

The economic advantages of chain organization

Brett Hollenbeck*

This article considers the rapid spread of chain firms in many industries. The conventional explanation is that chains generate economies of scale in costs. Alternatively, the structure of chains may enhance demand by helping firms develop reputations, among other reasons. I quantify the value of these explanations empirically with a large, detailed data set on the hotel industry, combining a reduced-form analysis of revenues with a structural estimation of firm costs. Revenue analysis shows substantial evidence of a large chain premium. Cost estimation shows that after accounting for unobserved heterogeneity, chain-affiliated firms receive no cost advantage relative to independent firms.

1. Introduction

■ Over the past several decades, the chain business model has come to dominate most retail and service industries. A striking growth in shares of sales, employment, and number of establishments organized as chains has been well documented (Foster, Haltiwanger, and Krizan, 2006; Jarmin, Klimek, and Miranda, 2009; Foster et al., 2015).¹ This phenomenon is not just widespread but has been shown by economists to matter for issues ranging from employment and wages, to competition, to aggregate productivity.² This broad scope reflects the fact that organizational form matters for a variety of economic outcomes. The success of the horizontal chain business model has been striking, especially considering the large fees paid to affiliate with chains, but what

*University of California, Los Angeles; brett.hollenbeck@gmail.com.

This article has previously circulated under the title “The spread of horizontal chains: efficiency or information?” I would like to thank Eugenio Miravete and Stephen Ryan for their guidance and support. I also thank Allan Collard-Wexler, Stephanie Houghton, Mike Mazzeo, Sanjog Misra, Peter Rossi, David Sibley, Haiqing Xu, and Rostislav Bogoslovskiy for their help as well as all participants in the UT-Austin weekly IO Seminar and seminar participants at CalTech, UCLA, Chicago-Booth School of Management, FTC-Bureau of Economics, DOJ Antitrust Division, Stanford GSB, the University of British Columbia, the Stanford Institute for Theoretical Economics, and the 2015 Yale Marketing-Industrial Organization Conference.

¹ I use “chains” to mean any business that operates multiple outlets offering similar goods or services under the same banner. The spread of chains can be seen in Figure 1, taken from Foster et al. (2015), which uses a slightly more restrictive definition.

² Examples of work on wages and employment include Basker (2005) and Jarmin, Klimek, and Miranda (2009). For work on competition, see Basker and Noel (2009) and Jia (2008), among others. Research on aggregate productivity includes Foster, Haltiwanger, and Krizan (2006) and Bloom, Sadun, and Reenan (2013). This is just a sample of the available literature, for an overview of work related to chains and franchising, see Kosova and Lafontaine (2012).

explains this success? The conventional explanation is that firms form chains to take advantage of economies of scale and operate with lower costs. An alternative explanation is that the advantage to chain affiliation operates through revenues and that the chain structure itself provides this revenue advantage. These effects have different potential implications for welfare and for the future of competition in these industries. This article seeks to examine the revenue effects of chain affiliation in comparison to cost-side effects in the context of the hotel industry in Texas.

A natural explanation for why independent businesses would join together into a chain network is that this allows them to exploit economies of scale to lower costs and enhance efficiency. Fixed costs can be spread over more firms, inputs can be purchased with greater bargaining power and distributed using an efficient network. Firms may have substantially lower capital costs and insurance costs when affiliated with a known and successful chain. These types of economies of scale have been shown to be important drivers of the “big box” retail chains that have been the focus of the extant literature on the economics of chain firms.³

Alternatively, for a given level of costs, chain affiliation may benefit firms by increasing revenues. In many industries, particularly industries with little repeat business or that sell experience goods, consumers have very little prior information on the quality of the product or service that firms offer. If consumers are risk averse, they favor firms whose reputation they know and may find search and experimentation costly.⁴ Firms in settings where little information is available may affiliate with one another and operate under one banner to generate a reputation for consistent quality to attract risk-averse consumers. Firms affiliated with a well-known chain could thus earn a substantial premium over independent firms offering uncertain quality even if the underlying product, and hence costs, is identical. In essence, chains may be offering a solution to a lemons problem by facilitating repeat interactions that could not otherwise occur by providing uniform services in many settings. This is, in a way, economies of scale on the demand side, as the larger a chain is, the more consumers have the opportunity to interact with it, increasing its value. The same effect could be seen as firms invest in national advertising to develop a reputation over time. Finally, information learned by firms on revenue management strategies and shared within chains might also increase revenues, as could access to centralized booking systems.

This article is the first to consider in depth this demand-side explanation for chain affiliation. I examine both cost and revenue explanations empirically over time within the same industry and attempt to answer the question: which of these is primarily responsible for the success and spread of the chain model in this nonretail setting? The answer is of interest for several reasons. First, although cost-side advantages have been shown in large scale retail, it is unclear if the results from that literature hold outside that industry. Second, the implications for consumers may differ. Success due to lower costs are unambiguously positive for consumers, improving competition and lowering prices. Where the chain model is successful mainly due to revenue effects, however, the effects are more ambiguous. Consumers may find themselves paying higher prices for the same quality good, but they may also find more high quality goods available.

Third, the answer has implications for competition in the future. Over the past decade, online review and rating websites have allowed consumers to document and share their experiences with almost every conceivable type of firm.⁵ Many industries are therefore transitioning from a very low-to a relatively high-information environment, and the path this transition will take depends on what has driven the chain model’s success to this point.

³ See, for instance, Holmes (2011) and Jia (2008), who show the importance of distribution costs to Wal-Mart’s expansion strategy.

⁴ Throughout, I use the term “reputation” to denote the signal a firm’s brand name provides to consumers. This signal should at minimum reduce variance in a consumer’s prior over firm quality, something risk-averse consumers value. It may also increase or decrease mean expected quality. In a seminal article, Kreps (1990) discussed the importance of firm reputation, Horner (2002) further examines the role of firm reputation in competition, and Cai and Obara (2009) explicitly model reputation building in low-information environments as an incentive for horizontal expansion.

⁵ According to the Local Consumer Review Survey 2012, 85% of consumers checked online reviews before making purchasing decisions in 2012.

This article uses a unique firm-level data set on the hotel industry to examine empirically the nature of chain affiliation and the relative contribution of each factor to firm profits. Two features of these data are particularly useful. I observe the full population of firms; this allows a direct comparison between chain-affiliated and independent firms rather than simply relying on observations of a single chain like McDonald's or Walmart. In addition, I observe firm revenue, making it possible to separate demand-side factors from cost-side factors. These data is supplemented with online reviews data and market-level demand shifters.

The lodging industry is an ideal setting to examine the questions presented above for several reasons. First, hotels compete in a large number of geographically distinct markets. Second, unlike retailers or firms in other service industries, they offer close to a single product, a night's stay in a room. This product is differentiated between firms almost entirely on universally agreed-on quality and not other factors. These result in a relatively straightforward problem with few confounding factors. Third, the trend toward chain firms in this industry closely matches the aggregate trend.⁶ Finally, franchising and the hotel industry are significant sectors of the economy in their own right, with hotels generating \$196 billion in sales in 2012 and employing 2 million individuals according to the 2012 Economic Census, and franchises were responsible for \$1.3 trillion sales (9.2% of Gross Domestic Product) and 7.9 million employed (Kosova and Lafontaine, 2012).

The empirical strategy combines a structural estimation of firm costs with a detailed, reduced-form examination of hotel revenues to quantify the value to firms of chain affiliation, and the share of this value that comes through lower costs versus higher revenues. To do so, I model the dynamic game of market entry, exit, and type choice, and use these decisions to identify sunk and fixed costs. In the first-stage estimation of revenue, I examine the impact on revenues of hotels that add or drop affiliation during the sample period. I find that conditional on firm and market characteristics, chain-affiliated properties earn 20% higher revenue per room than independent firms. This is a substantial advantage and it is robust to a variety of specification tests.

I then examine the conventional explanation that the chain-affiliated firms are more efficient than independent firms by examining their operating costs. These costs are unobserved; instead, I estimate a dynamic model in the style of Arcidiacono and Miller (2011) to recover the cost structure of the different firm types to examine what cost advantage is associated with chain affiliation. This is one of the first applications of the Arcidiacono-Miller estimator, which allows for flexible and persistent unobserved heterogeneity, and I show that allowing for this significantly changes the results from a more restrictive estimator.⁷ The dynamic model produces realistic estimates of firm operating costs. They suggest, however, that chain firms gain no cost advantage from their affiliation after controlling for quality and unobserved market-level heterogeneity. I solve and simulate the model using the estimated parameters and find that it fits well.

I next examine patterns in revenue to consider how organizing as chains generates this revenue advantage. Potential sources of the increase in revenue include reputation effects, loyalty effects, and centralized bookings. Tests of revenues are consistent with a variety of predictions of a model where chains signal quality to consumers with low information. The chain premium declines over the past decade as online reputation mechanisms become more widely used. The chain premium appears immediately when a firm joins a chain, as opposed to slowly phasing in, and it is positively correlated with chain size. Finally, online reviews data show strong correlations between customer information and independent firm success, and among firms with large numbers of online reviews, the chain premium disappears completely. Although other explanations such as loyalty programs or centralized bookings likely play a role as well, these patterns suggest at minimum that a large share of the revenue premium is due to reputation effects.

A growing literature addresses the spread of chains. In part this reflects the success of large retailers such as Walmart (Basker, 2005; Foster, Haltiwanger, and Krizan, 2006; Holmes, 2011; Jia, 2008; Ellickson, Houghton, and Timmins, 2013; Foster et al., 2015; Zheng, 2016).

⁶ This can be seen by comparing Figures 1 and 4.

⁷ Scott (2013) also employs this estimator in a study of agricultural land use.

This literature also takes advantage of newly developed methods to estimate structural models of competition from entry and location decisions.⁸ These articles consider entry and location decisions, sometimes combined with structural demand models, and study their implications on firms' underlying cost structures and the nature of competition. Stahl (2016) uses a similar empirical strategy as this article, employing a structural framework to deconstruct costs and revenues in the telecoms industry in order to study the effects of deregulation and consolidation. Suzuki (2013) uses the same data source as this article and similar methods in a study of land-use regulations.⁹

A great deal of work has also been done in the vertical integration and franchising literature on why and when chain firms own and manage their own outlets and when they franchise them out. Kosova and Lafontaine (2012) and Lafontaine and Shaw (2005) summarize the state of this literature. This vertical decision is interesting, but is in many respects secondary to the initial horizontal spread of the chain structure and the underlying motivation thereof, which this literature approaches only indirectly. Similarly, the franchise literature often implicitly recognizes the importance of quality reputation effects across chains; see, for instance, Blair and Lafontaine (2005) and Pashigian and Bowen (1994). Despite this, no previous work has fully explored the broader demand-side explanation for chain organization.

Mazzeo (2004) is one of the few studies that considers the same margin as this article, the firm's decision of whether or not to join a chain. That article also considers the hotel industry. Using cross-sectional data on hotel locations in rural markets, he finds that chain affiliation is positively correlated with a measure of economic uncertainty and is more common in markets off major highways. He interprets this as potentially resulting from different shares of repeat business customers, where fewer repeat business customers increases the importance of chain affiliation, although it is difficult to say in general how the customer composition varies over highway markets and nonhighway markets without data on the subject. One notable recent work that also studies the effect of online reviews on firm outcomes is Luca (2011), who studies the effects of Yelp.com reviews on Seattle restaurants. He finds ratings primarily affect independent restaurants, and that during the time period studied, chain restaurant market share is declining. Both results are consistent with this article's findings.

2. Model

□ **Overview.** The empirical strategy presented here begins with a model of the industry as a dynamic discrete game being played out in a large number of local markets. Firms decide whether to remain active in the market or exit, and potential entrants decide whether to enter or stay out, as well as what type of firm to open. By combining observed entry and exit behavior with observed revenues, it is possible to identify operating costs as the set of costs that best rationalize these two pieces of data.

These cost estimates will be used to quantify potential economies of scale in costs in this industry. Because estimates of firm revenues are the key input to the estimation of this model and are of interest in their own right, I describe them at length in the next section.

□ **Environment.** I model the hotel industry as a dynamic discrete game, where in each period firms compete against one another in local markets. The game is modelled following Ericson and Pakes (1998). Each market contains a set of actors, differentiated along three dimensions: their

⁸ Recent examples include Thomadsen (2005), Toivanen and Waterson (2005), Aguirregabiria (2009), Igami and Yang (2016), and Blevins, Khwaja, and Yang (2015) on fast-food; Davis (2006) on movie theaters; Seim (2006) on video rental stores; Misra, Nair, and Ellickson (2012) and Villas-Boas (2007) on supermarkets; and Suzuki (2013) and Mazzeo (2002) on hotels.

⁹ Suzuki (2013) also estimates a dynamic model using Texas hotel data to test if land-use regulations raise firm costs. The key differences between Suzuki (2013) and this article are that his focus is on total entry costs, which I do not estimate, and he restricts attention to the firms in the six largest chains.

quality type (one, two, or three stars), whether or not they are affiliated with a chain, and whether they are an incumbent or a potential entrant.¹⁰ Each firm in each period chooses whether or not to be active in the market and determines its type.

Each market is described by a vector of state variables that determine payoffs. Denote the common state vector x_{it} . This includes endogenous variables, namely, the number of each type of firm participating in the market. Denote the number of firms of each type as $\{n_q^c, n_q^u\}$, where $q \in \{1, 2, 3\}$. The endogenous component of x_{it} consists of a 6×1 vector containing n_q^c and n_q^u . The vector x_{it} also contains own type and market characteristics such as traffic levels.

In addition to this, firms observe private information that they use in making their decisions. Specifically, they observe a private signal about their profits in the coming year and use that signal in part when they decide whether or not to stay active. This signal can represent demand conditions, cost conditions, or both. This is modelled as a vector of one period, i.i.d. shock to profits, defined as $\epsilon_{it}(a_{it})$, where firm actions are represented by $a_{it} \in \{0, c, u\}$. Here, 0 denotes exit, c denotes operating as a chain, and u denotes operating as an unaffiliated or independent firm.

Time is discrete over an infinite horizon, and the timing is as follows. At the beginning of each period, all active firms draw their private payoff values ϵ_{it} and decide whether or not to remain active in the market or exit irreversibly. Continuing firms decide if they wish to switch types between chain and independent. At the same time, a set of potential entrants observe the state of the market x_{it} and a draw of private values ϵ_{it} , and decide whether to enter the market or not. If they do not, they are replaced by a new set of potential entrants in the following period.

Once these decisions have been made, firms compete and earn revenues $R(x_{it}, a_{it}, a_{-i,t}; \theta_R)$ and incur operating costs $C(x_{it}, a_{it}; \theta_C)$.¹¹ The vectors (θ_C^c, θ_R) parameterize the cost and revenue functions. Note that although revenues depend on the actions of ones' rivals, costs do not.¹² The private information component of payoffs is additively separable. Per-period payoffs can thus be written:

$$\pi(x_{it}, a_{it}, a_{-i,t}; \theta) = R(x_{it}, a_{it}, a_{-i,t}; \theta_R) - C(x_{it}, a_{it}; \theta_C) + \epsilon(a_{it}). \quad (1)$$

At the time of entry, potential entrants jointly decide whether or not to affiliate or remain independent, and which quality level to operate at. When entering, the firm pays an entry cost, $EC(c_i, q_i)$, that is a function of quality and affiliation. For potential entrants, private information shocks help determine not just whether the firm will be active but also what type and quality level they will choose. Note that entrants choose both their quality level and chain affiliation type. Quality level is then fixed, but incumbents can switch from chain to independent, or vice versa.

Throughout this article, all decisions are assumed to be made by local business owners. Despite operating under the brand of a national chain, that chain's headquarters makes neither the exit nor entry decision. This is how the market operates in reality, with few exceptions. Hotels belonging to a chain must uphold certain quality standards and pay a share of revenues or flat franchise fee, but are otherwise independent.¹³ This fact also results in a much simpler game than if a central body was making entry and exit decisions across a large number of markets, in that it removes the need to estimate a model of network formation and location choice by each chain. This feature is distinct from the literature on large-scale retail chains that have focused much attention on network formation. See, for instance, Holmes (2011).

In reality, local entrepreneurs decide to open a hotel and then choose which chain to affiliate with, if any. Although exclusive territory agreements given to existing franchisees are an important

¹⁰ Because there is such a small number of four-star firms, and none enter or exit during the sample period, I exclude them from this analysis.

¹¹ I choose not to model the stage game that generates revenue, instead, taking revenue as a flexible function of firm characteristics, market characteristics, and the number and type of competitors faced by each firm.

¹² I model all costs as fixed costs. Industry analyses suggest that greater than 90% of costs are considered fixed. This is particularly true when aggregating to the period of a year, as I do.

¹³ In some cases, hotel chains do employ strategies with respect to entry, but the focus of these is on "showcase" hotels in large markets. In the types of markets I focus on here, the process is initiated and controlled by local entrepreneurs.

restriction on an entrant’s choice set, this should not be a factor once I aggregate up to the level of chain versus independent. Whereas an entrant may be foreclosed from opening a Holiday Inn, for instance, the number of other potential chain partners is so high that this should not affect a new firm’s overall type choice.

□ **Equilibrium.** Following Ericson and Pakes (1998), firms’ strategies are restricted to be anonymous, symmetric, and Markovian. Firms thus consider the current state vector of payoff-relevant variables when making their decisions, and all firms facing the same state behave the same way. Denote their strategies $\sigma_i : (x_i, \epsilon_i) \rightarrow a_i$. Given these strategies, the incumbent firm’s problem can be summarized:

$$\begin{aligned}
 V^I(x_i, \epsilon_i, \sigma_i; \theta) = & \max_{a_i} \{ \epsilon_i(a_i = 0), \epsilon_i(a_i = u) + \mathbb{E}_{a_{-i}} [\pi(x_i, a_i = u, a_{-i}) \\
 & + \beta \mathbb{E}_{x'_i, \epsilon'_i} V^I(x'_i, \epsilon'_i, \sigma_i; \theta)] \epsilon_i(a_i = c) + \mathbb{E}_{a_{-i}} [\pi(x_i, a_i = c, a_{-i}) \\
 & + \beta \mathbb{E}_{x'_i, \epsilon'_i} V^I(x'_i, \epsilon'_i, \sigma_i; \theta)] \}. \tag{2}
 \end{aligned}$$

The value of being an incumbent in state (x_i, ϵ_i) is either the value of exiting or continuing in the market, earning expected profits plus a continuation value, whichever is higher.

The choice of firm type is made upon entry. Thus, entrants face a somewhat more complex problem. First, firms decide whether or not to enter, and then whether to operate as a chain or independently, and at what quality level. Let $q_i \in \{1, 2, 3\}$ denote quality level. The entrant’s second-stage decision can be summarized as:

$$\begin{aligned}
 V_2^E(x_i, \sigma_i, \epsilon_i^2; \theta) = & \max_{[c,u], q_i} \{ -\theta_c^{EC}(q_i) + \epsilon_i^2(a_i, c, q_i) + \beta \mathbb{E}_{x'_i} V_c^I(x'_i, \sigma_i; \theta^c), \\
 & - \theta_u^{EC}(q_i) + \epsilon_i^2(a_i, u, q_i) + \beta \mathbb{E}_{x'_i} V_u^I(x'_i, \sigma_i; \theta^u) \}, \tag{3}
 \end{aligned}$$

and the entrant’s first-stage decision is:

$$V_1^E(x_i, \epsilon_i^1, \sigma_i; \theta) = \max_{a_i} \left\{ \epsilon_i^1(a_i = 0), \epsilon_i^1(a_i = 1) + \mathbb{E}_{\epsilon_i^2} V_2^E(x_i, \epsilon_i^2, \sigma_i; \theta) \right\}. \tag{4}$$

The value of being a potential entrant in state (x_i, ϵ_i) is the higher of the values of staying out or entering, where the firm chooses to be a chain or unaffiliated firm of any quality level optimally post-entry. Entering entails paying a one-time cost that depends on type choice and then becoming an incumbent firm in the next period.

These firm value functions are indexed by the strategy functions $\sigma(x)$, which firms use to forecast their rivals’ behavior and their own future behavior. The strategies $\sigma(x)$ form a Markov-Perfect Nash Equilibrium (MPNE) if for all $V(\cdot)$ above and all possible alternatives $\tilde{\sigma}(x)$:

$$V(x_i, \sigma(x), \epsilon_i) \geq V(x_i, \tilde{\sigma}(x), \epsilon_i). \tag{5}$$

The presence of private information guarantees the existence of at least one pure strategy-MPNE, as shown by Doraszelski and Satterthwaite (2010). There is no way to guarantee uniqueness, however.

3. Revenue analysis

■ In this section, I will document a revenue premium associated with chain affiliation. Ideally, we would observe prices and quantities directly, but only data on overall revenue are available, and thus demand parameters cannot be directly estimated. In any event, there is generally no single price for a hotel room. Prices vary over the time of year, day of week, and even method of

purchase.¹⁴ I instead focus on a firm revenue per room to measure performance. This is the most widely used measure of demand in the hotel industry. This section focuses on establishing and measuring the revenue effects of chain organization. In Section 6, I examine revenue in greater depth to consider the sources of this advantage, but the initial revenue estimates are themselves of interest for two reasons. First, they establish the fact that organizing as a chain directly produces a revenue advantage, which is an important result for explaining the dramatic success of chains in recent decades. Second, the revenue estimates are the primary input into the structural model used to estimate costs in Section 4.

□ **Data.** The state of Texas collects a hotel occupancy tax. Consequently, the full quarterly revenues of all Texas lodging establishments is available from the Texas Comptroller of Public Accounts at the property level. Tax revenue data is particularly trustworthy, because incorrectly reporting it is considered unlawful tax evasion.¹⁵ The revenue data is quarterly from 2000–2012, although data for 2000 are incomplete, and therefore dropped from most of what follows. For each hotel, I also observe location, capacity, and a measure of age. This information, along with chain affiliation, was cross-checked with a number of sources, including the American Automobile Association's *AAA TourBook*, published each year, and various hotels booking websites. The *AAA TourBook* also provides a standardized measure of quality, giving a rating of one through four stars for each hotel listed. Of the hotels in our sample, 57% of affiliated firms have been rated.¹⁶ My analysis focuses on rural markets, in which the bulk of hotels are one or two stars. The full distribution of star ratings breaks down as follows: 52.9% one star or unrated, 28.9% two stars, 17.9% three stars, and 0.2% four stars.

I also collect data from TripAdvisor.com, the world's largest travel review website. Users rate firms on a five-star scale and leave detailed reviews. I use a December 2012 cross-section containing average user rating, the number and distribution of reviews, each firm's ranking within their market, and the number of reviews by reviewer type (business, family, etc.). I also calculate the standard deviation of user ratings from the ratings distribution.

The analysis here is largely restricted to rural markets, where a market is defined as rural if there is no other market within 20 miles of it.^{17,18} This is for three reasons. First, large cities contain a very large number of hotels. These hotels are more horizontally differentiated in a number of unobservable ways, some cater to business travelers, others to recreational, and within a large, sprawling city such as Houston, location is a key attribute. It is not necessarily clear, therefore, which firms are competing with whom, and thus how to define a market. More importantly, there is probably a large degree of unobserved heterogeneity. Second, hotel chains often own and operate a small number of their properties themselves. These are concentrated in large markets, whereas in rural markets, nearly 100% of chain hotels are franchised out. This matters because chain affiliation will be assumed to be endogenous at the level of the hotel for some of what

¹⁴ In this industry and others with franchising arrangements, prices are set by franchisees and not franchisors. Chains still maintain some influence over prices by using advertising. For more on chain pricing, see Ater and Oren (2015) and Kalnins (2016), and the references therein.

¹⁵ The unit of analysis throughout this article is at the level of the individual property. Similarly, I define "firm" throughout at the level of the property because hotels are almost universally owned and managed by independent franchisees in these data.

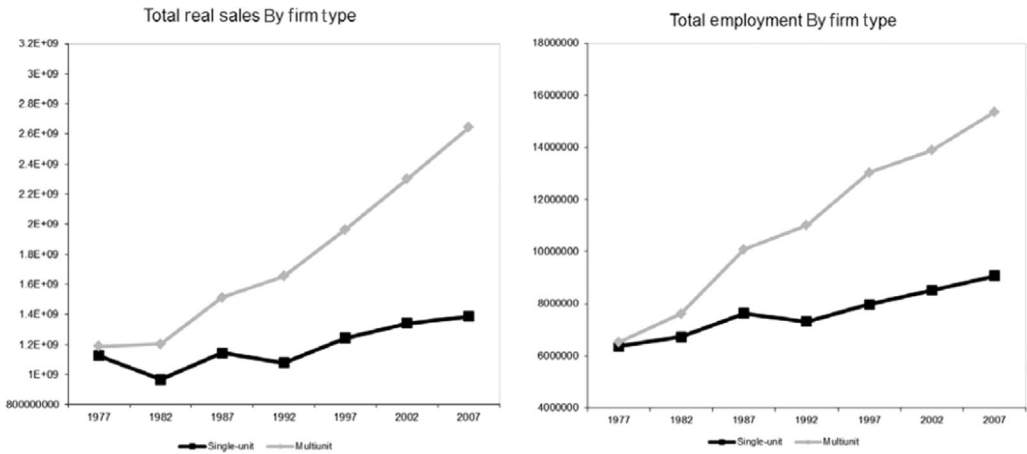
¹⁶ As AAA does not rate firms below a minimum quality standard, unrated firms are assigned a score of one star. For unrated chain-affiliated firms, I assign the modal star rating of its chain partners. There is very little within-chain variation in star ratings. For almost all the analysis that follows, I have examined the results when using only the set of AAA-rated firms to test the importance of this assumption, and in no case are the results significantly different.

¹⁷ For all results that follow, I define "market" as the nearest city, but also include firms in the same county among potential competitors. In addition, because hotel customers are frequently highway travelers, there is still potential substitution across markets. As a result, for markets on major highways, I test inclusion of the firms in adjacent counties as potential competitors. I find that including them has no significant effect on results.

¹⁸ Excluded markets are highlighted in Figures 2 and 3.

FIGURE 1

THE SPREAD OF CHAIN FIRMS



Note: This figure is taken from Figure 2 of Foster, Haltiwanger, Klimek, Krizan, and Ohlmacher (2015). Data comes from the Longitudinal Business Database. Here, SU stands for single-unit and MU stands for multiunit (chain).

TABLE 1 Hotel Summary Statistics

	Chain	Independent
N	676	627
Rooms		
Total	42,306	27,246
Mean	62.4	43.3
Minimum	20	20
Maximum	492	445
Mean RevPar		
Total	32.07	19.07
1 Star	20.36	17.44
2 Stars	28.82	24.04
3 Stars	42.21	34.08
4 Stars		44.00

Note: This table presents summary statistics on hotel size and revenue per available room per day (RevPar) for small market hotels. Star ratings are based on the *AAA TourBook*.

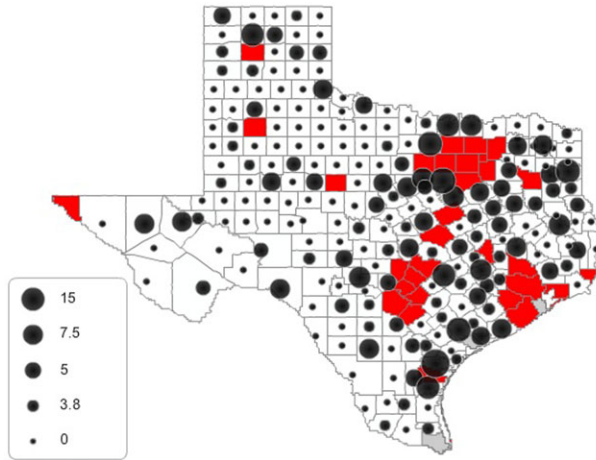
follows.¹⁹ Fortunately, Texas contains a great many rural and isolated markets. After restricting attention to these markets, our sample contains 353 markets with 1465 hotels. The mean market had two chain and 2.77 independent hotels active in 2012. As seen in Figure 4, the Texas rural hotel industry displays the same dynamics seen nationally.

The measure of performance I will focus on is daily revenue per available room, or “RevPar.” This is computed simply as revenue divided by capacity and the number of days. For much of what follows, I use annual means, aggregating up from quarterly because most market data are annual. For the full sample, mean chain RevPar is \$32.07 and mean independent RevPar is \$19.07. Summary statistics can be seen in Table 1. The physical distribution of chain and independent firms is mapped in Figures 2 and 3, with excluded counties highlighted.

¹⁹ The organizational structure of chain-affiliated hotels can vary. In some cases, the chain both owns and operates the property, in other cases, a franchisee owns the property but the chain has a contract to manage it. In nearly all firms in

FIGURE 2

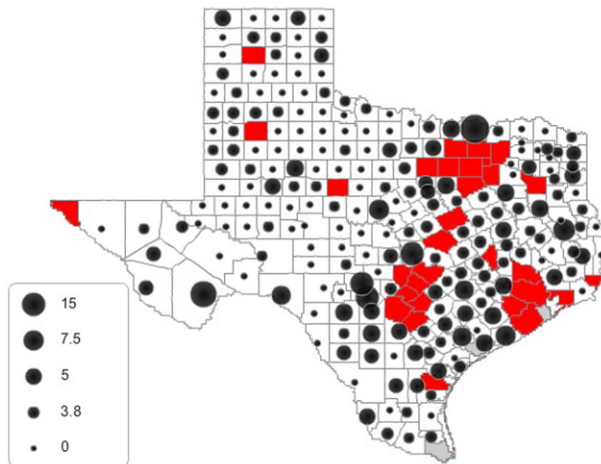
GEOGRAPHIC DISTRIBUTION OF CHAIN FIRMS [Color figure can be viewed at wileyonlinelibrary.com]



Note: This figure shows a map with the number of chain hotels represented at the county level. Shaded counties are major metropolitan areas that are excluded from the analysis.

FIGURE 3

GEOGRAPHIC DISTRIBUTION OF INDEPENDENT FIRMS [Color figure can be viewed at wileyonlinelibrary.com]



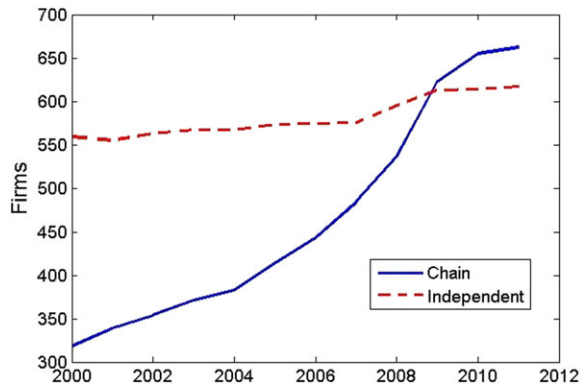
Note: This figure shows a map with the number of independent hotels represented at the county level. Shaded counties are major metropolitan areas that are excluded from the analysis.

Data show that along with being more likely to have a high quality rating, chain-affiliated hotels are more likely to be active in larger and more attractive markets. To account for demand-side factors that influence firm revenue and market structure, I collect data on each market. From the Census Bureau, I collect data on county unemployment rate and population, and total county retail sales as measures of market size or business activity. From the Texas Railroad Commission,

this sample, the chain neither owns nor operates the property. Instead, these roles are taken by local franchisees and the chain simply licenses its branding and provides the franchisee with an operating manual.

FIGURE 4

GROWTH OF CHAINS IN RURAL TEXAS HOTELS [Color figure can be viewed at wileyonlinelibrary.com]



Note: This figure shows the total number of chain and independent hotels over time for small markets.

I add data on the number of currently producing wells for both oil and natural gas in each county. I also gather Texas Department of Transportation data on average daily traffic passing through each market. This measure is a key determinant of demand in the rural roadside hotel industry. Summary statistics on these data can be seen in Table 2. Together, variation across time and markets in these factors should help capture exogenous shifts in demand. In particular, the growth of the natural gas industry in Texas over the past decade has had a significant impact on hotel demand and is exogenous to hotel performance. Examples of this impact on revenue and entry patterns will be described below.

□ **Revenue estimates.** In this section, I show that chain firms earn a substantial revenue premium over otherwise identical independent firms. Table 1 shows that, on average, chain firms earn higher revenues, but this could reflect a number of factors. To test if chain-affiliated firms earn a revenue premium after controlling for firm and market characteristics, I regress RevPar on these market-specific factors as well as the number of chain and independent competitors in the market. Specifically, I consider the model

$$RevPar_{imt} = x_{imt}\beta_1 + firm_{it}\beta_2 + c_{it}\delta^c + market_m + time_t + \epsilon_{imt}, \quad (6)$$

where x_{imt} are data on market characteristics such as population, as well as the number and type of competitors in each market, $firm_{it}$ are other firm characteristics such as AAA rating and TripAdvisor.com rating, c_{it} indicates whether a firm is a member of a chain in period t , and $market$ and $time$ are year and market dummies. The ultimate object of interest is δ^c , which is the remaining effect on revenue of chain affiliation after controlling for firm and market characteristics.

Although market structure variables are endogenous with respect to the same demand conditions that partially determine revenue, I am not concerned with this causing bias in estimation. Opening or closing a firm is a long-term decision, with new firms having a time to build of over a year. As a result, short-term fluctuations in demand conditions should have little effect on market structure. I also include market fixed effects, so highly persistent unobserved demand conditions are accounted for. To control for medium-term fluctuations, I include the aforementioned six demand shifters as well as two variables measuring their rates of change.

With year dummies omitted for space, results are shown in Table 4. Because the dependent variable here is RevPar, coefficients can be interpreted as the effect on dollars per room per day.

TABLE 2 Regressor Summary Statistics

	Mean	Standard Deviation	Minimum	Maximum	Frequency	Unit of Observation	Source
County firms							
Chains	4.24	3.76	0	19	Monthly	County	Texas Comptroller
Independents	4.15	3.41	0	17	Monthly	County	Texas Comptroller
Demand shifters							
Daily traffic	14,289	13,756	0	100,000	Annual	Highway	TxDOT
Population	20,130	279,734	880	423,970	Annual	County	Census
Total sales (\$ billion)	3.41	20.6	0.03	438	Annual	County	Census
Gas wells	326.6	737.3	0	6155	Annual	County	Texas Railroad Commission
Oil wells	461.9	843.5	0	8261	Annual	County	Texas Railroad Commission
Unemployment	5.72	2.27	1.9	17.8	Annual	County	Census
Firm variables							
AAA rating	1.68	.89	1	4	Annual	Firm	AAA
Chain affiliation	.41	.49	0	1	Monthly	Firm	Combination of sources
TripAdvisor rating	2.03	1.87	0	5	Cross-section	Firm	TripAdvisor
Number of reviews	11.76	24.83	0	483	Cross-section	Firm	TripAdvisor
RevPar	23.89	21.86	0	550.6	Monthly	Firm	Texas Comptroller

Note: This table presents summary statistics on market and firm characteristics, where market refers to county and firm refers to a property operated continually at a specific address.

After controlling for quality and other factors, we see that affiliating with a national chain is associated with an average premium of \$5.78 of daily RevPar. This represents a 27.7% premium, or roughly \$120,000 per year for a firm with 60 rooms.

□ **Switchers.** There is an important potential source of bias in the above analysis. Chain affiliation may be correlated with unobserved factors that increase revenue. This would be true, for instance, if chain hotels were more likely to be built on the best locations. This would cause upward bias in the estimate of the chain premium.

Ideally, we could measure the counterfactual revenues of the same hotels with and without a chain affiliation. Although this is impossible, 104 hotels do add or drop chain affiliation in the sample period.²⁰ Here, I measure the effect of this change on revenue in a firm fixed effects context, using only within-firm variation in revenue to identify the chain premium. The econometric model for RevPar that I consider is:

$$RevPar_{imt} = x_{imt}\beta_1 + firm_{it}\beta_2 + c_{it}\delta^c + market_m + time_t + firm_i + \epsilon_{imt}, \tag{7}$$

where $firm_i$ is a property-level, unobserved, time-invariant determinant of revenue. Because we observe switchers, I can estimate a demeaned version of equation (7) to eliminate $firm_i$ and estimate δ^c off the subpopulation of switchers. No firms that add or drop chain affiliation

²⁰ Roughly twice as many hotels add a chain affiliation as drop it.

TABLE 3 Chain Summary Statistics

Chain	Number of Outlets	Mean Rooms	Mean RevPar	AAA Rating
Best Western	275	66.2	37.7	2
Holiday Inn Express	224	80.5	54.0	3
La Quinta	213	99.8	41.8	3
Comfort Inn	204	68.3	42.1	3
Super 8	159	57.6	29.9	1
Hampton Inn	157	87.7	58.9	2
Days Inn	142	66.5	26.7	2
Motel 6	136	94.1	24.3	1
America's Best Value Inn	88	62.8	20.3	2
Courtyard by Marriott	74	130.9	64.1	3
Quality Inn	72	90.5	31.1	2
Hilton Garden Inn	68	230.0	77.1	2
Holiday Inn	67	181.2	50.9	3
Ramada Inn	60	111.5	21.8	2
Fairfield Inn	60	94.9	53.1	3
Regency Inn	55	50.5	17.0	2
Candlewood Suites	53	92.7	43.2	3
Residence Inn	53	112.7	77.9	3
Econo Lodge	47	51.9	22.4	1
Extended Stay	46	99.9	33.7	2
Hawthorn Suites	45	81.5	38.6	2
Marriott Hotel	44	385.8	86.9	3
Homewood Suites	43	103.9	73.0	3
Studio 6	41	105.2	27.6	2
Sleep Inn	37	66.8	37.1	2

Note: This table presents summary statistics on hotel size and revenue RevPar for the 25 most common chains in all Texas markets. Star ratings is the mode for each chain.

change their star rating, indicating that switches occur within fairly well-defined quality tiers, rather than accompanying a significant change in underlying firm quality. Along with superficial changes in branding, joining a chain requires following a set of standardized operating procedures.²¹

Results from this estimation are in column 3 of Table 4. Time-constant explanatory variables are eliminated, and most estimates are similar to the previous results. The fixed effect regression provides an estimate of the chain premium of \$4.35, lower than the estimate without firm fixed effects, but still a substantial advantage. This is equivalent to a 21.1% premium, or roughly \$95,000 per year for a 60-room hotel.

By adding dummies for the number of years before or after the switch occurred, we can trace out the timing of the revenue boost switchers receive and test whether the chain premium results from selection on a trend in unobservables. This would be the case, for instance, if chains were dropping underperforming firms or adding independent firms that had recently improved their quality. If this were the case, a one-year lead of the switch should pick up this reverse effect and

²¹ Each chain maintains an operating manual that is loaned to the franchisee upon opening or converting a new hotel. This document is confidential. Franchise agreements are typically not confidential and lay out a set of obligations for the franchisee to follow. Based on those that I have reviewed and the anecdotal responses of hoteliers I have spoken to, the franchisee must agree to the possibility of a preopening inspection, and the franchisor maintains the right to conduct periodic inspections through the length of the agreement. These inspections might be random or might be prompted by guest satisfaction surveys, depending on the chain. If a property is found to be deficient in some way, they may be charged a fee by the franchisor or may be required to correct the deficiency. Some switches are accompanied by changes in room number or ownership. The estimates from switches that do and do not accompany these changes show no significant differences, suggesting the result is not due to significant renovations or management improvements.

TABLE 4 Revenue Estimates

	Pooled OLS	Market FE	Firm FE
Chain status	6.450*** (.526)	5.973*** (1.009)	4.347*** (1.015)
Number of competitors	-.052 (.042)	-.945*** (.254)	-.964*** (.261)
Mkt. chain share	-6.626*** (.709)	-9.189*** (2.178)	-6.639*** (2.002)
Log county revenue	.291 (.197)	3.718*** (.857)	3.888*** (.851)
Log capacity	-.043 (.371)	-2.008* (1.081)	-18.519*** (2.46)
Log traffic	-1.129*** (.227)	.594 (1.348)	2.632 (2.551)
Log gas wells	.288*** (.066)	-.125 (.421)	-.211 (.612)
Δ Gas wells	3.356*** (1.219)	2.302*** (.796)	2.120*** (.578)
Log oil wells	-.030 (.065)	.009 (.343)	-.027 (.618)
Δ Oil wells	-1.924 (1.334)	-1.219 (1.076)	-.731 (.705)
Log population	.059 (.243)	-3.574*** (1.233)	-9.725* (5.511)
Unemployment	-.484*** (.099)	-1.768*** (.479)	-1.582*** (.418)
2 Stars	2.489*** (.684)	3.145* (1.659)	8.998 (17.642)
3 Stars	18.398*** (.959)	20.169*** (3.346)	7.548 (18.613)
4 Stars	41.569*** (2.834)	33.691*** (11.474)	
1 Star*user rating	2.574*** (.169)	2.645*** (.494)	
2 Stars*user rating	3.309*** (.187)	3.393*** (.478)	
3 Stars*user rating	1.572*** (.235)	1.697** (.768)	
σ (Ratings)	-7.459*** (.790)	-6.669*** (1.962)	
Year dummies	Yes	Yes	Yes
R ²	.32	.39	.76
N	12,807	12,807	12,807

Notes: The dependent variable is RevPar. Standard errors are in parentheses, they are robust and clustered at the market level for all specifications. All data is annualized and Δ denotes the one-year change.

eliminate the estimated chain premium. I perform this test using regression equation:

$$\begin{aligned}
 RevPar_{imt} = & x_{imt}\beta_1 + firm_{it}\beta_2 + \chi\{pbs_{it} > 0\}pbs_{it}\delta_{\tau}^c + \chi\{pas_{it} > 0\}pas_{it}\delta_{\tau}^c + time_t \\
 & + firm_i + \epsilon_{imt},
 \end{aligned}
 \tag{8}$$

where the variables pbs_{it} and pas_{it} are the number of periods before switching and after switching, respectively, and $\chi\{\cdot\}$ are indicator functions. The coefficient on these variables is the effect on revenue today of switching chain affiliation status some number of periods in the past or future. I perform tests of this sort in Table 5. In column 4, we see the coefficient on a one-year lead of joining a chain is positive but not statistically significant and when it is included, the chain premium is \$4.07 and highly significant.

TABLE 5 Revenue Effects of Adding or Dropping Affiliation

Specification	All Switchers	C → U	U → C	All Switchers	All Switchers
Chain status	4.39*** (.73)	.74 (1.58)	6.24*** (1.08)	4.07*** (1.28)	
Lead of chain status				1.82 (1.27)	
Number of competitors	-.99*** (.11)	.43 (.60)	.08 (.42)	-.93*** (.11)	-.87*** (.11)
Market chain share	-6.82*** (1.09)	-2.07 (5.75)	-9.78*** (2.70)	-6.75*** (1.07)	-4.51*** (1.02)
Log county revenue	3.62*** (.32)	.63 (1.95)	7.98*** (1.20)	2.30 (.32)	
Log capacity	-18.84*** (1.06)	-17.45*** (3.85)	-16.25*** (2.51)		-18.19*** (1.04)
Unemployment	-1.69*** (.15)	-2.64** (1.11)	-.32 (.66)	-1.04*** (.15)	-1.81*** (.07)
Log traffic	2.62*** (.97)	-3.38	8.64** (3.79)	2.52*** (.92)	.36 (.93)
Log gas wells	-.15 (.22)	1.41 (2.22)	-4.77*** (1.24)	.18 (.22)	.08 (.22)
Log oil wells	-.18 (.51)	1.02 (.40)	-6.74** (2.66)	.32 (.33)	.45 (.34)
Log population	-12.28*** (2.45)	4.79*** (1.12)	-3.27 (7.87)	-2.81*** (.91)	-1.18 (.79)
2 Stars	1.67 (3.22)	5.77 (13.30)	-2.19 (4.13)	-4.10 (3.73)	2.71 (3.12)
3 Stars	9.48*** (3.55)			2.02 (5.87)	9.22*** (3.49)
Years before switch					
11					2.61 (3.59)
10					3.83 (3.10)
9					3.95 (2.94)
8					3.86 (2.36)
7					4.42** (2.11)
6					3.60* (2.00)
5					1.42 (1.88)
4					1.05 (1.78)
3					.007 (1.72)
2					1.24 (1.68)
1					.40 (1.65)
Years after switch					
1					4.06** (1.82)
2					2.75 (1.69)
3					2.97* (1.70)

(Continued)

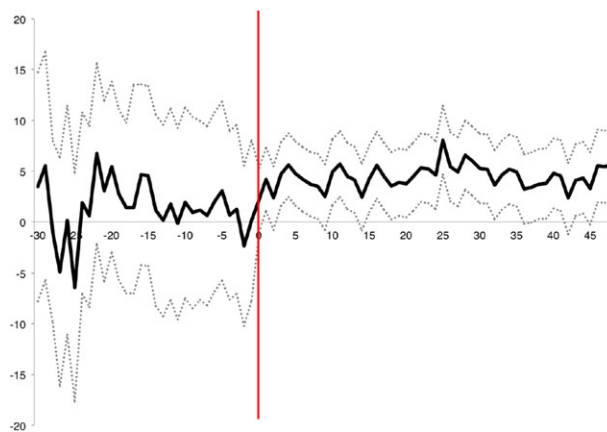
TABLE 5 Continued

Specification	All Switchers	C → U	U → C	All Switchers	All Switchers
4					3.77** (1.73)
5					4.62*** (1.80)
6					6.63*** (1.94)
7					3.50* (2.01)
8					-.20 (2.22)
9					-1.21 (2.57)
10					.59 (2.88)
Year dummies	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes
R ²	.37	.36	.35	.14	.14
N	1124	385	739	1124	1124

Notes: This table analyzes the effects of adding or dropping chain affiliation with firm fixed effects. The dependent variable is RevPar. Specifications labelled “C → U” use firms that switch from chain to unaffiliated, and vice versa.

FIGURE 5

CHAIN PREMIUM BEFORE AND AFTER SWITCHING [Color figure can be viewed at wileyonlinelibrary.com]



Note: This figure shows detailed revenue estimates for firms that add chain affiliation during the sample period. Dummies for the number of months before and after the switch occurred are included in revenue regressions, and the coefficients on those are plotted. Month 0 corresponds to the first full month after adding affiliation. Dashed lines represent a 95% confidence interval.

I show visually a full set of monthly leads and lags in Figure 5. Two things stand out: first, there is no increase in revenue associated as a firm approaches adding a chain affiliation, and thus the chain premium is not just a product of national chains selecting high quality firms to allow into their chain or vice versa. Second, the chain premium shows up at its mean level immediately, it does not need time to phase in. Table 6 presents results from a comparison of revenues for the quarter before a switch occurs and the quarter after it occurs for firms adding or dropping chain

TABLE 6 One Quarter Revenue Comparison for Switchers

	Firm FE
Chain status	3.51** (1.70)
Number of competitors	-3.14 (2.15)
Log county revenue	13.71* (8.21)
Log traffic	24.13 (19.12)
Log gas wells	-3.37 (7.60)
Log oil wells	6.03 (11.04)
Log population	113.32 (74.18)
Unemployment	-9.88** (3.92)
Year dummies	Yes
Quarter dummies	Yes
R^2	.85
N	267

Notes: The dependent variable is RevPar. The data are firms that add or drop chain affiliation and contain the quarter prior to the switch and the quarter after. The quarter during which the switch occurs is dropped. Standard errors are in parentheses, they are robust and clustered at the market level for all specifications.

affiliation.²² The estimated chain premium falls slightly, to \$3.51, but is still close to the mean value in the sample containing all periods. This is further evidence that the premium appears immediately and does not result from quality improvements that phase in over time.

When estimated separately, the chain premium measured off firms switching from independent to chain is substantially higher than that measured just off firms switching from chain to independent. We would expect to see this result if there is heterogeneity in the value of affiliation over properties or markets. Properties with a high value are more likely to operate as a chain or switch from independent to chain, and properties with a low value are likelier to operate independently or switch from chain to independent. This is partly why it is necessary to include property-level fixed effects and measure the chain premium off of switchers. We might not expect, then, to see a large fall in revenue for firms dropping chain affiliation if they are the firms with low chain values, that is, the benefits of branding, loyalty, etc., are lower for them and thus, they feel the benefits to them of chain affiliation do not justify paying the franchise fees. This is also why it is important to use the aggregate measure of the chain premium that includes both types of switchers, as I do throughout, to capture the average value.

This chain premium is the average over the 12 years in the sample. We can also look at how the revenue premium varies from year to year by interacting chain with year dummies. This results from the following regression:

$$\log(\text{RevPar}_{imt}) = x_{imt}\beta_1 + \text{firm}_{it}\beta_2 + c_{it}\delta_i^c + \text{time}_t + \text{firm}_i + \epsilon_{imt}. \quad (9)$$

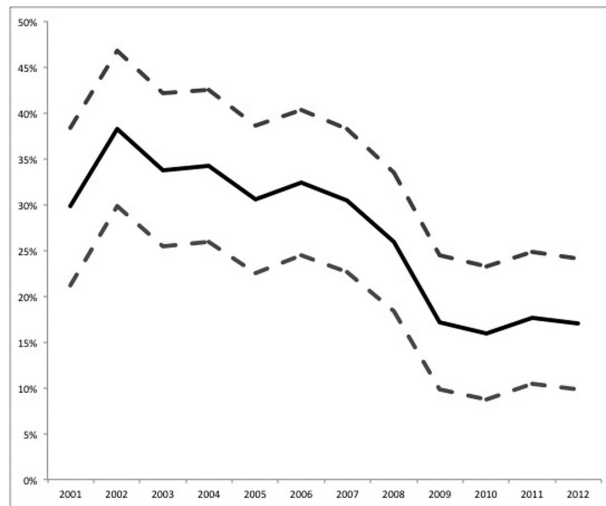
The results can be seen in Figure 6. The chain revenue premium, expressed as a percentage, steadily decreases over the decade, from around 30% in 2002 to about 15% in 2012.²³ This

²² This uses the same specification in equation (7), excluding data from all periods except the quarters before and after a chain affiliation switch.

²³ This result does not appear to be caused by increasing numbers of chain competitors, as it is robust to different specifications that include the number of chain firms in each market, the share of chain firms, and the full-type distribution in revenue estimates. In addition, although the United States as a whole experienced a large recession in 2008–2009, large

FIGURE 6

CHAIN PREMIUM (PERCENT) OVER TIME



Note: This figure presents the estimated chain premium as a percent of total revenue after controlling for firm and market characteristics. Dashed lines represent the 95% confidence interval.

decrease suggests the advantage to operating as a chain is declining. I discuss this result in greater detail in Section 6 below.

4. Recovering costs

■ In this section, I will describe the empirical strategy for recovering the cost structure of the industry. Why are we interested in these costs? I have already shown that organizing as a chain is associated with a significant revenue premium, but this does not give the full story. Whether these firms also derive an efficiency or cost advantage from their affiliation matters for how we should think about them. If they do, and it is substantial, the implications for social welfare are different than if the only advantage derives from consumer demand, and the implications for the future of the industry depends on whether chains will continue to thrive if the premium they are able to charge declines as consumers have access to greater information.

Unfortunately, these costs are not observable. Recovering them requires estimating a structural model. This methodology follows a recent tradition in empirical work of using two-stage methods to recover the structural parameters of settings of firm competition that can be characterized as dynamic games, beginning with Aguirregabiria and Mira (2007) and Bajari, Benkard, and Levin (2007), extended more recently by Arcidiacono and Miller (2011). These methods have recently been applied to a number of questions in industrial organization, including environmental regulation (Ryan, 2012), land use regulations (Suzuki, 2013), production spillovers (Gallant, Hong, and Khwaja, 2017), demand fluctuations (Collard-Wexler, 2013), repositioning costs (Misra, Nair, and Ellickson, 2012), dynamic effects of Medicare hospital regulations (Gowrisankaran et al., 2013), and telecoms mergers (Stahl, 2016). A good recent overview of these methods is by Bajari, Hong, and Nekipelov (2013).

regions in Texas were experiencing an economic boom at this time, due to new gas production. Comparing this figure for counties with and without falling employment show no significant difference, suggesting business cycle dynamics do not play a major role in this result.

By combining observed entry and exit behavior with observed revenues, it is possible to identify operating costs as the set of costs that best rationalize these two pieces of data.²⁴ These are the costs we are concerned with when considering the proposition that a chain-affiliated firm's success is due to economies of scale. This would result in reduced costs on bulk purchases of inputs and advertising, as well as lower costs of capital and insurance. In addition, it is well known from prior work and from the revenue estimation in Section 2 that unobserved heterogeneity plays a large role in driving both revenue and profits, and so it is crucial to account for this in estimation. In this section, I demonstrate one of the first applications of the method described in Arcidiacono and Miller (2011) for allowing flexible unobserved heterogeneity in both revenue and costs, and show that this makes an important difference in results.

□ **Empirical strategy.** In this section, I discuss the variables and assumptions I will use to take the above model to the data. Until recently, estimating the underlying parameters of dynamic discrete games has been considered too difficult to be practical. The reason is that solving for an equilibrium of the game is computationally demanding and must be done for every set of parameters considered in solving a maximum-likelihood problem. Beginning with Aguirregabiria and Mira (2007) and Bajari, Benkard, and Levin (2007), however, two-step methods have been developed to avoid fully solving for equilibrium at every parametric evaluation. Instead, reduced-form policy functions governing entry, exit, and type choice are estimated directly from the data, and are assumed to reflect equilibrium play. These are then used to estimate underlying structural parameters of the dynamic games.

The strategy followed here follows this tradition. Reduced-form policy functions governing entry, exit, and switching behavior are estimated, and then these are used to estimate choice-specific value functions directly. Resolving the firm's discrete choice problem using estimated revenues and future values allows us to recover per-period costs.

Revenue adjustments. Two adjustments must be made to observed revenues at this stage. First, I subtract franchise fees paid by chain-affiliated firms. In the previous results on revenue, I do not adjust for these fees. In that section, I am primarily concerned with identifying and explaining a revenue premium earned by chain-firms, and am not interested in how this premium is divided between the franchisee, and franchisor. In the current section, I model the entry and exit decision of the franchisee, and so it is necessary to remove these fees. Fortunately, hotel chains charge uniform fees across members, and these are collected and published by *Hotel Management*, a trade publication. The standard contract consists of a flat initial fee ranging from \$30,000 to \$100,000, followed by 7%–10% of revenue thereafter. In the first stage, I subtract out the royalty component from annual revenues. The fixed costs I estimate in this section are thus the net costs before fees, or the underlying economic costs of operation. The full costs of operation are these costs plus the franchise fees paid.

Second, I adjust revenue for selection on entry and exit. We observe the revenues of firms who do not exit, and model the entry/exit process as a function of revenue shocks. Revenue thus needs to be adjusted for selection on these shocks. This can be done using a control function approach, as described in Ellickson and Misra (2012). Specifically, assuming the private component of revenue is distributed according to a Type 1-Extreme Value (TIEV) distribution, the expected value of this for active firms is

$$\mathbb{E}(\epsilon_i | x_i, a_i = 1) = \gamma - \ln(1 - p(\widehat{a_i = 0} | x_i)), \quad (10)$$

where $a_i = 0$ indicates exit.

After making these adjustments, revenues are estimated mostly as described in previous sections with two differences. First, for the purposes of our second-stage estimation, revenues are estimated over the state variables of the dynamic discrete game, that is, no market dummies.

²⁴ During the sample period, 143 firms exit and 603 firms enter.

Second, to keep the specification flexible, I include quadratic terms on market characteristics and interactions between firm and market characteristics.

Policy functions. The first step is approximation of firm policy functions determining entry, exit, and type switching decisions. To do so, I estimate choice probabilities in reduced form. I represent possible firm actions as $a_i \in \{0, c, u\}$, where 0 indicates exit, c indicates continuing as a chain firm, and u indicates continuing as an unaffiliated or independent firm. For incumbent firms, I set choice probabilities to take the logit form where the probability of remaining active is:

$$p(a_i = 1|x_{it}) = \frac{\exp(x_{it}\beta)}{1 + \exp(x_{it}\beta)}. \tag{11}$$

The state x_{it} includes the full state of the market, that is, $\{n_q^c, n_q^u\} \forall q$, as well as own type and market characteristics, quadratic terms of these characteristics, and interactions between firm type and market characteristics. The goal is simply to flexibly capture firm choice probabilities. I test robustness on the logit assumption in the next section. Aguirregabiria and Mira (2007) show that under general conditions, this Conditional Choice Probability (CCP) is equivalent to the firm’s policy function.

Estimating total costs of entry requires making an assumption on the number of potential entrants in each market, each period. Because this is unknowable, overall entry costs are in general, not well identified. Because the focus of this article is on the difference between costs for chain and independent firms, this itself is not a problem, however. I estimate a two-stage process where a set of potential entrants decides to enter the market or not, using a simple logit. This produces an estimate of base-level entry cost, and is not identified without assuming a number of potential entrants. I then estimate the second-stage type choice as a multinomial logit across all types, chain and unaffiliated, and each quality level. This produces estimates of the different entry costs for each type, normalizing one-star independent firm costs to 0, which are identified off the type choices of the 603 entries that occur in the sample.

Value function inversion. Here, I describe how firm value functions are decomposed into revenue, cost, and continuation values, and how these continuation values are constructed using our estimates of policy functions. Because revenue is observable, once continuation values are estimated, it becomes straightforward to estimate the remaining piece, the firm’s cost function.

First, I give some notation. Denote the choice-specific value function:

$$v(x_i, a_i) = \pi(x_i, a_i) + \beta \mathbb{E}_{x'_i} \bar{V}(x'_i), \tag{12}$$

where $\bar{V}(x_i)$ is the value of being in state x_i before realization of ϵ_i draws, which I will refer to as the *ex ante* value function. I find it by integrating $V(x_i, \epsilon_i)$ over the distribution of ϵ_i ,

$$\bar{V} \equiv \int V(x_i, \epsilon_i) g(\epsilon_i) d\epsilon_i. \tag{13}$$

The model implies:

$$a_i^*(x_i, \epsilon_i) = \operatorname{argmax}_{a_i} \{v(x_i, a_i) + \epsilon_i(a_i)\}. \tag{14}$$

If I assume ϵ_i is distributed T1EV with dispersion parameter σ_ϵ , then the probability of remaining active is:

$$p(a_i|x_{it}) = \frac{\exp\left(\frac{1}{\sigma_\epsilon} v(x_i, a_i)\right)}{1 + \exp\left(\frac{1}{\sigma_\epsilon} v(x_i, a_i)\right)}, \tag{15}$$

where the value of exit $v(x_i, 0)$ has been normalized to 0. Hotz and Miller (1993) provided the insight that specifying the choice in this way can allows us to invert the above equation and write the choice-specific value function as a function of estimable choice probabilities:

$$v(x_i, a_i) = \ln(p(a_i|x_{it})) - \ln(p(a_i = 0|x_{it})). \tag{16}$$

Using the first-stage estimates of $\widehat{p(a_i|x_{it})}$ from equation (11), I can thus form estimates of choice-specific value functions directly off the data. I can also use the policy function estimates to calculate the conditional state transition function. Because private information shocks are assumed to be independent across firms, the probability distribution of a firm's rivals choosing a_{-it} is $P(a_{-i,t}|x_t) = \prod_{j \neq i} p(a_{j,t}|x_t)$. Denote the transition of the state vector $F(x'_i|x_i, a_i, a_{-i})$. The transition kernel a firm thus faces is:

$$f(x'_i|x_i, a_i) = \sum_{a_{-i}} P(a_{-i}|x)F(x'_i|x_i, a_i, a_{-i}), \tag{17}$$

which can be calculated using our first-stage policy function estimates and estimates of transition processes for exogenous state variables.

Arcidiacono and Miller (2011) show how, due to the T1EV assumption on ϵ_i , and due to the fact that exit is a terminal state, the *ex ante* value function can be expressed solely as a function of the CCP's. Because the probability of exit is a function of the relative values of exit and remaining active, all information about the value of being active is contained in the probability of exit at this state. Specifically,

$$\int \overline{V}(x'_i)f(x'_i|x_i, a_i)dx'_i = \sigma_\epsilon \left(\gamma - \int \ln(p(0|x'_i))f(x'_i|x_i, a_i)dx'_i \right), \tag{18}$$

where γ is Euler's Constant.²⁵ To simplify notation, I will refer to this item as:

$$\widehat{\mathcal{V}}(x_i, a_i) \equiv \sigma_\epsilon \left(\gamma - \int \ln(\widehat{p(0|x'_i)})f(x'_i|x_i, a_i)dx'_i \right), \tag{19}$$

which I can estimate directly from our CCP estimates by simulating over the distribution $f(x'_i|x_i, a_i)$ a large number of times and calculating $\ln(p(0|x'_i))$ at each draw.²⁶ The choice-specific value function in equation (12) can now be written as:

$$v(x_i, a_i; \theta) = R(\widehat{x_i, a_i; \theta_R}) - C(x_i, a_i; \theta_C) - SC(a_i, a'_i)\chi\{a'_i \neq a_i\} + \beta\widehat{\mathcal{V}}(x_i, a_i), \tag{20}$$

where $SC(a_i, a'_i)$ is a switching cost paid for changing affiliation types, and $\chi\{a'_i \neq a_i\}$ is an indicator for if a switch occurs. By incorporating estimated parameters into a nonlinear second stage, this estimation is an example of the "forbidden regression." With 12,348 observations and highly flexible first-stage estimation procedures, however, results should be sufficiently accurate as to minimize any asymptotic bias this might cause.

The structural form of the discrete choice problem in equation (11) can now be solved, offset with estimates of $\widehat{R(\cdot)}$, and used to find the parameters of the cost function.²⁷ The firm's discrete choice is solved with fitted revenues in the current state and the expected one period ahead value. This approach has the appeal of being computationally straightforward, and thus can accommodate a large number of state variables that would otherwise pose a significant computational burden.

□ **Unobserved heterogeneity.** The preceding assumes there is no persistent unobserved heterogeneity across firms or markets. The reduced-form estimates presented earlier, however, provide evidence of unobservables influencing key revenue parameters. It is important, therefore, to allow for this when estimating the structural model.

In general, it is difficult to account for persistent unobservable effects in dynamic game models due to the highly nonlinear nature of estimators. Aguirregabiria and Mira (2007) show

²⁵ For a full derivation of this, see Arcidiacono and Miller (2011).

²⁶ The dispersion parameter σ_ϵ is unknown and is treated as an extra parameter in final-stage estimation. Although parameter estimates in discrete choice models are typically scaled by this unknown dispersion parameter, because revenues are observed in dollars, costs parameters can also be expressed in dollars.

²⁷ Specifically, because a firm's choice probability is defined in equation (11), I resolve a logit model on firm type and continuation value, offsetting revenue estimates.

a method for accounting for permanent unobserved characteristics that influence payoffs in stationary dynamic games. Arcidiacono and Miller (2011) propose a method for estimating models with potentially time-varying unobservable factors that affect both payoffs and state transitions. This is the approach I follow here.

I assume markets are in some unobserved state s , drawn from a discrete, finite support S . The unobserved state can affect firm profits, firm choice probabilities, and other state transition probabilities included in x_{it} , such as traffic or population.

For a given market state s_m , the terms in equation (20) become:

$$v(x_i, a_i, s_m; \theta) = R(x_i, \widehat{a_i, s_m}; \theta_R) - C(x_i, a_i, s_m; \theta_C) - SC(a_i, a'_i) \chi \{a'_i \neq a_i\} + \beta \mathcal{V}(\widehat{x_i, s_m, a_i}), \tag{21}$$

where

$$R(x_i, \widehat{a_i, s_m}; \theta_R) = x_i \beta_0^R + \beta_1^R s_m \tag{22}$$

and $\mathcal{V}(\widehat{x_i, s_m, a_i})$ is constructed from $p(0|x'_i, s_m) = \Lambda(x_i \beta_1^E + \beta_1^E s_m)$. In other words, the unobserved state is a market-level constant that enters additively into revenue as well as the index inside the probability function for entry, exit, and type choice. This allows the unobserved state to affect either costs or revenues. The algorithm then iterates over two steps. In the first, the conditional probability of each observation being in each unobserved state is calculated using data and the assumptions of the model. In the second, these distributions are taken as given, and parameters governing payoffs and transitions are estimated conditional on them, treating them as weights.

In essence, in this procedure, we are adding structure to what is unobserved by comparing what our model predicts with the outcome in the data. For example, if for market m , most firms in most periods earn revenues that are higher than their predicted value, and fewer firms exit than our model predicts, it indicates this market likely has a high value of the unobserved state.

Dynamic games like that modelled here have the potential for multiple equilibria. This poses an empirical challenge for full-solution maximum likelihood methods, as there is no way to know which equilibrium outcome to use in evaluating the likelihood, even if all equilibria could be found. Two-step estimators avoid this problem by conditioning on the equilibrium found in the data and assuming that only one equilibrium is played there. This still allows for different equilibria to be played in different markets as long as those market types correspond to observable market characteristics. This assumption is thus further weakened if the estimation allows for unobserved heterogeneity in market types, which could capture equilibrium selection if multiple equilibria were present. In addition, although the estimator used here relies on an iterative procedure to identify the unobserved heterogeneity, the iteration is done within the first stage. This produces consistent first-stage estimates of CCPs that still rely on the data to select the equilibrium as in a standard two-step method.

The assumption used for estimation here would produce inconsistent first-stage estimates if there were multiple equilibria in the data that were not captured by the observable and unobserved market types. This is a difficult condition to test, but in the first subsection of Section 5, I informally test for consistency by comparing the model's predictions regarding entry, exit, and type choice to the outcomes observed in the data, and I find they fit well.

5. Cost results

■ In this section, I present estimates of firm operating costs as well as estimates of entry and exit behavior.

For incumbent firms, I estimate policy functions using logistic regression, where the dependent variable is whether or not the firm stays active, and the independent variables are market characteristics, firm characteristics, and interactions between them. Table 9 shows summary

statistics on the 143 firms that exit in the sample period. Reduced-form estimates of an incumbent firm's probability of remaining active are presented in Table 10. I test a variety of logit specifications as well as probit and poisson. We see that chains and higher quality firms are less likely to exit. Specification III, which includes quadratic and interaction terms, is used in what follows.

In the first stage, entry is modelled as a flexible poisson process. Entrants of each firm type arrive at independent rates that vary with market characteristics. Results are shown in Table 11. Although results differ across firm types, the two strongest predictors of entry for almost all firm types are natural gas production and interest rates. Markets whose existing firms are older on average also attract more entry. These poisson rates of entry are what are used to simulate forward the market state to calculate incumbent continuation values.

I estimate county unemployment rate, total sales, population, gas wells, and traffic as independent autoregressive processes of order 1. These transition probabilities are used along with firm policy functions to forward simulate and estimate continuation values at each state. Once continuation values are constructed, I form equation (20) for each firm. The common discount factor is set at .92.

The unknowns in this equation are the parameters of $C(\cdot)$. I model $C(\cdot)$ as linear in firm type and market characteristics. These are estimated using logistic regression, where fitted values of revenues and continuation values are included. Final results with market heterogeneity accounted for are presented in Table 12. Results are for market random effects distributed over a uniformly spaced grid with 10 points. Increasing this number up to 100 was considered and showed little improvement.

To correspond to our earlier measure of RevPar, costs are presented in terms of dollars per room per day. Different functional form choices for policy functions and continuation values give nearly identical results.²⁸ Bootstrap standard errors appear below in parentheses.²⁹ Full results including market-level unobserved heterogeneity are presented in columns 2–5.

As we expect, higher quality firms have correspondingly higher costs. Two-star firms' operating costs are roughly \$7 higher per room per day than one-star firms, and three-star firms' costs are \$23 higher per room per day. Fixed costs per room are also increasing slightly in number of rooms and traffic, although these are not statistically significant.

The most notable result is that chain firms have no significant cost advantage over independent firms, after conditioning on quality. Costs are estimated at around \$.5 less and this is not statistically significant. If larger chains are better able to achieve economies of scale, we would expect that operating in a larger chain would result in lower costs. Interacting chain status with dummies corresponding to chain size shows no significant evidence that operating in a larger chain lowers costs. Across many specifications, these results argue against the hypothesis that chains are successful because they are more efficient or productive than nonchain firms.

By comparing column 1, which does not control for market-level random effects, to columns 2–5, which do, we can evaluate the impact of accounting for unobserved heterogeneity. First, we see that estimates with unobserved heterogeneity show higher average costs for all firms via the constant term and through coefficients on both two- and three-star hotels. In addition, the key parameter of interest, the coefficient on chain affiliation, not only changes magnitude but changes sign when accounting for market unobservables. Without this adjustment, the result is that chains have somewhat higher costs, although not statistically significant, as opposed to a slightly negative but near zero difference. We would expect a discrepancy between the standard estimates and those with unobserved heterogeneity in cases where accounting for this heterogeneity improves the fit of first-stage estimates and where unobservable factors are related to observed characteristics. These results suggest first-stage estimates perform better and that chains are more likely than

²⁸ See Table 13, which compares the resulting cost estimates using different first-stage specifications of policy functions.

²⁹ These are determined by drawing whole market histories from the set of 353 markets, with replacement, and repeating the full procedure described above 100 times.

TABLE 7 Chain-Only Revenue Estimates

	Market FE	Firm FE
Log chain size	4.238*** (.251)	2.367*** (.595)
Number of independents	1.516*** (.341)	.807*** (.289)
Number of chains	-.732*** (.216)	-1.038*** (.184)
Log county revenue	8.738*** (.554)	8.153*** (.460)
Log capacity	-9.490*** (.645)	-22.930*** (1.877)
New firm	-5.145*** (.550)	-8.446*** (.618)
Unemployment	-1.832*** (.149)	-1.839*** (.123)
Log traffic	3.611* (2.025)	2.301 (1.677)
Log wells	-.387 (.438)	-.464 (.389)
Log population	11.158*** (.170)	10.494*** (3.677)
User rating	2.633*** (.170)	
2 Stars	3.744*** (.786)	
3 Stars	17.160*** (.819)	
N	5669	5669
R ²	.35	.17

Note: This table presents results on the sample of chain firms. The dependent variable is RevPar. Chain size is calculated as the number of chain members observed in the revenue data in 2012.

TABLE 8 Hotel Reviews Summary Statistics

AAA Rating	*	**	***
Chain			
Mean number of reviews	13.25	16.14	24.99
Average rating	2.72	3.18	3.79
Mean σ (ratings)	4.30	4.78	11.78
Mean reviews per room	.16	.25	.35
Independent			
Mean number of reviews	13.28	13.82	35.27
Average rating	2.93	3.04	3.70
Mean σ (ratings)	4.53	3.72	16.58
Mean reviews per room	.10	.22	.58

Note: This table presents summary information on TripAdvisor.com consumer ratings taken from a December 2012 cross-section, broken down by firm type.

independents to locate in markets that have higher costs that are not fully captured by observed characteristics like population, traffic, etc.

Chain firms have higher estimated entry costs than independent firms. These differences are roughly \$92,000, \$90,000, and \$65,000 at the one-two-and, three-star level, respectively. These differences are not statistically significant, and these costs include the initial franchise fee payment, which is typically between \$30,000–\$100,000. When the franchise payment is subtracted out, there is little difference in entry costs between independent and chain firms.

TABLE 9 Exiting Firm Summary Statistics

	Total	Standard Deviation
Number of firms		
Chain	25	
Independent	118	
Total	143	
AAA rating		
*	116	
**	20	
***	7	
Capacity	46.81	29.70
Age (years)	11.99	8.04
2001	7	
2002	9	
2003	14	
2004	9	
2005	14	
2006	15	
2007	11	
2008	10	
2009	15	
2010	10	
2011	15	
2012	14	

Note: This table presents summary statistics on the number of hotels exiting by type and year.

Estimated costs of affiliation switching are somewhat high relative to entry costs. Probably, this results from a general “inertia” in decision making of the sort that generates high estimates of switching costs in other contexts, for instance, insurance plan choice (Handel, 2013). Note that although total switching costs of chain to independent switches are higher than for independent to chain switches, the per room cost is similar. Hotels joining chains happen to be somewhat larger, at an average of 61.3 rooms compared to an average of 49.2 rooms for hotels leaving chains.

To gauge the accuracy of these results, I turn to industry survey data on costs. Smith Travel Research is an independent research firm that operates a large database on hotel industry data, producing annual reports on RevPar, gross operating profits, and operating costs of hotels. Their annual Hotel Operating Statistics Study combines data from roughly 6100 hotels in the United States. In Table 14, I present this summary data on average operating expenses for the limited service budget sector of hotels. The average reported daily per room operating cost is \$31.00. For a two-star chain firm, my estimates suggest an average daily cost of \$30.32, \$28.51, \$31.18, \$30.06, depending on specification. The operating costs I estimate thus fit very closely with average survey responses.

□ **Model analysis.** In this section, I examine how well the estimated model fits the data. There is no formal test for models of this sort, but solving for the model equilibrium with estimated parameters, simulating outcomes, and comparing these to actual observed outcomes is a useful way of evaluating the assumptions embedded in the modelling framework, including those regarding information, timing, functional forms, and so on.

I proceed by using the estimated parameters to solve the game described above for policy functions that form a MPNE. I then forward simulate from the true market structure in 2000 and compare the resulting outcome to the true outcome. Solving dynamic industry game models amounts to solving for a fixed point of a large system of nonlinear equations. This is too complex to solve analytically and so must be done computationally. To do this, I begin by forming firm profits

TABLE 10 Incumbent Firm's Policy Function

Specification	1	2	3
Chain Status	.360 (.330)	.409 (.331)	8.209** (4.09)
Number of independents	.091* (.053)	.024 (.054)	.029 (.054)
Number of chains	-.032 (.038)	-.015 (.038)	-.014 (.038)
Log county sales	.017 (.105)	-.103 (1.32)	-.644 (1.37)
Unemployment	-.065** (.039)	.082 (.140)	-.008 (.152)
Log traffic	-.282** (.126)	2.234** (.941)	2.056** (1.020)
Log wells	.077** (.032)	.128 (.107)	.137 (.107)
Log population	.180 (.136)	3.34*** (1.13)	3.632*** (1.143)
2 Stars	1.047*** (.347)	1.069*** (.346)	1.069*** (.349)
3 Stars	.885** (.409)	.893** (.411)	.903* (.413)
(Log sales) ²		.003 (.032)	.018 (.033)
(Unemployment) ²		-.010 (.008)	-.006 (.009)
(Log traffic) ²		-.145*** (.052)	-.136*** (.057)
(Log wells) ²		-.012 (.015)	-.011 (.015)
(Log population) ²		-.149*** (.053)	-.166*** (.053)
Chain*(Log sales)			-.556** (.271)
Chain*(Unemployment)			.187 (.127)
Chain*(Log traffic)			-.242 (.375)
Chain*(Log wells)			-.083 (.080)
Chain*(Log population)			.521 (.326)
N	12,270	12,270	12,270
R ²	.04	.06	.07

Note: Each specification is a logistic regression. The dependent variable is 1 if the firm is active that period and 0 otherwise.

for firm types in all market states using the estimated cost and revenue parameters discussed in the previous section. I then solve for a fixed point in firm value functions and policy functions using the stochastic algorithm of Pakes and McGuire (2001). With equilibrium firm policy functions governing exit, entry, and type choice, I can then simulate the distribution of firms in different markets.

I start at the true firm distribution in year 2000 and forward simulate 11 years. I repeat this many times for 100 randomly selected markets. In Table 15, I compare the true 2011 firm distribution with the results of model simulations, and find that the model performs well, predicting slightly too many one- and two-star chain firms, but matching the number and distribution of independent firms quite closely.

TABLE 11 Poisson Entry Probability By Firm Type

	Independent			Chain		
	1 Star	2 Stars	3 Stars	1 Star	2 Stars	3 Stars
Demand shifters						
Log county sales	.744*** (.23)	.111 (1.99)	-.119 (.569)	7.235** (3.68)	.284 (.41)	.190 (.29)
Unemployment	-.104 (.07)	-.070 (.429)	.035 (.234)	.228 (.48)	.005 (.10)	-.067 (.07)
Log traffic	-.033 (.28)	10.152 (6.26)	.217 (.555)	-.006 (1.59)	1.080* (.63)	-.873*** (.32)
Log gas wells	.247 (.19)	3.632** (1.84)	.221 (.204)	.199 (1.23)	.069 (.19)	.899*** (.21)
Log population	-.538 (.44)	11.155 (9.40)	-.027 (.105)	-29.389 (18.12)	.017 (.75)	1.269** (.54)
Interest rate	-.314** (.28)	-1.359* (.85)	.423 (.608)	-.733 (.88)	-.582*** (.21)	-.848*** (.14)
Mean market age	.168*** (.05)	.047 (.27)	.027 (.105)	.134 (.24)	.108 (.63)	.228*** (.05)
Number of existing firms						
Independents						
1 Star	-.781*** (.12)	1.707** (.84)	.173 (.224)	.855 (.159)	.440** (.20)	.190 (.28)
2 Stars	.492 (.32)	-6.207*** (1.67)	.134 (.859)	-2.208 (2.61)	-.859** (.38)	.634* (.30)
3 Stars	-.880 (.68)	-4.257 (3475)	2.227 (.763)	.335 (.712)	.746 (.68)	-.667 (.50)
Chains						
1 Star	-.220 (.41)	2.416 (2.06)	-16.481 (2087)	-4.826*** (1.65)	.250 (.49)	-.193 (.35)
2 Stars	-.255 (.20)	-3.436*** (1.16)	.057 (.411)	-1.978* (.205)	-1.905*** (.25)	.406** (.16)
3 Stars	-.302** (.13)	.495 (.90)	-.515 (.645)	.558 (.75)	.544*** (.15)	-1.074*** (.10)
N	2729	2729	2729	2729	2729	2729

Notes: Each column represents a poisson regression on the number of entrants of each firm type in each market each year. Independent variables are market characteristics and the number of firms already in the market by type of firm.

6. Sources of revenue advantage

■ The previous section has documented that chains earn a revenue premium of 20% after controlling for firm and market characteristics. In this section, I consider potential explanations for how organizing into chains generates this premium. One potential source of the increase in revenues is a branding effect, whereby firms use chain affiliation to credibly signal quality in low-information settings. This signalling results from repeat interactions with different members of the same chain, as well as national advertising. Additional sources of a revenue advantage include loyalty programs and access to centralized booking systems. The observed revenue premium is likely the result of a combination of these factors, but they make different empirical predictions in certain cases, and therefore, a close analysis of the revenue data can still be suggestive about their relative magnitudes. In particular, although it is impossible to directly observe how informed consumers are, that a substantial share of the chain premium is the result of signalling to low-information consumers is well supported by the data.

Two of these empirical patterns have already been introduced in Section 3. First, if the chain premium results from poor information, it should be declining over time, particularly since 2005, as online reputation mechanisms have developed and become an important part of the hotel

TABLE 12 Estimated Operating Costs (\$ per room per day)

	No Market RE	Market RE			
	1	2	3	4	5
Operating costs					
Constant	17.77*** (5.71)	23.55** (10.21)	21.82** (11.59)	20.53** (11.73)	23.76*** (9.88)
Chain	1.46 (1.18)	-.49 (2.76)	-.53 (1.23)	-.73 (2.62)	
2 Stars	6.04*** (1.67)	7.32*** (2.94)	7.21*** (2.74)	7.13** (3.33)	7.42** (3.06)
3 Stars	19.78*** (1.17)	23.35*** (4.21)	23.23*** (4.21)	23.08*** (3.76)	23.35*** (4.32)
Log(traffic)			.13 (.44)		
Log(capacity)				.79 (.99)	
Chain size < 100					1.67 (2.94)
100 < Chain size < 200					-3.07 (4.11)
200 < Chain size					.37 (3.92)
Entry costs					
Independent					
2 Stars		153417.12*** (52522.42)			
3 Stars		191026.22* (110100.10)			
Chain					
1 Star		91910.36 (60724.19)			
2 Stars		244177.83** (107545.04)			
3 Stars		256519.70* (155587.19)			
Switching costs					
Independent → Chain		129845.8*** (16762.3)			
Chain → Independent		152721.6*** (30646.1)			

Notes: This table presents final-stage estimates of firm operating costs, expressed in terms of dollars per room per day, with bootstrap standard errors. Chain size refers to the number of chain partners in Texas. Column 1 shows estimates without market-level random effects. Columns 2–5 calculate market random effects using the Arcidiacono and Miller (2011) procedure. Entry costs and switching costs are one-time costs, and so are not normalized to be per room per day. They are estimated in reference to the entry cost of a one-star independent firm, which is normalized to zero.

industry.³⁰ This is indeed the case, as Figure 6 demonstrates, the chain premium is significantly lower at the end of the decade than at the beginning. This result is robust to specifications of the competitive environment and does not seem to be business cycle-related. If the chain premium were driven by advertising, we would not expect to see this pattern, and in particular, if it was driven by the use of loyalty programs, we would see the opposite result, as loyalty programs have increased in use during this period.

³⁰ For background on the impact of search costs and consumer information on firm behavior and outcomes, see Baye, Morgan, and Scholten (2006). For my purposes, I consider low-information and high-search costs to be analogous.

TABLE 13 Estimates Using Different First-Stage Specifications

	Logit	Probit	LPM
Constant	23.55** (10.21)	24.98** (10.93)	25.34** (11.46)
Chain	-.49 (2.76)	-.56 (2.70)	-.39 (2.66)
2 Stars	7.32*** (2.94)	7.52*** (2.91)	6.48** (3.33)
3 Stars	23.35*** (4.21)	23.77*** (3.48)	24.12*** (3.96)

Notes: This table compares the results of final-stage costs estimation under different first-stage specifications of policy functions. I model entry/exit decisions as logit, probit, and linear processes, use these to form continuation values, and solve the final stage using each as inputs while incorporating market RE as described above.

TABLE 14 STR Operating Costs (\$ per room per day)

Administrative	3.31
Marketing	.64
Utilities	2.66
Maintenance	2.24
Total operating expenses	8.85
Franchise fees	1.99
Management fees	.65
Property taxes	1.99
Insurance	.65
Debt service and other	14.18
Payroll and related	7.97
Total	31.00

Notes: Data comes from Smith Travel Research 2012 Hotel Operating Statistics Study, which compiles survey data from 3593 limited service hotels across the United States in the preceding year. Operating costs are for limited service hotels in the budget category.

TABLE 15 Comparison of Model Simulations to Data

Firm Type	True 2001 Distribution	True 2011 Distribution	Simulated 2011 Distribution
Independent			
1 Star	1.80	3.89	3.74
2 Stars	.22	.49	.48
3 Stars	.11	.14	.14
Chain			
1 Star	.21	.10	.40
2 Stars	1.16	1.26	2.08
3 Stars	.59	2.17	2.08

Notes: The left two columns show the distribution of firm types in 2001 and 2011 in the data. The right column shows the results of model simulation. For each market, the model is started at the true 2000 distribution and forward simulated 10,000 times.

Second, if chain affiliation serves as a signal of quality, the chain premium should appear immediately when a previously unaffiliated firm joins a chain. Under that explanation, as soon as the chain sign is raised, the advantage should be present. Figure 5 shows that this is the case. If the premium were driven by subtle improvements in quality or management, we would expect it to take time to phase in. If the premium were just about the informational value of the name and sign, it should appear immediately, as it does.

For the next test, it is helpful to briefly discuss the theoretical underpinnings of the asymmetric information explanation. Typically, consumers and firms overcome the problem of low

information with reputations developed over repeated interactions. Chains may facilitate this process by operating firms in different markets with uniform standards of quality, increasing the potential for repeat interactions.

The implication of this is that the value of a chain's reputation should therefore depend on the number of potential interactions consumers can have with it. This prediction can be tested by testing the relationship between firm revenue and the number of chain outlets. The theory predicts a chain premium increasing in chain size. I test this by regressing:

$$RevPar_{imt} = x_{imt}\beta_1 + firm_{it}\beta_2 + CS_{it} + time_t + firm_i + \epsilon_{imt}, \quad (23)$$

where the new variable CS_{it} is chain size. Chain size is calculated as the number of members of each chain observed in the revenue data in 2012 and is not time-varying. Results can be seen in Table 7. These estimates exclude independent firms. We see a strong, significant, positive correlation between chain size and the chain premium. This lends support to the information-based hypothesis of chain success. This is the case overall, and when firm-level fixed effects are included. In this case, I estimate the effect of chain size on firms that switch chains during the sample period. For these firms, the effect is smaller but still significantly positive. There is a potential selection bias in both of these results, but nevertheless, we see that firms switching to larger chains see an increase in their revenue.

Next, I directly consider results from TripAdvisor.com customer reviews data. The number and consistency of online reviews can act as a measure of consumer information about different firms. Although the number of reviews a firm receives is clearly endogenous, as it is a function of the firm's past history of demand shocks, the mean rating and standard deviation of ratings are not. These data are inherently subjective, however, as consumers' ratings may relate to prior expectations, treating four-star firms differently from one-star firms, for instance. So long as they do not treat chain and independent firms differently, this does not present a problem for any of our results. Table 8 shows summary statistics on reviews, and Figure 7 shows their distribution. There appear to be no systematic differences in mean rating, standard deviation, or number of reviews between chain and independent firms.

To test the effect of number of reviews, I reestimate the chain premium for the subset of firms for whom the number of reviews is large. I use the same regression as in equation (7) but with subsamples of the data. In Figure 8, I display the estimated chain premium for a range of cutoff points. Below this is the number of firms of each type included in the truncated sample. We see that as the number of reviews increases, the resulting chain premium becomes statistically insignificant after 10 reviews and close to 0 after 20 reviews. This suggests that as the amount of information available to consumers increases, the chain premium might disappear completely.³¹ If the premium was driven by a difference in advertising spending between chain and independent firms or centralized bookings, we would not see these patterns. Although those factors undoubtedly do benefit chain hotels, this empirical pattern suggests a large role for asymmetric information in driving the revenue premium.

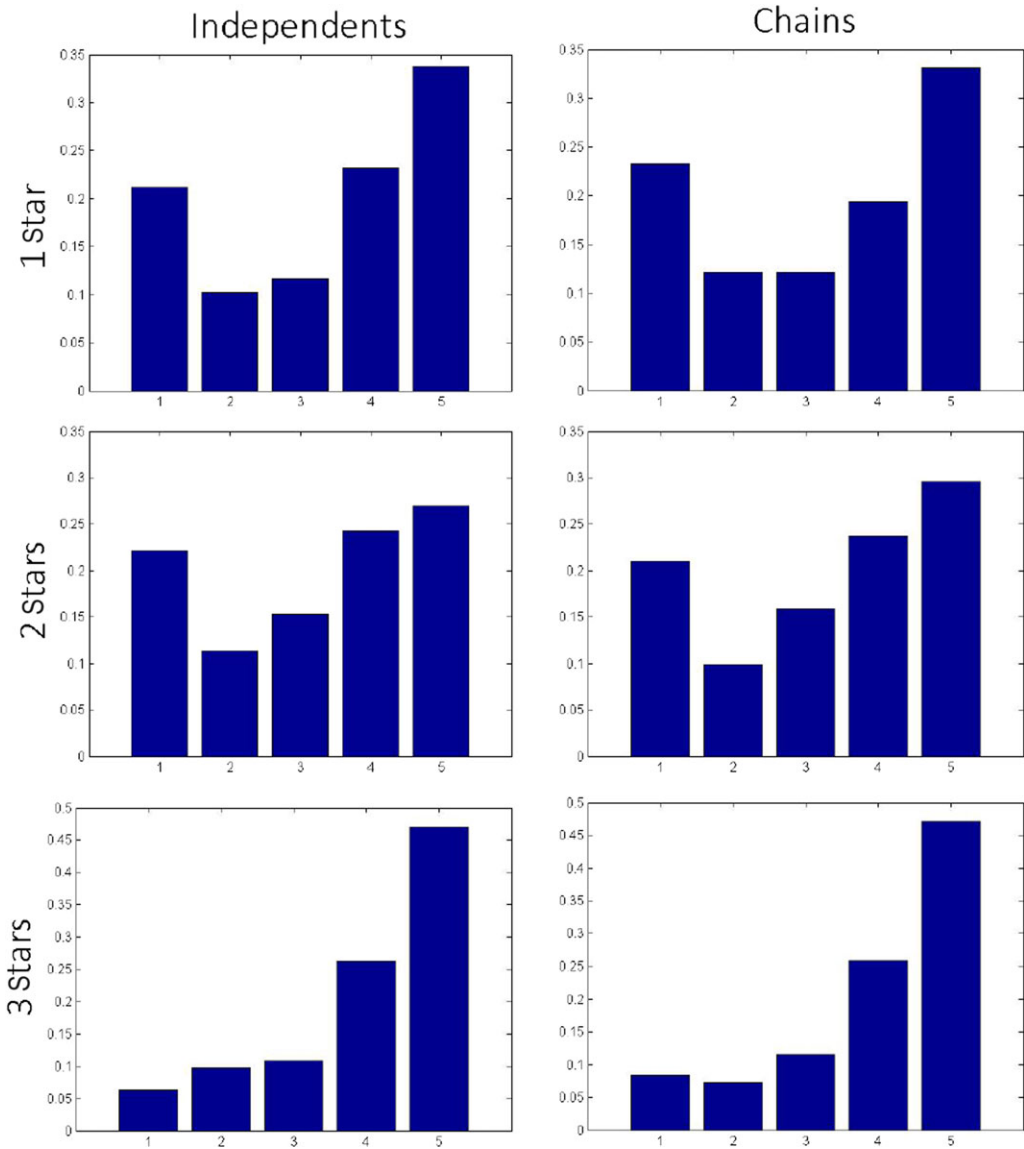
Overall, chains likely receive a revenue premium from several factors. Each of the empirical patterns here fit the predictions of an explanation based on chains providing an informative signal of quality to consumers with low information. Some, but not all, are also consistent with explanations based on loyalty or centralized bookings. It is impossible to directly measure the sources of the chain premium distinctly, but these results suggest that the structure of chain organizations and the way their structure alleviates asymmetric information play a major role.

This branding or reputation effect could come from two distinct sources. One is broader reach of advertising, or else higher levels of advertising as a corporate policy of certain chains relative to

³¹ There is a potential selection effect here, in that the most reviewed firms could have unobservable characteristics that lead them to show a lower chain premium. I test for this type of selection bias by performing a series of tests where, instead of truncating the sample on number of reviews, I truncate it separately on traffic level, age, and number of rooms, and in none of these cases do I find a similar pattern.

FIGURE 7

DISTRIBUTION OF RATINGS BY FIRM TYPE [Color figure can be viewed at wileyonlinelibrary.com]

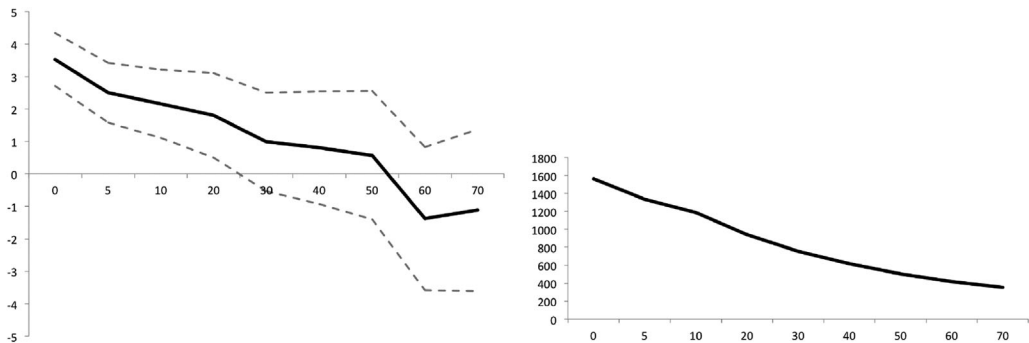


Note: This figure shows histograms of consumer ratings on TripAdvisor.com separated by firm type, where firm type is either chain or independent, and stars refer to AAA quality ratings.

independent hotels. Another is that chains offering similar services in different locations facilitate repeat interactions that otherwise would not have occurred, and in this way, generate a reputation across the firm as a whole. This second explanation relies more on low-information consumers, as these repeat interactions would not be necessary in a high-information setting where consumers have good knowledge of individual hotel quality. It is difficult to empirically distinguish between an advertising effect and a general reputation effect, but the decline in the chain premium over time corresponds to the growth of online reputation mechanisms, and the chain premium goes to zero among firms with large numbers of online reviews. These two pieces of evidence are suggestive that low information, and not advertising, are largely responsible for the reputation

FIGURE 8

CHAIN PREMIUM, TRUNCATED BY NUMBER OF REVIEWS



Note: The left chart presents the estimated chain premium in dollars on the y-axis, where the estimates are on samples truncated by just considering firms with a minimum number number of reviews, shown on the x-axis. The right chart shows the number of firms in each truncated sample.

effect component of the chain premium, although if chains view online reviews and advertising as close substitutes, these same patterns might appear. Ultimately, this question merits further study.

7. Conclusion

■ The rapid growth of chains over the past few decades has attracted the attention of policy makers and economists in a variety of fields. The success of this method of organizing has been documented in a number of articles, and has led some economists to employ structural techniques to measure the sources of its value. This literature has focused on “big box” retail and has thus found a number of cost advantages. This article also addresses the question of why firms organize into chains at all, but examines a nonretail industry and explores revenue-based advantages in addition to cost advantages. I use high quality firm-level data on chain and independent firms in the hotel industry to quantify the value to firms of chain organization, both through costs and through revenues.

As a first step in empirical analysis of these channels, I perform a careful analysis of hotel property-level revenue. Reduced-form analysis of these revenues show that chains earn a significant and quite large premium, a roughly 20% advantage over similar independent firms. Further analysis shows that this chain revenue premium is declining over time, and it appears immediately when a firm joins a chain. These revenue estimates are then used in the first stage in a structural estimation of industry operating costs. Contrary to the extant literature on chains in retail settings, estimation of firm costs in this industry suggest firms derive no cost advantage from chain affiliation. These cost estimates are identified by combining observed revenues with patterns in entry and exit across firm types, after accounting for market-level unobserved heterogeneity. This method produces estimates that correspond closely to survey data on operating costs, are robust across specifications, and fit the observed firm distribution well. The evidence thus suggests that the growth of chains in this industry results from substantial economic advantages, but that these advantages come through revenues and not through economies of scale in costs.

I next consider the source of the revenue advantage and how it might be generated by the chain structure itself. Several empirical patterns stand out: first, the substantial decrease in the chain revenue premium over time. Second, the fact that it appears immediately when a firm joins a chain. Third, the chain premium is larger for chains with more outlets. Fourth, among firms with large numbers of online reviews, the premium essentially disappears. Altogether, these patterns

support the view that chain firms enjoy this revenue premium at least in part from having a known reputation in settings where consumers have little information about product quality. This value could result from greater reach of advertising or from the chain structure itself, which facilitates repeat interactions that otherwise couldn't occur, and thereby creates a type of economies of scale on the demand side. This question merits further study.

References

- AGUIRREGABIRIA, V. "Estimation of Dynamic Discrete Games Using the Nested Pseudo Likelihood Algorithm: Code and Application." MPRA Paper no. 17329, University Library of Munich, Germany, 2009.
- AGUIRREGABIRIA, V. AND MIRA, P. "Sequential Estimation of Dynamic Discrete Games." *Econometrica*, Vol. 75 (2007), pp. 1519–1543.
- ARCIDIACONO, P. AND MILLER, R. "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity." *Econometrica*, Vol. 79 (2011), pp. 1823–1867.
- ATER, I. AND OREN, R. "Price Control and Advertising in Franchising Chains." *Strategic Management Journal*, Vol. 36 (2015), pp. 148–158.
- BAJARI, P., BENKARD, C., AND LEVIN, J. "Estimating Dynamic Models of Imperfect Competition." *Econometrica*, Vol. 75 (2007), pp. 1331–1370.
- BAJARI, P., HONG, H., AND NEKIPELOV, D. "Game Theory and Econometrics: A Survey of Some Recent Research." *Advances in Economics and Econometrics*, (2013).
- BASKER, E. "Job Creation or Destruction? Labor-Market Effects of Wal-Mart Expansion." *The Review of Economics and Statistics*, Vol. 87 (2005), pp. 174–183.
- BASKER, E. AND NOEL, M. "The Evolving Food Chain: Competitive Effects of Wal-Mart's Entry into the Supermarket Industry." *Journal of Economics and Management Strategy*, Vol. 18 (2009), pp. 977–1009.
- BAYE, M., MORGAN, J., AND SCHOLTEN, P. "Information, Search, and Price Dispersion." *Handbook in Economics and Information Systems*, 1 (2006), pp. 323–77.
- BLAIR, R. AND LAFONTAINE, F. *The Economics of Franchising*. Cambridge, England: Cambridge University Press, 2005.
- BLEVINS, J., KHAWAJA, A., AND YANG, N. "Firm Expansion, Size Spillovers and Market Dominance in Retail Chain Dynamics." *Management Science*, (2015).
- BLOOM, N., SADUN, R., AND REENAN, J.V. "Americans Do IT Better: U.S. Multinationals and the Productivity Miracle." *American Economic Review*, (2013).
- CAI, H. AND OBARA, I. "Firm Reputation and Horizontal Integration." *RAND Journal of Economics*, Vol. 40 (2009), pp. 340–363.
- COLLARD-WEXLER, A. "Demand Fluctuations and Plant Turnover in the Ready-Mix Concrete Industry." *Econometrica*, (2013).
- DAVIS, P. "Spatial Competition in Retail Markets: Movie Theaters." *RAND Journal of Economics*, Vol. 37 (2006), pp. 964–982.
- DORASZELSKI, U. AND SATTERTHWAITTE, M. "Computable Markov-Perfect Industry Dynamics." *RAND Journal of Economics*, Vol. 41 (2010), pp. 215–243.
- ELLICKSON, P., HOUGHTON, S., AND TIMMINS, C. "Estimating Network Economies in Retail Chains: A Revealed Preference Approach." *RAND Journal of Economics*, (2013), pp. 169–193.
- ELLICKSON, P. AND MISRA, S. "Enriching Interactions: Incorporating Revenue and Cost Data into Static Discrete Games." *Quantitative Marketing and Economics*, Vol. 10 (2012), pp. 1–26.
- ERICSON, R. AND PAKES, A. "Empirical Implications of Alternative Models of Firm Dynamics." *Journal of Economic Theory*, Vol. 79 (1998), pp. 1–45.
- FOSTER, L., HALTIWANGER, J., KLIMEK, S., KRIZAN, C., AND OHLMACHER, S. "The Evolution of National Retail Chains: How We Got Here." *U.S. Census Bureau Center for Economic Studies*, (2015), CES 15–10.
- FOSTER, L., HALTIWANGER, J., AND KRIZAN, C. "Market Selection, Reallocation, and Restructuring in the U.S. Retail Trade Sector in the 1990s." *Review of Economics and Statistics*, (2006), pp. 748–758.
- GALLANT, R., HONG, H., AND KHAWAJA, A. "The Dynamic Spillovers of Entry: An Application to the Generic Drug Industry." *Management Science*, (2017).
- GOWRISANKARAN, G., LUCARELLI, C., SCHMIDT-DENGLER, P., AND TOWN, R. "The Impact of the Medicare Rural Hospital Flexibility Program on Patient Choice." *International Journal of Industrial Organization*, Vol. 29 (2013), pp. 342–344.
- HANDEL, B. "Adverse Selection and Switching Costs in Health Insurance Markets: When Nudging Hurts." *American Economic Review*, Vol. 103 (2013), pp. 2643–2682.
- HOLMES, T. "The Diffusion of Wal-Mart and Economics of Density." *Econometrica*, Vol. 79 (2011), pp. 253–302.
- HORNER, J. "Reputation and Competition." *American Economic Review*, Vol. 92 (2002), pp. 644–663.
- HOTZ, V. AND MILLER, R. "Conditional Choice Probabilities and the Estimation of Dynamic Models." *Review of Economic Studies*, Vol. 60 (1993), pp. 497–529.

- IGAMI, M. AND YANG, N. "Unobserved Heterogeneity in Dynamic Games: Cannibalization and Preemptive Entry of Hamburger Chains in Canada." *Quantitative Economics*, Vol. 7 (2016).
- JARMIN, R., KLIMEK, S., AND MIRANDA, J. "The Role of Retail Chains: National, Regional, and Industry Results." In *Producer dynamics: New evidence from micro data* (pp. 237–262). Chicago: University of Chicago Press, 2009.
- JIA, P. "What Happens When Wal-Mart Comes to Town? An Empirical Analysis of the Discount Retailing Industry." *Econometrica*, Vol. 76 (2008), pp. 1263–1316.
- KALNINS, A. "Pricing Variation within Dual-Distribution Chains: The Different Implications of Externalities and Signaling for High- and Low-Quality Brands." *Management Science*, (2016).
- KOSOVA, R. AND LAFONTAINE, F. "Much Ado About Chains: A Research Agenda." *International Journal of Industrial Organization*, Vol. 30 (2012), pp. 303–308.
- KREPS, D. "Corporate Culture and Economic Theory." *Perspectives on Positive Political Economy*, 1990.
- LAFONTAINE, F. AND SHAW, K. "Targeting Managerial Control: Evidence from Franchising." *RAND Journal of Economics*, Vol. 36 (2005), pp. 131–150.
- LUCA, M. "Reviews, Reputation and Revenue: The Case of Yelp.com." Harvard Business School Working Paper no. 012-16, 2011.
- MAZZEO, M.J. "Competitive Outcomes in Product-Differentiated Oligopoly." *The Review of Economics and Statistics*, Vol. 84 (2002), pp. 719–728.
- . "Retail Contracting and Organizational Form: Alternatives to Chain Affiliation in the Motel Industry." *Journal of Economics and Management Strategy*, Vol. 13 (2004), pp. 599–615.
- MISRA, S., NAIR, H., AND ELLICKSON, P. "Repositioning Dynamics and Pricing Strategy." *Journal of Marketing Research*, Vol. 49 (2012).
- PAKES, A. AND MCGUIRE, P. "Stochastic Algorithms, Symmetric Markov Perfect Equilibrium, and the Curse of Dimensionality." *Econometrica*, Vol. 69 (2001), pp. 1261–1281.
- PASHIGIAN, P. AND BOWEN, B. "The Rising Cost of Time of Females, the Growth of National Brands and the Supply of Retail Services." *Economic Inquiry*, Vol. 32 (1994), pp. 33–65.
- RYAN, S. "The Costs of Environmental Regulation in a Concentrated Industry." *Econometrica*, Vol. 80 (2012), pp. 1019–1062.
- SCOTT, P. T. "Dynamic Discrete Choice Estimation of Agricultural Land Use." Toulouse School of Economics Working Paper 526, 2013.
- SEIM, K. "An Empirical Model of Firm Entry with Endogenous Product-Type Choices." *RAND Journal of Economics*, Vol. 37 (2006), pp. 619–640.
- STAHL, J. "Effects of Deregulation and Consolidation of the Broadcast Television Industry." *American Economic Review*, Vol. 106 (2016).
- SUZUKI, J. "Land Use Regulation as a Barrier to Entry: Evidence from the Texas Lodging Industry." *International Economic Review*, Vol. 54 (2013), pp. 495–523.
- THOMADSEN, R. "The Effect of Ownership Structure on Prices in Geographically Differentiated Industries." *RAND Journal of Economics*, Vol. 25 (2005), pp. 908–929.
- TOIVANEN, O. AND WATERSON, M. "Market Structure and Entry: Where's the Beef?" *RAND Journal of Economics*, Vol. 36 (2005), pp. 680–699.
- VILLAS-BOAS, S. "Vertical Relationships between Manufacturers and Retailers: Inference with Limited Data." *Review of Economic Studies*, Vol. 74 (2007), pp. 403–419.
- ZHENG, F. "Spatial Competition and Preemptive Entry in the Discount Retail Industry." Columbia Business School Research Paper No. 16–37, 2016.