

Too Far Away? The Effect of Distance to Headquarters on Business Establishment Performance[†]

By ARTURS KALNINS AND FRANCINE LAFONTAINE*

In the population of over 1.7 million Texan sales-tax collecting business establishments, we show that greater distance to owner headquarters is 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 argue that this detrimental distance effect is robust and causal, arising even when we control for the potential endogeneity of HQ distance using instrumental-variable and matched-pair analyses. We interpret this as evidence of monitoring and local information asymmetry problems for distant owners. (JEL D82, G34, L25, R32)

Does geographic distance matter for the day-to-day operations of firms? Research has shown that distance constrains the flows of goods, capital, and information across and within countries, and both between and within firms (e.g., Jaffe, Trajtenberg, and Henderson 1993; Leamer and Levinsohn 1995; Lerner 1995; Adams and Jaffe 1996; Coval and Moskowitz 2001; Portes and Rey 2005). Studies also have shown that distance affects organizational form decisions (e.g., Brickley and Dark 1987; Lafontaine 1992; Lafontaine and Shaw 2005; Kalnins and Lafontaine 2004) and consumer behavior (Bronnenberg, Dhar, and Dubé 2009; Hortaçsu, Martinez-Jerez, and Douglas 2009; and references therein). In the international trade literature, authors have argued that distance reduces trade because of the cost of transporting goods. However, Rauch (2001) and Anderson and van Wincoop (2004) emphasize informational frictions, such as the increased costs of monitoring distant subsidiaries' behavior and the lack of local knowledge as major factors impeding trade. Allen (2012) develops a heterogeneous firm model with information frictions and shows evidence that the vast majority

* Kalnins: School of Hotel Administration, Cornell University, 545D Statler Hall, Ithaca, NY 14853-6902 (e-mail: atk23@cornell.edu); Lafontaine: Stephen M. Ross School of Business, University of Michigan, 701 Tappan St., Ann Arbor, MI 48109-1234 (e-mail: laf@umich.edu). We thank Juan Alcacer, Myles Shaver, and Jordan Siegel, as well as two anonymous referees, the editor and seminar participants at Baruch College, Carnegie Mellon University, Cornell University, Lehigh University, MIT, Tilburg University, Tsinghua University, University of Maryland, University of Michigan, University of Missouri, University of Montreal, and University of Toronto for their comments. We are also grateful to Robert Picard for his expert assistance with the data, and to our respective institutions for their support. The usual caveat applies.

[†] Go to <http://dx.doi.org/10.1257/mic.5.3.157> to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

of the negative effect of distance on trade flows is attributable to such frictions. Similarly, agency theory and the business press alike suggest that informational asymmetries between a local establishment and a distant owner headquarter (HQ) affect establishment performance. In his autobiography, Ray Kroc, the founder of McDonald's, writes:

"One thing 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!"

— Kroc (1977, 141)

The fact that this type of monitoring could still affect the effort of a manager and his employees in a business as standardized as a McDonald's restaurant speaks volumes about the potential importance of distance to HQ as a factor affecting establishment performance.

In this paper, we analyze data from the entire population of 1.7 million sales-tax collecting (i.e., retail and small-scale service) businesses founded between 1990 and 2003 in the state of Texas. We show that, despite owners' incentives to choose their headquarters and establishment locations to minimize this effect, a business establishment's geographic distance from headquarters significantly reduces its survival duration. We address selection issues—both owner HQ and establishment location selection—and other potential sources of bias using instrumental variable estimation and matched "twin" and "sibling" establishment analyses. We also rely on specially chosen subpopulations to address other potential sources of bias. Across all our analyses, we find a strong and statistically significant negative relationship between distance to HQ and establishment survival for all but the very largest owners in our data. For the lodging industry, where we have an 18-year quarterly panel of revenue data, we show that increases in distance to owner HQ due to HQ moves or hotel acquisitions are associated with reductions in revenues per available room, the standard top-line measure of performance in the industry. This finding further reinforces our conclusions regarding the detrimental effect of distance to HQ, and contradicts the idea that the longer survival duration for establishments near the HQ might be due to managerial favoritism, as in Landier, Nair, and Wulf (2009).

The paper is organized as follows. In Section I, we briefly discuss several theoretical arguments relating geographic distance from HQ to the performance of establishments. In Section II, we describe our population of sales-tax collecting businesses, and the revenue panel data for the lodging industry. We present our results in Section III, and conclude in Section IV.

I. Why Distance to Headquarters Might Affect Establishment Performance

A. Agency Theory, Commuting, and Employee Monitoring

The most developed literature on the implications of distance to headquarters for establishment performance relates to the idea that business owner effort is needed

to monitor and motivate employees.¹ The literature on franchising for example discusses how franchising may reduce employee moral hazard by allocating ownership rights to a local business person. 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. Kalnins and Lafontaine (2004) show that franchisors are more likely to allocate new establishments to existing franchisees if these franchisees have a nearby headquarters. When franchising is not used, this literature implies, the managers and employees at a business distant from the owner's HQ are likely to exert lower effort.²

In a dataset containing a large number of sole proprietors, evidence that more distant businesses exit more rapidly might reflect decisions to reduce commuting costs 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.

B. Knowledge of Local Market Conditions

Owners with nearby HQs are expected to enjoy informational advantages beyond those related to employee incentives and monitoring. Hayek (1945, 524) conjectured that the “man on the spot” has the best knowledge of local conditions and can therefore make the best decisions. Greening, Barringer, and Macy (1996) found that distant entrepreneurs have difficulty collecting the information required to develop potential outlet sites and to recruit quality managers. Further, search that is geographically remote from a firm's existing locations is said to result in information that is often misinterpreted even if it can be collected (Baum, Li, and Usher 2000). Finally, customers and suppliers may have better information about the quality of business establishments belonging to owners with HQs in their community. Such information advantages can result in useful local networks, information sharing and increased patronage (Freeman and Audia 2006; Granovetter 2005), and, thus, longer survival durations for locally owned establishments. Consistent with such arguments, Dahl and Sorenson (2012) find that owners who establish businesses in regions where they have lived for a long time see their businesses survive longer.

C. Market Power

If an owner has several establishments near her headquarters, and a reduced density at farther away locations, we might observe superior performance near the HQ

¹ 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. Also see Giroud (forthcoming) on the effect of changes in travel costs to plants on investments into plants and plant productivity in manufacturing.

² See also Krueger (1991) on differences in pay schemes for employees across these two forms of organization.

not because of increased agency or information asymmetry at distant locations, but because of market power near the HQ.³ We examine how distance to HQ affects the survival of owners' very first establishments to address this possibility.

D. Possible Endogeneity of Distance to Headquarters

Owners choose both the location of their headquarters and that of their establishments. In particular, 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.

High-quality owners also may be systematically more likely to venture farther 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 chains, where many establishments are far from the chains' HQs. We address this by holding owner constant in matched pair and other subsample analyses.

If entrepreneurs who find a lucrative establishment location also choose to designate that location as their HQ, while owners of less profitable establishments are less systematic in their choice of HQ location, we could observe a negative distance-to-HQ effect that would have no causal interpretation. We address this empirically by focusing attention on establishments with preexisting HQs.

Finally, a negative "within owner" relationship between HQ distance and performance might result from a sequential development process where entrepreneurs commence operations at the best of many alternative locations, and also locate their HQ there or near that "best" location. Businesses opened later and farther from HQ would then be at less lucrative sites. Below, we consider how distance to HQ affects the survival of owners' very first establishments, thereby excluding any effect that could arise from such a sequential development process. In addition, we hold founding dates and order of entry constant in some analyses, thereby controlling for the sequencing of establishments.

E. Possible Endogenous Establishment Closure Criterion

Individuals may prefer, for nonpecuniary reasons, to locate and maintain in operation those establishments that are close to a personal residence, where the residence also may function as the HQ (Katona and Morgan 1952). Indeed, Landier, Nair, and Wulf (2009) find evidence that plants near headquarters are less likely to be shut down, even if they under-perform other plants. They argue that managers obtain a social perquisite from treating their local communities better, and do so at the expense of shareholders.

Since the vast majority of our establishments are proprietorships, there is far less scope for managerial perquisite taking. However, we can directly ascertain whether

³ In a Hotelling model, for example, the owner of two adjacent establishments should be able to set higher prices and generate more profits than two owners each controlling one such establishment (e.g., Levy and Reitzes 1992).

distance from HQ is associated with true performance differences, rather than simply longer survival, by combining survival and revenue analyses in our hotel dataset. If owners or managers prefer nearby businesses to remain open for nonpecuniary reasons, we should observe a negative relationship between distance and survival, but not between distance and revenue.

II. Data and Basic Data Patterns

A. *The Texas Sales Tax and Hotel Revenue Tax Datasets*

Our dataset of retail and service businesses in Texas is uniquely comprehensive. We identify every entity that collected sales taxes from consumers in the state—and hence every establishment that sold goods or services to end consumers—between 1990 and 2003. 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. After grouping records relating to the same business establishment, and eliminating those lacking valid SIC codes (primarily due to the 2004 switch to NAICS), those with out-of-state headquarters (because these owners often have many establishments in other states), those founded before 1990 (because of possible survivor bias), and some with missing data, we are left with records for 1,713,602 separate establishments founded between 1990 and 2003 for which we have all the required information for our analyses below. (See the Appendix for more details).⁴

For each establishment, we know the identity and address of the owner and the name and address of the business. The owner's stated address, which we refer to as its headquarters or HQ, can be one of three types. For 840,710 establishments, or 49.1 percent of them, the HQ address is that of an establishment of this owner—the one under consideration or any other one. For another 526,206 establishments (30.7 percent), the owner HQ address is a street address unrelated to any establishment. This could be a residential home address or a dedicated office address, but we cannot distinguish these. For 346,686 (20.2 percent) establishments, the owner address is a post office box. Of course, no administrative activity can take place at the latter. More generally, we cannot ascertain how much administrative activity takes place at any of the HQ addresses provided by owners. But if the addresses do not function as good proxies for owners' true administrative bases, analyses based on them will be noisy, reducing the likelihood that we observe any effect.

For each business, we also know the date at which it began operations under its owner, which we refer to as the founding date. We treat the earliest outlet that we observe in the data—whether it exists already in 1990 or is opened later—as the owner's first. A few owners might have opened and closed outlets prior to 1990, but we cannot ascertain that.

If it occurs before the end of our data, i.e., before October 2006, we also know when the establishment goes out of business, or is acquired if applicable. The fact

⁴ Unfortunately, we do not observe the level of taxes paid, and thus cannot infer revenues for these businesses.

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 108,337, or only 6.3 percent, of the 1.71 million establishments in our data.

The data also include information about each owner’s legal ownership form. Proprietorships are the form of choice for owners of 1,186,384 establishments, or 69.2 percent of them. Texan corporations own another 303,621 (17.7 percent) of establishments, while Texan limited liability companies own 26,602 of them (1.6 percent). Another 150,445 (8.8 percent) are operated by general partnerships, and 25,819 (1.5 percent) by limited partnerships. A few establishments (20,731 or 1.2 percent of them) belong to foreign (non-Texan) corporations with HQs in Texas.⁵ Typically these owners are public companies based in and operated from Texas, but registered in another state.

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, so we do not include establishments founded in 2004 or after in any of our analyses.⁶ However, we 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, quarterly revenue and size (number of rooms) data in addition to all the information described above, with the exception of legal ownership form. This allows us to analyze revenue effects using an unbalanced panel of 191,513 quarterly observations from the 5,457 hotels of in-state owners in operation in Texas between 1990 and 2008. Many hotels are excluded from the larger sales tax dataset because they do not collect sales taxes for room rentals.

For all the businesses in our data, we geocoded latitude and longitude coordinates for the business and owner HQ address, more than three million addresses total. We then calculated the distance in miles from each of the businesses to its owner’s HQ.⁷ Table 1 shows the distribution of these distances for the establishments in the sales tax dataset. Three main patterns emerge. First, a substantial number of owners open even their first few establishments some distance away from their HQ: 20 percent, or 233,211 of owners’ very first establishments, are opened eight or more miles away from the owner’s HQ. The mileage distance for the eightieth percentile jumps to 11.0 miles (19.0 miles) for owners’ second (third through fifth) establishments.

Second, distance to HQ is typically smallest for an owner’s first establishment. Owners often use their first establishment as the HQ location, as indicated by the more than 40 percent of such establishments that have a 0.0 distance from the owner

⁵ Recall that we limit our data to establishments with Texan HQs. All our results remain the same if we exclude the establishments of non-Texan corporations with such HQs.

⁶ In 2004, the source 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 businesses started after our 2004 data download (in July).

⁷ 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. 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—DISTANCES (IN MILES) TO HQ FOR ALL BUSINESSES FOUNDED 1990 TO 2003

Percentile	All establishments	Owners' 1st establishment	Owners' 2nd establishments	Owners' 3rd–5th	Owners' 6th–10th	Owners' 11th +
10	0.0	0.0	0.0	0.0	0.2	6.0
20	0.0	0.0	0.0	0.0	2.4	12.8
30	0.0	0.0	0.0	0.3	5.3	22.2
40	0.0	0.0	0.4	1.7	9.2	46.3
50	1.3	0.6	1.7	3.4	14.6	109.3
60	2.9	2.1	3.3	6.0	24.2	162.7
70	5.6	4.2	6.0	10.5	48.4	201.9
80	10.7	8.1	11.0	19.0	127.9	238.7
90	26.3	17.8	25.1	61.8	225.7	292.0
100	760.2	760.2	741.3	741.3	745.8	735.1
Total	1,713,602	1,166,055	334,952	155,238	22,266	35,091

HQ. This proportion decreases monotonically for owners' later establishments.⁸ Similarly, the median distance of an establishment to its owner's HQ increases monotonically, from 0.6 miles for an owner's first to 109.3 miles for the 11th+. The increase is large because the final column includes many establishments of the large wholly owned chains with Texan HQs.

Third, the table shows that 1.17 out of 1.71 million establishments are opened by owners with no previously opened establishment. Another 334,952 establishments are opened by owners who have only one previous establishment. So while large chains have grown tremendously in recent years, as described in Foster, Haltiwanger, and Krizan (2006) and Basker, Klimek, and Hoang Van (2012), the vast majority of sales-tax collecting businesses are owned by individuals or small corporations that own at most one other establishment.⁹

III. Methods and Results

A. Duration Analyses

We analyze the survival of the 1,713,602 establishments in our data using a Weibull duration model. In this model, the hazard, or instantaneous transition from origin (active business) to destination (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. A popular choice for the nonnegative function of the covariates, $g(X)$, is $e^{X\beta}$, where X is the vector of independent variables, and β is a vector of coefficients.

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.

⁸ The 0.0 distance to HQ for establishments opened after an owner's first arises from (i) new establishments opened at the same address as the first establishment (e.g., at a mall), and operated simultaneously with the first; (ii) new establishments opened nearby, in cases where geocoding lacks perfect accuracy; or (iii) owners' moves to a new HQ.

⁹ Within franchised chains, the owners are the franchisees. They, too, tend to have only one or few outlets.

TABLE 2—DURATION REGRESSIONS; DESTINATION STATE IS ESTABLISHMENT EXIT

	Preexisting HQs by:					
					3 years	5 years
Log distance to HQ (in miles + 1)	1.09*** (0.001)	1.10*** (0.001)	1.09*** (0.001)	1.10*** (0.001)	1.17*** (0.002)	1.16*** (0.003)
Log (number of previous estabs + 1)		0.86*** (0.001)		0.87*** (0.001)	0.89*** (0.004)	0.91*** (0.004)
Dummy variables:						
SIC codes			522	522	522	521
Establishment zip codes			1,503	1,503	1,503	1,503
Number of establishments	1,713,602	1,713,602	1,713,602	1,713,602	182,789	119,899

Notes: Coefficients are exponentiated. Standard errors and significance levels are for original coefficients. Regressions also include dummy variables for HQ geocode quality and for establishment geocode quality, for organization type, and for year of founding.

*** Significant at the 1 percent level.

Specifically, if the exponentiated coefficient b is greater (smaller) than one, the difference $(b-1)*100$ indicates the percentage by which a one-unit increase in the explanatory variable would increase (decrease) the hazard of exit. Because exponentiated coefficients are more easily interpreted, we show these in Tables 2 and 3. As is standard in survival analyses, we present standard errors and show statistical significance for the original 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 6.3 percent of establishments that are acquired),¹⁰ or from January 1, 2007, one day after the last end date in the data, if the establishment is still open at that time (right-censored observations).

In Table 2, we regress duration on the log of (distance in miles +1) between the establishment and its owner's HQ. In columns 2, 4, and 6, we include the (log of) number of establishments previously opened by the owner. To control for various dimensions of heterogeneity that could affect exit rates, all regressions include dummy variables for the six forms of organization (proprietorships, limited and general partnerships, Texan limited liability, and Texan and non-Texan corporations), the year of founding, which control for macroeconomic conditions at the time of founding, and for the seven geocoding quality for the establishment and also for the owner HQ addresses.¹¹ In columns 3 to 6, we add dummy variables for each four-digit SIC code and each establishment zip code in the data.

Results in Table 2 clearly show the importance of distance to HQ in relation to establishment exit. Every additional unit of the log of (distance to HQ +1) results in a 9 to 10 percent increase in the hazard of exit. This effect is quite consistent across

¹⁰ As we mention in Section III E, where we discuss robustness, treating these as exits does not affect any of our results.

¹¹ 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. We have verified that our results are robust to examining only the subset of establishments and owners whose locations are geocoded at a high-quality level.

all our specifications.¹² Moreover, the distance effects are economically important: in our most flexible specification (column 4), a one standard deviation increase in the log of (distance to HQ +1) from the mean value of 1.275, to 2.78, reduces the probability that the establishment survives for 3 (5) years from 45 percent to 40 percent (28 percent to 23 percent). Note that when distance is zero, the probability of an establishment surviving 3 (5) years is 49 percent (33 percent).

One potential explanation for the negative distance-to-HQ effect is that the most promising establishment might be the most desirable place for an owner to locate his or her HQ. In the last two columns of Table 2, we include only establishments whose owners' first and only HQ address predates the establishment under observation by at least three and five years. Results for this subsample confirm that there is a strong distance effect, even for these establishments whose owner HQ locations can be considered predetermined. We conclude that the distance-to-HQ effect is not due to locating HQs ex post in or near the establishments with the most potential. In fact, the distance-to-HQ effects for this group are larger than those for the overall population, at 16 to 17 percent per unit increase in the (log of) distance: results in the last column imply that the probability of surviving 3 (5) years goes from 63 percent to 54 percent (48 percent to 38 percent) as the log distance measure is increased by one standard deviation from its mean for this sample, i.e., from 1.72 to 3.53.

In Table 3, we show duration results for subpopulations defined on the basis of owners' previously opened number of establishments. We do this using the most flexible specification from Table 2, namely the one in column 4. We show regression results in columns 1 and 2, followed by the mean and standard deviation of distance to HQ in columns 3 and 4. In columns 5 to 7, we show the effect of setting the log(distance +1) to zero, and then to its mean value, and finally to its mean plus one standard deviation, while holding all other variables at their means (where means and standard deviations are specific to each sample). The number of establishments in each subset is reported in the last column.

The results from these analyses allow us to refute three of the mechanisms discussed in Section I 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. Second, if the negative distance-to-HQ effect were due only to "within-owner" sequential expansion processes from most- to least-desirable locations, we again should not find a negative distance-to-HQ effect within the sample of first establishments. Third, our results for establishments that are not the first suggest that monitoring and information asymmetric issues, as opposed to commuting, are an important driver of the distance-to-HQ result in our data. Owners of two or more establishments must rely on employees and salaried managers in at least some of their establishments. For such owners, commuting costs can only be invoked as a reason to close far away

¹² Also, including HQ zip code dummy variables leads to slightly higher coefficients for distance (see online Appendix). As our goal with the zip code dummy variables is to control for differences in the markets where the establishments are located, we focus on the more conservative results obtained without HQ zip code dummy variables.

TABLE 3—DURATION REGRESSIONS, BY NUMBER OF ESTABLISHMENTS OPENED PREVIOUSLY

Establishment is owner's:	Log distance to HQ (miles +1)		Log distance descriptive stat.		Estimated probabilities of surviving 3+ years (all other vars. at mean)			Number of estabs
	Exp. coeff.	SE	Mean	SD	At zero distance	At mean distance	At mean +1 SD	
First	1.08***	0.001	1.13	1.41	0.45	0.42	0.38	1,191,036
2nd	1.13***	0.001	1.33	1.47	0.54	0.49	0.43	329,584
3rd	1.14***	0.003	1.54	1.55	0.56	0.49	0.42	98,411
4th	1.14***	0.005	1.82	1.65	0.57	0.49	0.41	34,404
5th	1.12***	0.007	2.18	1.72	0.60	0.52	0.45	14,634
6th–10th	1.12***	0.007	2.75	1.76	0.65	0.56	0.49	18,624
11th–25th	1.06***	0.010	3.67	1.61	0.79	0.74	0.72	10,828
26th–50th	1.07***	0.018	4.14	1.53	0.89	0.85	0.84	5,348
51st+	1.01	0.013	4.51	1.36	0.85	0.84	0.84	10,733

Notes: Coefficients in column 1 are exponentiated. Standard errors and significance levels are for original coefficients. Regressions in the last four rows include the (log of one plus) number of establishments previously opened (this variable is constant in the first five rows). All regressions also include dummy variables for HQ address geocode quality, for establishment address geocode quality, for organization type, for year of founding, and for every establishment SIC and zip code.

*** Significant at the 1 percent level.

establishments because they increase the cost of monitoring employee behavior and keeping track of the local market. The fact that distance matters more for owners' second through tenth establishments than for their first further supports the idea that monitoring and information asymmetry issues—which are more acute when the owner cannot be continuously on site—are an important factor behind the detrimental effect of distance to HQ.

Two other aspects of the results in Table 3 also are noteworthy. First, the estimated probabilities of survival at different distances (zero, mean and one standard deviation above the mean) increase quite systematically with the number of establishments opened earlier by the owner, confirming the benefit of more establishments.

Second, the distance between the establishment and HQ has a strikingly similar effect on exit rates for the 495,657 establishments (28.9 percent of the 1,713,602 establishments) that are an owner's second through tenth. This is true in terms of coefficients, and magnitudes as well. This suggests a very consistent effect of HQ distance across five establishment subpopulations with no overlap. The distance-to-HQ penalty diminishes after an owner has opened ten establishments, but it is only after an owner has opened 50 establishments that the penalty disappears. The lack of distance effect for the largest owners suggests that they rely on some mechanisms, such as relative performance evaluation across stores (Carmichael 1983), better monitoring technologies, and/or regional monitoring locations (Brickley and Dark 1987) that allow them to achieve good outcomes in their many stores despite distance. Of course, relative performance evaluation becomes more accurate with more stores to compare, while both monitoring technologies and regional monitoring locations may become profitable only when spread across a large enough number of stores.

In sum, our duration analyses show a clear and important link between HQ distance and survival. They also allow us to rule out HQ location choices, market power, and commuting as factors generating the distance-to-HQ effect in our data.

TABLE 4—PROBIT AND IV PROBIT REGRESSIONS

	Survival duration					
	3 years			5 years		
	Probit	IV Probit Stage 1	IV Probit Stage 2	Probit	IV Probit Stage 1	IV Probit Stage 2
Log distance to HQ (miles + 1)	-0.068*** [-0.026] (0.001)		-0.078*** [-0.030] (0.010)	-0.082*** [-0.026] (0.001)		-0.102*** [-0.032] (0.012)
Log (previous estabs opened + 1)	0.123*** [0.047] (0.002)	0.475*** (0.002)	0.128*** [0.049] (0.005)	0.122*** [0.038] (0.002)	0.485*** (0.002)	0.132*** [0.041] (0.006)
Proportion of workers with commute < 15 minutes		-1.50*** (0.015)			-1.48*** (0.017)	
Dummy variables:						
SIC Codes	522	522	522	522	522	522
Establishment zip codes	1,503	1,503	1,503	1,502	1,502	1,502
Total establishments (<i>N</i>)	1,306,725	1,306,725	1,306,725	1,101,180	1,101,180	1,101,180
Proportion that survives	0.384		0.384	0.242		0.242
Wald test of exogeneity			1.11			3.08

Notes: Marginal effects, dy/dx , in square brackets, are calculated at mean values, where y is the proportion of establishments that survive the specified number of years. The samples for these regressions exclude establishments founded in 1990 to avoid any possibility that owners contribute to the measure of commuting times. All regressions include dummy variables for HQ geocode quality, for establishment geocode quality, for organization type, and for year of founding. Standard errors are in parentheses.

*** Significant at the 1 percent level.

Still, unobserved owner and business location characteristics and associated endogeneity and selection issues might be driving our results. We rely on two distinct approaches to address these. First, we estimate binary probit regressions where we instrument for distance from HQ. Second, we follow the spirit of the large literatures on human twins (e.g., Ashenfelter and Krueger 1994) and patents (e.g., Jaffe, Trajtenberg, and Henderson 1993) and present a series of matched pairs analyses of twin and sibling establishments. Such analyses have the advantage, relative to duration or binary choice regression models, that the owners are held constant. This is infeasible in duration analyses given the size of our populations of establishments and owners.

B. IV Probit Regressions

We present results from binary probit survival analyses, where the dependent variables are whether an establishment has survived for at least three or five years, in Table 4. The main advantage of the probit model is that, via the control function approach (e.g., Rivers and Vuong 1988), we can rely on instrumental variables to address the issue of HQ distance endogeneity.

We use data on commuting, 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 percent of whom are wage-earning employees, whose commute to their workplace from home is less

than 15 minutes.¹³ In these analyses, we focus on establishments founded in 1991 or later to ensure that the 1990 census data are predetermined.¹⁴ Our instrument captures the tendency of the population living in the vicinity of the owner's HQ address to work nearby. We expect owners to have a greater tendency to locate their establishments farther from their HQ in locales where long commutes are the norm, and conversely to keep distance to HQ smaller in those areas where people do not commute far to work. 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, surrounding neighborhoods or cities. This, in turn, leads to the high (negative) 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 owner's HQ is located to be systematically related to the survival of establishments, in part because these may or may not be near that location.

The first and fourth columns in Table 4 show results from binary probit regressions that treat distance to HQ as exogenous. Since the dependent variable in these regressions is survival of some minimum duration, a negative coefficient denotes a reduction in survival times. Thus results in this table are consistent with our duration results, indicating again that distance to HQ has a negative effect on survival. We show marginal effects, calculated at the means, in square brackets. We find that for an establishment with the mean characteristics, a one standard deviation increase in the log of distance to HQ (miles +1), i.e., an increase from 1.20 to 2.69 in this sample, decreases the likelihood that the establishment survives at least three years from 0.38 to 0.35. The five-year survival rate decreases from 0.24 to 0.21. These effects are similar in magnitude to those from our duration analyses in Table 2.

We show instrumental variable probit results in the remaining columns of Table 4. The first stage, in columns 2 and 5, shows the high (negative) correlation between the proportion of low commuting times and distance to HQ. The *t*-statistics of the commuting time variable are -98 and -89 , respectively, for the three- and five-year regressions, implying a very strong instrument. The second-stage results, in columns 3 and 6, are very similar to the corresponding probit results in columns 1 and 4.¹⁵ The similarity of the results suggests that endogeneity issues may not be a major concern in our setting. Indeed, we cannot reject (at the 5 percent level) the hypothesis that distance between an establishment and its owner's HQ can be treated as exogenous according to the Wald test statistics reported at the bottom of Table 4. We expect this is due to the nature of retail and small-scale service industries, where owners are constrained, when it comes to choosing the location of

¹³ 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, subtracted them from one, and divided by the total number of commuters in the zip code.

¹⁴ We do this to ensure that the 8 percent of respondents that are business owners, some of whom may own businesses in our data, cannot affect the reported commuting times.

¹⁵ See also the online Appendix for a comparison of our probit results to standard OLS and linear IV regressions that ignore the binary nature of the dependent variable, as suggested by Wooldridge (2002, 472).

their establishment in particular, by both zoning laws and the need to go where their customers are.

C. Matched Pairs of Twins and Siblings

In this section, we address the issue of owner heterogeneity by focusing on twin and consecutive non-twin sibling pairs of establishments. The simple outcome that we examine is which establishment of a pair survives longer. 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 four-digit SIC code, and founded, or more typically 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 owner characteristics, such as their quality and reservation wage, but also for HQ type and HQ location, which are fixed at the owner level in these analyses.

For each multi-establishment owner, we sample randomly one pair of sequential non-twin siblings and/or twins that exhibits within-pair variation in distance to HQ and survival time.¹⁶ For siblings, we only include those establishments of owners whose ownership form and HQ address remain identical for both. We analyze twins separately from 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. Still, sibling results are valuable because they represent a much larger, and yet still within-owner sample. They also allow us to address directly how the order of opening affects survival separately from distance. Since our analyses of twins and siblings do not directly address potentially endogenous HQ location, as in Table 2, we also analyze subsamples of establishments whose owners' HQs were set well before the establishments themselves.

Results in Table 5 show a strong survival benefit from being closer to the owner's HQ. For 58.1 percent of the 9,457 pairs of twins, and for 57.7 percent of the 181,283 pairs of non-twin siblings, the closer establishment remains in business longer. Rows 1a and 1b in Table 5 address 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 is opened earlier or later. Complementing the duration result for owners' first establishments from Table 3,

¹⁶ Sampling 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, ten 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, and the results remained equivalent to those in Table 5.

TABLE 5—RELATIVE SURVIVAL WITHIN PAIRS OF TWINS AND NON-TWIN SIBLINGS

	TWINS: percent of cases where twin closer to HQ survives longer	SIBLINGS: percent of cases where non-twin sibling closer to HQ survives longer
1. All pairs	58.1 percent of 9,457***	57.7 percent of 181,283***
1a. First sibling closer	NA	60.2 percent of 56,725***
1b. Second sibling closer	NA	56.5 percent of 124,558***
2. Siblings/twins are first establishments	60.9 percent of 5,041***	58.4 percent of 131,761***
3. One previous establishment	57.7 percent of 1,361***	55.8 percent of 34,995***
4. 2–4 previous establishments	57.3 percent of 1,181***	55.9 percent of 12,420***
5. 5+ previous establishments	51.5 percent of 1,874*	53.3 percent of 2,107***
6. Same HQ exists > 3 years	54.0 percent of 2,037***	59.8 percent of 13,438***
7. Same HQ exists > 5 years	53.7 percent of 1,545***	59.6 percent of 8,581***

*** Significant at the 1 percent level using a binomial distribution test with $H_0 = 50$ percent.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

this result alleviates the concern that survival advantages might be driven by the sequence of entry rather than distance per se.

The remainder of Table 5 displays the relative likelihood of longer survival for the closer establishment for subsamples constructed to reproduce analyses in Tables 2 and 3. Rows 2–5 separate the twin and sibling pairs by the number of establishments previously opened by their owner. Consistent with our results in Table 3, we find a decrease in the importance of HQ distance with increases in the number of prior establishments. This again is in line with the existence and success of large chains in the retail and small-scale service sectors. Rows 6–7 present results for pairs with predetermined HQ locations, as we did in the last two columns of Table 2. As before, we find that the advantage of being closer to the HQ is strong in these subsamples as well.

D. HQ Moves, Acquisitions, and Revenues in the Lodging Industry

In this section, we present results for our lodging industry data, the one industry for which the Texas Comptroller makes establishment-level revenue data public. Our examination of revenues complements our survival analyses in two main ways. First, the HQ-moving hotel owners provide another setting where the owner and establishment can be held constant while analyzing distance effects. Specifically, we include fixed effects for every hotel/owner combination in our HQ-mover analysis, and for every hotel in our acquisitions analysis. Thus, the identification of the distance-to-HQ effect is achieved from changes due to owners moving their headquarters or from changes in ownership. Second, we can shed further light on the cause of the longer survival periods for establishments close to their owners' HQs, in particular we can assess whether nonpecuniary or perquisite-taking motives might explain the longer survival of establishments that are closer to their owner's HQ.

The dependent variable we focus on is total room revenues per calendar quarter per available room (RevPAR), the standard top-line performance metric used in the industry. The median RevPAR per calendar quarter among these hotels is \$1,718

and the mean is \$2,366. Our independent variable of interest again is the distance from the establishment (the hotel) to the owner's HQ. Between 1990 and 2008, 667 in-state owner HQ moves (i.e., changes in HQ location without a change in ownership) and 3,601 acquisitions by in-state owners (i.e., changes in HQ location that accompany an ownership change) took place among the 5,457 properties.¹⁷

In addition to hotel or hotel/owner fixed effects, we include calendar quarter fixed effects to control for macroeconomic fluctuations and potential seasonality, as well as fixed effects for all main chains (e.g., Holiday Inn, Econo Lodge) in the data. A total of 2,622 hotels operated under a major 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 hotels switch brand affiliation, or go from affiliated to independent operations or vice versa. Note that, as in our overall sample analyses, our owner HQ locations are always those of the Texas-based franchisees and not those of the franchisor.

Before analyzing the effects of changes in distance to HQ on revenues, we first verified that the same detrimental effect of distance on survival occurs in our hotel data. Using data on the population of 5,457 hotel properties (all establishments with more than 20 rooms, a standard industry cutoff) of in-state owners in operation in Texas between 1990 and 2008, we found that whether we used duration or probit analyses, distance to HQ had a statistically significant negative effect on the survival of hotels just as it did in our overall data.

Turning now to our revenue analyses, we report regression results for acquisitions in the first two columns of Table 6. In these regressions, we rely only on the 5,207 periods of ownership by owners who never moved their HQs so that the acquisition results are net of the potential HQ-mover effects shown in columns 3 and 4. As before, we show results when we control for owner size, using the log of the number of hotels of the owner in the calendar quarter, in the second column. In the third and fourth columns of Table 6, we focus on the sample of hotels where the owner remains the same, but has moved its headquarters. Again we use fixed effects for the 667 hotels, and include dummy variables for brands and calendar quarters.

Results in Table 6 show a negative effect of distance to HQ on revenues: all coefficients of the distance to HQ variable are negative, and except for column 1, they are all significant. In column 1, an acquiring owner with a more distant HQ experiences a loss of \$15.54 for each unit increase in the (log of) mileage distance to HQ. When we control for the number of establishments of the owner in the calendar quarter, as we should, the reduction in RevPAR is even larger, at \$22.79. These represent reductions of 0.7 to 1.0 percent, given average RevPAR of \$2355 in this subsample. Of the 5,207 hotels included in analyses in the first two columns, a total of 3,121 never changed hands over the time we observe them, 1,166 changed ownership once, 533 twice, 247 three times, and 140 more than three times. Estimating separate regressions for the hotels with only one, or with two, or three or more acquisitions (the

¹⁷ 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. Finally, some properties are acquired more than once, as described below.

TABLE 6—QUARTERLY REVENUE PER AVAILABLE ROOM REGRESSIONS: 1990–2005

	Acquisitions [†]		HQ movers (owner remains same)	
Log distance to HQ (in miles)	-15.54 (12.81)	-22.79* (12.89)	-54.48*** (18.97)	-59.10*** (19.38)
Log number of Establishments		123.31*** (42.28)		-249.51* (138.63)
Fixed effects:				
Hotels	5207	5207	667	667
Brands	69	69	18	18
Calendar quarters	73	73	73	73
Observations	163,772	163,772	27,741	27,741

Note: Standard errors clustered at the hotel level in parentheses.

[†] We exclude periods of ownership by HQ-moving owners so that the HQ-mover and Acquisition subpopulations have no overlap.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

equivalents of Table 6, columns 1 and 2) also yields negative effects of distance to HQ.

In the last two columns of Table 6, we find larger monetary losses in quarterly RevPAR associated with increased distance to HQ due to an owner move, namely \$54.48 and \$59.10 for each unit increase in the (log of) mileage distance to HQ. In other words, holding owner and establishment characteristics constant, changes in distance to HQ have a clear detrimental effect on top-line performance. The average RevPAR in this sample is similar to that above, at \$2,435. Thus the effects of distance for these owners amount to reductions of more than 2 percent of top-line results. It would appear then that owners that move their HQs are particularly susceptible to the type of issues that arise with distance. Moreover, the fact that “within owners” an increase in the number of establishments has a negative effect on RevPAR suggests that the owners also have difficulties managing larger holdings. In the acquisition subsample, as in our earlier analyses, larger holdings are associated with larger RevPAR. In other words, being purchased by an owner with more establishments has a positive effect on a given hotel’s RevPAR, but holding owner constant, more establishments means more demands on the owner’s time, which translates into poorer performance in these data. This suggests that the number of businesses of an owner in our cross-sectional analyses captures some version of owner quality, as discussed earlier.

In sum, our revenue results confirm the negative performance consequences of increases in distance to HQ. Since our estimates may be biased towards zero because HQ-movers and acquiring owners would be unlikely to move their HQs to a distant location if they expected that this would severely hamper their operations, we find it noteworthy that we observe a negative HQ-distance effect on RevPAR at all. Moreover, because we do, we can conclude that perquisite taking by managers (Landier, Nair, and Wulf 2009) either is not at play in our context, or is outweighed by the real performance benefits induced by more hands-on management by local owners.

E. Robustness Tests

We have conducted a number of robustness tests for all regressions and pair comparisons above, including reproducing results from Table 2 and the first rows of Table 5 for a variety of subpopulations. We discuss some of these here. More details are available in the online Appendix.

First, we handled the 6.3 percent of establishments that are acquired in our overall sample 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 far less than predicted future cash flows would warrant (see, e.g., Fraser 1999) when they sell their businesses. However, our results are unchanged if we use the final exit date to measure survival spans for all establishments. Further, our results remain the same if we simply exclude from our analyses all establishments that are acquired. Our results even hold for the set of 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 sets of establishments founded each year between 1990 and 2003. We found that distance to HQ was an important factor throughout, but there was no evident trend downward in this effect—as perhaps the availability of new technologies or methods of doing business might warrant—or in the survival times of establishments. If anything, we found some evidence of an increasing effect of distance to HQ over time in our data. We conclude that technological advances, such as Internet-based or electronic reporting, had either not yet been adopted much by the type of small firms we focus on at the time of our data, or have not attenuated the effect of distance to HQ on survival.

Third, we separated the observations for retail (SICs of 5200-5999), services (SICs of 7000-7999), and an “other” category. We found that the negative distance-to-HQ effect was important in all three, but larger among service sector establishments than in retail. This result is as one would expect if the role of the owner in monitoring employees and overseeing activities locally is even more important in service oriented establishments than in retail.¹⁸

Fourth, the results also hold separately for all six major ownership forms, although they are smaller for general partnerships (8.8 percent of establishments) and the non-Texan corporations (1.2 percent of establishments). 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 smaller effect is consistent with results in Table 3.

Fifth, we found that distance to HQ had a negative effect within all three subsamples based on HQ address types (address of an establishment, PO Box, or separate

¹⁸ We also carried out separate analyses for many sectors that were both well defined (e.g., not miscellaneous categories) and well represented in our data. These analyses confirmed that the distance-to-HQ effect was important in all of them, even as it differed somewhat across industries. See the online Appendix for details.

street address). The distance-to-HQ penalty is largest when the HQ is established in an establishment. It remains sizable even for the PO Box subpopulation, which suggests that, for the 19.7 percent of establishments with a PO Box as “HQ address,” the owner’s indication of a HQ address separate from its business address remains a meaningful proxy for the location of the owner’s activities.

Sixth, we assessed the effect of distance separately for establishments located in rural and urban areas, using both duration and probit analyses. We found that the negative distance-to-HQ effect was important in both sets, and the magnitudes were similar as well. We further analyzed how the combination of locations, HQ and establishment, as rural and urban might affect the size of the distance effect. We found that distance had an important negative effect on survival in all cases, but that the effect was largest for owners with urban HQs and their urban establishments.¹⁹

Finally, in honor of Ray Kroc’s statement mentioned in the introduction, we carried out an analysis of the effect of distance to owner (i.e., franchisee) HQ among McDonald’s establishments. Despite the very standardized business format and controls under which they operate, we again found a negative and statistically significant effect of distance to owner HQ for the establishments of McDonald’s Texan franchisees, who on average own about three establishments each (see also Kalnins and Lafontaine 2004).

IV. Summary and Conclusion

In this paper, we have investigated the effect of geographic separation of headquarters and establishment location on performance in the retail and service sectors, sectors that contribute many more jobs to the US economy than does manufacturing.²⁰ 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 (i) its survival duration under that owner, for all but the very largest owners, using the entire population of retail and service businesses in Texas that serve the end consumer, and (ii) its revenues under that owner, using data on the Texas lodging industry. These results arise 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. We view the robust negative distance effect in our data as strong evidence of the type of monitoring and informational asymmetries emphasized by Hayek (1945) in his seminal work, and by the more recent theory of the firm literature.

We did find evidence that the advantages of proximity to HQ dissipate for owners of fifty establishments or more. This suggests that owners of retail and service chains have found mechanisms—including perhaps capital-intensive monitoring

¹⁹ We also verified that our results hold if we use a different measure of distance, namely the minimum number of zip codes that one would have to traverse to get from the HQ zip code centroid to the establishment’s zip code centroid, a measure formally known as the “order of contiguity” among economic geographers. This measure provides an alternative way to capture the fact that the same distance in miles may not have the same impact in rural and urban areas. See the online Appendix, Table A7.

²⁰ According to the Bureau of Labor Statistics, in 2004, 15 of the 110 million employees in the United States worked in manufacturing, while 15 million worked in the retail, and 13 million in the leisure and hospitality sectors (table B-3: employees on nonfarm payrolls by major industry sector and selected industry detail).

technologies or regional monitoring locations whose presence would not be captured in our data—allowing them to operate efficiently outside the immediate vicinity of their HQ. However, we have few large owners in our data. The vast majority of establishments, more than 1.50 of the 1.71 million establishments, are owned by individuals with only one or two of them. The performance of these establishments is very much influenced by their distance to their owners' HQ. Assuming that on average these establishments rely on the labor of the owner and perhaps one or two other individuals, their success or failure will have affected the livelihood of a probable three to four million individuals in Texas at some point during our data period.

Firms with a potentially lucrative business concept, in part, can address the problem of distance via franchising, i.e., by choosing franchisee owners specifically because they have HQs close to the site of an establishment (e.g., Brickley and Dark 1987; Kalnins and Lafontaine 2004). Our results in that sense are consistent with the large literature on franchising summarized by Lafontaine and Slade (2007), which finds that outlets located further away from a chain's HQ are much more likely to be franchised.

Our results can be reconciled with the growth of chains in the retail and small-scale service sectors, in the sense that many of the chains whose success is noted in the literature are franchised. Still, we view explaining the lack of a distance-to-HQ effect for large owners in our data, and explaining the success of large chains such as Starbucks or Home Depot that do not rely on franchising, as important areas of future research. Indeed, the presence of such corporately owned and operated chains demonstrates that large firms can manage the economically important monitoring and informational costs associated with increased distance to HQ. A better understanding of the mechanisms they rely on to achieve this, we believe, could help alleviate the turnover costs experienced by smaller owners and sole proprietors, and by their employees and consumers alike, throughout these sectors.

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 dates shown in Table A1.

The state of Texas purges its inactive dataset of older tax permits. In 1995, it kept permits for five years after the out-of-business date. After that, permits were purged

TABLE A1—DATA ACQUISITION INFORMATION

	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

Notes: 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.

within four years and then three 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 datasets, we end up with a list of all businesses establishments in operation at some time between 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 owners whose 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 grouped establishment records under the same owner and establishment identification number if we found that the establishment name and address were the same. We set the establishment start date of the remaining 2,930,309 potential business establishments to the earliest first-sale date associated with the business establishment, and the end date to the latest end-of-business date associated with the same.

Having identified our basic establishments, we proceeded to delete 182,205 of them that appeared at some point in some active files, but never appeared again in a subsequent active or 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. We removed another 193,859 establishments because the establishment or HQ did not have a valid Texas zip code. The majority of these were establishments with out-of-state owners. We also removed 44,155 establishments that were of organizational forms other than the large six (proprietorships, Texan corporations, Texan limited liability companies, foreign corporations, general partnerships, and limited partnerships). Most of these were nonprofits, and government, religious, social, and professional organizations. Finally, we removed establishments founded after 2003, and some whose zip codes were outside Texas. These subtractions led to a set of 2,087,982 different business establishments founded in the state of Texas before or in 2003 that we could rely on to measure owner characteristics, such as order of founding.

As for our analyses, we further reduce the set that we consider to include only those founded after 1990. We also lose some observations due to lack of data on location or industry, or because the SIC or zip codes were too rare (likely typos). Removing these establishments also made our analyses with SIC and zip codes dummy variables manageable. The final result is a dataset of 1,713,602 establishments founded in Texas between 1990 and 2003.

REFERENCES

- Adams, James D., and Adam B. Jaffe.** 1996. "Bounding the Effects of R&D: An Investigation Using Matched Establishment-Firm Data." *RAND Journal of Economics* 27 (4): 700–721.
- Allen, Treb.** 2012. "Information Frictions in Trade." <https://dl.dropbox.com/s/bq3hdyv2x7qrgq4/Allen%20-%20Information%20Frictions%20in%20Trade.pdf>.
- Anderson, James E., and Eric van Wincoop.** 2004. "Trade Costs." *Journal of Economic Literature* 42 (3): 691–751.
- Ashenfelter, Orley C., and Alan B. Krueger.** 1994. "Estimates of the Economic Return to Schooling from a New Sample of Twins." *American Economic Review* 84 (5): 1157–73.
- Basker, Emek, Shawn Klimek, and Pham Hoang Van.** 2012. "Supersize It: The Growth of Retail Chains and the Rise of the 'Big-Box' Store." *Journal of Economics and Management Strategy* 21 (3): 541–82.

- Baum, Joel A. C., Stan Xiao Li, and John M. Usher. 2000. "Making the Next Move: How Experiential and Vicarious Learning Shape the Locations of Chains' Acquisitions." *Administrative Science Quarterly* 45 (4): 766–801.
- Brickley, James A., and Frederick H. Dark. 1987. "The choice of organizational form The case of franchising." *Journal of Financial Economics* 18 (2): 401–20.
- Bronnenberg, Bart J., Sanjay K. Dhar, and Jean-Pierre H. Dubé. 2009. "Brand History, Geography, and the Persistence of Brand Shares." *Journal of Political Economy* 117 (1): 87–115.
- Carmichael, H. Lorne. 1983. "The Agent-Agents Problem: Payment by Relative Output." *Journal of Labor Economics* 1 (1): 50–65.
- Coval, Joshua, and Tobias Moskowitz. 2001. "The Geography of Investment: Informed Trading and Asset Prices." *Journal of Political Economy* 109 (4): 811–41.
- Dahl, Michael S., and Olav Sorenson. 2012. "Home Sweet Home: Entrepreneurs' Location Choices and the Performance of Their Ventures." *Management Science* 58 (6): 1059–71.
- Enz, Cathy A. 2004. "Issues of Concern for Restaurant Owners and Managers." *Cornell Hospitality Quarterly* 45 (4): 315–32.
- Foster, Lucia, John Haltiwanger, and C. J. Krizan. 2006. "Market Selection, Reallocation, and Restructuring in the U.S. Retail Trade Sector in the 1990s." *Review of Economics and Statistics* 88 (4): 748–58.
- Fraser, Jill Andresky. 1999. "When It's a Seller's Market." *Inc.*, July 1. Accessed May 18, 2013. <http://www.inc.com/magazine/19990701/823.html>.
- Freeman, John H., and Pino G. Audia. 2006. "Community Ecology and the Sociology of Organizations." *Annual Review of Sociology* 32 (1): 145–69.
- Giroud, Xavier. Forthcoming. "Proximity and Investment: Evidence from Plant-Level Data." *Quarterly Journal of Economics*.
- Granovetter, Mark. 2005. "The Impact of Social Structure on Economic Outcomes." *Journal of Economic Perspectives* 19 (1): 33–50.
- Greening, Daniel W., Bruce R. Barringer, and Granger Macy. 1996. "A qualitative study of managerial challenges facing small business geographic expansion." *Journal of Business Venturing* 11 (4): 233–56.
- Hayek, F. A. 1945. "The Use of Knowledge in Society." *American Economic Review* 35 (4): 519–30.
- 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 (1): 53–74.
- Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics* 108 (3): 577–98.
- Kalnins, Arturs, and Francine Lafontaine. 2004. "Multi-unit ownership in franchising: evidence from the fast-food industry in Texas." *RAND Journal of Economics* 35 (4): 747–61.
- Kalnins, Arturs, and Francine Lafontaine. 2013. "Too Far Away? The Effect of Distance to Headquarters on Business Establishment Performance: Dataset." *American Economic Journal: Microeconomics*. <http://dx.doi.org/10.1257/mic.5.3.157>.
- Katona, George, and James N. Morgan. 1952. "The Quantitative Study of Factors Determining Business Decisions." *Quarterly Journal of Economics* 66 (1): 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 Franchising in the Fast Food Industry." *Quarterly Journal of Economics* 106 (1): 75–101.
- Lafontaine, Francine. 1992. "Agency theory and franchising: some empirical results." *RAND Journal of Economics* 23 (2): 263–83.
- Lafontaine, Francine, and Kathryn L. Shaw. 2005. "Targeting managerial control: evidence from franchising." *RAND Journal of Economics* 36 (1): 131–50.
- Lafontaine, Francine, and Margaret Slade. 2007. "Vertical Integration and Firm Boundaries: The Evidence." *Journal of Economic Literature* 45 (3): 629–85.
- Landier, Augustin, Vinay B. Nair, and Julie Wulf. 2009. "Trade-offs in Staying Close: Corporate Decision Making and Geographic Dispersion." *Review of Financial Studies* 22 (3): 1119–48.
- Leamer, Edward E., and Levinsohn, James. 1995. "International Trade Theory: The Evidence." In *Handbook of International Economics*, Vol. 3, edited by Gene M. Grossman and Kenneth Rogoff, 1339–94. Amsterdam: Elsevier Science.
- Lerner, Josh. 1995. "Venture Capitalists and the Oversight of Private Firms." *Journal of Finance* 50 (1): 301–18.

- Levy, David T., and James D. Reitzes.** 1992. "Anticompetitive Effects of Mergers in Markets with Localized Competition." *Journal of Law, Economics, and Organization* 8 (2): 427–40.
- Portes, Richard, and Hélène Rey.** 2005. "The determinants of cross-border equity flows." *Journal of International Economics* 65 (2): 269–96.
- Rauch, James E.** 2001. "Business and Social Networks in International Trade." *Journal of Economic Literature* 39 (4): 1177–1203.
- Rivers, Douglas, and Quang H. Vuong.** 1988. "Limited information estimators and exogeneity tests for simultaneous probit models." *Journal of Econometrics* 39 (3): 347–66.
- Wooldridge, Jeffrey M.** 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.