerships). Most of these were non-

profits, and government, religious, social and professional organizations. We were left with 2,125,374 establishments.

We then removed 341,375 establishments that were founded before 1990 and thus could not be used in all our analyses, as well as the 68,317 establishments founded during the first few months of 2004. Finally, we removed 7911 establishments and 24118 establishments for SIC codes and zip codes, respectively, that had fewer than 100 establishments total opened between 1990 and 2003. Many of these "very small" SICs and zip codes appeared to be typographical errors because the DBA name appeared quite inconsistent with the SIC designation, or because the city appeared inconsistent with the zip code. Removing these establishments made our analyses with SIC and zip codes dummy variables manageable. The final result was a data set of 1,683,653 establishments founded between 1990 and 2003.