

Communication, Coordination, and Market Structure in the Lodging Industry: A Structural Examination*

Zhou Yang[†]

Virginia Tech

January 2009

Abstract

This paper adopts a structural approach to examine the impact of pre-play communication on market structure in the context of the lodging industry. We identify such impact by comparing the hotels' entry behavior under correlated equilibrium with that under Nash equilibrium. Since a correlated equilibrium incorporates pre-play communication, while Nash equilibrium does not, such comparison allows us to assess the impact of pre-play communication. We first obtain the parameter estimates of the entry game based on correlated equilibrium, and then simulate the counterfactual where each potential entrant has to make decisions simultaneously and independently, that is, by playing Nash equilibrium. The comparison across the true and counterfactual regimes suggests that pre-entry communication does not have a significant impact on the number of firms in a market. However, the communication does seem to increase within-market differentiation. In addition, pre-play communication increases industry profit.

*I would like to thank Patrick Bajari, Kai-Uwe Kuhn, Francine Lafontaine, Serena Ng, Steve Salant, and Lones Smith for their insightful comments. All remaining errors are my own.

[†]Email: zyang74@vt.edu

1 Introduction

A fundamental question in Industrial Organization is to understand the determinants of market structure, especially the number of firms and their size distribution in a market. Among the many factors that potentially affect market structure, demand size, cost structure, and post-entry strategic interactions have attracted much attention from I.O. economists. Another important factor, pre-entry communication among potential entrants, however, has been largely ignored in the literature. In industries with non negligible sunk costs, potential entrants have strong incentives to communicate their plans among themselves before committing such costs. One important function of such pre-entry communication activities is that they can help reduce the strategic uncertainty and enable potential entrants to achieve greater level of coordination.

One interesting question then is how important empirically such pre-entry communications are in affecting the entry behavior and market structure. The challenge in answering this question first lies in the fact that, in a given industry, there is little variation in the extent of pre-entry communication. Furthermore, no data exist that can be used to measure the extent of pre-entry communication. Given these two challenges, we propose to measure the communication effects using a structural approach in the present paper. We first estimate an entry game that involves pre-play communication. We then conduct the counterfactual analysis by comparing the market configuration in the true regime, where potential entrants communicate, and the configuration in the counterfactual regime, where pre-play communication activities are eliminated.

The primary tool we rely on is the solution concept of correlated equilibrium of simultaneous-move games. By the revelation principle in the sense of Myerson (1982), any equilibrium of a simultaneous-move game embedded with any communication mechanism must be a correlated equilibrium of this game. Hence, in order to incorporate communication into an entry game, it is without loss of generality for us to focus on the correlated equilibria of this entry game. On the other hand, if no communication is allowed, potential entrants will play Nash. Therefore, to examine the impact of communication on entry behavior and market structure, we need

only compare the correlated and Nash equilibria of the same entry game.

We implement this idea in the context of the lodging industry. We construct a data set that covers all economy and midscale hotel chains operating in the state of Texas. This data set fits our needs well for the following reasons. First, entry into a new market involves nontrivial sunk cost. Second, hotel chains compete with each other in many markets, and they frequently communicate among themselves through numerous trade associations, annual conferences, and forums.¹ In our entry model, we allow the chains to decide whether or not to enter a market as well as what capacity level to choose if they enter. We specify the profit function in the spirit of Mazzeo (2002).

We estimate the payoff parameters based on the correlated equilibrium set of this entry game. After obtaining the parameter estimates, we simulate the counterfactual where each potential entrant has to make entry and capacity decisions simultaneously and independently by playing Nash equilibrium. The comparison across the true and counterfactual regimes suggests that pre-entry communication does not have a significant impact on the number of firms in a market. However, the communication does seem to increase the **within-market dispersion in capacities**. In addition, pre-play communication increases industry profit.

To the best of our knowledge, this paper is the first attempt in the literature to empirically measure the effect of pre-entry communication on market structure. It is closely related to several strands of literature. First, there has been a large theoretical literature in I.O. that studies information exchange across firms. Even though there are a few papers in this literature that look at communication and entry, such as Farrell (1987), most of this literature focuses on exchange of information regarding demand and cost uncertainty, which can be viewed as post-entry payoff uncertainty. In terms of the research question, the closest paper to ours is by Christensen and Caves (1997), who investigate the impact of cheap talk communication on capacity expansion in the context of the pulp and paper industry.

Our paper is also related to a growing body of literature on the estimation of simultaneous-move games, especially entry games. Mazzeo (2002) studies entry and product differentiation in the motel industry. Ciliberto and Tamer (2006) examines market structure in the airline industry

¹See next section for examples.

using a bounds estimation approach. Bajari, Hong and Ryan (2006) explicitly model the equilibrium selection mechanism and allow the equilibrium strategy to be mixed, a possibility that had been ignored heretofore in the literature.

The remainder of this paper is organized as follows. Section 2 offers some industry background and describes the data. Section 3 provides a theoretical discussion of market entry with communication. Section 4 sets up the structural model. Section 5 discusses the estimation approach. Section 6 presents the estimation results and counterfactual analysis. Section 7 concludes.

2 Industry and Data

2.1 The Lodging Industry

The lodging industry is one of the largest among all service industries in the U.S. economy, with an average annual revenue greater than \$60 billion (in 2000 U. S. dollars, same below) over the last decade (Smith Travel Research, 2005). Among the five different subsectors² (*i.e.*, Bed & Breakfast, Economy, Midscale, Upscale, Upper-upscale/Luxury) in this industry, the economy and midscale subsectors have been the fastest growing over the past several decades. Table 1.1 contains some descriptive statistics about the annual revenues and number of establishments for these two subsectors as well as the whole lodging industry in the U.S. for the years 2000 and 2006. During this time period, the whole lodging industry grew in annual sales revenue from US \$62.16 billion to US \$78.72 billion, for an average annual growth rate of 4%. In comparison, the annual sales revenue increased from US \$17.76 billion to US \$31.92 billion for the economy and midscale segments, at a significantly higher average annual rate of 12%. The growth in the number of establishments reveals a similar pattern.

There are two types of establishments operating in the sectors of economy and midscale hotels:

²The exact categorization of the lodging industry into these different segments sometimes is a little different across studies. The categorization method adopted in this paper follows Smith Travel Research, which groups hotels into different segments according to amenities and services provided, as well as the price range. This grouping method is used most often by both researchers and practitioners.

chain-affiliated and independent hotels and motels. Establishments affiliated with a chain are usually owned or franchised by the chain, and operated under a widely recognized brand name. Examples of such chains include Best Western, Comfort Inn, Days Inn, Holiday Inn, Super 8, etc. Independent hotels, on the other hand, usually do not have a widely recognized brand name, and are often owned and operated by a small-business person (or, in some cases, by a private company), and are more “local.” Although there are also a large number of independent hotels operating in the sectors of economy and midscale hotels, the chains clearly play a dominant role in shaping the landscape of the lodging market: the ten largest brand owners hold more than 50% of the U.S. market (Hotel and Motel Management, 2004).

Within the economy or midscale sector of the lodging industry, chains clearly have certain advantages over independent hotels. First, on the cost side, an obvious advantage of a chain over an independent hotel is that a chain can reduce the purchasing and marketing expenses by exploiting the chain-wide economy of scale. Second, on the demand side, as an experience good, the lodging service is often plagued by the informational asymmetry between the hotel and consumers regarding the quality of the services. This problem is especially acute in remote rural areas, where the customers are more likely to be one-time service purchasers. A chain’s brand name can mitigate this asymmetric-information problem by acting as a commitment device, since bad service in one location will certainly have a negative externality on the establishments affiliated with the same brand in another location.

Another potentially less well understood but still important advantage of chains relative to independent hotels is that they typically have access to more channels through which to communicate with their rivals. Because they frequently compete in many markets, and more importantly, because investments on new establishments are typically lumpy, the chains have incentives to “talk” with each other before making their decisions on entry into new markets. Indeed, besides the trade publications circulated by the American Hotel & Motel Association and various other trade groups, there are numerous conferences and forums held annually where the large chains can get together and potentially exchange ideas and plans about hotel development and investment. Examples of such forums include the annual Strategic Conference on Hospitality, the Lodging Conference, and the Lodging Industry Investment Council, all of

which cite “serving as a communication network” as a primary objective. Albeit (sometimes) soft information exchange, these communication activities to some extent do help coordinate their decisions by reducing strategic uncertainty.

2.2 Data Sources

Our data consist of information on the hotel establishments that operate in the economy or midscale sector in each Texas cities or town, and come primarily from three sources. First, we assemble the data set on Texas Economy or Midscale hotels based on the hotel database maintained by the Texas Comptroller’s Office. This database contains detailed information on each hotel establishment operating in the state of Texas, including its chain affiliation, capacity (number of rooms), and address. We then extract the subset of chain-affiliated establishments that operate in the Economy/Midscale sector by referring to the complete list of hotel chains operating in this segment provided by Smith Travel Research, an independent consulting firm specializing in the lodging industry. From this Texas hotel database, we also count the number of independent hotels operating in each city or town in Texas.

To supplement the hotel information, we also collect demographic and economic data describing the conditions in each market. The demographic variables, which include population, per capita income, and number of housing units, are downloaded from the web site of U.S. Census Bureau. Finally, for the data describing the economic conditions, we assemble information from the Zip Business Patterns on the counts of employment in retail establishments³, and at gasoline stations⁴. The Zip Business Patterns Database assigns an employment scale category for each NAICS industry at the zip code level. In order to obtain these two employment counts at the city level, we aggregate them to each city or town that contains the corresponding zip codes. Finally, we create a dummy variable to indicate whether a market is a tourist attraction based on the information provided by the Texas State Data Center.

³Previous papers, such as Chung and Kalnins (2001) use the counts of establishments in the related industries. We use the counts of personnel employed by these industries to take into consideration the fact that some establishments may be larger than the others.

⁴Following Chung and Kalnins (2001), we use the employment size by gas stations to proxy for the highway traffic flow, which potentially affects the demand for hotel service.

2.3 Market Definition and Sample Selection

Since this paper focuses on the Economy and Midscale chains in the hotel industry—which we believe have a low direct competition with the Upscale or Luxury hotels—the remaining task in defining a market is to find an appropriate geographic unit so that the extent of competition between two hotels across different locations is likely to be low as well. Depending on data availability, there are several possible ways to define a geographic market in the hotel industry. In principle, one could define a geographic market at the zip code, city, or county level; or alternatively, some other mutually exclusive geographic units, such as highway exits. For example, Chung and Kalnins (2001) define a market to be at the level of five-digit zip code; Conlin and Kadiyali (1999, 2006) consider a city as a geographic market; and Mazzeo (2002) treats as a market the cluster of motels close to an exit along the U.S. interstate highways.

Among these, Mazzeo’s (2002) approach specifically limits the extent to which motels in one market compete with those in another market. However, this market definition is likely to capture only part of the demand that comes from passing-by consumers, and therefore further reduces the market segment to one that is narrower than what he intends to study. Although it can be used to exploit the variation across markets to the largest possible extent, the market definition adopted by Chung and Kalnins (2001) risks allowing hotels in one zip code to compete directly with others in another zip code. In this paper, we follow Conlin and Kadiyali (1999, 2006) in defining a market at the level of city (or town). While it is less than ideal as well, we do believe this geographic scope is both large enough to encompass all the potential demand for the market segment considered in this paper, while still substantially limiting the extent to which hotels from different geographic units are likely to compete with each other.

After defining a market to be the economy/midscale sector of a city or town in Texas, we focus on those cities (or townships) with a population size between 1,000 and 300,000 according to the census 2000 data.⁵ According to a report by Smith Travel Research (2005), it is very unlikely for a hotel chain to enter a market with population size less than 1,000. Also, we remove from the data the cities with population greater than 300,000 based on our belief that

⁵We also tried 200,000 and 400,000 as the top cutoff population. The sample we obtained are essentially the same as when we set 200,000 to be the cutoff. .

the market configurations as well as the strategic interactions among hotels in these large cities are necessarily very complicated and beyond the scope of this paper.

2.4 Descriptive Statistics

After imposing the above selection criteria, we obtain a sample of 828 chain-affiliated establishments that belong to the Economy/Midscale sector and that operate in 437 Texas cities (or townships). These 828 establishments are each affiliated with one of the 33 different national or regional chains. Table 2 lists the number of establishments affiliated with these chains (in terms of the number of establishments) as well as the number of markets in which these chains are present. Among the 33 chains, the six largest in Texas (in terms of the number of establishments affiliated) operate a total of 664 establishments, or more than 80% of all the chain-affiliated establishments; the other 27 chains operate 164 establishments in total, and none of them have more than 30 establishments.⁶ We therefore focus on the six largest chains when modeling the market structures observed in the data.

Within the top 6 chains, Best Western has the largest number of establishments, with 175 affiliated establishments operating in 167 markets, of which only 10 markets contain multiple Best Western establishments; Motel 6 has the smallest number of establishments, 64. Since most market entry in the data features a single establishment, we will abstract from the choice of number of establishments in our specification.

Another important choice a chain has to make when deciding whether or not to enter a specific market is its capacity level.⁷ Table 3 summarizes the capacity distribution for the establishments affiliated with each of the top 6 chains. Motel 6 and Holiday Inn have the highest mean capacities, with 93 and 90 rooms, respectively, and the other 6 chains each have an average capacity approximately equal to 60. Within each chain, there is a wide dispersion in capacity level across markets. For instance, Holiday Inn has a capacity level as high as 245 and as low

⁶Among these 27 chains, Econo Lodge is the sixth largest chain, with only 28 establishments. The remaining 26 chains each have fewer than 15 establishments.

⁷To see how sensitive of the results are to the choice of threshold, we tried the threshold of 55 and 65 as well in the estimation, and the results did not change much. Personal communications with the consultants at Smith Travel Research also suggest that 60 is a reasonable threshold.

as 40. To simplify the structural modeling and estimation, we we assign an establishment to the high- or low-capacity type according to whether it has more than 60 rooms, based on the fact that both the mean and median capacities across all the chain-affiliated establishments in our data set are roughly equal to 60.⁸

To take a first look at the market configurations of the top 6 chains in our data, we display the frequencies of the observed number of chain-affiliated establishments for these six chains in Table 4. From this table, we can see that there are only 21 markets that have all these six chains present, while there are a total of 187 markets where none of the above six chains are present. The mean population for the markets having one out of the six chains present equals 76656.08, while those markets that have none of the six on average are much smaller, with a population size equal to 5034.227.

To take a closer look at the market configurations of the 6 chains, we present in Table 5 the breakdown of markets by the market configurations in each market. Each row of the table displays the market configurations that share the same number of low-capacity chain establishments, and each column presents the configurations sharing the same number of high-capacity chain establishments. The number in each cell represents the frequency of the corresponding market configuration. For example, in 25 out of the 437 markets, a single chain with high capacity is present. Market configurations having the same number of firms are on the same diagonal running from the bottom left corner to the top right corner. A close inspection of the table reveals one pattern: among the different market configurations having the same number of firms, more “balanced” (in the sense that the number of low- and high-capacity chain establishments is similar) configurations seem to appear more often. While only suggestive, the frequencies listed in the table do seem to be consistent with the notion that hotel chains operating in the same market try to differentiate in capacities, other things equal.

In Table 6, we present the summary statistics for the data. The mean population across all the markets is 17.43 thousand, and the median is 5.87 thousand, suggesting that the distribution

⁸To see how sensitive of the results are to the choice of threshold, we tried the threshold of 55 and 65 as well in the estimation, and the results did not change much. Personal communications with the consultants at Smith Travel Research also suggest that 60 is a reasonable threshold.

of the population across the markets is right skewed. The two independent variables used to proxy the economic conditions of the market—retail employment and the employment by gas stations, have a mean of 2.44 thousand and 0.30 thousand, respectively. Furthermore, about 9% of the markets contain some tourist attractions. Finally, the average number of establishments (across markets) affiliated with the top 6 chains is 1.42, and on average there are 2.61 independent establishments in each market.

3 Market Entry with Communication: Theoretical Investigation

In this section, we set up a general entry model that incorporates communication possibilities among the chains who consider whether or not to enter a market as well as what capacity level to choose if they decide to enter. The framework discussed here serves several purposes. First, it provides a guideline for our structural modeling in section 4, which will then be taken to data. In particular, we discuss the revelation principle (Myerson, 1982) in the context of entry games. As will be clear in section 4, this technique can simplify the structural modeling and estimation substantially. Secondly, we use the theoretical framework here to qualitatively assess the impact of pre-play communication on the equilibrium market configuration. In section 4, we will quantify this impact using a structural model adapted to the lodging industry.

Modeling an entry game with pre-play communications poses several nontrivial challenges. The main challenge stems from the fact that, in principle, potential entrants can communicate with each other in many different ways. The communications can be, among other things, a cheap-talk announcement saying “I will enter this market”, an observation of some public signals about future market conditions, or a costly signal such as building some facilities that may later be used in production (*i.e.*, establishing a supply chain).

We follow Myerson (1982) in formulating the possible communications among potential entrants as taking the following structure: Each potential entrant $i \in I$ is endowed with a *message space* M_i that contains all the possible messages player i could receive, and a *report space* R_i containing all the possible reports i could send. For any profile of reports $r = (r_i)_{i \in I}$,

and any profile of messages $m = (m_i)_{i \in I}$, let $v(m|r)$ denote the conditional probability that m would be the profile of messages received by the various players if profile of reports r were sent out by them. Hence, $v(m|r)$ constitutes a mathematical characterization of an arbitrary *communication system* that could be used by the players to facilitate their communications. Given this mathematical characterization of the possible communications among the potential entrants, we are now ready to describe the model.

3.1 A Model of Entry with Pre-Play Communication

Consider an otherwise stylized entry game embedded with pre-play communication activities. In stage 1, each potential entrant $i \in I$ simultaneously sends out report r_i to the communication system, which in turn sends out messages $m = (m_i)_{i \in I}$ (message m_i to player i) according to conditional probability distribution $v(m|r)$. In stage 2, after receiving message m_i , each potential entrant i decides simultaneously whether or not to enter by choosing $a_i \in A_i = \{N, H, L\}$, with $a_i = N$ denoting “do not enter”, and $a_i = H$ and $a_i = L$ denoting “enter with high capacity” and “enter with low capacity”, respectively. Those firms that have chosen $a_i = H$ or L compete with each other and the post-entry profits are realized. Since our focus is on the impact of communication on entry behavior, we do not explicitly model the post-entry competition process, but rather assume each player i makes profit equal to $\pi_i(a_i, a_{-i})$ when i 's entry decision is a_i , and the profile of entry decisions made by all the other firms is a_{-i} . We normalize i 's payoff $\pi_i(a_i, a_{-i})$ to be 0 if $a_i = N$.

In the above communication-embedded entry game, a pure strategy for player i is described by (r_i, δ_i) , where r_i denotes the report to send and $\delta_i : M_i \rightarrow A_i$ a mapping determining how the entry decision a_i , depends on the possible message received, m_i . Let $B_i = \{(r_i, \delta_i) | r_i \in R_i, \delta_i : M_i \rightarrow A_i\}$ denote the set of pure strategies available to player i . Given any profile of pure strategies $(r_j, \delta_j)_{j \in I} \in B = \times_{j \in I} B_j$, player i 's payoff is

$$\bar{\pi}_i \left[(r_j, \delta_j)_{j \in I} \right] = \sum_{m \in M} v(m|r) \cdot \pi_i \left\{ [\delta_j(m_j)]_{j \in I} \right\}. \quad (1)$$

In the above game with communication, we impose the rationality condition that the players

behave according to some Nash equilibrium $\sigma = (\sigma_i)_{i \in I}$, where $\sigma_i \in \Delta(B_i)$ ⁹ denotes a generic mixed strategy available to player i .

The model we are considering here is rather general. The specification of the communication stage includes many familiar settings as special cases. For instance, a “sunspot” would correspond to the situation where $v(m|r)$ does not depend on the reports sent by the players, and where it is common knowledge that each player receives the same message generated by the public randomization via $v(m|r)$. Another special case is that potential entrants make announcements regarding their intention of entry before making actual entry decisions, as in Farrell (1987), who studies how pre-play announcements about intentions of entry can help overcome coordination failure in entry decisions. It is also worthwhile to note the difference between this model and the literature on information exchange regarding demand or cost uncertainty. The payoff structure in this model is common knowledge, and the purpose of communication here is to solve strategic uncertainty.

Our model does, however, rule out certain communication mechanisms, such as costly reporting. To see this, note that player i 's report r_i affects its profit only indirectly—through the possible impact on the other firms entry behavior. One may well argue that an announcement of entry that later turns out to be false would hurt the potential entrant's credibility, therefore, the argument goes, the report r_i should directly enter i 's profit function $\pi_i(\cdot, \cdot)$, along with a_i and a_{-i} . This is certainly an important possibility, for example, in the software industry, or in other durable goods industries, where the announcement of an intended entry into certain product segments will change consumers' expectations. In the hotel industry, however, we believe that this type of credibility-hurting mechanism is less relevant, and that costless communications can capture fairly well the reality of the industry, in which communications such as announcing entry plans is often regarded as “soft” information (Kuhn, 2001).

3.2 Revelation Principle

In order to take the above entry game with communication to data, one might consider solving for all the Nash equilibria of this game. Taking this route raises at least two issues. First, con-

⁹ $\Delta(B_i)$ denotes the set of probability distributions over the set B_i .

ceptually, we would need to explicitly specify the whole structure of the communication stage, including the message spaces $M_i (\forall i \in I)$, the report spaces $R_i (\forall i \in I)$, and the conditional probability $v(m|r)$. But the lack of knowledge about the fine details of the communication activities in fact prevents us from doing so. Second, methodologically, even if we are comfortable with making strong assumptions about the communication activities, solving for the Nash equilibrium set for a game having nontrivial scale—which is the case in the current empirical exercise—is computationally infeasible.

The key to the approach that we take in this paper is to characterize the properties that *any* Nash equilibrium of a communication-included game corresponding to *any* communication system must satisfy—which is equivalent to saying, we want to construct a set each element of which is a Nash equilibrium of a communication-included game obtained by adding some communication system to the original entry game. The appropriate tool that enables us to achieve this characterization is a revelation principle for complete-information games with communication (Myerson, 1982).

Theorem 1 *Consider the above communication-embedded entry game corresponding to a communication system $v(m|r)$, $\Gamma_v = [I, (B_i)_{i \in I}, (\bar{\pi}_i)_{i \in I}]$ ¹⁰, where I is the set of potential entrants, B_i is the set of pure strategies available to player i , and $\bar{\pi}_i(\cdot)$ is the payoff function for i , as given in (1). Then the probability distribution $\mu = (\mu(a))_{a \in A} \in \Delta(A)$ over the set of possible profiles of entry decisions is a correlated equilibrium of the original entry game without communication $\Gamma = [I, (A_i)_{i \in I}, (\pi_i)_{i \in I}]$; Conversely, any correlated equilibrium of the entry game without communication corresponds to a Nash equilibrium of some communication-embedded entry game obtained by adding an appropriate communication system to the original entry game.*

This theorem suggests that the correlated equilibrium set of the original entry game without communication is analytically equivalent to the union of all the Nash equilibrium sets each of which corresponds to a communication-embedded entry game obtained from adding an

¹⁰Throughout the paper, we use Γ_v and Γ to denote entry games with and without communication, respectively. The subscript v here denotes the communication system v .

appropriate communication system to the original entry game. In other words, we can use a correlated equilibrium of the entry game without communication to “simulate” the equilibrium distribution over the outcomes in any communication-embedded entry game. In the next section (section 4), we will invoke this theorem when constructing the structural model to be brought to data. Myerson (1982) contains the proof of this theorem in a more general setting.

3.3 Impact of Communication on Market Structure

In order to isolate the impact of pre-play communication on the equilibrium market configuration in our entry game, let us consider the following thought experiment. In the first scenario, the potential entrants have to make entry decisions simultaneously and *independently*. That is, the potential entrants behave according to a Nash equilibrium of the original entry game $\Gamma = [I, (A_i)_{i \in I}, (\pi_i)_{i \in I}]$. In the second scenario, the potential entrants are allowed to communicate with each other through the communication system $v(m|r)$. In other words, the potential entrants now are playing a communication-included game, $\Gamma_v = [I, (B_i)_{i \in I}, (\bar{\pi}_i)_{i \in I}]$. Comparing the equilibrium market configurations across the above two scenarios will shed lights on the impact of communication on market structure.

By the revelation principle above, for any possible communication system $v(m|r)$, all the Nash equilibria of Γ_v must be a correlated equilibrium of Γ . Thus, by working with the correlated equilibrium set of the original entry game Γ without communication, we automatically allow the rich communication possibilities among the potential entrants. Therefore, comparing the market configurations across the above two scenarios boils down to comparing the Nash equilibrium set and the correlated equilibrium set, both of Γ .

In particular, as we shall do in the counterfactual experiment later, we can choose a correlated equilibrium randomly from the Pareto-efficient frontier of the correlated equilibrium set and the Nash equilibrium that exhibits the highest level of coordination among the Nash equilibrium set. The comparison of market configurations across these two equilibria will provide a lower bound for the impact of communication on market configuration. The result of the comparison, however, critically depends on the underlying payoff structure of the original entry game, and thus is an empirical problem. Therefore, in order to measure the impact of communication on

market structure, we need to empirically model the payoff structure, which we turn to next.

4 Structural Modeling

4.1 Model Setup

We follow the theoretical discussions above as guidelines in formulating the empirical model on the market structure of lodging industry while also taking into consideration the fact that hotels are naturally divided into two groups: chains and independent hotels. To be specific, the model developed here is a two-stage game. In the first stage, hotel chains simultaneously decide whether or not to enter a market, as well as what capacity level to choose if they enter, using a *correlated strategy*. They take into account the possible reaction from the small, independent hotels (competitive fringe) when making their entry choices. In the second stage, small, independent hotels observe the chains' choices and then make their own entry decisions.¹¹ We assume they enter the market until the profit for one extra independent hotel becomes negative. Hotels (both the chains and the independent ones) in the same market then compete with each other and the profits are subsequently realized. All the hotels' profitability and payoff structures are assumed to be common knowledge among the firms.

Note that in this setup, we do not explicitly model the communication stage, but instead allow the potential entrants to play a correlated strategy, which, according to the revelation principle discussed above, has automatically taken into consideration the possible pre-play communication. Another modeling choice we have made is that we treat the small, independent hotels as competitive fringe and assume they move as a Stackelberg follower. Hence, as soon as chains have made their entry and capacity decisions, the whole market structure is determined. One might raise concerns with this approach by referring to the fact that, in reality, some small hotels have long existed in a market before chains enter, therefore, a three-stage game might be more appropriate. Jia (2006) encounters the same problem when formulating the interactions between discount chains and small retail stores. She also adopts a two-stage approach and treats the small retail stores as a Stackelberg follower. She justifies this approach by arguing

¹¹In the present setup, we do not explicitly model independent hotels' capacity decisions due to tractability.

that, even if they have made the entry decisions before the chains do, small stores will be likely to change their decisions—due to the lack of ability to make credible commitment—after chains’ entry. We believe this justification applies to our context as well.

4.2 Profit Functions

In order to take our model to data, we need to specify the parametric forms of the profit functions for both the chains and independent hotels. Let M stand for the set of markets and I ¹² the set of top six chains in the data. For each market $m \in M$, the number of small, independent hotels is denoted by N_{sm} ¹³. Furthermore, we denote chain i ’s entry decision in market m by a ternary variable $a_{im} \in A_i = \{N, H, L\}$, with $a_{im} = N$ if i decides to stay out of this market, $a_{im} = H$ if i decides to enter with high capacity, and $a_{im} = L$ otherwise. Finally, we let X_m and η_m represent the observed and unobserved (to the econometrician) market-specific characteristics, respectively, and let v_{Hm} stand for the unobserved profit shock if i chooses to enter with high capacity.

Similar in spirit to Mazzeo (2002), we specify chain i ’s entry profit as follows if it decides to enter with capacity type $T \in \{H, L\}$:

$$\pi_{Tm} = X_m \beta_T + \delta_{TH} \sum_{j \in I \setminus \{i\}} 1 \{a_{jm} = H\} + \delta_{TL} \sum_{j \in I \setminus \{i\}} 1 \{a_{jm} = L\} + \delta_{Ts} N_{sm} + \rho_T \eta_m + \sqrt{1 - \rho_T^2} v_{Tm}. \quad (2)$$

For independent hotels, we impose the symmetry assumption and let each independent hotel s have an identical profit function:

$$\pi_{sm} = X_m \beta_s + \delta_{sH} \sum_{i \in I} 1 \{a_{im} = H\} + \delta_{sL} \sum_{i \in I} 1 \{a_{im} = L\} + \delta_{ss} N_{s,m} + \rho_s \eta_m + \sqrt{1 - \rho_s^2} v_{sm}. \quad (3)$$

We normalize the profits from staying outside the market to 0 for both chains and independent hotels. The parameters to be estimated are $\theta \equiv \{\beta_k, \delta_{k,l}, \rho_k\}$, $k, l \in \{H, L, s\}$.

¹²In Section 3.3, we use I to denote the set of potential entrants. Here, with a little abuse of notation, we use it to denote the set of top 6 chains.

¹³Note that the set of chains, I , does not have m as its subscript, because all the chains are potential entrants into each market. In contrast, the independent hotels are “local” to a specific market.

Both the high or low capacity chain establishment's profit π_{Tm} and the independent hotel s 's profit π_{sm} are composed of three basic components: observed market size, competition effect, and the unobserved profit shock. For a chain establishment of type T , the market size $X_m\beta_T$ parameterizes the market-specific demand shifters, such as population size, etc. The competition effect from the rival chains is captured by the term $\delta_{TH} \sum_{j \in I \setminus \{i\}} 1\{a_j = H\} + \delta_{TL} \sum_{j \in I \setminus \{i\}} 1\{a_j = L\}$, and that from the independent hotels is by $\delta_{Ts}N_{sm}$.

The unobserved profit shock $\rho_T\eta_m + \sqrt{1 - \rho_T^2}v_{Tm}$ is a weighted average between the heterogeneity at the market-level, η_m , and that at the chain-level, v_{Tm} , with ρ_T reflecting the relative importance of the former. $\rho_T\eta_m + \sqrt{1 - \rho_T^2}v_{Tm}$ is assumed to be unobserved to the econometrician but common knowledge among all the hotels in the market. We will discuss this assumption further in the next subsection. We assume η_m is *i.i.d.* across markets and v_{Tm} is *i.i.d.* across both markets and capacity types. Furthermore, we specify both η_m and v_{Tm} to be standard normal, which, combined with the weights we put on these two error terms, implies that the variance of the unobserved profit shock is normalized to 1. This normalization is necessary because the scale of the parameters and the variance of the error terms are not separately identified, which is a standard issue in the discrete choice literature.

Finally, in the profit function π_{sm} for independent hotels, the competition effect from the chains is parameterized by the term $\delta_{sH} \sum_{i \in I} 1\{a_{im} = H\} + \delta_{sL} \sum_{i \in I} 1\{a_{im} = L\}$, and that from the rival independent hotels is captured by the term $\delta_{ss}N_{s,m}$. Since we assume independent hotels act like a Stackelberg follower, in equilibrium, the number of these hotels is a function of the chains' entry and capacity decisions made earlier. In the following section on estimation, we will mainly focus on chains' entry decisions, since as soon as chains have made their entry decisions, the number of small hotels will be determined.

4.3 Discussion of the Modeling Choices

The approach taken here to model chains' entry decisions is now standard in the empirical IO literature on modeling market structures. I.O. economists have used various versions of (2) to study markets for differentiated products. For example, Seim (2006) adopts a similar specification of profit function when studying video stores entry and location choices. The closest

specification of profit functions to ours is Mazzeo (2002), who models the market structure in the motel industry, where the motel can choose low or high service quality besides its entry decision. While he considers all the motels, chain affiliated or independent, our focus, is on the hotel establishments affiliated with large chains, since the goal of the paper is to examine the impact of communications among these chains on market structure. Therefore, we abstract from the differences among small, independent hotels by treating them as identical competitive fringe.

We assume the entire structure of the game is common knowledge among all the potential entrants (both chains and independent hotels). That is, the functional forms of the profits, all the market- and firm-specific characteristics, all the parameters, and all the unobserved profit shocks (at both the market and the firm levels) are common knowledge. In other words, the game is one of complete information. On the other hand, the econometrician knows the functional forms of the profit functions and observes the realization of the covariates but not that of the unobserved market- and firm- specific characteristics. Her goal is to make inferences about the parameters.

One may ask why we impose the complete information assumption on the structure of the game—a game of incomplete information is substantially simpler to estimate, after all (Bajari, Hong, and Ryan, 2006). Indeed, Seim (2006) studies the location choice and entry behavior in video rental industry by specifying a Bayesian game. Our considerations are two fold. First, the simultaneous-move framework adopted to model entry decisions in the I.O. literature is better understood to describe the market structure in long run equilibrium, rather than the actual entry behaviors. Therefore, the unobserved profitability—which could be the possible sources of asymmetric information—should be viewed as the “trend” of the unobserved profitability. Therefore, we believe the informational asymmetry regarding the profitability among the potential entrants is not an important issue in our application. Secondly, from an identification point of view, we would have to impose strong assumptions on the information structure if we want to model the asymmetric information, otherwise the observed choices are not sufficient for us to separately identify the parameters in the profit function and those in the information structure. Based on the above two considerations, we model our entry game

as one of complete information.

The major departure of our paper from the previous literature is that we assume chains' entry and capacity decisions are determined by some correlated equilibrium, rather than impose the more restrictive solution concept of Nash equilibrium, which assumes the strategic independence and precludes pre-play communications among players. Using correlated equilibrium in the current context has several advantages. First, and most importantly, while modeling the communication activities directly would pose many challenges, the solution concept of correlated equilibrium provides an indirect and elegant approach to incorporating pre-play communication into the entry game without having to pay attention to the fine details of the communication activities. (Recall the revelation principle in section 3.3.)

Secondly, the correlated equilibrium set is mathematically straightforward to characterize. It is a well known fact that simultaneous-move games of complete information typically have multiple equilibria. Hence, in order to put theoretically consistent restrictions on data, we have to find all the equilibria (including the mixed strategy equilibria)—or at least characterize the common properties across all the possible equilibria—for each game with deterministic payoffs corresponding to each possible realization of the covariates and econometric errors. One may then use these restrictions to *set* estimate the model (Ciliberto and Tamer, 2006; Yang, 2007¹⁴). Depending on the research question to be answered, one may choose to further impose some equilibrium selection rule to be able to *point* estimate the model (Bajari, Hong and Ryan, 2006). Whichever of the above two approaches is warranted by the research question, finding a mathematically tractable characterization of the equilibrium set remains key. However, characterizing the Nash equilibrium set over and over again for a game having nontrivial scale is computationally challenging. In contrast, exactly characterizing the correlated equilibrium set is computationally trivial. We will explain these issues more clearly when we discuss the estimation approach in section 5.

¹⁴He uses the relationship between Nash and correlated equilibrium to impose restrictions on data and obtain the bounds estimates of the parameters in static games of complete information, where players are assumed to behave according to some Nash equilibrium.

5 Estimation

For each market m , the econometrician observes the entry and capacity decisions of the chains, $(a_i)_{i \in I} \equiv a$, and the number of independent hotels, N_{sm} ¹⁵, as well as the market size variables X_m . The econometrician seeks to make inferences about the underlying parameters θ by assuming all the players behave according to some equilibrium of the structural model as specified above. The natural approach, which is a standard one in estimating structural models, is first to compute the conditional choice probabilities $\Pr(a|X_m, \theta)$ ¹⁶ that are implied by the equilibrium conditions of the structural model and then to proceed with either Maximum Likelihood (MLE) Estimator or General Method of Moments (GMM) approach, where the conditional choice probabilities $\Pr(a|X_m, \theta)$ are defined as follows:

$$\Pr(a|X_m, \theta) = \int_{S(\epsilon_m)} \mu_m(a|X_m, \theta, \epsilon_m) \cdot F(d\epsilon_m), \quad (4)$$

where $\epsilon_m \equiv (\eta_m, \nu_{Hm}, \nu_{Lm}, \nu_{sm})$ denotes the vector of unobservables for market m ; $F(\cdot)$ and $S(\epsilon_m)$ stand for the distribution and support of ϵ_m , respectively; and $\mu_m(\cdot|X_m, \theta, \epsilon_m)$ denote the correlated equilibrium (to be discussed below) for the game corresponding to $(X_m, \theta, \epsilon_m)$.

There are, however, two potential complications that may render it impossible to compute (4). First, the structural model specified in last section almost surely has multiple correlated equilibria, which implies that $\mu_m(a|X_m, \theta, \epsilon_m)$ is not a well-defined function of the covariates, parameters and unobservables, but instead a correspondence. Secondly, even if these conditional probabilities are a well-defined function, the integral in (4) typically does not have a closed-form solution. To tackle these two issues, we 1) impose some economically reasonable selection rule that helps pin down a unique correlated equilibrium out of the equilibrium set; and 2) apply the simulation methods developed by McFadden (1989) and Pakes and Pollard (1989) to obtain an estimate of the conditional probabilities $\Pr(a|X_m, \theta)$. In the following, subsections 5.1 and 5.2 characterize the correlated equilibrium set and then select a particular correlated equilibrium using some selection rules; Subsection 5.3 describes the estimation

¹⁵Since we independent hotels are symmetric, the number of these hotels choosing to enter is a sufficient statistic of the entry decisions of all the independent hotels who are potential entrants.

¹⁶Since the number of independent hotels N_{sm} is uniquely determined by the chains' entry and capacity decisions, we suppress N_{sm} in the expression.

algorithm, with a special attention paid to the simulation methods.

5.1 Characterization of the Equilibrium Set

In order to compute the conditional probabilities $\Pr(a|X_m, \theta)$, we shall first compute the equilibrium set of the two-stage game as specified in section 4. Before characterizing the equilibrium set for this game, let us formally define the equilibrium notion of this game. To make explicit its dependence on $(X_m, \theta, \epsilon_m)$, we denote this two-stage game by $\Gamma^e(X_m, \theta, \epsilon_m)$ ¹⁷. Furthermore, let $A \equiv \times_{i \in I} A_i$ represent the set of all action profiles for the chains, with a generic element $a = (a_i)_{i \in I}$, and $\Delta(A)$ denote the set of probability distributions over the set A .

A correlated equilibrium of the game $\Gamma^e(X_m, \theta, \epsilon_m)$ then consists of a probability distribution $\mu_m \in \Delta(A)$ together with a function¹⁸ $N_{sm} : A \rightarrow \mathbb{N}$ mapping from the set of chains' action profiles A to the set of nonnegative integers \mathbb{N} , such that the following conditions hold:

1. For any possible action profile $a = (a_i)_{i \in I}$ taken by the chains, $N_{sm}(a)$ is the maximum integer number such that an independent hotel's profit $\pi_{sm}(a, N_{s,m}) \geq 0$, where the dependence of π_{sm} on the other variables as well as on the parameters has been suppressed. If for all $N_{s,m} \in \mathbb{N}$, each independent hotel earns profit $\pi_{s,m}(a, N_{s,m}) < 0$, then the number of independent hotels $N_{s,m}(a)$ equals 0;
2. Chains have no incentives to deviate from whatever action profile realized according to the distribution μ_m . That is, for $\forall i \in I, \forall a_i \in A_i, \forall d_i \in A_i$,

$$\sum_{a_{-i} \in A_{-i}} \mu_m(a) \{ \pi_{im}[a_{-i}, a_i, N_{sm}(a_{-i}, a_i)] - \pi_{im}[a_{-i}, d_i, N_{sm}(a_{-i}, d_i)] \} \geq 0, \quad (5)$$

where we suppress the dependence of π_{im} on the other variables or the parameters.

The equilibrium conditions above are defined by backwards induction. Condition 1 describes the equilibrium entry behavior for small, independent hotels, who are assumed to be symmetric

¹⁷Following the convention in game theory literature, we use Γ to denote normal-form games and Γ^e to denote extensive-form games.

¹⁸In this paper, we use $N_{s,m}$ to denote both the function that describes the competitive fringes' strategy and the value (*i.e.*, the number of independent hotels that enter the market) this function takes.

and decide whether or not to enter after observing chains' decisions. This condition requires that, for a market with N_{sm} small hotels, each of them earns nonnegative profits $\pi_{sm}(a, N_{sm}) \geq 0$, but with one extra small hotel, they will each earn a negative profit $\pi_{sm}(a, N_{sm} + 1) < 0$. Furthermore, for markets without any small hotels, condition 1 requires that $\pi_{sm}(a, N_{sm}) < 0$ for any nonnegative integer N_{sm} . Note that $N_{sm}(a)$ can be easily calculated $\forall a \in A$.

Condition 2 is the equilibrium condition for correlated equilibrium. To understand the incentive compatibility requirements as stated in the inequality system (5), imagine a mediator who randomly draws an action profile a of entry decisions, from the set A with probability $\mu_m(a)$. The mediator then tells each chain *privately* to make the entry decision as specified in the action profile a . Based on the probability distribution μ_m and the instruction given to it, each chain can calculate its expected payoff from obeying or disobeying the instruction. Hence, a correlated equilibrium μ_m is defined as a probability distribution such that no player has an incentive to deviate from the instruction given to him according to μ_m .

Of the above two conditions, condition 2 is key. In the discussions that follow, we will focus on the incentive compatibility requirements as stated in (5) by treating $N_{sm}(\cdot)$ as given. Note, however, that generically there are multiple correlated equilibria since there are multiple μ'_m 's that satisfy the inequality system (5). In a companion paper, Yang (2007) uses (5) to form inequality restrictions on the data and then looks for bounds estimates of the parameters. The approach taken in this paper is a little different, since our goal for the structural modeling and estimation is to obtain the point estimates of the parameters and then simulate the counterfactual where the chains make entry and capacity decisions independently by playing some Nash equilibrium. Finally, observe that the correlated equilibrium set is convex, and that the system in (5) is linear in μ_m , a property that makes characterizing the correlated equilibrium set computationally trivial compared to the Nash equilibrium set. We will exploit the convexity and linearity extensively below. Next subsection will consider several different selection mechanisms that enable us to pin down one particular correlated equilibrium.

The profit function in (2) for the chains exhibit certain symmetry. In other words, if two chains choose to enter with the same capacity level, then their expected entry profits are the same. Therefore, it is reasonable to focus on symmetric correlated equilibria. That is, for any given

market, we assume that two market configurations a and a' are assigned the same probability whenever they entail the same number of low- and high-capacity entrants. More formally, $\mu(a) = \mu(a')$ if and only the following holds:

$$\sum_{i \in I} 1\{a_i = H\} = \sum_{i \in I} 1\{a'_i = H\}, \quad \sum_{i \in I} 1\{a_i = L\} = \sum_{i \in I} 1\{a'_i = L\}.$$

In our structural model, there are a total of 6 chains, each of which has three actions available. Without imposing the symmetry assumption, a correlated equilibrium μ would be a vector of dimension equal to $3^6 = 729$. However, after imposing the symmetry requirement, we can reduce the number of dimensions drastically to $\frac{(6+1)(6+2)}{2} = 28$.

Let N_L and N_H denote the number of low- and high-capacity chain establishments determined by market configuration a , then by this symmetry assumption, we can write the conditional choice probabilities in (4) as follows:

$$\Pr[(N_{Lm}, N_{Hm}) | X_m, \theta] = \int_{S(\epsilon_m)} \mu_m[(N_L, N_H) | X_m, \theta, \epsilon_m] \cdot F(d\epsilon_m), \quad (4')$$

where N_{Lm} and N_{Hm} are the number of the low- and high-capacity chain hotels in market m , respectively. We will base our estimation on (4').

5.2 Equilibrium Selection

Ever since the seminal work by Bresnahan and Reiss (1990, 1991), economists have recognized the nontrivial challenge posed by multiple equilibria to the structural estimation of game-theoretic models. Also, since Berry (1992), economists have proposed various methods to pick a most “reasonable equilibrium” out of the equilibrium set. For example, Berry (1992) suggests that more profitable firms may move first, and hence focus on the subgame perfect equilibrium associated with this order. Bajari, Hong and Ryan (2006) explicitly model the equilibrium selection mechanism and try to identify it from data. They model the equilibrium selection mechanism as such that an equilibrium is more likely if it is trembling-hand perfect, is in pure strategies, or maximizes industry profit.

There are, however, no generally accepted rules that help select a correlated equilibrium—the game theory literature on the selection of correlated equilibrium is much less developed than

that on the selection of Nash equilibrium. Following Schelling (1960), we believe that some correlated equilibria are nevertheless more “focal” than others. One such a focal point, for example, is the Pareto-efficient frontier of the correlated equilibrium set. In the current context, hotel chains are allowed to communicate with each other before making their entry decisions. Hence, it is rather reasonable to assume that chains never play a correlated equilibrium that is Pareto dominated by another correlated equilibrium. By focusing on the Pareto-efficient frontier of the correlated equilibrium set, we can drastically reduce the possible correlated equilibria being played.

However, in order to obtain the conditional choice probabilities $\Pr [(N_{Lm}, N_{Hm}) | X_m, \theta]$ in (4'), we have to impose further selection criteria to help us pin down a unique correlated equilibrium among the Pareto-efficient frontier. We consider two alternative selection criteria. The first one is that each correlated equilibrium in the Pareto-efficient frontier is selected uniformly (*i.e.*, with equal probability). The second possible selection rule pins down the particular correlated equilibrium that maximizes the expected sum of the players' entry profits.

Either of the above mechanisms selects a particular correlated equilibrium from the Pareto-efficient frontier. Furthermore, note that selection rule 1 allows less coordination among chains than rule 2 does. In the actual estimation, we estimate the structural model under each of these two alternative selection rules.

Computing one particular correlated equilibrium under rule 1 is straightforward, as we can first construct the Pareto-efficient frontier, and then select a correlated equilibrium randomly according to the uniform distribution on the Pareto-efficient frontier. Selection rule 2 can also be implemented using linear programming. Specifically, for a game corresponding to $(X_m, \theta, \epsilon_m)$, we look for a symmetric correlated equilibrium $\mu_m [\cdot | X_m, \theta, \epsilon_m]$ that maximizes the following expected sum of the chains' entry profits:

$$\max_{\mu_m \in \Delta(A)} \sum_{i \in I} \left[\sum_{a \in A} \mu_m [N_{Lm}(a), N_{Hm}(a) | X_m, \theta, \epsilon_m] \pi_{im}(a; X_m, \theta, \epsilon_m) \right].$$

5.3 An MSM Estimator

In principle, we could use either maximum likelihood estimator (MSL) or the method of simulated moments (MSM) to estimate the parameters θ . One advantage of MSM over MSL is that it provides an unbiased and consistent estimate for a fixed number of simulation draws. Hence, in this paper, we use the MSM approach. The MSM approach is similar to GMM except that we have to replace the “exact” moment functions used in GMM by the corresponding simulated ones.

To see how GMM would work in our setting, note that we would first need to come up with a vector of moment functions $g(X_m)$ such that:

$$E[g(X_m, \theta_0)] = 0.$$

We then minimize a weighted quadratic form of the sample analogue:

$$\min_{\theta \in \Theta} \left[\frac{1}{M} \sum_{m=1}^M g(X_m, \theta) \right]' A \left[\frac{1}{M} \sum_{m=1}^M g(X_m, \theta) \right], \quad (6)$$

where A is a positive definite weighting matrix.

The key, then, is to find such moment functions $g(\cdot)$. In our setting, we could obtain such $g(\cdot)$ by interacting the prediction errors of the model and the exogenous variables, where the prediction errors are the differences between the model predicted probability of observing a market configuration and its empirical counterpart. Specifically, a market configuration is described by a pair of nonnegative integers (N_L, N_H) , where N_L and N_H denote the number of low- and high-capacity establishments affiliated with one of the top six chains.¹⁹ Since $N_L + N_H \leq 6$, we have a total of $\frac{(6+1)(6+2)}{2} = 28$ possible market configurations. If we enumerate all the possible market configurations by $k = 1, 2, \dots, 28$, then the k^{th} moment function, which corresponds to the k^{th} market configuration, is:

$$g^k(X_m, \theta) = [1 \{(N_{Lm}, N_{Hm}) = k\} - \Pr(k|X_m, \theta)] \oplus f(X_m), \quad (7)$$

where (N_{Lm}, N_{Hm}) denotes the actual market configuration in market m , $1\{\cdot\}$ is an indicator function, and $f(X_m)$ is any function of the exogenous data.

¹⁹Recall that as soon as N_L and N_H are determined, the number of small hotels determined as well, since small hotels are Stakelberg followers.

For the estimation, we cannot obtain a closed-form expression for the model predicted probability of observing the k^{th} configuration, $\Pr(k|X_m, \theta)$, $k = 1, \dots, 28$. We can, however, replace $\Pr(k|X_m, \theta)$ by its simulated counterpart, $\widehat{\Pr}(k|X_m, \theta)$. To do this, for each of the S rounds of simulations, $s = 1, \dots, S$, we randomly draw the following four independent vectors from the standard normal distribution: a vector of market-wide structural error $\{\eta_m^{(s)}\}_{m=1}^M$, and three vectors of type-specific errors: $\{v_{Lm}^{(s)}\}_{m=1}^M$, $\{v_{Hm}^{(s)}\}_{m=1}^M$, and $\{v_{sm}^{(s)}\}_{m=1}^M$. During each round of simulation, we obtain the simulated profits and solve for the probability of the market configuration. We then take the average over these S simulation draws to obtain $\widehat{\Pr}(k|X_m, \theta)$. Finally, we obtain an unbiased estimate $\widehat{g}^k(X_m, \theta)$ of the k^{th} moment function by replacing $\Pr(k|X_m, \theta)$ with $\widehat{\Pr}(k|X_m, \theta)$ in (7).

During the estimation procedure, we evaluate the objective function (6) using $\widehat{g}^k(X_m, \theta)$, and search for the parameter values that minimize the objective function. In the search over the parameter space, we use the same set of simulation draws for all values of θ . For the weighting matrix, we first obtain a preliminary estimate for θ by setting A equal to the identity matrix, and then compute the optimal weighting matrix based on this preliminary estimate. In minimizing the objective function in (6), we use the bounds estimates as suggested in Yang (2007) to set the starting values.

6 Estimation Results and Counterfactual Experiment

6.1 Parameter Estimates

In the estimation we include four explanatory variables to proxy the market size in the profit functions of the chain and independent hotels: log of population, log of retail employment, log of employment in gasoline stations, and a dummy variable indicating whether the market contains popular tourist attractions. The inclusion of population is standard in structural entry models and is meant to capture the pure size effect, since there are usually more hotel establishments in a market with larger population. Retail employment proxies for the size of the retail industry, which is closely related to the demand for hotel service in a market. Following a similar approach adopted by Chung and Kalnins (2001), we use employment in

gasoline stations in a market to represent the source of demand from highway traffic flow passing through the market. Finally, some markets might contain various tourist attractions and this would drive up the demand for hotel services. We take this source of demand into consideration by including the tourism dummy into the set of explanatory variables.

Table 7 presents the parameter estimates under the two alternative selection rules that can be used to pick a specific correlated equilibrium. There are three sets of parameters: $\beta_H, \delta_{HH}, \delta_{HL}, \delta_{Hs}, \rho_H$ measure the effects of demand-side variables as well as competition on the profits of a chain that chooses to enter with high capacity; $\beta_L, \delta_{LH}, \delta_{LL}, \delta_{LS}, \rho_L$ and $\beta_s, \delta_{sH}, \delta_{sL}, \delta_{ss}, \rho_s$ are the corresponding sets of coefficients for low-capacity chain establishments and independent hotels, respectively. Most of the coefficients are statistically significant and have the expected signs. Among the coefficients for the demand-side variables, log employment in gasoline stations is not significant for chain-affiliated establishments with high capacity, somehow suggesting that highway traffic flow does not significantly affect such establishments' profitability. The coefficients measuring the competition effects are significant and economically intuitive, except for the effects of independent hotels on the chains' profitability, which is not statistically significant. Among the three coefficients (ρ_H, ρ_L, ρ_S) that are used to measure the relative importance of unobserved market-specific vs. firm-specific factors, both ρ_H and ρ_L are statistically significant.

We report the estimation results under both selection rules. Recall that, under the first selection rule, any Pareto-efficient correlated equilibrium is selected with equal probability; and under the second selection rule, the correlated equilibrium that maximizes the expected sum of chains' entry profits is selected from the correlated equilibrium set. The parameter estimates for all four market size variables are, however, very similar across these two selection rules. The difference in estimates for the competition effects is rather small as well. This finding suggests that the parameter estimates are more or less robust to either of these two selection rules used, which suggests that the set of Pareto-efficient frontier itself might be enough to identify the model parameters. In the paragraphs that follow, we will discuss and interpret these parameter estimates corresponding to the first selection rule.

Population has a significant and positive effect on the entry profits for all the three types

of hotels. However, the relative magnitude of this effect across different types implies that a chain that plans to enter a market with larger population size will tend to choose high capacity, other things equal. In the hotel industry, establishments with higher capacity might be able to provide more amenities and better services to consumers. Therefore, the results here suggest that larger markets tend to have more high-quality hotels, which is consistent with Sutton's theory (Sutton, 1991, 1998). The interpretation for the effects of the other two demand-side variables—retail employment and the employment in gasoline stations—is similar, even though the coefficients associated with the latter are not precisely estimated for chain establishments. The tourism dummy has a significantly positive impact on the entry profits for both types of chain establishments, and for independent hotels as well.

The δ -parameters displayed in Table 7 reveal the competitive effects among the firms. First of all, all the δ 's are estimated to be negative, consistent with the intuition that entry profits tend to decline when other firms enter the market as well. The difference in size among these δ 's, however, suggests a significant heterogeneity among the competitive effects. For a chain that chooses to enter with a high capacity, the entry of another high-capacity chain establishment has approximately twice (-3.0459 vs. -1.5083) the impact compared to that of a low-capacity chain establishment, while the entry of an independent hotel has no statistically significant impact. The inspection of the middle bottom part of the table reveals a similar pattern—a low-capacity chain is negatively affected more by a rival that also chooses to enter with low capacity. These parameter estimates suggest that chains have strong incentives to differentiate among themselves in terms of the capacity.

Finally, the three ρ -parameters— ρ_H , ρ_L , and ρ_S —tell us the relative importance of unobserved market-specific vs. type-specific profit shocks for high-capacity chains, low-capacity chains, and independent hotels, respectively. The estimates for these three parameters are somehow counterintuitive, since, *a priori*, we would expect independent hotels to be relatively more likely to be affected by market-wide shocks. But the estimates of these three ρ -parameters suggest that this is not the case: ρ_S is significantly smaller than both ρ_H and ρ_L (0.5579 vs. 0.6639 and 0.7235).

6.2 Effects of Communication on Market Structure: Counterfactual Analysis

One of the advantages of structural modeling and estimation is that it enables us to simulate the counterfactual market configurations without being subject to the Lucas critique (Lucas, 1976). In this subsection, we return to the primary motivation of the paper and measure the impact of pre-play communication on market configuration among chain hotels. To achieve this goal, we first use the data and the parameters obtained above to compute the market configurations in the counterfactual world, where potential entrants make entry and capacity decisions independently without communicating with each other, and then compare them with the actual configurations where we assume that potential entrants communicate with each other before making entry and capacity decisions. The differences in market configurations across these two scenarios inform us about the magnitude of the effect of communication on market structure.

Specifically, let the profit functions be as specified earlier and the parameters therein take the values as reported in last subsection. We now consider two regimes. In the true regime, potential entrants play a *correlated equilibrium*, and hence make their entry decisions after communicating with each other. In the counterfactual regime, potential entrants play a *Nash equilibrium*, and therefore make their entry decisions independently. To see how market structures differ in these two regimes, we simulate the equilibrium market configurations for the 437 markets in our sample under both regimes. In so doing, we use the parameter estimates and the market condition variables to compute the expected number of high- and low-capacity chains under both regimes.

There are, however, two potential obstacles in performing the simulation. First, during the simulation, some games may have multiple equilibria. We tackle this issue as follows. Under the true regime, we choose a correlated equilibrium randomly from the Pareto-efficient frontier of the correlated equilibrium set, as we have done in the estimation exercise under the first selection rule; Under the counterfactual regime, we choose the Nash equilibrium that exhibits the highest level of coordination among the Nash equilibrium set, *i.e.*, the one that maximizes

the industry profits among the Nash equilibrium set.²⁰ Hence, our comparison will provide a lower bound for the impact of communication on market configuration. The second obstacle is that the Nash equilibrium selected in the counterfactual regime or the correlated equilibrium in the true regime may be a mixed-strategy equilibrium. If this is the case, we shall compute the expected number of high- or low- capacity chains predicted by the equilibrium, with the expectation taken under the distribution induced by the equilibrium.

Table 8 presents the summary statistics for the market configurations under both the counterfactual and true regimes. In the left panel of Table 8, we report the average number of low- and high-capacity establishments affiliated with the six largest chains, as well as the average number of independent hotels. The average number of low-capacity establishments affiliated with the top 6 chains across the 437 markets under the counterfactual regime is equal to 0.63, while this number is 0.65 under the true regime. The average number of high-capacity establishments under the counterfactual and true regimes are 0.92 and 0.90, respectively. On average there are 2.03 independent establishments under the true regime, lower than its counterpart in the counterfactual regime (3.02). But overall, the average number for each of the three types of establishments does not significantly differ across the true and counterfactual regimes. The right panel of Table 8 breaks down the 437 markets by the number of top 6 chain-affiliated establishments in each. The frequency of each possible category listed in the right panel of Table 8, however, does not significantly differ across the two regimes.

To take a closer look at the difference in market configurations across the two regimes, we break down the markets further according to the capacity distribution in each market. Table 9 presents the conditional probability for each possible market configuration. The configurations in each row share the same number of firms, $N_H + N_L$, and each column correspond to the set of configurations that have the same value for $N_H - N_L$. The two numbers (The one in parenthesis corresponds to the counterfactual regime.) in each cell represent the proportions of markets corresponding to this cell, conditional on $N_H + N_L$. For example, under the counterfactual regime, across all the markets where there are 2 out of the top 6 operating, the

²⁰Due to the symmetric structure of the entry game in our setting, computing the Nash equilibrium set is made significantly easier. We also make the computation parallel, since we have already obtained the parameter estimates.

proportions of the markets with $N_H - N_L = 2, 0, -2$ equal 0.30, 0.48, and 0.22, respectively. Let us compare these conditional probabilities across the two regimes for $N_H + N_L = 2, 3, 4, 5, 6$. In each of these five cases, more balanced market configurations—the market configurations with $N_H - N_L$ close to 0—appear more frequently in the true regime than in the counterfactual regime. If we interpret a more balanced structure as exhibiting a higher level of coordination among the entrants in making capacity choices, the results in Table 9 suggests that pre-entry communication can help firms achieve greater coordination in choosing capacities.

Tables 8 and 9 together suggest that pre-entry communication does not have significant impact on the number of firms operating in a market, but does affect the within-market capacity distribution, conditional on the number of firms. As the final part of our counterfactual analysis, we calculate the average profit for both low- and high-capacity hotels under the two regimes, based on Equations (2) and (3). Table 10 reports the results of such calculation. The results indicate that, if pre-entry communication were not allowed, the average profit for a high-capacity chain establishment would decrease from 1.29 to 1.08, or by 16.28%. The average profit for a low-capacity chain-affiliated establishment would be reduced by 8.57% if no pre-entry communication is allowed.

7 Conclusion

This paper has adopted a structural approach to examine the impact of pre-play communication on the market structure in the context of lodging industry. We identify such impact by comparing the hotels' entry behavior under correlated equilibrium with that under Nash equilibrium. Since a correlated equilibrium incorporates pre-play communication, while Nash equilibrium does not, such comparison can tell us the impact of pre-play communication. We first obtain the parameter estimates of the entry game based on correlated equilibrium, and then simulate the counterfactual where each potential entrant has to make decisions simultaneously and independently by playing Nash equilibrium. The comparison across the true and counterfactual regimes suggest that pre-entry communication does not have a significant impact on the number of firms in a market. However, the communication does seem to in-

crease the within-market differentiation in capacity. In addition, pre-play communication also increases average profit for both high- and low- capacity establishments.

The methodology developed in this paper can be applied to many contexts where pre-play communication is important in reducing strategic uncertainty. For example, one could use a similar approach to study the impact of communication among radio stations, where they make both the entry and format decisions. Furthermore, in the current paper, there is no asymmetric information about the payoff structure among the players. In other settings such as auctions, there is asymmetric information among bidders and they may communicate with each other prior to bidding. It would be interesting to see whether the method used in this paper can be extended to those settings as well.

References

- Akerberg, D., and G. Gowrisankaran (2006), "Quantifying Equilibrium Network Externalities in the Ach Banking Industry," UCLA and Washington Working Paper.
- Andrews, D., S. Berry, and P. Jia (2005), "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria with An Application to Discount Chain Store Location," Yale University Working Paper.
- Aumann, R. (1974), "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1, 67-96.
- Bajari, P., H. Hong, and S. Ryan (2006), "Identification and Estimation of a Discrete Game of Complete Information," University of Minnesota Working Paper.
- Bajari, P., H. Hong, J. Kraner, and D. Nekipelov, "Estimating Static Models of Strategic Interactions," University of Minnesota Working Paper.
- Berry, S. (1992), "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 60, 889-917.
- Bjorn, P., and Q. Vuong (1984), "Simultaneous Equations Models for Dummy Endogenous Variables: A Game-Theoretic Formulation with Application to Labor Force Participation," Caltech Working Paper.
- Bresnahan, T., and P. Reiss (1990), "Entry in Monopoly Markets," *Review of Economic Studies*, 57, 531-553.
- Bresnahan, T., and P. Reiss (1991), "Empirical Models of Discrete Games," *Journal of Econometrics*, 81, 57-81.
- Christensen, L., and R. Caves (1997), "Cheap Talk and Investment Rivalry in the Pulp and Paper Industry," *The Journal of Industrial Economics*, 45, 47-73.
- Chung, W., and A. Kalnins (2001), "Cheap Talk and Investment Rivalry in the Pulp and Paper Industry," *Strategic Management Journal*, 22, 969-988.
- Chwe, M. (2004), "Statistical Game Theory" Department of Political Science, U.C.L.A. Work-

ing Paper.

Ciliberto, F., and E. Tamer (2006), "Market Structure and Multiple Equilibria in Airline Markets," University of Virginia and Northwestern University Working Paper.

Farrell, J. (1987), "Cheap Talk, Coordination, and Entry," *The Rand Journal of Economics*, 18, 34-39.

Hart, S., and D. Schmeidler (1989), "Existence of Correlated Equilibria," *Mathematics of Operations Research*, 14,18-25.

Heckman, J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931-959.

Kalnins, A. (2007), "The U.S. Lodging Industry," *Journal of Economic Perspectives*, 20, 203-218

Jia, P. (2006), "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retail Industry," M.I.T. Working Paper.

Kuhn, K. (2001), "Fighting Collusion by Regulating Communication Between Firms," *Economic Policy*, 16, 169-204

Mazzeo, M. (2002), "Product Choice and Oligopoly Market Structure," *Rand Journal of Economics*, 32, 1-22.

McLennan, A., and R. McKelvey (1996), "The Computation of Equilibrium in Finite Games," *Handbook of Computational Economics*, Vol.1, 87-142, Elsevier.

Murty, K. (1976), *Linear and Combinatorial Programming*, John Wiley & Sons.

Myerson, R. (1982) "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," *Journal of Mathematical Economics*, 10, 67-81.

Myerson, R. (1986), "Acceptable and Predominant Correlated Equilibria," *International Journal of Game Theory*, 15, 133-154.

Myerson, R. (1991), *Game Theory*, Harvard University Press, Cambridge, MA.

Nau, R., S. Canovas, and P. Hansen (2003), "On the Geometry of Nash Equilibria and Correlated Equilibria," *International Journal of Game Theory*, 32, 443-453.

Reiss, P. (1996), "Empirical Models of Discrete Strategic Choices," *American Economic Review*, 86, 421-426.

Seim, K. (2001), "Spatial Differentiation and Market Structure: the Video Rental Industry," Yale University Ph.D. Dissertation.

Sposito, V. (1989), *Linear Programming with Statistical Applications*, Iowa State University Press, Iowa.

Tamer, E. (2003), "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70, 147-165.

Table 1: The U.S. Lodging Industry and Its Economy and Midscale Sectors

Year	Sales (Billion)		No. of Establishments (Thousand)	
	Industry Total	Sectors	Industry Total	Sectors
	2000	62.16	17.76	45.86
2006	78.72	31.92	47.21	21.58

Source: Smith Travel Research. The sales Revenue are in 2000 US dollars

Table 2: Affiliation and Number of Establishments

Affiliation	Number of Establishments
Best Western	175
Holiday Inn	123
Comfort Inn	109
Super 8	99
Days Inn	94
Motel 6	64
Other Chains (27)	164
Independents	1140
Total	1968

Note: the other chains (27 in total) are not listed in the table, and there are a total of 164 establishments affiliated with these chains

Table 3: Capacities of the Top 6 Chains

Affiliation	Mean Capacity	Min Capacity	Max Capacity
Best Western	60	27	193
Holiday Inn	90	40	245
Comfort Inn	62	36	126
Super 8	58	26	150
Days Inn	66	24	152
Motel 6	93	40	169

Table 4: Market Configuration of the Top 6 Chains(I)

	No. of Markets	Percent of Total Markets
6	21	4.8
5	21	4.8
4	24	5.5
3	29	6.6
2	52	11.9
1	103	23.6
0	187	42.8
Total	437	100

Note: This table breaks down the markets by the number of top six chains (see Table 2 for the list of these chains) operating in each market.

Table 5: Market Configuration of the Top 6 Chains(II)

	High Capacity						
Low Capacity	0	1	2	3	4	5	6
0	187	25	5	5	1	2	1
1	76	22	9	8	3	6	-
2	25	11	9	4	4	-	-
3	4	4	5	7	-	-	-
4	1	2	2	-	-	-	-
5	4	1	-	-	-	-	-
6	0	-	-	-	-	-	-

Note: This table breaks down the markets in the data by the number of low- and high- capacity chain affiliated establishments (our of the top six). For example, the number 25 corresponding to the cell (0,2) means that there are 25 markets (out of 437) in which a single chain with high capacity is operating.

Table 6: Summary Statistics

Variable	Mean	Std.	Median	Minimum	Maximum
Population (Thousand)	17.43	29.75	5.87	1.02	21.99
Employ-retail (Thousand)	2.44	4.60	0.64	0.00	38.92
Employ-gas (Thousand)	0.30	0.40	0.15	0.00	2.67
Tourism	0.09	0.29	0.00	0.00	1.00
No. of Top 6 Chain Establishments	1.42	1.76	0.00	1.00	8.00
No. of Independent Establishments	2.61	3.55	2.00	0.00	27.00

Source: Population is from the 2000 U.S. census data; Employment by the retail sector as well as by the gas stations is obtained from the Zip Business Patterns Database. The tourism dummy is constructed based on the information provided by the Texas State Data Center.

Table 7: Parameter Estimates

Variable	High-Capacity Establishments		Low-Capacity Establishments		Independents			
	Selection 1	Selection 2	Variable	Selection 1	Selection 2	Variable	Selection 1	Selection 2
Log(Pop)	0.72 (0.37)	0.82 (0.30)	Log(Pop)	0.37 (0.18)	0.73 (0.14)	Log(Pop)	0.54 (0.12)	0.77 (0.40)
Log(Emp-ret)	0.33 (0.19)	0.13 (0.20)	Log(Emp-ret)	0.25 (0.34)	0.33 (0.15)	Log(Emp-ret)	0.30 (0.08)	0.32 (0.09)
Log(Emp-gas)	0.16 (0.19)	0.24 (0.23)	Log(Emp-gas)	0.25 (0.23)	0.33 (0.12)	Log(Emp-gas)	0.34 (0.10)	0.45 (0.20)
Tourism	0.27 (0.14)	0.45 (0.24)	Tourism	0.20 (0.00)	0.22 (0.19)	Tourism	0.22 (0.38)	0.33 (0.13)
Constant	3.20 (0.19)	0.31 (0.21)	Constant	2.52 (1.03)	2.23 (1.12)	Constant	1.98 (0.87)	2.09 (0.78)
δ_{HH}	-3.05 (0.12)	-1.22 (0.44)	δ_{LH}	-1.20 (0.33)	-1.30 (0.44)	δ_{sH}	-0.10 (0.07)	-0.12 (0.31)
δ_{HL}	-1.51 (0.70)	-2.28 (0.82)	δ_{LL}	-0.22 (0.10)	-0.27 (0.11)	δ_{sL}	-1.30 (0.05)	-1.50 (0.06)
δ_{Hs}	-0.09 (0.31)	-1.13 (2.85)	δ_{Ls}	-0.10 (0.30)	-0.16 (0.02)	δ_{ss}	-2.33 (0.33)	-2.45 (0.31)
ρ_H	0.66 (0.31)	0.50 (0.23)	ρ_L	0.72 (0.39)	0.82 (0.40)	ρ_s	0.56 (0.45)	0.61 (0.50)
Function Value	70.32	85.19						

Note: Standard errors are in parentheses. They are obtained based on simulation. $\delta_{j,k}$ denotes the competition effect of type k on type j , where j, k can be type H, L , or s .

Table 8: Summary Statistics for True and Counterfactual Configurations

Regime	Mean No. of		Frequency of Top 6 Chain Establishments							
	L-Capacity Establishments	H-Capacity Establishments	Independent Establishments	0	1	2	3	4	5	6
True	0.65	0.90	2.03	165	113	46	35	34	26	18
Counterfactual	0.63	0.92	3.02	160	115	53	32	31	22	24

Note: For the true regime, I use the parameter estimates (reported in Table 7) and the data to compute the expected market configuration for each market, under the assumption that potential entrants behave according to correlated equilibria. The first row reports the summary statistics of the results under the true regime. I do the same thing for the counterfactual regime, except that the potential entrants now behave according to Nash equilibria. The second row displays the summary statistics of the results under the counterfactual regime.

Table 9: True and Counterfactual (in Parentheses) Configurations

	6	5	4	3	2	1	0	-1	-2	-3	-4	-5	-6
$N_H - N_L$	-	-	-	-	-	-	1	-	-	-	-	-	-
$N_H + N_L$	-	-	-	-	-	-	(1)	-	-	-	-	-	-
0	-	-	-	-	-	0.25	-	0.75	-	-	-	-	-
1	-	-	-	-	-	(0.30)	-	(0.70)	-	-	-	-	-
2	-	-	-	-	0	-	0.51	-	0.49	-	-	-	-
3	-	-	-	-	(0.30)	-	(0.48)	-	(0.22)	-	-	-	-
4	-	-	-	0.12	-	0.32	-	0.48	-	0.08	-	-	-
5	-	-	-	(0.21)	-	(0.29)	-	(0.40)	-	(0.10)	-	-	-
6	-	-	0.08	-	0.24	-	0.44	-	0.16	-	0.08	-	-
	-	-	(0.12)	-	(0.30)	-	(0.50)	-	(0.05)	-	(0.03)	-	-
	-	0.07	-	0	-	0.27	-	0.53	-	0.07	-	0.07	-
	-	(0.03)	-	(0.21)	-	(0.23)	-	(0.22)	-	(0.12)	-	(0.19)	-
	0.05	-	0.09	-	0.09	-	0.41	-	0.27	-	0.09	-	0
	(0.05)	-	(0.08)	-	(0.02)	-	(0.34)	-	(0.30)	-	(0.11)	-	(0.10)

Note: This table breaks down the markets by the capacity distribution in each market for both regimes. The columns represent the differences between the numbers of high- and low-capacity firms in the market. The two numbers in each cell represent the proportions of markets corresponding to this cell, conditional on $N_H + N_L$. For example, for markets with a total of 3 firms, there are four possible capacity configurations: $N_H - N_L = 3$; $N_H - N_L = 1$; $N_H - N_L = -1$; $N_H - N_L = -3$. Under the true(counterfactual) regime, the conditional probabilities (conditional on $N_H + N_L = 3$) for these four configurations are 0.12(0.21), 0.32(0.29), 0.48(0.40), and 0.08(0.10), respectively.

Table 10: Average Profits Under True and Counterfactual Regimes

Regime	H-Capacity Establishments	L-Capacity Establishments
True	1.29	0.70
Counterfactual	1.08	0.64