1-1-2012

Empirical Studies Of Revenue Management Practices: Understand Your Competition And Customers

Jun Li
University of Pennsylvania, lijun1@wharton.upenn.edu

Empirical Studies Of Revenue Management Practices: Understand Your Competition And Customers

Abstract
This dissertation empirically examines factors that challenge revenue management practices in travel industry --- air-travel and lodging. In particular, it focuses on strategic interactions among firms and strategic interactions between firms and customers. While most traditional revenue management focuses on single firm problems, better understanding competition and customers recently become two emerging themes in both theories and practices of revenue management. Meanwhile, with 20 years' successful implementation of sophisticated revenue management systems in both airline and hotel industries, they have accumulated rich data to better understand threats and opportunities currently facing both industries. Furthermore, with the proliferation of online distribution channels, extensive information has been made available to both customers and competitors. How to utilize such opportunity to understand customers and competition remains a question to both industry professionals and academic researchers.

This dissertation contains three parts. The first part studies implications of strategic alliances in the airline industry. Airlines in the same alliance are competitors and partners at the same time. After alliances are formed, airlines' networks are expected to be consolidated and capacity redundancies would be eliminated, as intensity of competition decreases among alliance partners. However, we find that alliance partners seek to overlap more in their networks. We also find evidence that average prices increase by about $11 per one-way segment coupon in markets where two partners are both present. After ruling out other plausible competing mechanisms, we conclude that these findings are most likely driven by multimarket competition. The second part of the dissertation studies travelers' strategic decision to delay purchases in anticipation of price decreases when purchasing air-tickets. By estimating a structural model on booking and posted fare data, we find that 4.9% to 44.9% of the population are strategic, and that incorporating such strategic customer behavior will increase revenues by 3% to 5% in certain city-pair markets. The third part of the dissertation bridges the two themes by applying a consumer-centric lens to better understand competition in hotel industry. Using online search and clickstream data, we propose a methodology to identify key competitors based on which hotels customers have compared. This approach also provides a network view of localized competition structure. We also find that there is approximately 50% mismatch between competition sets perceived by customers and hoteliers. Independent hotels and distant hotels are usually left out of competition sets.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Operations & Information Management

First Advisor
Gerard P. Cachon

Keywords
Competition, Consumer Behavior, Empirical Research, Revenue Management

Subject Categories
Business | Business Administration, Management, and Operations | Management Sciences and Quantitative Methods | Operational Research

EMPIRICAL STUDIES OF REVENUE MANAGEMENT PRACTICES: UNDERSTAND YOUR COMPETITION AND CUSTOMERS

Jun Li

A DISSERTATION

in

Operations and Information Management

For the Graduate Group in Managerial Science and Applied Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2012

Supervisor of Dissertation

____

Gérard P. Cachon, Fred R. Sullivan Professor of Operations and Information Management

Graduate Group Chairperson

____

Eric T. Bradlow, Professor of Marketing, Statistics, and Education

Dissertation Committee

Serguei Netessine, Timken Chaired Professor of Global Technology and Innovation, INSEAD

Gérard P. Cachon, Fred R. Sullivan Professor of Operations and Information Management

Sergei Savin, Associate Professor of Operations and Information Management

Nelson Granados, Assistant Professor of Information Systems, Pepperdine University

To My Family

Acknowledgements

First, I would like to thank my advisor Serguei Netessine for his guidance, support, and most importantly, trust, without which this thesis would not have been possible, and without which the duration of my doctoral study would not be as joyful and productive. I would also thank Gerard Cachon and Sergei Savin for providing valuable support, raising thoughtful and challenging questions and being inspiring examples. Thanks to Nelson Granados for his knowledgable input throughout and especially in the second study of this dissertation. I am also very grateful to all corporate sponsors, who would remain anonymous, for providing data and insights to studies involved in this dissertation.

I would take this opportunity to thank my parents WANG Chengfang and LI Xiangdong for being loving and supportive and for always holding faith in me.

I would also express my appreciation for my colleagues at Penn — faculty, staff and all my fellow doctoral students at the OPIM department, and many colleagues who provided valuable comments during presentations and discussions. I am grateful to my friends — Vibhanshu Abhishek, Alessandro Arlotto, Hessam Bavafa, Andres Catalan, John Cui, Pnina Feldman, Santiago Gallino, Jose Guajardo, Caiyan Li, Toni Moreno, Jaelynn Oh, Fazil Pac, Fangyun Tan, Necati Tereyagoglu, Dan Yang, and all my fellow dancers at Penn Chinese Dance Club, who had made my past five years in Philadelphia most enjoyable period of my life. Many thanks to WU Di for holding my hands and believing in me.

ABSTRACT

EMPIRICAL STUDIES OF REVENUE MANAGEMENT PRACTICES: UNDERSTAND YOUR COMPETITION AND CUSTOMERS

Jun Li

Gerard Cachon

This dissertation empirically examines factors that challenge revenue management practices in travel industry — air-travel and lodging. In particular, it focuses on strategic interactions among firms and strategic interactions between firms and customers. While most traditional revenue management focuses on single firm problems, better understanding competition and customers recently become two emerging themes in both theories and practices of revenue management. Meanwhile, with 20 years' successful implementation of sophisticated revenue management systems in both airline and hotel industries, they have accumulated rich data to better understand threats and opportunities currently facing both industries. Furthermore, with the proliferation of online distribution channels, extensive information has been made available to both customers and competitors. How to utilize such opportunity to understand customers and competition remains a question to both industry professionals and academic researchers.

This dissertation contains three parts. The first part studies implications of strategic alliances in the airline industry. Airlines in the same alliance are competitors and partners at the same time. After alliances are formed, airlines' networks are expected to be consolidated and capacity redundancies would be eliminated, as intensity of competition decreases among alliance partners. However, we find that alliance partners seek to overlap more in their networks. We also find evidence that average prices increase by about $11 per one-way segment coupon in markets where two partners are both present. After ruling out other plausible competing mechanisms, we conclude that these findings are most likely driven by multimarket competition. The second part of the dissertation studies travelers' strategic decision to delay purchases in anticipation of price decreases when purchasing air-tickets. By estimating a structural model on booking and posted fare data, we find that 4.9% to 44.9% of the population are strategic, and that incorporating such strategic customer behavior will increase revenues by 3% to 5% in certain city-pair markets. The third part of the dissertation bridges the two themes by applying a consumer-centric lens to better understand competition in hotel industry. Using online search and clickstream data, we propose a methodology to identify key competitors based on which hotels customers have compared. This approach also provides a network view of localized competition structure. We also find that there is approximately 50% mismatch between competition sets perceived by customers and hoteliers. Independent hotels and distant hotels are usually left out of competition sets.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

This dissertation investigates factors that challenges revenue management practices in the travel industry (air-travel and lodging industries). Particularly, it focuses on strategic interactions among firms, and strategic interactions between firms and consumers. In the past 20 years, revenue management has become an active area, but there are relatively few empirical studies in this field and considerable gap between theories and practices. However, those industries most sophisticated in revenue management, i.e., airline and hotel, also present rich opportunities for empirical research, as automated electronic systems have been keeping good record of various types of data including transaction, price and inventory data. Meanwhile, development of online distribution channels recording consumer data can potentially be combined with the above mentioned datasets as well.

Started from the inventory control problem for one single firm (Belobaba 1989, Gallego and van Ryzin 1994), practices and studies of revenue management are moving towards incorporating competition (Netessine and Shumsky 2005, Perakis and Sood 2006) and customer choices (Vulcano et al. 2010, Shen and Su 2007, Netessine and Tang 2009) into the big picture. This dissertation consists of three parts revolving exactly around these two aspects: competitive environment and consumer behaviors in revenue management practices in air-travel and lodging industries. Part 1 looks at how competition have been altered by strategic partnership in the airline industry and its long-run revenue implications. Part 2 investigates the presence of strategic consumers and its implications on revenue strategies in the air-travel industry. Part 3 bridges the two focuses by applying a consumer-centric lens to better understand the competitive environment in the lodging industry. I summarize hereafter the key findings from each of these three empirical studies.

1.2 Summary of Results

1.2.1 Partnering with Competitors — An Empirical Analysis of Airline Alliances and Multimarket Competition

Competition has become an important theme in the operations management literature and, according to recent theoretical and empirical work, the key finding is that firms tend to overstock or overproduce under competition. Following this prediction, one would expect that, after airlines start a multifaceted collaboration by forming an alliance, their networks would be consolidated and capacity redundancies would be eliminated, as intensity of competition decreases among alliance partners. Surprisingly, we find exactly the opposite: in the post-alliance era, alliance partners seek to overlap their networks more and they increase capacities on the markets in which two partners are already present. At the same time, average prices in those markets increase by about \$11 per one-way segment coupon. We explain these results using predictions based on the theory of multimarket competition: as firms seek out opportunities to establish multimarket contact to strengthen mutual forbearance, they have incentives to increase overlap even though this decision may not seem optimal or efficient locally or in the short term. We examine other plausible competing mechanisms built on theories of capacity and service competition and commonly cited benefits of airline alliances but ultimately we conclude that our findings are most likely driven by the multimarket competition. This paper therefore underscores the importance of going beyond simple bilateral competition models whose predictions may not hold when firms compete operationally in multiple markets, a phenomenon which is widespread in many operations-intensive industries.

1.2.2 Are Consumers Strategic? Structural Estimation from the Air-Travel Industry

Many consumers have learned to delay purchases, anticipating that prices might decrease. Such *strategic* or *forward-looking* behavior has attracted increasing attention from various disciplines, including operations management, information systems, marketing, and economics. However, there is currently no empirical evidence of the extent to which this strategic decision-making actually takes place. Combining two unique data sources from the air-travel industry (posted fares data and booking data), we use a structural model to estimate the fraction of strategic consumers in the population, assuming different levels of sophistication in consumers' perception of future prices: perfect foresight and rational expectations. We find that 4.9% to 44.9% of the population are strategic across markets, measured by the 5th and 95th percentiles. Using a non-parametric approach, we further find that most strategic consumers arrive either at the beginning of the booking horizon or close to departure. Finally, our counterfactual analysis shows that, contrary to conventional wisdom, the presence of strategic consumers does not necessarily hurt revenues. Rather, the impact varies by market — more likely to be negative on business markets, and positive on leisure markets. As a result, commitment to a non-decreasing pricing strategy benefits business markets but not leisure markets. Among markets benefiting from this pricing strategy, the median revenue improvement is 3.5%, and the quartiles are 1.8% and 5.6%.

1.2.3 Who are My Competitors? Let the Customer Tell

Identifying competition set is a challenging question to many competitive industries – hotel industry being one of them. Despite its commonly agreed importance to strategy development and daily operations, little data-driven analytical approach has been developed to address this question. We propose in this paper a simple and intuitive methodology to identify true competitors from customers' perspective using online search and clickstream data. We build up a competition network based on customer comparison of hotels. To illustrate the value of a customer-centric approach in competitive analysis, we further contrast the customer-based competition network with a hotelier-based competition network developed from price matching patterns. We found that there is about 50% mismatch between the two networks, and that hoteliers tend to ignore independent and distant hotels while at the same time they tend to pay too much attention to branded hotels and hotels in close proximity. This result is robust to many alternative measures of competition. Finally, we note that the proposed methodology can easily be applied to many other industries to aid business in identifying their key competitors and to enrich our understanding of networked competition.

Chapter 2

Partnering with Competitors — An Empirical Analysis of Airline Alliances and Multimarket Competition

2.1 Introduction

In March 2002, American Airlines raised the advance purchase requirement of discounted business-travel tickets from three days to seven days, which is equivalent to an estimated 10% price increase. However, most airlines refused to follow this suit except for Continental. As a result, American rolled back the fare increase in most markets, and shot back by putting \$199 one-way fares in 10 markets flown by Northwest, United, Delta and US Airways each, while excluding Continental from this revenge. In turn, Northwest fought back by offering similar fares in 20 markets flown non-stop by American, which triggered another round of fight where American expanded its cheap fares to 20 Northwest markets, and Northwest escalated the war to 160 American markets[1]. This type of price wars is not an exception but rather a rule in the airline industry. For instance, Busse (2002) identified 31 major price wars for the 14 largest airlines during the period of 1985-1992 using Wall Street Journal Index. A common feature of these fare wars is that multiple markets are generally involved, as airlines fight back not only in the local market where the price war is initiated, but also in multiple markets where they compete with the focal rival. The consequence of this multimarket retaliation is significant: Morrison and Winston (1996) find that a particular price war lead to a 32.4% price decrease, on average. As a result,

> *[Firms that compete against each other in many markets] may hesitate to fight vigorously because the prospects of local gain are not worth the risk of general warfare.... A prospect of advantage from vigorous competition in one market may be weighed against the danger of retaliatory forays by the competitor in other markets.* — Edwards (1955)

The above examples are just the tip of the iceberg. Ever since the deregulation in 1978, competition has increased dramatically in the airline industry, often manifesting

[1] Latest airfare battle turns into "street fight"; Fliers win as struggling airlines duke it out. USA Today. 19 March 2002.

itself through entries and price cuts which have significantly reduced the profitability of this industry and led to many bankruptcies. In response, airlines initiated a wave of alliance partnerships in the late 1990s and the early 2000s, aimed at achieving efficiencies and synergies through collaboration on frequent flier programs, lounges, facility utilization, information technology and procurement.

According to operations management theories of firms competing on operational decisions such as inventory, capacity or production levels (e.g., Lippman and McCardle 1997), competition results in overstocking, overproduction and overbuilding of capacity. Therefore, as cooperation among alliance partners increases, one would expect that they would consolidate flight networks and reduce capacity redundancies in markets in which they both operate. While competition within an alliance is tempered, competition between alliances accelerates: Christian Klick, a Star Alliance vice president, commented that,

> *Competition used to be strictly between airlines, but competition is really happening between alliances now*[2].

As a result, we would expect airlines to compete more vigorously with competitors from different alliances through aggressive entries and capacity (over)investment.

Surprisingly, what we find in this paper is *exactly* the opposite: in the post-alliance era, airlines are more likely to operate and install higher capacity in the markets where

[2]The Middle Seat: Shopping for Perks Among the Big Airline Alliances. Wall Street Journal. (Eastern Edition). New York, N.Y.:Jul 8, 2010. p. D.1.

their alliance partners are also active, and less likely to operate and install capacity in those markets where non-allied carriers are active. This result is robust to multiple alternative model specifications. We also find that airlines charge higher prices in markets in which they operate with partners. This effect is highly economically significant: in a typical duopoly market after an alliance, airlines are able to charge a $11 premium on a one-way segment coupon in markets shared with partners rather than with competitors. Both of these results are also contrary to what airlines claim would happen after the alliance and these results also support the general concerns of policy makers: "[allied airlines might] compete less aggressively in price or capacity in overlapping markets"[3].

Although surprising, our findings are consistent with predictions of the multimarket contact theory. In order to impose a credible threat of retaliation, airlines establish and strengthen their multimarket contact. By strategically overlapping and increasing capacity on routes where their partners are already present, airlines strengthen their ties with partner airlines and solidify the partnership. On the other hand, alliances also facilitate the process of building multimarket contact. Note that airlines are still competing with their partners — an alliance is not a merger, but it is in between of a merger and a perfectly competitive environment. While one would normally expect firms to consolidate their activities after merging, effects of alliance on operation and capacity decisions are more subtle. The fact that firms compete with each other in multiple markets in this industry further complicates what we would expect to see from competing

[3]Transportation Research Board report. Entry and Competition in the U.S. Airline Industry: Issues and Opportunities. 1999.

airlines with and without an alliance partnership. Our work therefore underscores the importance of incorporating the perspective of the multimarket competition into operations management models to offer new and additional insights, and it calls for more empirical research to understand the intricacies of multimarket competition. We also draw attention of the regulators and industry managers to the effects of airline alliances that damage consumers through higher prices, the actions that airlines explicitly deny before alliances are approved by the regulators.

2.2 Literature and Hypotheses Development

Competition has become an important theme in the operations management literature in the past two decades. Studies in this field investigated effects of competition on inventory and production (Lippman and McCardle 1997), on supply chain coordination (Cachon 2001), on the joint decisions for inventories and prices (Zhao and Atkins 2008), on technology decisions and capacity investment (Goyal and Netessine 2007), on service quality (Allon and Federgruen 2009), and on pricing strategies (Perakis and Sood 2006), just to name a few. One rather general finding of this literature is that firms behave suboptimally under competition, that is, firms tend to overstock or overproduce under competition as compared to the centralized scenario. Lippman and McCardle (1997), Mahajan and van Ryzin (2001), Netessine and Rudi (2003) demonstrate this result for newsvendors who compete on inventory or production levels given exogenous retail prices. The problem becomes more complicated when firms compete on both inventory and prices. Zhao and Atkins (2008) show that competition leads to higher inventory stock level and lower retail prices. Netessine and Shumsky (2005) study airline revenue management competition and show that more seats are protected for higher-fare passengers under competition. An empirical study by Cachon and Olivares (2009) confirms the aforementioned predictions using automotive dealership inventory data — they find that competition leads to higher service level or equivalently higher stocking level at dealerships. A few recent empirical studies examine the trade-offs between prices and service quality in service competition (Allon et al. 2011, Guajardo et al. 2011, Buell et al. 2011). Based on these theoretical and empirical findings, we would expect to observe capacity consolidation and network segregation among alliance partners, as the degree of competition among partners decreases post-alliance. We therefore form our first hypothesis as follows:

Hypothesis 1A*: In the post-alliance era, airlines are more likely to reduce overlap and decrease capacity in the markets in which their alliance partners possess strong market power, as compared to markets dominated by competitors engaged in alternative alliances.*

As we noted in the introduction, one important feature of the airline industry is that firms compete in multiple markets. Unfortunately, there is a dearth of research in operations management on the topic of multimarket competition and the majority of the literature that we cite models a single competitive market, or it sidesteps issues around multimarket competition. At the same time, multimarket competition is quite common in practice (Greve and Baum 2001). It is the norm in multiple industries — airlines, telephone and cable, banks and retail chains, to name a few. As we will show shortly, incorporating the multimarket competition perspective may offer new insights into operational strategies. As firms seek out opportunities to establish multimarket contact to strengthen mutual forbearance, "a situation in which two firms understand each other's motives and strategies and implicitly coordinate to avoid competing intensely" (Jayachandran et al. 1999), they have incentives to seek overlap even though this decision may not seem optimal or efficient in the local market or in the short run.

Dating back to Edwards (1955) and Bain (1956), multimarket contact has been an active area of research in industrial organization economics and strategy fields. Multimarket competition is generally considered to increase mutual forbearance and temper rivalry, and a high market concentration is necessary (if not sufficient) for tacit collusive behaviors such as mutual forbearance. This view is largely supported by the empirical evidence. Studies found that multimarket contact leads to higher prices, greater profits and more stable competition structure, that is, lower rates of entry and exit (see Jayachandran et al. 1999 for a review). For instance, Baum and Korn (1996) find that multimarket contact is associated with lower entry and exit rates using data from California-based commuter air carriers from 1979 to 1984. Gimeno and Woo (1996) find that multimarket contact significantly increases prices. Most of the studies in this literature, however, take multimarket contact as an independent variable and focus on the effects of multimarket contact on outcome variables such as prices, profits and market turnovers. However, relatively little is known about the evolution of multimarket contact and the corresponding mechanisms through which multimarket contact is established. In fact, the rise of multimarket contact is itself a dilemma in the sense that "[In order to deter aggressive actions by rivals], firms must enter each other's markets [first], which is just the kind of action that the deterrent is supposed to limit" (Stephan et al. 2003).

Our paper recognizes alliances as a potential facilitator of this process. Alliances reinforce the credibility of the "friendly intension" of initial entries, which would otherwise be seen as an aggressive action and hence induce fierce retaliation. The two main processes through which multimarket contact enables mutual forbearance are: familiarity and deterrence (Jayachandran et al. 1999). That is, when firms are familiar with the capabilities and strategies of their rivals, or when firms are able to prevent their rivals from initiating aggressive actions, mutual forbearance is enhanced. Note that an alliance may also increase familiarity and facilitate deterrence among its members. Alliance members are commonly involved with activities such as facility and personnel sharing, joint marketing programs, reciprocal frequent flier programs, and joint purchasing, etc., which offer opportunities for alliance members to become familiar with each other's capabilities and strategies. As they enhance interactions, members have strong incentives to restrain from aggressive behaviors, or otherwise they may be punished in a number of ways. However, a priori we are not certain whether alliances serve as a substitute or a complement to multimarket contact. The most closely related research that we are able to identify is that firms with multimarket contact are more likely to collaborate on R&D partnerships (e.g., Scott 1988). Based on the literature on multimarket competition, we form the following competing hypothesis:

Hypothesis 1B: *In the post-alliance era, airlines are more likely to increase overlap and capacity in the markets in which their alliance partners possess strong market power, as compared to markets dominated by competitors engaged in alternative alliances.*

As an aside, if the last hypothesis were true, this result would also reconcile with the inverted U-shaped effect of multimarket contact on entries found in the literature (e.g., Baum and Korn 1999). It would also concur with (Greve 2006) in that "firms appeared to avoid entry into markets in which the competitive reactions of the incumbents were unpredictable", as alliance partnership reduces this unpredictability. As suggested by this literature, concentration is critical to sustain mutual forbearance (e.g., Busse 2002, Jayachandran et al. 1999). Hence, in our analysis we will examine not only the presence of an alliance partner, but also its market share measured by passenger traffic volume, while at the same time controlling for the overall market concentration.

The two competing hypotheses offer completely opposite predictions regarding airline operational strategies after alliances. Which one has a more compelling support is an empirical question. By bridging the two streams of theories in operations management and strategy, our main contribution is to provide empirical evidence on how multimarket competition affects operational decisions. Our paper also contributes to the burgeoning empirical literature of operations and revenue management in the airline industry, an industry which has accumulated rich data in the past decades but received relatively little attention from empiricists. Recent papers have examined airline flight operations such as delays and cancelations (Li et al. 2010, Arikan and Deshpande 2010), capacity utilization

(Ramdas and Williams 2009, Cho et al. 2007) and revenue management practices (Cho et al. 2007, Vulcano et al. 2010, Newman et al. 2010). Among these topics, airline network structure starts to capture interest of researchers. Arikan et al. (2010) develop a stochastic model to measure the propagation effect of flights delays through the aviation network. Network-based revenue management is gaining popularity both theoretically (Talluri and van Ryzin 2004) and practically. Challenging problems arise as a result of airline alliances, such as maximizing the total revenue of the combined networks of partners and designing incentive-compatible revenue sharing schemes (Wright et al. 2010, Hu et al. 2010, Netessine and Shumsky 2005). While most of these papers examine a single airline network, multimarket competition has not been studied either theoretically or empirically.

Our findings also contribute to economics literature on airline alliances (Brueckner 2003, Armantier and Richard 2008) and airline entry (Berry 1992). For instance, Gayle (2008) find "conflicting" evidence of collusive behavior after domestic airline alliances. In general, economic literature on airline entry largely regards decisions across markets as independent and it has not considered implications of multimarket competition. Bajari et al. (2007) propose a two-stage algorithm to estimate the dynamic entry game, which is applied in Benkard et al. (2010) to simulate the long-term dynamics of the airline merger. Our approach to describing airline entry behavior is close to their first stage but our focus is on changes of equilibrium behavior before/after alliances and how they are associated with the identity of allied vs. non-allied players. Estimating dynamic games will not shed additional light on the question we aim to answer and is beyond the scope of this paper. In addition to entry, we also provide evidence on capacity decisions conditional on airlines' operating decisions.

2.3 Model

2.3.1 Entry Model

We model the segment presence, entry and exit using a Probit model. A segment is an airport-pair or a city-pair where airlines operate direct flights, and it is the most basic decision-making level for flight operation. An alternative approach would be to model origin-destination (O&D) presence, either through operations of direct or connecting flights. We choose segment over O&D because the presence decision at O&D level would involve the presence of multiple segments, and thus violating the independence assumption on observations. Although decisions at the segment level may also be correlated due to connecting possibilities, this correlation can be addressed more conveniently by controlling for positions of endpoints and of the segment in the airline networks. To this end, consider the following problem. A carrier, indexed by i, decides whether or not to operate a direct flight in a set of segments indexed by m, where $m = 1, 2, 3, ..., M$, and this decision is made at the beginning of every period of time (i.e., year). This decision is based on both the level of demand and of competition. Note that demand in a segment includes not only those passengers who travel on the O&D served directly by this segment, but also passengers who travel on connecting itineraries partially served by this segment. The level of competition is affected by incumbent carriers and potential entrants — not only by the level of overall competition but also by the identity of the competitors. Whether the incumbent is from the same or from a different alliance presumably makes a difference. For instance, after United formed a partnership with US Airways, United's strategy for a particular market might be affected by whether US Airways is currently operating in the focal market.

Specifically, consider the potential profit (which may include both immediate and long-term gains) y^*_{imt} from operating direct flights in segment m:

$$y^*_{imt} = X'_{im(t-1)}\beta + f\big(D_{it}, \mathit{PartnerShare}_{im(t-1)}, \mathit{CompetitorShare}_{im(t-1)}\big) + \lambda_t + \alpha_i + \epsilon_{imt}. \tag{2.1}$$

$X_{im(t-1)}$ represents the lagged control[4] variables for characteristics of the market and of the network. It includes 1) *segment features* such as distance, population and per-capita income of both end points, level of competition (including only direct flights), presence of low-cost carriers (LCC), level of congestion (i.e., the load-factor); 2) *network node features* (considering cities or airports as the nodes and connections between them as edges of the network) such as degree of centrality[5], competition level and LCC presence at both cities or airports; 3) *network edge features* such as connectivity (number of indirect paths) and level of competition at the city-pair or airport-pair level (See Table 2.1 for a complete description of variables included.). D_{it} is a $\{0, 1\}$ variable

[4] Using more lagged years does not add much explanatory power but causes collinearity problems. One year lag is also a common practice in related papers.

[5] One way to account for entries due to international connections is to include international gateways as a control. However, this variable is highly correlated with the degree of centrality.

indicating that carrier i is in an alliance partnership at time t if $D_{it} = 1$, 0 otherwise. $f\left(D_{it}, PartnerShare_{im(t-1)}, CompetitorShare_{im(t-1)}\right)$ denotes the effect of an alliance, and it will be dependent upon the partner's market share and competitor's market share, which we will elaborate shortly[6]. λ_t controls for the time trend that is common to all carriers (e.g., economic conditions). α_i controls for the time-invariant carrier effects. ϵ_{imt} is an idiosyncratic shock which is observable to decision-makers but not to econometricians. For now we assume that ϵ_{imt} are i.i.d. across i, m and t, and we will discuss and address the potential endogeneity concern subsequently.

Note that the underlying profit y^*_{imt} is not observable. Instead, what we observe is a $\{0, 1\}$ variable, y_{imt}, which indicates whether carrier i operates a direct flight on market m at time t. The relationship of the two variables can be formalized using the following threshold policy (Benkard et al. (2010)),

$$y_{imt} = \mathbf{1}\{y^*_{imt} \geq 0 | y_{im(t-1)} = 0\}, \tag{2.2}$$

$$y_{imt} = \mathbf{1}\{y^*_{imt} \geq -\gamma_{imt} | y_{im(t-1)} = 1\}. \tag{2.3}$$

where the threshold is higher for a potential entrant (Eq. 2.2) than for the incumbent (Eq. 2.3), that is, the potential entrant faces an entry barrier. Note that the threshold for the entrant is normalized to zero without loss of genearality. The threshold for the incumbents may also differ. As carriers with larger market power are usually more

[6] All major carriers that do not have codeshare partnership with the carrier are included as its competitors. We also tried an alternative modeling approach in which we define partner's and competitor's presence using {0,1} binary variable instead of using the market share but the results are consistent.

capable of surviving lower temporary profits, we allow the threshold to be dependent on the carrier's own market share $S_{im(t-1)}$, i.e., $\gamma_{imt} = \gamma S_{im(t-1)}$.

The data generating process can be summarized as follows,

$$
\begin{aligned}
y^*_{imt} &= \gamma S_{im(t-1)} + X'_{im(t-1)}\beta + f(D_{it}; i, m, t) + \lambda_t + \alpha_i + \epsilon_{imt}, && (2.4)\\
y_{imt} &= \mathbf{1}\{y^*_{imt} \geq 0\}. && (2.5)
\end{aligned}
$$

Now we take a closer look at the effect of the alliance and the identity of incumbents (partner vs. competitor). We suspect that the presence of a partner will affect the focal carrier's decision differently than the presence of a competitor. However, since the partner is not assigned randomly but chosen by airlines, one needs to be cautious about the potential *selection bias*: the fact that United chooses US Airways as a partner may reflect certain complementarities/similarities of their networks, e.g., for some reason they tend to receive correlated demand shocks in certain markets, which are not observable to econometricians. These possibilities make United more or less likely to operate in the market in which US Airways is present. If this is true, we would observe such correlation even before the alliance is formed between United and US Airways. To address this issue it is important to control for the intrinsic correlation (not induced by the alliance partnership) and this is done through the difference-in-difference approach.

$$f\left(D_{it}, PartnerShare_{im(t-1)}, CompetitorShare_{im(t-1)}\right) = \delta D_{it}$$
$$+ \delta_{p1} PartnerShare_{im(t-1)} * (1 - D_{it}) + \delta_{p2} PartnerShare_{im(t-1)} * D_{it}$$
$$+ \delta_{c1} CompetitorShare_{im(t-1)} * (1 - D_{it}) + \delta_{c2} CompetitorShare_{im(t-1)} * D_{it}, \qquad (2.6)$$

where δ is the direct effect of an alliance, δ_{p1} is the effect of the partner's market share on the entry probability pre-alliance, and δ_{p2} describes the same effect post-alliance. Similarly, δ_{c1} and δ_{c2} denote the effects of competitors' market share on the entry probability before and after an alliance[7], respectively. We are interested in the changes of the effects before and after an alliance is formed:

$$\begin{array}{rc} \textit{change of partner's effect} & \delta_{p2} - \delta_{p1}, \\ \textit{change of competitor's effect} & \delta_{c2} - \delta_{c1}, \\ \textit{difference-in-difference} & (\delta_{p2} - \delta_{p1}) - (\delta_{c2} - \delta_{c1},) \end{array}$$

where $\delta_{p2} - \delta_{p1}$ represents the change in the partner's influence on the carrier's entry decision after an alliance, and $\delta_{c2} - \delta_{c1}$ represents the change in competitors' influence. Ultimately, we want to know whether the changes have been different (in both direction and magnitude) for partners and competitors. A significantly negative estimate of the difference-in-difference term supports Hypothesis 1A, i.e., airlines reduce overlaps

[7] So far the partner/competitor effects are assumed to be the same for all carriers but we account for carrier-specific effects in the robustness test.

with partners after alliances, and a significantly positive estimate would support Hypothesis 1B, i.e., airlines increase overlaps with partners after alliances. Note that the *identification* of the difference-in-difference term comes from the fact that the change in probabilities of operating direct flights in a segment differs in two types of markets: those in which the partner is present vs. those in which competitors are present. The identification is not solely due to the variation of alliance status both over time and among carriers, although this adds further variation for identification purposes.

We summarize the model as follows:

$$
\begin{aligned}
y_{imt}^{*} &= \gamma S_{im(t-1)} + X'_{im(t-1)}\beta + \delta D_{it} + \delta_{p1} PartnerShare_{im(t-1)} + \delta_{c1} CompetitorShare_{im(t-1)} \\
&\quad + (\delta_{p2} - \delta_{p1}) PartnerShare_{im(t-1)} * D_{it} + (\delta_{c2} - \delta_{c1}) CompetitorShare_{im(t-1)} * D_{it} \\
&\quad + \lambda_t + \alpha_i + \epsilon_{imt}, \qquad (2.7) \\
y_{imt} &= \mathbf{1}\{y_{imt}^{*} \geq 0\}. \qquad (2.8)
\end{aligned}
$$

2.3.2 Capacity Model

We further investigate how capacity decision is adjusted after an alliance, conditional on the airline deciding to stay in the market. We want to see whether this adjustment in capacity differs for markets operated together with partners vs. competitors. We use a model that is similar to the entry model above but with appropriate changes:

$$
\begin{aligned}
K_{imt} = {} & \gamma S_{im(t-1)} + X'_{im(t-1)}\beta + \delta D_{it} \\
& + \delta_{p1} PartnerShare_{im(t-1)} + \delta_{c1} CompetitorShare_{im(t-1)} \\
& + (\delta_{p2} - \delta_{p1}) PartnerShare_{im(t-1)} * D_{it} + (\delta_{c2} - \delta_{c1}) CompetitorShare_{im(t-1)} * D_{it} \\
& + \lambda_t + \alpha_i + \epsilon_{imt}, \qquad (2.9) \\
\epsilon_{imt} = {} & \mu_{im} + \xi_{imt}, \qquad (2.10)
\end{aligned}
$$

where K_{imt} is the capacity measured as the logarithm of the carrier's number of seats supplied in the segment annually. The rest of the variables are as previously defined. Specifications of the error terms will be discussed in the endogeneity section that follows.

2.3.3 Endogeneity

Correlated Random Effects Model. Even though we use lagged market share among the explanatory variables and we control for as many relevant covariates as possible, lagged market share may still be correlated with the unobserved profitability. For instance, if profitability shocks are correlated over time, some markets may be more or less profitable for some specific carriers, or if demand and supply shocks are autocorrelated over time, lagged market share will still be correlated with the current profitability shock. In linear panel data models (such as the capacity model), this type of endogeneity is commonly addressed by allowing for correlation between the fixed effects and other covariates, by allowing for serial correlation in shocks, and by using lagged first-difference of independent variable as instruments. However, addressing this endogeneity is more complicated in nonlinear panel data models (such as the Probit entry model we present here). Chamberlain (1980) – Mundlak (1978) developed a Correlated Random Effects Probit Model to address endogeneity problems in dynamic nonlinear panel data models.

Following this classical approach, we decompose the error term into two parts: an unobserved carrier-market specific term and an idiosyncratic shock $\epsilon_{imt} = c_{im} + \varepsilon_{imt}$, where c_{im} can be regarded as the unobserved component of the carrier-market specific profitability shock. Traditional random effects model would require strict exogeneity $E(c_{im}|W_{imt}) = 0$, where W_{imt} represents all the explanatory variables. However, this assumption might be violated as airlines, based on their experiences in the market, may have some knowledge about the market profitability specific to itself, and furthermore, this shock may be correlated with its partner's or competitor's profitability in the same market and hence correlated with even the lagged market shares. That is, c_{im} can be correlated with the market share of the focal airline, its partners and competitors, biasing our estimation of the effects of partner's market share and competitor's market share. To allow for correlation between the carrier-market specific profitability shocks, the essential idea is to explicitly model the correlation between c_{im} and W_{imt} in a specific way (see Wooldridge (2010)),

$$c_{im} \sim \text{Normal}(\psi + \overline{W}_{im}\xi, \sigma_c^2), \tag{2.11}$$

where $\overline{W}_{im}$ is the average of W_{imt} over time. It turns out that this estimation can be done within the traditional random effects framework by adding $\overline{W}_{im}$ to the original estimation. By allowing this carrier-market specific effect, we account for the subject-specific heterogeneity (which can be endogenous).

Another way to allow for this correlation over time is to specify serial correlation through AR(1) process. Such models have been developed under the Generalized Estimating Equation framework (see Wooldridge (2010)). We will also present estimation results under this specification.

Similar endogeneity concerns arise in the capacity model as well but in the linear panel data model this endogeneity can be addressed more easily. To do so, one would want to use random effects and fixed effects models. One complication is that, once we include the lagged independent variable among explanatory variables, for example, we may want to include lagged capacity level in right-hand side of the capacity model, but the traditional fixed effects model will still produce biased estimates. The most recent approach to get around this issue is a GMM estimator proposed by Arellano and Bond (1991). We will also show the estimation result using the Arellano-Bond estimator to account for potential correlation between the error term and explanatory variables $K_{im(t-1)}, X_{im(t-1)}$.

Endogeneity Concerns regarding the Identification of the Difference-in-Difference Effect. The classical difference-in-difference identification strategy is based on a few implicit assumptions: 1) Without the alliance treatment, the effects of the partner's and competitors' share on the entry probability would have followed the same trend over time. To check robustness of the results, we allow for different trends by adding separate yearly shocks to the partner's effect (δ_{p1}) and the competitor's effect (δ_{c1}), which is the equivalent of adding interaction terms between yearly dummies and the partner/competitor market share. 2) Effects of partner's/competitors' market share are the same for all airlines. To check robustness, we allow for carrier-specific "attitudes" towards partners and competitors by including interaction terms between carrier dummies and partner/competitor market share. 3) Without the alliance, the change of partner's/competitors' effects would have followed the same trend for every carrier for both treated (allied) and untreated (non-allied) airlines. We relax this assumption by including a carrier-specific linear trend in the effects of partner's/competitor's market shares (similar to Besley and Burgess (2004)) which allow carriers to follow different trends in a limited but revealing manner.

2.4 Data

The principal data sources for our study are the Bureau of Transportation Statistics' T-100 Domestic Segment Data and Airline Origin and Destination Survey (DB1B) which we supplement with population and income data from the Bureau of Economic Analysis. The T-100 Domestic Segment Data provides quarterly information on seat capacity, number of enplaned passengers and the load-factor. The DB1B data is a 10% quarterly sample of all airline tickets in the United States, which includes price information. These are standard data sources for the closely related studies.

To accurately measure the impact of the treatment we need a proper time window. The span of our study runs from 1998 through 2006, which is equivalent to 8 years of data as we use lagged control variables. We choose this particular time period to balance the 'before' and after' periods around the major alliance events that took place in 2003.[8]. We do not use years far ahead because airlines' strategy might have changed over a long time-frame due to policy/technology/management/economy changes. Moreover, we estimated the decay of the effect using longer horizons, and found that one to two years after the alliance is the period in which most route adjustments are made[9]. Although quarterly data is available in our databases, we use yearly data because this is a more appropriate time-frame in the airline industry to make entry and exit decisions and yearly aggregation corrects for the seasonal effects. The carriers of interest are major domestic airlines including AA (American), UA (United), US (US Airways), CO (Continental), DL (Delta) and NW (Northwest). Since we use yearly data, a carrier is defined as present in a market if it operates a direct flight on the market throughout the year. We consider CO and NW in partnership starting from 1999 (officially approved in November

[8]We adjust for other major changes in the airline industry during the period of study. 1)*September 11 Effects.* We use year dummies to account for industry-wide effect and 9·11-UA/AA dummies. We also replace markets that experience temporary exit in 2002 and re-entry in 2003 as being active in 2002. 2)*Acquisition and merger.* American Airlines acquired Trans World Airlines in 2001. The routes taken over from Trans World are not counted as entries. Two national airlines, US Airways and America West, merged in 2005. However, America West continued reporting under its code until 2007. 3)*Bankruptcy.* All major airlines experienced financial difficulties from 2002 to 2004. Four filed for bankruptcy protection while continuing their operations. Although we could control for financial performances in the model, it does not affect the results as the financial shocks are generic to the airline overall, but they are not market-specific – there is no particular reason why markets operated with partners or competitors should be affected more.

[9]We also tried to extend the study to longer horizons and the results remained qualitatively unchanged.

1998), UA and US in partnership starting from 2003, and DL and CO, DL and NW in partnership starting from 2004[10].

We also made the effort to replace regional airlines by their parental major airlines. This modification is necessary because during the past decades major airlines gradually gave up direct presence in many smaller markets and instead contracted with regional partners to operate on these routes. This does not mean that the major airlines have ceased operations in these markets; they simply started operating in a more efficient way by utilizing smaller aircrafts in smaller markets. Moreover, consumers still buy tickets to these destinations under the brand name of the major airlines and the major airlines control pricing and revenue management systems of the smaller carriers. Without accounting for these shifts, we would have counted many more exits. Regional airlines accounted for no less than 20% of all the tickets sold. Details of this correction procedure are found in Appendix Table 5.1. Note that the same regional airline may have operated for different major carriers at different times of its history, e.g., Air Wisconsin started the transition from serving United Airlines to US Airways in 2003 when United Airlines filed for bankruptcy. Also note that the same regional airline may contract with two or more major airlines at the same time and even on the same markets. We use DB1B data to help us correctly identify these markets as well as the percentage of capacity contracted for each major airline. For example, if both major airlines A and B sell tickets on the same flight operated by the regional airline X, there can be two possibilities. One is that X only contracts with airline A, but airline B can also sell tickets on flights operated

[10] Officially approved in June 2003. The results are not sensitive to this specification

by X through codesharing agreements with A. The other possibility is that X contracts with both A and B. To distinguish these two cases, we look at the percentage of tickets sold on A and B. In the former case, A sells the majority of the tickets. In the latter case, A and B sell comparable portions. Practically, we use 80% as the dividing point. The results are not sensitive to this specification.

Following standard strategies used with this data (e.g., Benkard et al. 2010), we select the 75 largest U.S. airports, where size is defined by the enplaned passenger traffic. We then map the 75 airports to Metropolitan Statistical Areas (MSAs). We use Composite Statistical Area (CSA) or Metropolitan Division when necessary. For example, airports serving the New York area, JFK, LGA, EWR and ISP are grouped to New York-Newark-Bridgeport CSA. This grouping accounts for spatial correlation among these airports since airports close to each other usually have correlated demand and supply shocks. This grouping gives us 62 MSAs (see Appendix Table 5.2 for details). We supplement the airline data with annual population estimates and per-capita incomes for these MSAs from the Bureau of Economic Analysis.

We construct the set of markets containing all possible directional combinations[11] of these 62 MSAs, which gives us $3,782$ markets. Thus, we have a panel data of $22,696$ market-carrier dyads for 8 years. To account for occasional redirection of flights due to unforseen events such as severe weather, we only count an airline as operating the market in a particular year if it carried more than 36000 passengers in that market-year

[11]Results are similar when using non-directional market definitions.

(as in Benkard et al. (2010)), which roughly corresponds to one flight per day[12]. To focus on the most common types of trips, we place the following restrictions on the raw data to obtain average market fares, following Ito and Lee (2007) : 1) we restrict our analysis to round-trip, coach class tickets; 2) we limit our analysis to tickets with no more than two coupons per directional leg; 3) we exclude itineraries with fares per person less than $25 or greater than $1,500 since they might represent employee tickets or frequent flyer miles tickets or data errors; 4) we exclude itineraries on which the marketing carrier of either segment was a non-U.S. carrier. All of these are standard data transformations which are commonly used in the literature utilizing the same data sources.

Table 2.1 describes the key variables and Table 2.2 presents the summary statistics and the correlation information. The left-hand side variables include the operating status, capacity, average fare, traffic and the load-factor of each carrier on each market-year. The variables of interest include market shares (for the partner and for the competitor), the alliance status, and the interaction terms between these two sets of variables. The covariates included in this study fall into the following three categories: market characteristics, network nodal features, and network edge features. As in most other airline industry studies, we include market control characteristics such as distance, demographics at both endpoints, the level of competition and the low-cost carrier's market share. In addition, we believe that a key operational measure, i.e., the load-factor, is a critical indication in the airline entry decision and the price level, since the load-factor reflects

[12]We also tried different cut-off values such as 3600 as used in Berry (1992) and Borenstein and Rose. (1994), which corresponds to one regional jet per week. Our results are not sensitive to this cut-off value.

the congestion level of the market, carriers' operating costs, and their ability to manage demand uncertainty. Inspired by the network perspective adopted in the alliance formation literature, we add flight network features into this study[13]. The second category, i.e., network nodal features, include carriers' origin and destination degree, market share, level of competition and presence of low-cost carriers. Airlines' decisions to offer direct services on a market also depend on all the connecting possibilities through transferring at the origin and destination airports. In a hub-and-spoke network, origin and destination degree of centrality capture these possibilities, and this metric defines to a great extent the position of a route in the airline's entire network[14]. We choose degree centrality measures over hub-or-spoke measures since the former describe network positions more accurately, allowing for the emergence of subhubs during the time of study. The third category of covariates includes network edge measures, i.e., the number of one-stop connecting routes between the origin and the destination since airlines' entry decisions are also dependent on the alternative existing connecting services. Finally, we include measures of competition and low-cost carrier presence when accounting for these connecting possibilities. These network features have not been traditionally included in the related literature, though recently Benkard et al. (2010) started to adopt similar measures. We believe that inclusion of these measures is critical in recognizing that flight operation and capacity decisions are deeply embedded in the structure of the

[13]Flight networks are different from the relational networks in the management literature. Nodes are represented by airports in the former and by airlines in the latter.

[14]We also tried to include other types of network centrality measures, such as closeness centrality and betweenness centrality. However, these more sophisticated centrality measures did not add much value on top of degree centrality. We believe the reason is that in a hub-spoke network, degree centrality already contains most information.

entire network. We note that our summary statistics are consistent with those in related studies of the airline industry.

Table 2.1: Description of Variables

Category		Variables	Description
	(1)	own	{0,1} variable indicating whether the airline is operating in the market
	(2)	capacity	number of seats operated by operating airline
	(3)	passenger	number of passengers enplaned on operating airline's flights
	(4)	average fare	average one-way directional leg fare (obtained from itinerary fare, distance-weighted)
	(5)	alliance	{0,1} variable indicating whether the operating carrier has formed any alliance in a particular year
	(6)	own share	market share of the operating carrier on the directional leg (including only direct flights)
	(7)	partner share	market share of the alliance partner of the operating carrier on the directional leg
	(8)	competitor share	market share of the out-of-alliance competitors of the operating carrier on the directional leg
	(9)	partner share x alliance	interaction term of partner share and alliance
	(10)	competitor share x alliance	interaction term of competitor share and alliance
Segment Features	(11)	log(distance)	log of non-stop distance between origin and destination
	(12)	log sqrt(pop1 * pop2)	log of geometric average of population at the origin and the destination MSAs
	(13)	log(income)	log of average of per capita income at the origin and the destination MSAs
	(14)	loadfactor	average congestion level of the leg (total # of passengers / total # of seats)
	(15)	HHI	Herfindahl-Hirschman Index of the competition level on the leg
	(16)	lcc	market share of the low-cost carriers on the leg
	(17)	own loadfactor	congestion level of the operating carrier on the leg
Node Features	(18)	origin HHI	Herfindahl-Hirschman Index of the competition level at the origin
	(19)	dest HHI	Herfindahl-Hirschman Index of the competition level at the destination
	(20)	origin lcc	market share of the low-cost carriers at the origin
	(21)	dest lcc	market share of the low-cost carriers at the destination
	(22)	own origin degree	average of indegree and outdegree of the operating carrier at the origin city (degree is defined by the number of direct flights)
	(23)	own dest degree	average of indegree and outdegree of the operating carrier at the destination city (degree is defined by the number of direct flights)
	(24)	own origin share	market share of the operating carrier at the origin
	(25)	own dest share	market share of the operating carrier at the destination
Edge Features	(26)	# of indirect paths	number of connecting routes by the operating carrier on the O&D market
	(27)	O&D HHI	Herfindahl-Hirschman Index of the competition level at the O&D market (including both direct and connecting routes)
	(28)	O&D lcc	market share of the low-cost carrier at citypair market (including both direct and connecting routes)

Table 2.2: Summary Statistics and Correlation Table

		Mean	Std. Dev.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1)	own	0.078	0.268	1.00													
(2)	capacity	19982	97628	0.68	1.00												
(3)	passenger	14111	69826	0.67	0.99	1.00											
(4)	average fare	150.30	76.96	-0.07	-0.05	-0.05	1.00										
(5)	alliance	0.563	0.496	-0.03	-0.05	-0.04	-0.07	1.00									
(6)	own share?	0.632	0.320	0.88	0.60	0.60	-0.06	0.00	1.00								
(7)	partner share†	0.701	0.304	-0.12	-0.09	-0.09	-0.12	0.16	-0.14	1.00							
(8)	competitor share†	0.735	0.284	-0.10	-0.07	-0.07	-0.03	-0.11	-0.18	-0.18	1.00						
(9)	partner share * codeshare†	0.387	0.410	-0.09	-0.07	-0.07	-0.11	0.27	-0.10	0.72	-0.14	1.00					
(10)	competitor share * codeshare†	0.420	0.416	-0.09	-0.07	-0.07	-0.04	0.37	-0.12	-0.09	0.58	-0.05	1.00				
(11)	log(distance)	6.820	0.748	-0.12	-0.09	-0.07	0.55	0.00	-0.11	-0.09	-0.07	-0.07	-0.02	1.00			
(12)	log_sqrt(pop1x pop2)	14.384	0.665	0.20	0.26	0.25	0.05	0.00	0.07	0.08	0.29	0.04	0.23	0.10	1.00		
(13)	log(income)	10.394	0.128	0.10	0.11	0.11	0.01	0.28	0.05	0.08	0.12	0.15	0.20	0.08	0.44	1.00	
(14)	loadfactor†	0.508	0.335	0.27	0.18	0.19	-0.21	0.03	0.24	0.22	0.36	0.17	0.23	-0.16	0.33	0.18	1.00
(15)	HHI	0.615	0.385	-0.03	-0.06	-0.06	-0.05	-0.01	0.14	0.12	0.03	0.10	-0.03	-0.08	-0.19	-0.07	0.19
(16)	lcc	0.218	0.360	-0.11	-0.07	-0.06	-0.10	0.05	-0.16	-0.15	-0.22	-0.10	-0.11	0.06	-0.01	-0.05	0.06
(17)	own loadfactor†	0.726	0.104	0.99	0.68	0.68	-0.07	-0.03	0.87	-0.13	-0.11	-0.09	-0.09	-0.09	0.19	0.11	0.28
(18)	origin HHI	0.292	0.168	0.13	0.08	0.08	-0.02	0.00	0.17	0.13	0.12	0.09	0.04	-0.11	-0.04	-0.07	0.15
(19)	dest HHI	0.292	0.168	0.11	0.07	0.06	-0.05	-0.02	0.16	0.16	0.15	0.09	0.05	-0.13	-0.05	-0.07	0.16
(20)	origin lcc	0.289	0.191	-0.11	-0.09	-0.08	0.00	0.10	-0.12	-0.11	-0.17	-0.04	-0.05	0.17	-0.05	0.00	-0.07
(21)	dest lcc	0.289	0.191	-0.11	-0.08	-0.07	0.01	0.11	-0.11	-0.13	-0.19	-0.05	-0.07	0.18	-0.04	0.00	-0.08
(22)	own origin degree	15.748	16.312	0.42	0.39	0.38	0.02	-0.14	0.35	-0.09	0.00	-0.09	-0.05	0.01	0.30	0.15	0.17
(23)	own origin degree	15.748	16.312	0.42	0.38	0.38	0.02	-0.13	0.35	-0.10	0.00	-0.09	-0.05	0.00	0.30	0.15	0.15
(24)	own origin share	0.115	0.178	0.02	-0.01	-0.02	0.01	-0.07	0.06	0.02	0.03	0.00	0.01	-0.12	-0.03	-0.03	-0.05
(25)	own dest share	0.115	0.178	0.02	-0.01	-0.02	0.00	-0.07	0.05	0.03	0.06	0.00	0.02	-0.13	-0.03	-0.02	-0.03
(26)	# of indirect paths	0.778	1.263	0.02	0.03	0.03	0.10	-0.02	-0.01	0.01	0.10	-0.03	0.07	0.25	0.22	0.18	0.11
(27)	O&D HHI	0.345	0.270	0.18	0.13	0.12	-0.25	-0.01	0.24	0.20	0.23	0.14	0.11	-0.34	0.07	0.05	0.43
(28)	O&D lcc	0.182	0.257	-0.04	-0.01	0.00	-0.18	0.07	-0.12	-0.11	-0.15	-0.05	-0.05	-0.03	0.04	0.00	0.20
		Mean	Std. Dev.	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)
(15)	HHI	0.62	0.38	1.00													
(16)	lcc	0.22	0.36	-0.07	1.00												
(17)	own loadfactor†	0.05	0.19	-0.03	-0.11	1.00											
(18)	origin HHI	0.29	0.17	0.15	-0.19	0.12	1.00										
(19)	dest HHI	0.29	0.17	0.17	-0.21	0.10	-0.02	1.00									
(20)	origin lcc	0.29	0.19	-0.09	0.42	-0.10	-0.38	0.01	1.00								
(21)	dest lcc	0.29	0.19	-0.10	0.44	-0.09	0.01	-0.40	0.03	1.00							
(22)	own origin degree	15.75	16.31	-0.05	-0.03	0.41	0.00	-0.05	-0.17	0.06	1.00						
(23)	own origin degree	15.75	16.31	-0.06	-0.03	0.41	-0.04	-0.02	0.05	-0.16	-0.11	1.00					
(24)	own origin share	0.11	0.18	0.07	-0.16	0.01	0.17	0.01	-0.40	-0.04	-0.06	0.02	1.00				
(25)	own dest share	0.11	0.18	0.09	-0.16	0.01	0.00	0.21	-0.03	-0.40	0.02	-0.07	0.00	1.00			
(26)	# of indirect paths	0.78	1.26	-0.03	-0.01	0.03	-0.15	-0.13	-0.13	-0.10	0.08	0.07	0.20	0.19	1.00		
(27)	O&D HHI	0.34	0.27	0.36	-0.04	0.17	0.28	0.18	-0.19	-0.13	0.05	0.04	0.08	0.11	0.01	1.00	
(28)	O&D lcc	0.18	0.26	-0.13	0.77	-0.03	-0.16	-0.17	0.46	0.42	-0.01	0.01	-0.22	-0.22	-0.08	0.08	1.00

†: Summary statistics shown for these variables are conditional on operating.

2.5 Results

2.5.1 Estimation Results of the Entry Model

We begin our analysis by providing key summary statistics for exploratory purposes and in order to allow for initial understanding of the competitive landscape in the industry. Table 2.3 shows the entry and exit dynamics of the major airlines in the study period. We notice that, over the 8-year period, on average there has been approximately 10% turnover (entry/exit) each year. However, there was some turmoil in turnover right after the alliances were formed by most major airlines. United, US Air, Delta and Northwest all saw a sizable increase in the number of entries after the alliances in 2003 and 2004. One may suspect that this turmoil is caused by the events of September 11 and the financial crisis followed right after. If this is the driving force, the entire flight network should be affected and there seems to be no particular reason why markets previously operated together with certain airlines should be affected more or less than markets operated with others. To examine this issue, it is helpful to look into changes of overlap patterns with partners (or partners-to-be) vs. changes of overlap patterns with competitors over the same period of time. We provide relevant summary statistics in Table 2.4. Two airlines are considered to overlap in a segment if they both operate direct flights in it. United and US Airways have seen a notable increase in both the absolute number and the percentage of overlapping segments after their alliance, a change from 5.5% to 11.1%. Although the changes for Skyteam (Delta, Northwest and Continental) seem to be less obvious, a fair comparison is to contrast this with overlapping trends

Table 2.3: Airline Route Statistics 1999-2006

	UA				US				AA			
	stay	entry	exit	turnover	stay	entry	exit	turnover	stay	entry	exit	turnover
1999	282	3	4	0.024	338	18	9	0.078	255	26	2	0.109
2000	281	9	4	0.046	338	13	18	0.087	275	46	6	0.185
2001	278	5	12	0.059	331	1	20	0.060	297	11	24	0.109
2002	259	2	24	0.092	258	2	74	0.229	289	98†	19	0.380
2003	243	9	14	0.089	228	2	32	0.131	370	17	16	0.085
2004	246	25	6	0.123	221	25	9	0.148	334	19	53	0.186
2005	257	31	14	0.166	221	29	25	0.220	344	16	9	0.071
2006	264	10	24	0.118	241	17	9	0.104	346	7	14	0.058
	DL				CO				NW			
	stay	entry	exit	turnover	stay	entry	exit	turnover	stay	entry	exit	turnover
1999	388	12	12	0.060	209	11	5	0.075	234	8	2	0.042
2000	391	21	9	0.075	219	7	1	0.036	240	5	2	0.029
2001	380	5	32	0.090	226	11	0	0.049	244	4	1	0.020
2002	363	15	22	0.096	231	3	6	0.038	247	3	1	0.016
2003	344	6	32	0.101	229	2	5	0.030	239	3	11	0.056
2004	334	30	16	0.131	231	9	0	0.039	239	10	3	0.054
2005	316	33	48	0.223	237	6	3	0.038	241	13	8	0.084
2006	319	12	30	0.120	243	8	0	0.033	218	3	36	0.154

†: AA acquired Trans-World Airlines in 2001.

among competitors from different airlines. As we show in the third column of Table 2.4, cross-alliance overlapping significantly decreased in both absolute and relative terms, a change from 22.1% to 19.0%, which makes the change in overlaps between same-alliance carriers more prominent.

Although this preliminary evidence is already indicative, we now move to rigorous statistical analysis. The three columns in Table 2.5 represent results from Probit models using an increasing number of control variables: Model 1 has only demographic and segment-level controls, Model 2 adds some network features, and Model 3 has a full set of network controls to demonstrate robustness of our results. Standard collinearity tests indicate no multicollinearity problems, and the models are estimated with a good model fit, at R^2 around 0.88 (largely due to high persistency over time). The estimates of the control variables are consistent with existing literature both in magnitudes and

Table 2.4: U.S. Domestic Airlines Flight Network and Overlap 1998-2006

	UA & US			**DL,NW & CO**		
	overlapped # routes	total # routes	overlapped %	overlapped # routes	total # routes	overlapped %
1998	33	600	5.50%	70	780	8.97%
1999	39	602	6.48%	64	798	8.02%
2000	38	603	6.30%	67	816	8.21%
2001	33	582	5.67%	60	810	7.41%
2002	22	495	4.44%	60	800	7.50%
2003	24	458	5.24%	55	768	7.16%
2004	33	484	6.82%	64	789	8.11%
2005	44	495	8.89%	64	782	8.18%
2006	53	479	11.06%	62	741	8.37%
	Airlines from different alliance					
1998	291	1318	22.08%			
1999	304	1331	22.84%			
2000	334	1352	24.70%			
2001	323	1324	24.40%			
2002	282	1358	20.77%			
2003	266	1314	20.24%			
2004	285	1309	21.77%			
2005	263	1340	19.63%			
2006	245	1293	18.95%			

signs. Congestion level (both the overall load-factor and the focal carrier's own load-factor) is positively associated with more entries. The degree of competition, both at the segment and the nodal levels, are positively associated with entries, which can be explained by higher underlying profitability of the market. Competition from low-cost carriers, both at segment and nodal levels, poses a credible threat and leads to fewer entries. The network controls are mostly significant and contribute to the explanatory power. A high degree of centrality is associated with a higher probability of operating, which also reflects the fact that airline's market power at the endpoints is strongly associated with active entries (e.g.,Berry 1992). Note that the degree of competition on O&D level is negatively correlated with entry, while competition at segment level is positively correlated with entry.

We now move on to discuss the variables of main interest, including the change of the

Table 2.5: Probit Model of Entry/Stay/Exit 1998-2006

		(1)	(2)	(3)
Parameters of Interest	**alliance**	0.412***	0.871***	0.763***
		(0.053)	(0.079)	(0.081)
	own share	1.291***	0.446***	0.460***
		(0.072)	(0.060)	(0.061)
	partner share before alliance (δ_{p1})	-0.690***	-0.945***	-1.095***
		(0.090)	(0.113)	(0.114)
	competitor share before alliance (δ_{c1})	-0.204***	-0.267***	-0.409***
		(0.051)	(0.060)	(0.062)
	partner share after alliance (δ_{p2})	-0.474***	-0.531***	-0.668***
		(0.077)	(0.100)	(0.103)
	competitor share after alliance (δ_{c2})	-0.797***	-0.645***	-0.793***
		(0.060)	(0.074)	(0.076)
	change in partner effect ($\delta_{p2} - \delta_{p1}$)	0.216***	0.414***	0.427***
		(0.111)	(0.141)	(0.141)
	change in competitor effect ($\delta_{c2} - \delta_{c1}$)	-0.593***	-0.378***	-0.384***
		(0.070)	(0.082)	(0.082)
	difference-in-difference	0.809***	0.793***	0.811***
	($\delta_{p2} - \delta_{p1}$)-($\delta_{c2} - \delta_{c1}$)	(0.127)	(0.156)	(0.156)
Segment Features	**log(distance)**	-0.241***	-0.199***	-0.228***
		(0.015)	(0.018)	(0.021)
	log sqrt(pop1 * pop2)	0.303***	-0.089***	-0.083***
		(0.023)	(0.028)	(0.028)
	log(income)	0.652***	0.389***	0.246
		(0.156)	(0.181)	(0.184)
	loadfactor	0.779***	0.970***	0.985***
		(0.067)	(0.086)	(0.089)
	HHI	-1.056***	-1.154***	-1.330***
		(0.060)	(0.073)	(0.078)
	lcc	-0.839***	-0.273***	-0.314***
		(0.060)	(0.077)	(0.117)
	own loadfactor	4.601***	3.902***	3.898***
		(0.068)	(0.074)	(0.075)
Node Features	**origin HHI**		0.757***	0.822***
			(0.086)	(0.090)
	dest HHI		0.770***	0.912***
			(0.087)	(0.090)
	origin lcc		-0.410***	-0.211
			(0.117)	(0.122)
	dest lcc		-0.385***	-0.256
			(0.116)	(0.119)
	own origin degree		0.029***	0.030***
			(0.001)	(0.001)
	own dest degree		0.029***	0.030***
			(0.001)	(0.001)
	own origin share		0.240**	0.152*
			(0.079)	(0.082)
	own dest share		0.227**	0.092
			(0.080)	(0.083)
Edge Features	**# indirect paths**			0.060***
				(0.012)
	O&D HHI			0.568***
				(0.074)
	O&D lcc			-0.224
				(0.130)
	year dummy	yes	yes	yes
	carrier dummy	yes	yes	yes
	# obs	181536	181536	181536
	Pseudo R-Square	0.8703	0.8917	0.8927

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

partner effect, the change of the competitor effect and the difference-in-difference term. We focus on explaining the results from Model 3. In support of Hypothesis 1B, the change of partner effect is significantly positive (0.427), the change of competitor effect is significantly negative (-0.384), and the difference-in-difference term is positive and significant (0.811). The results are consistent after we control for potential endogeneity between lagged market share and profitability shock, as shown in Table 2.6. The changes of partner and competitor effects have been slightly biased up, the difference-in-difference estimate is almost the same (0.907 under correlated random effect model, and 0.784 under serial correlation model). This estimated difference-in-difference term is also at a similar scale when we relax different assumptions underlying the difference-in-difference identification[15]. These results demonstrate consistently that, post alliance, airlines are more likely to operate in a market in which their partners have a strong presence. In addition, we conduct placebo analysis using randomly chosen years, and there are no significant changes of entry strategy in those years other than years when the alliances were formed, which suggests that there appears to be no "fundamentally different" changes in the markets operated by partners due to reasons other than alliances.

Economic Significance. We next compute the marginal effects to understand the economic significance of these estimates. Recall that in the Probit model marginal effects

[15]Note that, after including interaction terms of the yearly dummy and the partner's/competitors' share, the identification of the change in partner's effect ($\delta_{p1} - \delta_{p2}$) comes from the differences in the timing of alliance formation. The identification of the change in the competitor's effect ($\delta_{c2} - \delta_{c1}$) comes from both the timing differences and the fact that AA never started a partnership with other major domestic airlines. As variation in timing is low (UA, US in 2003, and DL in 2004), it is not surprising that the change in partner's effect is absorbed by year-specific partner effect. However, the key conclusion stands as the difference-in-difference term is still significant and at the same scale.

Table 2.6: Results for the Entry Model with Corrections for Endogeneity

	Traditional RE	**Correlated RE**	**GEE**	**GEE-AR(1)**
change in partner effect	0.233***	0.343***	0.328***	0.275***
	(0.117)	(0.130)	(0.114)	(0.113)
change in competitor effect	-0.614***	-0.564***	-0.450***	-0.509***
	(0.073)	(0.081)	(0.074)	(0.074)
difference-in-difference	0.848***	0.907***	0.778***	0.784***
	(0.089)	(0.097)	(0.087)	(0.084)
# obs	181536	181536	181536	181536

GEE: Generalized Estimating Equation.

depend on the predicted probability of market presence at the point under consideration. We follow Anderson and Newell (2003) to calculate the marginal effects. Figures 4.3 and 4.4 illustrate the marginal effects at different market share levels of the operating carrier, of the partner and of competitors. The X, Y and Z-axes in both figures are the partner's share, competitor's share and the marginal effect on the probability of operating. Figure 4.3 demonstrates the marginal effect on entry probability (i.e., conditional on not operating previously). The change in partner's marginal effects and competitors' effects are depicted in red and blue, respectively. We show both the average and the 95% confidence interval of these effects. The clear gap between the two confidence intervals shows that the marginal effects of the "difference-in-difference" term is significant and positive. We make a conservative interpretation of this graph based on the point with the narrowest gap. A 1% increase in the partner's market share would induce an additional 0.005% in the entry probability, while the same change in competitor's market share would lead to a 0.007% decrease in entry probability. This change may look small at the absolute level, but it actually corresponds to 2% increase from the baseline entry probability (at 0.005 as inferred from the data). To make this more tangible, we compare two typical types of markets (i.e., average markets from the data) conditional on that the focal carrier does not operate in the market: one dominated by a partner (80% market share), the other by a competitor (80% market share). In the first market, the carrier's entry probability increases from 0.0057 to 0.0098 after the alliance, or it is almost doubledWe estimate that, on average, alliances are responsible for 4 more entries into partner-dominated markets and 18 fewer entries into competitor-dominated markets annually. The difference-in-difference is 22 entries per year, which is highly economically significant (20% of the baseline annual entries).

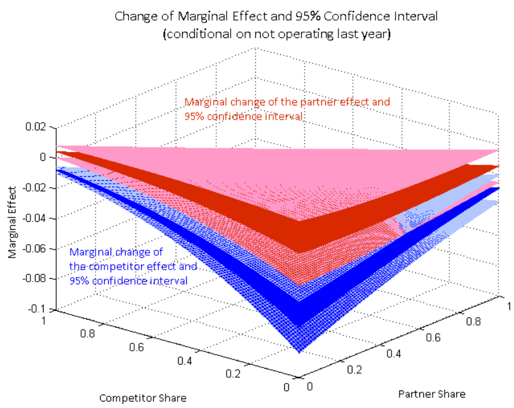


Figure 2.1: Marginal Effect and 95% Confidence Interval

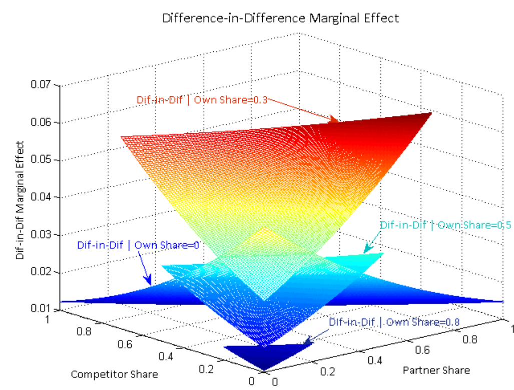


Figure 2.2: Marginal Effect at Different Levels of Market Share

In what types of markets based on the competition structure are we more likely to observe partner-favoring behaviors? We analyze this question based on four levels of market share by the operating carrier: 0%, 30%, 50% and 80%, which correspond to typical levels of market shares in the following four scenarios: when the focal airline is a potential entrant, or is an incumbent in oligopoly, duopoly, and monopoly settings, respectively. The marginal effects of these difference-in-difference terms are all positive and significant as shown in Figure 4.4. Note that the largest parter-favoring effect is observed in the oligopoly case (30%), where the airline is likely to compete with one partner and one competitor. In this scenario, the tension between competitors from different leagues makes it more valuable to gain additional market power. The effect is the smallest in the monopoly case (80%), in which the operating airline is most likely to operate in such a market regardless of the market power possessed by the partner. The effect for non-existing operating carriers (0%) and the duopoly case (50%) fall in between, as expected.

2.5.2 Estimation Results of the Capacity Model

We continue the discussion of our findings for the capacity model with result shown in Table 2.7. The model is estimated under Pooled OLS, Random Effects and Fixed Effects and under dynamic panel data model with Arellano-Bond estimator. Again, in support of Hypothesis 1B, the results are consistent with our earlier results for the market entry: airlines increase capacity in the markets where they compete with their partners, while decreasing capacity in those markets where they compete with competitors. That is, instead of reducing capacity redundancy, airlines actually expand capacity in those markets where their partners are also present. We focus on the Fixed Effects Model and Arellano-Bond estimators as they properly address the endogeneity issues. To assess the economic impact of these estimates, we compare two typical duopoly markets: one operated by the carrier of interest and its partner, the other by the same carrier and one

Table 2.7: Regression of Capacity conditional on Stay 1998-2006

	Without Lagged Capacity			With Lagged Capacity
	Pooled OLS	Random Effect	Fixed Effect	Arellano-Bond
change in partner effect	0.116*	0.153***	0.193***	0.175***
	(0.062)	(0.034)	(0.032)	(0.054)
change in competitor effect	0.055	-0.160**	-0.148***	-0.140***
	(0.037)	(0.021)	(0.021)	(0.029)
difference-in-difference	0.061	0.313***	0.341***	0.315***
	(0.071)	(0.039)	(0.038)	(0.060)
# observation	13351	13351	13351	13351
adj. R-square	0.576			-
within		0.212	0.244	
between		0.454	0.192	
overall		0.454	0.215	

of its competitors. Each carrier possesses 45% of the market share, and operates 250,000 seats annually (inferred empirically from the data). Note that, in these scenarios, the only difference is the "identity" of the other player (i.e., a partner or a competitor). In the market operated with a partner, after the alliance, each airline increases seat capacity by 8.69%, which corresponds to 21,700 more seats annually (418 seats weekly) – roughly 3 additional flights per week. However, if the market is operated with a competitor, each airline would reduce its annual seat capacity by 6.65%, or 16,625 fewer seats annually (319 fewer seats weekly) – 2 less flight per week. If we look at the difference-in-difference estimate, the capacity change in the partner's market over the competitor's market is 15%, or 5 flights per week. These capacity changes are clearly economically significant in addition to being statistically significant.

2.6 Examination of Competing Explanatory Mechanisms

Although so far we find compelling reasons to support Hypothesis 1B and the multimarket contact explanation behind it, there are also alternative explanations which might (at least partially) explain what we have observed. These explanations are mainly formed around two commonly cited benefits of alliances: one is cost reduction, and the other is customer acquisition. We examine the most plausible arguments of each.

2.6.1 Supply Side: Cost Reduction

It is commonly argued that airline alliances help their members take advantage of cost synergies through cost reduction activities such as facility and personnel sharing, joint marketing programs, and joint procurement. The cost reduction may come in two forms: entry cost reduction and operating costs reduction. Entry costs mainly include expenses associated with acquiring gate slots and purchasing additional aircrafts, if needed. At congested airports, operating barriers such as slot controls for takeoffs and landings and long-term exclusive-use gate leases make it difficult for most airlines to enter: entering airlines sometimes have to sublease gates from the big incumbent airlines (General Accounting Office 1998). In this scenario, entry barriers are lowered for partners who engage in gate sharing. As far as operating costs are concerned, fuel and labor costs are the two biggest components. Fuel costs rose from 15% to 36% of the total revenue since

2003, while labor costs are 25% of the total revenue over past 6 years (S&P Industry Report). Joint fuel procurement and hedging, personnel sharing and additional bargaining power with labor unions should help alliance members reduce operating costs. Based on the argument of cost reduction through alliances, two alternative explanations may arise,

Alternative Explanation 1: *Airline alliances lead to reduced entry costs, and hence they result in higher entry rates into markets in which the alliance partners operate.*

Though this argument may be plausible to explain the observed increase in entry rates, it offers no insight into the reasons for capacity expansion: conditional on being an incumbent, an airline's capacity decision is not so much dependent on the changes of entry costs. Thus, we do not think this explanation provides adequate support to our findings.

Alternative Explanation 2*: Airline alliances lead to reduced operating costs, which in turn lead to higher rates of entry and higher level of capacity in markets in which the alliance partners operate.*

If this argument were true, we would see consistent changes in both entry and capacity decisions as we have documented in this paper. However, if this argument were true, lower operating costs would also lead to lower average prices in the long run. In an industry as competitive as this, firms compete mainly on prices and frequent price wars are manifestations of this fact. The price competition is further escalated by constant invasions from low-cost carriers. Recently, the price competition has led airlines to “innovate” in new ways to cut base prices — such as charging for food and baggage. As a result, we note that in this industry any reduction in costs is most likely to be reflected in final prices. This point has also been emphasized by alliance officials: “we are committed to pass on these cost benefits to consumers”. Therefore, we would expect to observe lower prices as a result of lower operating costs. To examine this possibility, we analyze changes of prices in markets operated with partners vs. markets operated with competitors before and after alliances, using a similar framework to what we used for the entry and capacity decisions. We conduct the analysis using pooled OLS, random effects and fixed effects models, and obtain consistent results. We show results based on the fixed effects model which addresses potential endogeneity concerns in Table 2.8 Column 1. Contrary to what this explanation suggests, after formation of alliances, airlines actually charge \$4.2 *more* on average in segments operated together with their partners, compared to \$7.3 *less* in segments operated with competitors (we obtain these numbers using the previous duopoly market example). That is, an \$11.5 premium is charged for an average one-way segment coupon in markets operated with partners. Segment prices are taken as the distance-weighted average of itinerary prices (Dana and Orlov 2009), to keep the analysis at a consistent level as in the entry and capacity models. Being aware of potential misspecifications of this approach, we also conduct a similar estimation on O&D itinerary prices, and find that a \$22 premium is charged for a round-trip ticket in typical duopoly markets. These results regarding price changes seemingly contradict the explanation based on lower operating costs.

Moreover, we also examine changes in operating costs using changes in load-factors,

Table 2.8: Fixed Effects Model of Average Segment Fare 1998-2006

	(1)	(2) own service quality	(3) alliance service quality
difference-in-difference	25.571***	32.176***	28.996***
	(3.000)	(2.865)	(5.559)
flight frequency (ln(# of departures))		17.619***	18.113***
(ln(# of departures))		(1.043)	(1.047)
total capacity (ln(# of seats))		-40.346***	-15.454***
(ln(# of seats))		(1.144)	(2.364)
difference-in-difference			-4.623
in partner's flight frequency			(3.097)
difference-in-difference			3.020
in partner's total capacity			(1.878)
# obs	14084	14084	14084
R-sq	0.555	0.596	0.601

perhaps the most important driver of operating costs and profitability in the airline industry (S&P Industry Report). The higher the load-factor, the lower the operating costs (measured by cost per revenue-passenger-mile) will be. We examine changes in load-factors again using the same framework of difference-in-difference estimation, and find that load-factors have decreased dramatically in markets operated with partners, i.e., *down* by 3.5 percentage points, while increased in markets operated with competitors, i.e., *up* by 2.4 percentage points (again measured in the typical duopoly market example), as shown in Table 2.9 Column 1. These changes in load-factors correspond to significant changes in costs — a 4.7% increase in costs on markets with partners, everything else held constant, and a 3.0% decrease in costs on markets with competitors (computed at the load-factor level of 78%). Summarizing the findings from prices and load-factors, the argument of reduced operating costs does not seem to be a convincing explanation for what we observe in the data.

Table 2.9: Fixed Effects Models for Load-Factor, Total Traffic, and Full Fare Traffic 1998-2006

	load-factor	traffic	% of full fare traffic
change in partner effect	-0.077***	0.033	0.274***
	(0.024)	(0.028)	(0.093)
change in competitor effect	0.053***	-0.169***	0.297***
	(0.015)	(0.017)	(0.055)
difference-in-difference	-0.130***	0.202***	-0.023
	(0.028)	(0.031)	(0.105)
# obs	14084	14084	14084
Adj R-squared			
Within R-sq	0.7499	0.4053	0.6306

2.6.2 Demand Side: Customer Acquisition

Another type of activities that alliance members are generally involved in is related to customer acquisition: 1) alliances provide better services by offering more options, smoother connections and shared alliance lounges; and 2) alliances allow consumers to accumulate and redeem miles on partners' flights through their reciprocal frequent flier programs (though certain restrictions may apply). These improvements in service quality may increase customers' willingness to pay, or induce more high-value customers to purchase. That is, alliances may lead to higher prices for reasons other than mutual forbearance, and these higher margins may provide incentives for airlines to increase capacity. Based on this logic, we examine the following arguments on the demand side.

Alternative Explanation 3: *More flight options and higher frequencies are associated with better service quality. The potential of charging higher prices for better services induces airlines to add additional flights in the markets operated with their partners.*

While this explanation is consistent with our observations regarding both capacity expansion and price increases, we examine whether the observed price premium can be explained away through these potential changes in service quality. Effects of more frequent flights might be two-fold: first, by operating additional flights in the markets, airlines are able to charge higher prices due to the improvement of their own services; second, by cooperating with their alliance partners, they may be able to charge higher prices for their partners' services as well, since consumers can earn and use flying miles with any airline of the same alliance. We examine the contribution of both mechanisms to the price premium, and the results from the fixed effects models are shown in Table 2.8. Column 2 shows that airlines are indeed able to charge higher prices for more frequent services that they provide: the effect of the number of departures (a proxy of service frequency) is significantly positive, while controlling for the total number of seats supplied. However, this does not diminish the price premium charged in markets shared with partners. Meanwhile, Column 3 presents no evidence that airlines are charging additional premium for services provided by their partners: the price premium is still as high as \$13.0. Based on these results, we conclude that the quality argument does not explain away our findings regarding price premium. That is, even after controlling for the potential changes in service quality, partners still benefit from the additional pricing power developed from the multimarket contact.

Alternative Explanation 4: *After an alliance is formed, partners experience higher demand flowing through their networks, which leads airlines to increase network overlap, expand capacities, and charge higher prices.*

To see whether this explanation might be plausible, we examine changes in traffic volumes using the difference-in-difference framework. The results from fixed effects models are shown in Table 2.9. Contrary to the prediction of higher demand, we do not observe any significant increase in traffic volume in markets with partners. Recall that we documented a capacity increase of 420 seats per week. Nevertheless, there is no sizable increase in traffic (i.e., a statistically insignificant increase of 50 passengers per week). Conversely, in the markets in which competitors are present, traffic decreases by 260 passengers per week after capacity is reduced by 320 seats per week. Although the difference-in-difference term is significantly positive, note that we do not observe first-hand evidence of increasing demand in markets operated with partners. The significant difference-in-difference term can be explained by shrinking capacity in markets with competitors: a reduction of 320 seats per week corresponds to a decrease of 250 passengers per week at the average level of load-factor, i.e., 0.78. One may still suspect that potential increases in demand might have been offset by increases in prices. However, if we hypothesize that increasing demand is the driving force, it is unlikely that prices will offset all changes in demand, as demand changes are the first order effect. Summarizing these two points, we conclude that the demand argument does not provide a plausible explanation to our findings.

Alternative Explanation 5: *After an alliance is formed, the composition of demand changes. In markets jointly operated with partners, airlines attract more high-value customers with better services, which allows them to charge higher average prices.*

To study this explanation, we examine changes in the composition of travelers. The

DB1B fare data provides some limited information on fareclasses: restricted fare or unrestricted fare (i.e., full fare). We obtain the percentage of passengers traveling on full fares and adopt the difference-in-difference framework to examine how this percentage is affected by alliances. The results from the fixed effects model are presented in Table 2.9, Column 3. The difference-in-difference term is insignificant and negative, providing no support to the presumption that there is an increase in the proportion of high-value customers. We conclude that there is no evidence to support changes in demand composition.

To summarize, we examined five most plausible alternative explanations of our results based on both supply and demand effects of alliances. However, none of them provides compelling explanations to our findings.

2.7 Conclusion

In this paper, we study the changes of airlines' entry and capacity decisions after collaborating with other airlines through alliances. While theoretical models of operational decisions suggest that a decrease of competition (e.g., due to the alliance) will reduce the inventory or production level, we find exactly the opposite. Specifically, we find that, as the level of competition is reduced by the alliance, partners seek to overlap among themselves and increase capacities in markets in which they cooperate, while doing exactly the opposite in the markets operated by competitors from different alliances. These surprising findings are consistent with predictions of the multimarket competition theory — an important feature of the airline industry and of many other industries. Multimarket contact enables firms to form mutual forbearance and compete less aggressively. To enjoy the benefits of the multimarket contact, firms will strategically choose which market to operate in and how much capacity to install based on their competitors' network. This explains the changes in airlines' operational strategies that we observe post alliances: airlines seek to establish and strengthen multimarket contact with their alliance partners which, in turn, leads to less aggressive competition among alliance partners and allows them to charge an \$11 premium on average on a one-way segment coupon.

To confirm that multimarket competition is indeed the explanation for our findings, we carefully examine several competing explanations that are most plausible, building on theories of inventory and service competition and on the supply and demand side effects commonly cited as the main benefits of airline alliances. However, none of these alternatives is able to provide full support to what we observe. We therefore conclude that our findings are most likely driven by the multimarket competition.

Our findings have important implications both theoretically and empirically. Theoretically, we highlight the importance of incorporating the perspective of multimarket competition into analysis of operational decisions. Even though multimarket contact has been a well-studied topic in industrial organization economics and strategy, operations management community is yet to identify its implications on firms' operational strategies. Empirically, we also point out important research opportunities to study the impacts of multimarket competition in various industrial settings, such as airlines, banks, retail chains, and multi-plant manufacturers. Findings from various industries will help reconcile competing theories and offer insights to future theory developments. Finally, we draw attention of industry managers and regulators to possible anti-competitive effects of the airline alliances and the necessity to track evolution of the alliances after they are formed.

Chapter 3

Are Consumers Strategic? Structural Estimation from the Air-Travel Industry

3.1 Introduction

When was the last time you passed on an immediate purchase to wait for a sale? This can be a smart choice as inter-temporal price fluctuations are common across industries: fashion items are marked down several times towards the end of the season; storable goods are periodically put on sale; prices for high-tech gadgets with short life cycles dip soon after release; airlines, hotels, car rentals and theaters regularly revise prices or launch promotions. Of course, there are also consumers who do not strategize over the timing of purchase simply because they are not aware of the possibility that price decreases, have high waiting costs, or are not willing to forgo immediate gains in exchange for uncertain future gains. We will refer to *strategic* or *forward-looking* consumers as those who maximize long-run utility by strategically timing their purchases to obtain lower prices. We also refer to *myopic* and *non-strategic* consumers as those who maximize immediate utility and hence do not exhibit strategic behavior.

Studying strategic consumer behavior is important for several reasons. First, very little research to date provides direct and rigorous evidence of such behavior, relying instead on anecdotal accounts. Empirical evidence of strategic consumer behavior would not only enrich our knowledge of consumer behavior but also improve managerial decisions, such as pricing and inventory management. Second, most current demand forecasting models, including those used in airline revenue management systems, assume that consumer arrivals are exogenous and inter-temporally independent. However, correctly incorporating inter-temporal demand substitution may improve the accuracy of demand forecasting. Finally, revenue implications of strategic consumers could be significant. The common belief is that the presence of strategic consumers shifts demand from high to low prices, thus hurting revenues, but the extent of this effect is unclear: how to price and manage seat inventory in the presence of strategic consumers continues to puzzle industry professionals.

The air-travel industry is a particularly interesting test bed to study strategic consumer behavior. Unlike price fluctuations for seasonal or technological products, inter-temporal changes in airfares are much harder to predict. It is no longer surprising to learn that the person seated next to you on a plane paid much less for essentially the same seat. Revenue management, the practice of bucketing airfares into multiple classes and managing seat inventory dynamically, has revolutionized pricing (Talluri and van Ryzin 2005, Boyd 2007). Airlines constantly recompute protection levels and bid prices by taking into account the current number of available seats and latest demand forecasts, which results in constant opening, closing, and reopening of fare classes. Inventory managers can also override revenue management systems based on their expertise and local knowledge of demand. Moreover, the pricing department revises fares periodically in response to changes in demand patterns, operating costs, and competitors' actions. Sometimes, pricing managers will launch temporal sales, which may trigger industry-wide price wars and further add to price swings. Therefore, even the same inventory class may not always be priced at the same level. All of this is likely to trigger fare shopping on the consumer side.

Despite this volatile nature of airfares, consumers are not helpless. Proliferation of web tools, e.g., major online travel agents (e.g., Expedia, Orbitz), fare aggregators (e.g., Kayak, Sidestep), and opaque channels (e.g., Hotwire, Priceline) have made it much easier for consumers to compare prices. Microsoft Bing/Travel (formerly Farecast) even makes a suggestion to "wait" or "buy" based on the probability of prices going up or down within the next seven booking days, with 75% accuracy on average. Moreover, traditional offline agencies, which are often experienced and know the quirks of price movements, may help consumers obtain better deals in order to increase customer loyalty.

Based on our spot checks of various travel-related online forums, some travelers have reported savings of \$20 to \$150 (or 7% to 33%) by waiting for a better deal. Of course, consumers are not always that smart — there are also cases in which travelers reported that they had to pay more since the price actually went up instead of going down.

In this paper, we merge two separate data sources to investigate just how strategic consumers actually are. First, we have collected information on posted rather than transacted prices. This is important because in order to investigate strategic consumer behavior we need to know prices that were available. Second, many industries where one would suspect strategic consumer behavior have fragmented distribution channels. This normally limits access to sales information to a few retailers and makes it hard to obtain a complete picture of prices. In contrast, we are able to utilize booking information from Global Distribution Systems (GDSs), covering all bookings made through online and offline travel agents.

We use a simple yet flexible structural model to account for the dynamics caused by strategic consumers. Unlike most other empirical papers which incorporate forward-looking behaviors (e.g., Erdem and Keane 1996), our model does not impose the assumption of strategic consumer behavior a priori. We account for both static and dynamic price endogeneity, which cannot be easily addressed with reduced-form regressions. Moreover, our framework is flexible enough to incorporate many extensions with respect to consumer behavior, and it allows us to examine revenue implications through counterfactual analysis. Our structural estimation results suggest that 4.9% to 44.9%

of travelers are strategic across different city-pair markets, measured by the 5th and 95th percentiles. This fraction varies by the time to departure in a non-linear fashion — most strategic consumers arrive either at the very beginning of the booking horizon or close to departure. Based on the estimation results, we draw important implications for airline revenue management strategies. Surprisingly, it turns out that the presence of strategic consumers does not always hurt revenues, since firms get incremental sales from strategic consumers who otherwise would not buy if myopic. Strategic behavior tends to decrease revenues in business markets but may increase revenues in leisure markets. As a result, commitment to a non-decreasing pricing scheme turns out to be desirable in business markets but not necessarily in leisure markets. Among markets benefiting from this pricing strategy, the median revenue improvement is 3.5%, and the lower and the upper quartiles are 1.8% and 5.6%, comparable to revenue improvements commonly reported (see Talluri and van Ryzin 2005).

3.2 Literature Review

It is a common practice to assume an exogenous demand process in traditional revenue management research (e.g., Belobaba 1989, Gallego and van Ryzin 1994). Recently, however, there is a growing interest in studying operational decisions in the presence of strategic consumers (see Shen and Su 2007, Netessine and Tang 2009 for a review). Several papers have investigated the impact of strategic consumers on firms' strategies such as inventory procurement (Cachon and Swinney 2009), capacity rationing (Liu and van Ryzin 2008), and revenue management (Jerath et al. 2010), to name a few. One of the key findings from the modeling literature is that strategic consumers may exert a negative impact on firms' revenue: Anderson and Wilson (2003) find a 7% revenue loss; Aviv and Pazgal (2008) find a 20% revenue loss; Levin et al. (2009) find a 1% to 7% revenue loss. Nevertheless, other papers predict that strategic consumers can have either adverse or positive effects on seller revenue (Su 2007, Cho et al. 2008).

Despite the burgeoning interests from the modeling perspective, empirical evidence of strategic consumers is scarce and scattered in three streams of related literature. First, in the realm of revenue management, a few recent empirical papers study consumers' choices among multiple products (Vulcano et al. 2010, Newman et al. 2010) but without accounting for inter-temporal choice behavior. Second, in marketing and economics there is a stream of literature that incorporates forward-looking consumer behavior, starting with Erdem and Keane (1996). More recent papers include Hendel and Nevo (2006) and Nair (2007), to name a few. However, most of these papers build on the premise that consumers are forward-looking a priori, and the research objective is to uncover consumer preferences rather than to investigate how strategic consumers are. Third, behavioral economists have long been looking at inter-temporal choices and time-discounting. Recent work by Osadchiy and Bendoly (2010) reports that 38% of the subjects behave rationally in laboratory experiments by strategically timing their purchases.

To provide evidence of strategic consumer behavior, there are at least two potential estimation approaches. The first approach is to estimate a discount factor (or equivalently, a cost of waiting) as a continuous measure of consumer patience (e.g., Levin et al. 2009). There are two difficulties associated with this approach — identification and computational complexity. In the aforementioned empirical studies of dynamic models, utility functions cannot be identified (non-parametrically) when the discount factor is not fixed (see Rust 1994, Magnac and Thesmar 2002). Although parametric restrictions will allow utility functions and the discount factor to be identified jointly, the computational complexity usually makes it undesirable to do so. As a result, the common practice is to fix the value of the discount factor and conduct sensitivity analysis on this value, if necessary. The other approach is to segment the market into discrete consumer types — myopic and strategic (e.g.,Su 2007), and to estimate the fraction of strategic consumers. A similar idea has been applied in the latent class models in the marketing literature (Dillon and Kumar 1994). This approach greatly streamlines the estimation. Meanwhile, it is analogous to market segmentation and can be easily explained to and understood by industry professionals, so we adopt this approach.

Given the difficulty in identifying strategic behavior and the lack of appropriate field data, so far very little research provides direct evidence of strategic consumers. To the best of our knowledge, there are only two studies parallel to this paper. Chevalier and Goolsbee (2009) test whether textbook users are forward-looking by anticipating book revisions. They take the discount factor approach since the structure of the value function is much simpler in this setting. Another example is an unpublished paper by Hendel and Nevo (2011), who propose a demand anticipation model to account for consumer stockpiling behavior during temporary sales and obtain an estimate of the fraction of stockpilers. Our work is related to theirs as we also look at aggregate data and the model bears similar structure, but we differ in three aspects. First, the research objective is different. We start out with the purpose of documenting whether consumers are strategic or not, while their focus is the inter-temporal price discrimination as a mechanism of market segmentation. As a result, we ask different questions oriented towards strategic consumers: What factors affect the fraction of strategic consumers? What if more consumers become strategic? What pricing strategies may airlines adopt in response to strategic consumers? Second, prices are more volatile and harder to predict in the airline industry, and hence the existence of strategic consumer behavior is less obvious. Significant amount of price dispersion (Clemons et al. 2002, Chellappa et al. 2011) and inter-temporal fluctuation (Etzioni et al. 2003, Mantin and Gillen 2011) has been documented in the air-travel industry. Third, demand is relatively stable over time in their setting, and therefore price endogeneity is less of a concern. However, in our setting, due to the practice of revenue management, price endogeneity is an important issue that needs to be properly addressed.

To sum up, our study contributes to several streams of literature. Assembling field data to uncover inter-temporal choices is a significant challenge due to the level of detail and multiple data sources required. Our data is unique in this sense because it provides a detailed dynamic view of daily available prices and realized demand. Moreover, the structural model we propose is simple yet flexible, and at the same time it is consistent with aforementioned modeling literature, so our approach can be used to calibrate subsequent models.

3.3 Data

We use two main datasets with millions of records that provide information on airfares and bookings, respectively. First, fares were web-crawled from the major online travel agencies: Expedia, Travelocity, Orbitz, and Priceline. The dataset contains posted fares over a three-month booking period before each day of departure for each city-pair market. The departure dates are a random set of seven weeks including both peak and off-peak weeks in the spring travel season of 2005. On each booking day, we extract the three lowest round-trip (with seven-day length of stay) fares from each online travel agency for each particular market and departure date. A fare record contains: booking date, departure date, origin and destination airports, carrier, inventory class, fare type, fare amount, booking agency, and other details of the itinerary. We have also validated the web-crawled fare information by matching it with transaction prices obtained from a corporate sponsor in the same time period (matched with a correlation of 0.860). Second, booking data from the Marketing Information Data Transfer (MIDT) is obtained from an airline corporate sponsor. It contains U.S. point-of-sale reservations made in all major GDSs. GDSs distribute a large proportion of airline tickets via offline and online travel agents, accounting for 50% to 70% of all sales in the U.S., so we are able to ensure that sales information is available for most distribution channels. Markets are selected based on the largest passenger volume for the airline sponsor while ensuring wide geographic coverage of cities with large populations. These markets represent about 80% of market coverage for the airline. The dataset contains the following information on the outbound leg of all economy-class reservations: reservation date, departure date, origin and destination airports, carrier, inventory class, the booking agency, and number of passengers. The data provide a relatively comprehensive view of air-ticket reservations including their time stamps. Note that MIDT does not contain price information, so we merge it with the web-crawled price information. Whenever needed, we supplement this data using other publicly available sources, such as the DB1B 10% air ticket sample from the Department of Transportation and population and income data from the Bureau of Economic Analysis.

In sum, there are 5.9 million booking records and 4.5 million fare records. We merge the two datasets according to markets, departure date and booking date. Thus, the combined dataset provides a nearly complete picture of daily ticket sales and daily lowest fares for 111 city-pair markets (84 domestic), 45 departure dates, and a three-month booking period for each departure date. In this paper we focus on inter-temporal decisions and aggregate the data across airlines. We could explicitly consider substitution among different flights and carriers on the demand side and airline competition on the supply side. We do not possess data to fully implement this approach since we would need price histories associated with all outbound and inbound flight combinations of all airlines. Additionally, complexity of the model increases manyfold because one has to model and compute the competitive equilibrium among airlines in addition to modeling consumer expectations. We leave this for future research. Based on our preliminary analysis, however, we believe that modeling these aspects is secondary since our focus is on the inter-temporal substitution: we did attempt to control for the competition level in a crude way but found no significant effects, as we demonstrate later.

Even though we have taken great effort to assemble this massive database, note that this data is aggregate data in that we observe only the daily purchases and available prices, not individual search behavior. However, in most circumstances, firms have access only to aggregate data such as we have here, in which case structural estimations are appropriate to infer strategic consumer behavior. We further aggregate the data by booking week for the following reasons: First, this aggregation level is consistent with industry practices and with related papers (Mantin and Gillen 2011, Hendel and Nevo 2011, Granados et al. 2012): a common practice to construct a fare class is to require 21-/14-/7-day advance purchase. As a result, price varies more from week to week than from day to day as we will demonstrate. Second, one would expect that even myopic consumers may take a few days before making a purchase decision, and even the most strategic consumers cannot wait indefinitely. Letting consumers make purchase decisions at the weekly level allows us to account for air-travel planning horizons of both myopic and strategic consumers to a certain extent. Finally, for reasons to be explained later, aggregation at a lower level would make estimation prohibitively complex computationally.

We note significant amount of price variation from the data. The overall standard deviation of prices is about 80% of the mean, as shown in Table 3.1. To further inves-

Table 3.1: Summary Statistics of Demand and Price

	mean	std	min	max
total passengers by market departure date	90.1	110.0	0	1436
weekly passengers	6.4	13.3	0	309
weekly average fare (weighted)($)	370.35	288.58	110.33	4284.33

Table 3.2: Fare Fluctuations over Markets/Departure Dates/Booking Time

	decrease(fraction)	no change(fraction)	increase(fraction)
	daily lowest fare		
mean	0.279	0.472	0.249
	weekly fare (average of daily lowest fare)		
mean	0.395	0.093	0.512
range by market	[0.170, 0.561]	[0.004, 0.363]	[0.322, 0.630]
range by departure date	[0.300, 0.477]	[0.042, 0.153]	[0.408, 0.644]
by booking week 1	0.450†	0.175	0.375
2	0.499	0.135	0.367
3	0.474	0.138	0.388
4	0.413	0.133	0.454
5	0.449	0.106	0.445
6	0.435	0.084	0.481
7	0.446	0.071	0.483
8	0.425	0.076	0.499
9	0.460	0.080	0.460
10	0.456	0.073	0.471
11	0.198	0.035	0.767
12	0.035	0.006	0.959

† : *Fare decreases with 0.450 probability from week 1 to week 2.*

tigate the inter-temporal component of the price variation, we summarize frequencies of price trends in Table 3.2. On average, the weekly fare decreases in the subsequent week with 39.5% probability, increases with 51.2% probability, and remains constant with 9.3% probability. In comparison, most often (47.2% of the time) the daily fare stays unchanged. This probability varies significantly by market, departure date and booking week. This persistent inter-temporal variation underscores the uncertainty in prices faced by air-travel consumers, and also the potential opportunities for consumers to strategize on the timing of purchase.

3.4 Preliminary Results using Reduced-Form Regressions

Before formulating the structural model, we first look at preliminary results from reduced-form regressions, which highlight the potential endogeneity problems and lead us to the structural approach. In the absence of strategic consumers, demand d_t in period t is correlated with the price of that period p_t, but not with prices of past or future periods, everything else being properly controlled for. However, when some consumers strategically postpone their purchase decisions in anticipation of a price decrease, demand will be correlated with past and future prices. The lower the future prices or expected future prices, the more likely consumers are willing to delay their purchases. To test this, we estimate weekly demand using the following three models: 1) a basic model that accounts only for the price of the current week, 2) a model that accounts for realized future and past prices, and 3) a model that accounts for expected prices. To show the primary effects, we consider prices only one period (i.e., one week) ahead:

$$d_t = \alpha + \beta_0 p_t + X\gamma + \varepsilon_t, \tag{3.1}$$

$$d_t = \alpha + \beta_0 p_t + \beta_1 p_{t+1} + \beta_2 p_{t-1} + X\gamma + \varepsilon_t, \tag{3.2}$$

$$d_t = \alpha + \beta_0 p_t + \beta_1 \hat{p}_{t+1} + \beta_2 p_{t-1} + \beta_3 \hat{p}_t + X\gamma + \varepsilon_t, \tag{3.3}$$

where the X includes covariates such as polynomials of booking time and market departure date characteristics; see Table 3.3 for a description of these variables. Controlling for the realized future and past prices, Model 3.2 corresponds to the case in which strategic consumers have perfect knowledge about future prices. We expect β_1 and β_2 to be positive if there is inter-temporal substitution. However, consumers most likely do not have perfect knowledge of future prices, but rather make predictions of future prices based on historical prices and other information available. We control for the expected prices $\hat{p}_{t+1}$ and $\hat{p}_t$ in Model 3.3. Expected prices are obtained using four period lags of prices and other market covariates, which together explain up to 80% to 90% of the price variations. Strategic consumers who arrive at time $t-1$ will base their timing of purchase on price p_{t-1} and the expected price $\hat{p}_t$. The coefficient β_1, β_2 in Model 3.3 should have the same sign as in Model 3.2, while the coefficient β_3 of $\hat{p}_t$ should have a sign opposite to β_2.

Table 3.3: Variable Descriptions

Category	Variable
(1) Reduced-Form Regressions and Baseline Demand Model	
price	weekly price: average of daily lowest prices within a week
booking time	polynomials of booking week; final booking week dummy before departure
departure date characteristics	high-demand season; day-of-week dummies
(2) Predicting the Probability of Prices Falling for Weak-Form Rational Expectation	
relative prices	current_to_last: current week price divided by final week price
	current_to_initial: current week price divided by the first week price
	current_to_mktavg: current week price divided by the average market price
price volatility	the coefficient of variance of the daily prices in the current week
initial price	initial price
other measures	booking time and departure day characteristics
(3) Supply Model, Strong-Form Rational Expectation	
price history	price lagged by one period
demand history	cumulative demand
initial price	initial price
other measures	booking time and departure day characteristics

We run regressions on a sample market and present the results in Table 3.4. All signs are as expected except for that of the lagged price in Model 3.2, largely supporting our belief about the presence of strategic consumers. The results tell us how sensitive the current demand is to prices in adjacent time periods. Although this simple approach is indicative of the presence of strategic consumers, it suffers from several drawbacks. First is the price endogeneity commonly seen in static settings. The current price may be correlated with the current demand shock as firms usually have better knowledge of demand shocks than we do as econometricians, leading to positively biased price sensitivities. Second, the dynamics add an additional layer to price endogeneity. For example, in Model 3.2, the positive correlation between demand d_t and future price p_{t+1} can be explained either as strategic consumers waiting for prices to drop or as prices drop as a result of previously realized low demand, again biasing the price sensitivity positively. Using expected prices instead of realized prices may alleviate but not fully resolve this problem. The expected future price can still be correlated with demand shocks through the current price, which again introduces positive bias in price sensitivities. In fact, we do find many insignificant price coefficients for many markets. Since all price variables are endogenous, we would need multiple Instrument Variables (IVs), and finding IVs in this setting is not trivial because they would need to vary across the booking periods. Cost-based supply shifters (e.g., fuel prices and labor costs) are not useful because weekly prices do not respond to these cost factors in the short run. Other IVs used in airline studies, such as distance and demographics (Borenstein and Rose. 1994, Granados et al. 2012) are not applicable either since they do not vary over the booking period.

In addition to the endogeneity problem, it is also not clear how to explain the esti-

Table 3.4: Reduced-Form Regressions of Demand on (Expected) Future and Past Prices

	(1) current price	(2) realized future and past prices	(3) expected prices
price (p_t)	-0.044	-0.081	-0.177
	(0.021)	(0.041)	(0.046)
p_{t+1}		0.150	
		(0.026)	
p_{t-1}		-0.087	0.362
		(0.041)	(0.165)
expected price $\hat{p}_{t+1}$			0.300
			(0.040)
$\hat{p}_t$			-0.463
			(0.171)
booking time t	4.224	-1.187	9.783
	(2.695)	(2.748)	(4.762)
t^2	-0.985	0.147	-1.834
	(0.472)	(0.491)	(0.921)
t^3	0.088	0.022	0.122
	(0.024)	(0.026)	(0.050)
final week	25.113	84.738	145.341
	(5.222)	(12.166)	(17.367)
high season	4.450	3.419	5.875
	(2.102)	(2.050)	(2.358)
day-of-week dummies	yes	yes	yes
const	8.740	8.893	-16.116
	(6.681)	(7.287)	(8.989)
R-square	0.7300	0.7466	0.7554

Note: Standard errors are shown in parentheses.

mated coefficients (e.g., β_1, β_3): although one would get a sense of how sensitive demand is to adjacent prices, it is hard to get a sense of how strategic consumers are from these estimates, especially if we want to compare them across markets. In sum, although reduced-form regressions seem to provide some evidence of varying prices and strategic consumer behavior, these simple models cannot properly address multiple econometric issues. We turn next to the structural modeling approach.

3.5 The Structural Model

3.5.1 The Demand Model

We first consider the decision of a consumer who desires to travel in a particular city-pair market on a particular departure date. We assume that both the market and the day of departure are given exogenously. To focus on the inter-temporal substitution, we further ignore substitution between nearby airports and among adjacent departure dates[16]. In our model, needs for travel arise exogenously. A consumer arriving at booking time t can be either strategic (with probability θ) or myopic (with probability $1-\theta$). Our goal is to obtain an estimate of the fraction of strategic consumers (i.e., θ) in the population.

Myopic consumers are those who arrive at t and immediately make a purchase-or-not decision. If a myopic consumer decides not to buy at time t, he will never come back. Strategic consumers are those who arrive at t but may decide to postpone their purchase

[16] Capturing those types of substitutions would require different types of data, and is outside of the scope of this paper.

and come back later. For computational reasons, we make a conservative assumption that strategic consumers wait for at most one period of time[17]. Should she decide to wait, she will come back later and decide whether or not to purchase the ticket. The waiting decision of a strategic consumer depends on her expectation of future prices, which we discuss later. Consumers in our model are heterogeneous along the following three dimensions: 1) time of arrival (e.g., Su 2007, Aviv and Pazgal 2008); 2) strategic or myopic (e.g., Su 2007, Cachon and Swinney 2009); and 3) valuation of the products (e.g., Levin et al. 2009). Valuation does not affect waiting decisions in our model as we discuss shortly. Furthermore, we will allow for different price sensitivities between strategic and myopic consumers to incorporate heterogeneity in valuation distributions among them.

Knowing the consumers' problem, we model the aggregate demand observed in each booking period. The aggregate demand d_{mt} on a city-pair market and departure date dyad m and at booking time t is composed of three subgroups: 1) myopic consumers who arrive and decide to buy at time t; 2) strategic consumers who arrive and decide to buy at time t; 3) strategic consumers who arrive at time $t-1$ but wait for one more period and finally decide to buy at time t. Specifically,

$$d_{mt} = (1-\theta)q(p_{mt}, X_m, t) + \theta(1-z_{mt})q(p_{mt}, X_m, t) + \theta z_{m,t-1}q(p_{mt}, X_m, t-1) \quad (3.4)$$

[17]Estimation with two or more waiting periods is extremely expensive computationally. We tried allowing for two-period waiting under simpler expectation assumptions, perfect foresight and weak-form rational expectation. The estimated percentage is comparable while slightly smaller. However, the amount of strategic waiting, i.e., (amount of time waiting) * (number of people waiting), is almost the same.

where z_{mt} equals 1 if a strategic consumer decides to wait at time t, and 0 otherwise, and $q(\cdot)$ is the baseline demand that one would observe if all consumers were myopic. Note that the model itself does not impose the assumption of strategic behavior: only if θ is significantly different from zero will we have the evidence of strategic behavior. The baseline demand is a function of price p_{mt}, market and departure-date characteristics X_m, and the time of booking t. Note that the effective price for those consumers who arrive in the last period $t-1$ but wait until t is the current price p_{mt} rather than the last period's price $p_{m,t-1}$. That is, strategic consumers who decide to wait are not obligated to buy when they come back, and the decision will depend on the new price they see. Thus, the demand contribution from those who arrive in the previous period is $q(p_{mt}, X_m, t-1)$ rather than $q(p_{m,t-1}, X_m, t-1)$. In the following subsections, we discuss in detail the modeling alternatives, assumptions, and potential extensions of each input of this model.

Baseline Demand

Functional Form. The baseline demand $q(p_{mt}, X_m, t)$ represents the potential demand we will observe if all consumers are myopic. A common way to model demand is the *additive linear* demand model: $q(p_{mt}, X_m, t) = \alpha - \beta p_{mt} + f(t) + X_m\delta + \varepsilon_{mt}$, where $f(t)$ captures the time trend approximated, for instance, by a polynomial function of booking time t, and ε_{mt} is the demand shock. The main results of this paper will be based on the linear model, but we also present results with the nonlinear exponential demand model with multiplicative errors: $q(p_{mt}, X_m, t) = \exp\Big(\alpha - \beta p_{mt} + f(t) + X_m\delta + \varepsilon_{mt}\Big)$.

Endogeneity of Price and Structure of Demand Shocks. Even though we control for market-departure-date characteristics, prices may still be endogenous in this setting. First, pricing managers who monitor the demand and prices have better knowledge about the local demand than we do as econometricians. For example, when there is a special event, such as a conference or a convention, managers might adjust prices for that departure date and city accordingly. Second, pricing managers adjust prices based on previously realized demand shocks, and if demand shocks are auto-correlated, prices will be correlated with contemporaneous demand shocks as well. These particular features of the air-travel industry make price endogeneity a more prominent issue than in many other industries, such as the one studied in Hendel and Nevo (2011) where demand is more stable over time. Failure to address these endogeneity issues will result in biased estimates of price sensitivity, and hence the estimates of other parameters of the model such as the fraction of strategic consumers. The direction of this bias can go both ways. If price endogeneity is not properly accounted for, as usual, price sensitivities will be underestimated. During price drops, a part of the incremental demand caused by the price sensitivity of myopic consumers will be attributed to strategic consumers, so the fraction of strategic consumers will be overestimated. During price surges, however, the observed decrease in demand is smaller with strategic consumers than without them. Since the potential decrease in demand without strategic consumers is underestimated, the part attributed to the strategic consumers, that is, potential decrease minus observed decrease, will be underestimated. The overall effect is ambiguous. To address the endogeneity, we allow for the following structure of demand shocks:

$$\varepsilon_{mt} = \mu_m + \epsilon_{mt}, \tag{3.5}$$

$$\epsilon_{mt} = \rho\epsilon_{m,t-1} + \nu_{mt}, \tag{3.6}$$

where the demand shock ε_{mt} is decomposed to a market-departure-date specific shock, μ_m, and a serially correlated shock governed by AR(1) process, ϵ_{mt}. Price p_t is allowed to be correlated with the demand shock ε_{mt} through correlation with μ_m and $\epsilon_{m,t-1}$. The remaining part of the demand shock, ν_{mt}, is a pure noise, i.i.d across markets, departure dates and booking periods, and uncorrelated with other observables, including prices.

Heterogeneity in Price Sensitivities between Myopic and Strategic Consumers. One might expect strategic consumers to be more sensitive to prices than non-strategic consumers. To allow for this possibility, let β_n and β_s denote the price sensitivities of non-strategic and strategic consumers, respectively. The demand can be written as follows,

$$\begin{aligned}
d_{mt} &= (1-\theta)q_n(p_{mt}, X_m, t) + \theta(1-z_{mt})q_s(p_{mt}, X_m, t) \\
&\quad +\theta z_{m,t-1}q_s(p_{mt}, X_m, t-1), \\
q_n(p_{mt}, X_m, t) &= \alpha - \beta_n p_{mt} + f(t) + X_m\delta + \varepsilon_{mt}, \\
q_s(p_{mt}, X_m, t) &= \alpha - \beta_s p_{mt} + f(t) + X_m\delta + \varepsilon_{mt}.
\end{aligned}$$

Now we need to be careful about the interpretation of the fraction θ. In the previous case where strategic and non-strategic consumers have the same price sensitivities, the fraction of strategic consumers is constant regardless of prices. However, the observed fraction will change with prices if different price sensitivities are permitted. The higher the price, the fewer strategic consumers we expect to observe. Thus, the fraction θ represents the relative scale of strategic consumers at price zero. We define the observed fraction of strategic consumer $\theta_{obs,t}$ at time t as $\theta_{obs,t} = q_s(p_{mt}, X_m, t)/q_n(p_{mt}, X_m, t)$.

Consumer Expectations.

Strategic consumers' waiting decision is based on their beliefs about future prices. Note that price levels also manifest the risk of stock-out thanks to the practice of revenue management. Bid prices, as approximated by the lowest prices in our model, reflect the level of remaining seat inventory or the probability of stock-out at different booking occasions. We model consumers' waiting decision under three different circumstances with a decreasing level of consumer sophistication: perfect foresight, strong-form rational expectations and weak-form rational expectations. Under *perfect foresight* consumers predict future prices perfectly. Under *rational expectations*, consumers cannot predict the exact price individually, yet as a group they predict the future price distribution correctly. The distinction between strong form and weak form is that, under the *strong-form*, consumers consider airlines' pricing strategy when forming the expectation, while under the *weak-form*, the expectation is based on historical information only.

Perfect Foresight. At time t, strategic consumers know the exact future price $p_{m,t+1}$. Though unrealistic, this model can be used as a benchmark. Consider a consumer with utility of traveling denoted by ϕ. She will decide to wait if $\phi - p_{m,t+1} > \phi - p_{mt}$, or $p_{m,t+1} < p_{mt}$, that is, if price drops in the next period. Note that time discounting is negligible in this setting since we are looking at relatively short time periods such as days or weeks. Moreover, utility of air-travel does not depend on the time of purchase: the product is always consumed on the day of departure. As a result, a consumer's purchase-or-wait decision is not dependent on the value of the product ϕ. Moreover, when a consumer anticipates future prices perfectly, risk attitude will not be a determinant of the purchase-or-wait decision.

Weak-Form Rational Expectation. A strategic consumer i makes a prediction of the future price $p_{m,t+1}$ at time t based on information available to her at time t, i.e., I_{mt} and a personal shock, i.e., o_{imt}, $\tilde{p}_{i,m,t+1} = E[p_{m,t+1}|I_{mt}] + o_{i,mt}$, where $E[p_{m,t+1}|I_{mt}]$ is the expectation of future prices given information set I_{mt}. Information set I_{mt} includes, for instance, historical prices, price volatility, and market departure date characteristics. Under the rational expectation assumption, the distribution of consumer belief $\tilde{p}_{i,m,t+1}$ is the same as that of the true conditional distribution of future price $p_{m,t+1}|I_{mt}$. Therefore, the probability that strategic consumers wait will be the same as $Pr(p_{m,t+1} < p_{mt}|I_{mt})$, which can be estimated using a Logit or Probit model.

Strong-Form Rational Expectation. Consumers now consider airlines' pricing strategy when forming expectation of future prices. Consider the following game played by the airline and the consumers. Stage 1: at the beginning of time t, baseline demand shock ν_{mt} is realized, and consumers arrive and see the p_{mt}. Myopic consumers make their purchase based on p_{mt}. Each strategic consumer forms her own prediction of future price and makes the purchase-or-wait decision. Total demand d_{mt} is thus realized. Stage 2: at the beginning of time $t+1$, a supply shock denoted by $\omega_{m,t+1}$ is realized and airlines make a decision on price $p_{m,t+1}$ based on the previous demands $d_{mk}, k = 1, 2, ..., t$, the observed supply shock $\omega_{m,t+1}$ and other inputs.

In the rational expectations equilibrium, strategic consumers collectively anticipate the true price distribution. Consumers do not necessarily need to know the exact demand shock ν_{mt}. As long as it is realized, this information will not be wasted in an efficient market. The only source of variation for the price $p_{m,t+1}$ at the time when consumers form expectations is the unrealized supply shock. With this assumption, the equilibrium belief of future price distribution and probability of waiting can be found through estimating the supply model, which we describe in the next section.

Under either weak-form or strong-form rational expectation, once we obtain the probability of strategic consumers waiting Pr_{mt}, we can replace z_{mt} with Pr_{mt} in Equation 3.4. We note that the rational expectations model is likely to provide higher estimates of the fraction of strategic consumers than the perfect foresight model. The reason is that, under the perfect foresight, every strategic consumer anticipates price drops correctly and waits. Under the rational expectations, only a portion of strategic consumers correctly anticipate price drops. Thus, the same amount of demand shifting observed in the data caused by strategic waiting corresponds to more underlying strategic consumers under the rational expectations assumption.

Fraction of Strategic Consumers.

So far we have assumed the fraction of strategic consumers to be time-independent. However, strategic consumers may be more likely to arrive at certain stages over the booking horizon. We estimate a vector of fractions for the entire booking horizon non-parametrically, $\tilde{\theta} = [\theta_1, \theta_2, ..., \theta_{T-1}]$, where θ_t represents the fraction of strategic consumers arriving at the booking time t, and T is the final booking period. Note that no strategic consumers arrive in the last period before departure, i.e., $\theta_T = 0$, which is also true in the case of constant fraction.

3.5.2 The Supply Model

Recall that one reason to specify the supply model (or pricing strategy) is to estimate the demand model under the strong-form rational expectations. Another reason is that we need to know the firm's pricing strategy in order to compute new equilibrium in the counterfactual analysis as we change certain parameters. Ideally, the most desired supply model would mimic the exact network-based dynamic inventory control algorithms and demand forecasting strategies used in the airline industry, and take into account the effect of competition. Such a dynamic game is far too complex to be amenable both analytically and computationally within the scope of this paper. Moreover, this approach would invoke an assumption on firms' current conduct — that they are profit maximizers and already make the optimal decisions. This may not be the case (e.g., we know airlines do not currently take strategic consumers into account), and, more important, this assumption would make our pricing recommendations redundant. Instead, we take an empirical approach to approximate the firms' current equilibrium pricing strategy, as many other structural models do (e.g., Nair 2007). Note that short-term price fluctuations in the airline industry are not based on costs, as fixed costs are sunk and marginal costs are minimal. The key determinants of prices are the remaining seat inventory and demand forecasts (see Gallego and van Ryzin 1994). When restricted to one particular market, we can account for the remaining inventory level using the cumulative demand since the total capacity is largely fixed in the short term. We therefore model the price as a function of previous prices and demands, market and departure date characteristics, and supply shocks. We find that lag-1 price, initial price p_{m0}, and cumulative demand up to time t, together with other controls, are able to explain nearly 90% of the price variations for most markets. We therefore choose the following parsimonious model on the supply side:

$$p_{m,t+1} = \gamma_p p_{mt} + \gamma_d cum\_demand_{mt} + \gamma_0 p_{m0} + X_{m,t+1}\gamma + \omega_{m,t+1}, \omega_{m,t+1} \sim N(0, \sigma_{m,t+1}),$$

where $\omega_{m,t+1}$ is a random supply shock, which we assume to be uncorrelated with the demand shock, as the fixed component (μ_{mt}) and the autocorrelated component ($\rho\epsilon_{mt}$) of the demand shock are accounted for by lag-1 price, cumulative demand, and initial price. The remaining part of time $t+1$ demand shock ($\nu_{m,t+1}$) can be assumed to be uncorrelated with $\omega_{m,t+1}$. We do not find evidence of serial correlation in the residuals after controlling for the lag-1 price. However, we do find evidence of heteroscedasticity in the residuals over booking time, so we allow the variation of the supply shock (σ_{mt}) to change with booking time.

It is important to note at this point how the structural model outperforms the reduced-form regressions in addressing the econometric issues previously discussed. First, the mechanism by which expected prices affect current and future demands is explicitly specified in the structural model, rather than merely through correlations (β_1, β_3) in the reduced-form regressions. More important, the error generating process is explicitly specified as well. As the demand includes both strategic and myopic consumers, the demand shock is composed of shocks to both groups as well. The shock first affects the arrival of both strategic and myopic consumers, and then translates into the current and future demands dependent on the strategic behaviors. In the reduced-form regression, we have only one blended error term. Even if one can arbitrarily apply techniques to control for fixed effects and serial correlation in reduced-form regressions, it makes more sense to specify the underlying error generating process.

3.6 Identification and Estimation

3.6.1 Identification

As we outlined earlier, there is a significant amount of intertemporal price fluctuation present in the data: in addition to frequent changes in weekly fares, there are also cross-market, departure-date and booking-week variations. The identification of the fraction of strategic consumers θ is based on the variation of price trends. As we explain shortly, the estimation is conducted for each market separately, so the identification is based on the variation of price trajectories across booking periods and departure dates within the same market.

The presence of strategic consumers affects observed demand only when the price falls or is expected to fall. When the price increases or is expected to increase, strategic consumers behave in the same way as their myopic counterparts. In this case, variations in demand are attributed only to price elasticity. However, in the case of price reductions, variations in demand can be attributed to both price elasticity and strategic consumers. If we are able to quantify the changes in demand induced by price elasticity, the extra variations in demand can be attributed to strategic consumers. The question is: how do we identify price sensitivities and the fraction of strategic consumers separately? To illustrate this more precisely, recall the model under perfect foresight (Equation 3.4) and consider the following cases described by waiting decisions $z_{m,t-1}$ and z_{mt}, equivalent to price trends under perfect foresight:

$$\begin{aligned}
&\textit{Case 1: } z_{m,t-1} = 0, z_{mt} = 0 &\Rightarrow\quad & d_{mt} = q(p_{mt}, X_m, t),\\
&\textit{Case 2: } z_{m,t-1} = 0, z_{mt} = 1 &\Rightarrow\quad & d_{mt} = (1-\theta)q(p_{mt}, X_m, t),\\
&\textit{Case 3: } z_{m,t-1} = 1, z_{mt} = 0 &\Rightarrow\quad & d_{mt} = q(p_{mt}, X_m, t) + \theta q(p_{mt}, X_m, t-1),\\
&\textit{Case 4: } z_{m,t-1} = 1, z_{mt} = 1 &\Rightarrow\quad & d_{mt} = (1-\theta)q(p_{mt}, X_m, t) + \theta q(p_{mt}, X_m, t-1).
\end{aligned}$$

In Case 1, the price keeps rising from time $t-1$ to t and to $t+1$, so the observed sales consists of the baseline demand $d_{mt} = q(p_{mt}, X_m, t)$. Thus, $q(\cdot)$ can be identified. That is, price sensitivities are identified based on occasions with increasing prices. Once the baseline demand function is identified, each of the other three cases can help identify the fraction θ. Precisely, it is the variation in z, i.e., the price trend, that identifies the fraction of strategic consumers. Note, however, that the fraction θ is over-identified, since all three cases can help identification. This means that we can actually identify more parameters, which takes us to the identification of the time-variant fraction $\widetilde{\theta}_t, t = 1, 2, ..., T-1$. For example, Cases 2 and 3 can now be written as $d_{mt} = (1-\theta_t)q(p_{mt}, X_m, t)$ and $d_{mt} = q(p_{mt}, X_m, t) + \theta_{t-1}q(p_{mt}, X_m, t-1)$, respectively, which allows us to identify θ_t and θ_{t-1}.

To see if price sensitivities are still identified when we allow for different sensitivities between strategic and non-strategic consumers, we again look at different cases classified by the value of $z_{m,t-1}, z_{mt}$,

$$\textit{Case 1: } z_{m,t-1} = 0, z_{mt} = 0 \quad \Rightarrow \quad d_{mt} = -\big((1-\theta)\beta_n + \theta\beta_s\big)p_{mt} + X_{mt}\gamma,$$

$$\textit{Case 2: } z_{m,t-1} = 0, z_{mt} = 1 \quad \Rightarrow \quad d_{mt} = -(1-\theta)\beta_n p_{mt} + (1-\theta)X_{mt}\gamma.$$

From Case 2, we can identify $(1-\theta)\beta_n$, and $(1-\theta)\gamma$, but not θ, β_n, γ separately. From Case 1, we are able to identify γ from the variation of X_{mt}[18]. Together with $(1-\theta)\gamma$ identified in Case 2, we are able to identify θ. Since Case 1 also identifies $\big((1-\theta)\beta_n+\theta\beta_s\big)$,

[18]For simplicity, we include all variables other than price into X_{mt}, i.e., the constant, polynomials of t and market-departure-date characteristics X_m.

together with $(1-\theta)\beta_n$ and θ, we are able to identify β_s and β_n. Similarly, θ is again over-identified, since there are other cases based on $z_{m,t-1}, z_{mt}$ which help us identify θ as well. These over-identification conditions can again be utilized to identify the vector $\widetilde{\theta}_t, t = 1, 2, ..., T-1$.

So far we have discussed the identification for models under perfect foresight. In fact, the same logic also carries over to models under rational expectations since the only difference is that z_{mt} becomes a probability rather than a dichotomous variable. The variation in z (now Pr) still serves as the source of identification for θ. More precisely, under rational expectations, we first fit a model to predict strategic consumers' probability of waiting, and then use the variation in this probability to identify the fraction of strategic consumers. In summary, we obtain the identification because strategic consumers behave differently under different expectations of future prices.

3.6.2 Estimation

Note that our demand model is nonlinear in its *parameters* $(\theta, \beta, \gamma, \rho)$. The moment condition used in estimation has the mean independence property[19], $E[\nu_{mt}|X_m, p_{mt}] = 0$, which guarantees consistent estimators (see Wooldridge 2010). The key challenges in estimation are 1) finding the global optimal solution to a minimization problem where the first order conditions are not linear in parameters, and 2) accounting for fixed effects

[19] After controlling for fixed effects and serial correlation, it is relatively reasonable to assume that price is uncorrelated with the remaining part of the error term. We do not find evidence of correlation between the residuals ($\hat{\nu}_{mt}$) and prices (with a correlation less than 0.01). We also considered using previous cumulative demand and previous demand as instruments. However, they sometime show stronger correlation with the error terms than prices.

and serial correlations in error terms in the nonlinear minimization problem. We design an algorithm and describe it in the appendix. Briefly, the key idea is to first transform the nonlinear problem into a linear problem to address the error structure given a partial set of parameters, and then minimize over this parameter set.

We use bootstrapping to obtain the standard errors and confidence intervals.To guarantee that all booking periods for a particular departure date in a market are selected or not selected as a whole, we use clustered bootstrapping. That is, the resampling is done at the level of market-departure-date rather than booking period. Although adjacent departure dates in the same markets might not be completely independent, the correlation is of less concern compared to the correlation of booking periods for the same departure date. In our data, 250 rounds of bootstraps are sufficient to obtain convergence in standard error estimations. However, it takes substantially more rounds (1,000) to obtain an accurate estimate of the confidence interval. Since the estimates of the fraction of strategic consumers are bounded within $[0, 1]$ and the bootstrapping distributions of these fractions are largely asymmetric, it is important to bootstrap the percentile confidence intervals rather than to compute the confidence intervals from standard errors.

We perform estimation market by market since each market demonstrates significantly different patterns of time trends, seasonality, day-of-week effects and price sensitivities. Pooling all markets together without accounting for these differences would result in misspecification of the baseline demand model, which would further bias the estimation of the fraction of strategic consumers. To mitigate this effect we would need hundreds of market-specific coefficients. Instead, the estimation is more accurate and efficient when it is performed for each market separately. On average, it took one to five hours to run 1,000 bootstraps for each market, depending on the model specification. This amounts to 100 to 500 hours of computation time for each model specification run over all markets. The long running time is largely due to the non-linearity of the model, randomized initialization to guarantee global optimality, and the large number of bootstraps. We code the program in MATLAB and we run it on the Wharton Grid Computing Platform (a 20-node, 80-CPU Linux grid and cluster environment).

Variables

In this section, we discuss the variables used in the aforementioned models. We provide a list of these variables and their descriptions in Table 3.3.

Baseline Demand Model. The baseline demand is affected by the following factors. 1) Price. We use the lowest daily price among multiple candidates for the price measure. This is a good approximate price point since we care most about strategic consumers who are presumably in search of the lowest prices. On most booking days, there is no obvious deviation between the lowest price and, say, the average of the three lowest prices, except for a few days close to departure. In addition, lowest daily prices are also a commonly used measure in industry practices (e.g., Farecast) and related papers (e.g., Mantin and Gillen 2011). To aggregate daily price points to the weekly level, there are at least two options — the minimum and the average of the lowest daily prices within a week. Significant differences mostly appear in the final week before departure, when prices change rapidly from day to day. For this reason we use the average of the lowest daily prices as the measure of the weekly price; otherwise, the price sensitivity in the final weeks before departure is not well captured. 2) Booking time: booking week and its polynomials, and the final booking week before departure. We use booking week t, t^2, t^3 because the incremental explanatory power is minimal when adding higher polynomial terms. 3) Departure-date characteristics. Due to seasonal demand patterns, some departure dates have higher demand than others. For example, during the vacation seasons leisure destinations experience many more travelers than they normally do. We control for demand seasonality by identifying the high-demand season from the data, which corresponds roughly to the spring break period. Seasonality is more obvious for leisure destinations than for business destinations. We also control for the day-of-week effect of the departure date using dummies. Departures tend to cluster around weekdays in business markets, and around weekends in leisure markets.

Weak-Form Rational Expectations. In this case the predictors of future prices include historical prices and other information available to the decision makers. 1) Relative fares. The current fare relative to the fare of the last period, to the initial fare, and to the average market fare obtained from the DB1B dataset. 2) Price volatility. The coefficient of variation of daily fares within a week[20]. 3) The initial price. Initial prices are highly correlated with demands across departure dates, which reflects the fact that

[20]Using standard deviation and coefficient of variation with or without first-differencing yield similar results.

fare managers have some knowledge of demand when making the initial pricing decision. 4) Booking time and departure date characteristics.

The Supply Model, Strong-Form Rational Expectations. The variables include last period price, cumulative demand, the initial price, booking time and departure date characteristics.

3.7 Results

Our estimations provide consistent findings of strategic consumers across markets under various model specifications. We first illustrate these findings using different model specifications applied to two representative markets, and then we summarize them across all markets. The representative markets include one leisure market, labeled as L, with Orlando, Florida as the destination and one business market, labeled as B, with Atlanta, Georgia as the destination. The origins of both markets are disguised for confidentiality reasons.

3.7.1 Results under Different Baseline Demand Models

In Table 3.5 we compare three different baseline demand models under perfect consumer foresight: 1) the linear model, 2) the linear model with correction for price endogeneity, and 3) the nonlinear exponential model[21]. As we show in the table, all models lead

[21] We also estimate a nonlinear model with correction for price endogeneity. The performance of this model turns out to be almost the same as for the basic nonlinear model, but it takes four to five times longer to estimate due to the high level of non-linearity.

to similar estimates of the fraction of strategic consumers, about 3% to 5% for this particular market under perfect foresight. (3.6% under the linear model, 5.2% under the linear model with endogeneity corrections, and 2.5% under the nonlinear model.) We select the linear model with endogeneity correction as the basis for subsequent analysis since it addresses the potential biases associated with the basic linear model and is computationally more efficient than the nonlinear model. It turns out that the linear model with endogeneity correction has the best model fit as well in terms of R-square, explaining 39% of the variation on average, and in some markets as high as 60% to 70%. Comparing the results of the two linear models, we find that, as expected, price sensitivity is underestimated when the endogeneity issues are not addressed: without correction for endogeneity the price sensitivity is -0.174, while controlling for endogeneity it is -0.224. In this particular market, this further leads to an underestimation of the fraction of strategic consumers, i.e., 0.036 as compared to 0.052. In other markets, we observe overestimation of the fraction.

The signs of other coefficients are in line with our expectations. We see that the final week before departure has a strong positive effect on demand, especially in business markets. Departure date characteristics (seasonality and day-of-week effects) also have significant effects on total demand. In many markets, the estimated serial correlation between adjacent demand shocks is large and significant (mildly significant in this market though), highlighting the importance of controlling for serial correlation.

Table 3.5: Compare Different Baseline Demand Models (perfect foresight): Market L

	(1) **linear**	**(2)** **linear + fixed effect + AR1**	**(3)** **nonlinear**
fraction	0.036	0.052	0.025
95% confidence interval	[0.000, 0.114]	[0.001, 0.122]	[0.000, 0.112]
price	-0.174	-0.224	-0.005
	(0.038)	(0.031)	(0.001)
booking time t	-0.107	-10.293	0.196
	(3.300)	(3.611)	(0.136)
t^2	0.811	2.252	-0.003
	(0.572)	(0.618)	(0.025)
t^3	-0.047	-0.109	0.000
	(0.028)	(0.030)	(0.001)
final week	8.695	16.266	0.234
	(4.063)	(4.201)	(0.106)
high season	46.345		1.159
	(11.125)		(0.324)
day-of-week dummies	yes		yes
const	49.520		3.323
	(9.937)		(0.536)
AR1(ρ)		0.117	
		(0.064)	
R-square	0.2470	0.3937	0.2478

3.7.2 Results under Different Consumer Expectation Assumptions

Now we move on to compare the estimated fractions of strategic consumers under different assumptions about their expectations of future prices; we present the results in Table 3.6. In this market we find persistent evidence of strategic consumers regardless of the assumptions about expectations of future prices. Under the benchmark model, i.e., perfect foresight, the fraction of strategic consumers is significant at 5.2%. Under the rational expectations assumption, the estimates are higher, as we expected: 29.2% and 38.5% under strong-form and weak-form rational expectation, respectively. All estimates are statistically significant. Naturally, the results from the rational expectations

Table 3.6: Results under Different Consumer Expectation Assumptions: Market L

	(1) perfect foresight	(2) strong-form rational expectation	(3) weak-form rational expectation
fraction of strategic consumers	0.052	0.292	0.385
95% confidence interval	[0.001, 0.122]	[0.151, 0.541]	[0.198, 0.394]
price	-0.224	-0.224	-0.214
	(0.031)	(0.032)	(0.031)
booking time t	-10.293	-10.160	-6.814
	(3.611)	(3.612)	(3.542)
t^2	2.252	2.328	1.781
	(0.618)	(0.619)	(0.598)
t^3	-0.109	-0.117	-0.091
	(0.030)	(0.030)	(0.029)
final week	16.266	20.340	17.263
	(4.201)	(4.856)	(4.416)
AR1(ρ)	0.117	0.033	0.045
	(0.064)	(0.073)	(0.080)
R-square	0.3937	0.4073	0.4215

model are more realistic since it is not possible for consumers to predict future prices perfectly. One also needs to be cautious when comparing these fractions: although the estimate under rational expectations can be significantly higher than that under perfect foresight, it does not necessarily mean more strategic waiting. Ultimately, the demand shift observed due to strategic consumers is also dependent on consumers' ability to make predictions of future prices. Since some consumers are not able to correctly anticipate future prices, they would not behave "strategically" when they should. However, there is a possibility that with the proliferation of web tools such as fare charts and fare alerts, consumers are becoming better at making price predictions. The demand shifting effect will become more prominent in this case. Estimates of other control variables are consistent across three models and are of expected signs. The results of the prediction models used in the weak-form and the strong-form rational expectations are shown in Appendix Table 5.3.

3.7.3 Heterogeneity in Price Sensitivities

Strategic consumers may have different price sensitivity from myopic consumers. As we see in the identification section, estimations of price sensitivity and of the fraction of strategic consumers are closely related to each other. Therefore, there may be a concern that the fraction of strategic consumers we identify could be partially driven by the different price sensitivities of these two consumer groups. To investigate this possibility, we estimate different price sensitivities for strategic and myopic consumers. The results are shown in Table 3.7. For the leisure Market L, we do not observe significant differences in the price sensitivities among these two groups, and hence the estimates of strategic fractions do not change much. However, for the business Market B, we do observe a significant difference in price sensitivities among these two groups. As we discussed earlier, the observed fraction θ_{obs} of strategic consumers at a particular price point is not the same as the parameter θ. In Market L, the two estimates are similar since the difference between the two price sensitivities is small. In Market B, however, θ_{obs} is lower than θ.

Finally, we summarize the estimation results over all markets in Table 3.8 using histograms. For each market in our data, we estimate the strong-form rational expectation model under the same price sensitivity and different price sensitivities. We obtain on average 24% strategic consumers under the former and 17% under the latter. We also find a sizable heterogeneity across markets, with standard deviations of 12% and 11%, respectively. To investigate how strategic behavior is affected by the market character-

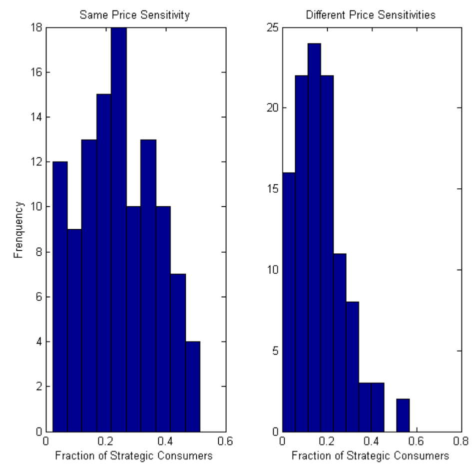
Table 3.7: Allowing for Different Price Sensitivities between Myopic and Strategic Consumers

	perfect foresight	**strong-form rational**	**weak-form rational**
		Market L	
fraction	0.094	0.275	0.378
95% conf. int.	[0.008, 0.182]	[0.106, 0.391]	[0.158, 0.548]
observed fraction	0.038	0.286	0.351
95% conf. int.	[0.004, 0.097]	[0.127, 0.435]	[0.174, 0.511]
price			
non-strategic (β_n)	-0.228	-0.229	-0.216
	(0.032)	(0.031)	(0.030)
strategic (β_s)	-0.311	-0.187	-0.206
	(0.335)	(0.125)	(0.066)
t	-9.773	-10.332	-6.805
	(3.609)	(3.745)	(3.243)
t^2	2.169	2.354	1.778
	(0.622)	(0.640)	(0.562)
t^3	-0.105	-0.118	-0.091
	(0.031)	(0.031)	(0.028)
last week	13.219	22.111	17.378
	(5.077)	(6.808)	(6.358)
const	0.094	0.044	0.050
	(0.062)	(0.063)	(0.074)
R-Square	0.3951	0.4076	0.4216
		Market B	
fraction	0.107	0.186	0.120
95% conf. int.	[0.002, 0.201]	[0.004, 0.367]	[0.007, 0.250]
observed fraction	0.036	0.048	0.026
95% conf. int.	[0.000, 0.093]	[0.000, 0.142]	[0.000, 0.105]
price			
non-strategic (β_n)	-0.013	-0.024	-0.026
	(0.032)	(0.017)	(0.016)
strategic (β_s)	-0.131	-0.201	-0.341
	(0.364)	(0.392)	(0.510)
t	6.365	4.474	4.285
	(2.135)	(1.476)	(1.354)
t^2	-1.219	-0.986	-0.971
	(0.335)	(0.254)	(0.241)
t^3	0.085	0.077	0.077
	(0.017)	(0.013)	(0.013)
last week	2.951	2.341	2.643
	(4.969)	(4.745)	(4.705)
const	0.488	0.572	0.587
	(0.102)	(0.065)	(0.059)
R-Square	0.5965	0.609	0.6108

Table 3.8: Fraction of Strategic Consumers over All Markets (strong-form rational expectation)

fraction	**same price sensitivities**	**different price sensitivities**
average (std)	0.243 (0.126)	0.171 (0.112)
range	[0.023, 0.514]	[0.003, 0.567]
[5th percentile, 95th percentile]	[0.049, 0.449]	[0.022, 0.368]
# of significant estimates	51	57
total # of markets	111	111

histogram



istics, we regressed the fraction (the logistic transformation of the fraction) on market characteristics such as level of competition, presence of low-cost carriers, distance and origin demographics, destination types (business vs. leisure), distribution channel (online vs. offline), and market size. The results are displayed in Table 3.9. We find that markets with shorter distance, higher income at the origin city, and smaller market size tend to have more strategic consumers. Given the number of markets we have, our ability to expand this analysis is limited.

Table 3.9: Explaining Strategic Behavior Using Market Characteristics

	same price elasticities	different price elasticities
business market (avg % of full fare)	1.907	1.572
	(1.358)	(1.410)
HHI	1.099*	1.023
	(0.597)	(0.619)
low cost carrier market share	-0.033	-0.170
	(0.792)	(0.822)
% of tickets distributed online	1.490*	1.443
	(0.845)	(0.877)
intl destination	0.356	0.500
	(0.317)	(0.329)
ln(distance)	-0.407**	-0.382**
	(0.153)	(0.159)
ln(origin population)	0.162	0.079
	(0.137)	(0.142)
ln(origin per capita income)	1.386*	1.652**
	(0.724)	(0.751)
ln(market size)	-0.240**	-0.227**
	(0.109)	(0.113)
constant	-14.503**	-16.333**
	(7.132)	(7.404)
# of obs (# of markets)	111	111
R-square	0.2372	0.2026

***:p-value*$<$ 0.05*;*:p-value*$<$ 0.1

3.7.4 Time-Variant Fraction of Strategic Consumers

To investigate the arrival pattern of strategic consumers over booking time, we estimate non-parametrically a time-variant vector of the fractions of strategic consumers. To show the general pattern, we group the results by destination, since the trend of the fraction over time for routes with the same destination are similar. Figure 3.1 shows the fraction of strategic consumers over twelve booking periods before the final booking period, estimated under strong-form rational expectation and with heterogeneity in price sensitivities. As we see in the figure, the fraction of strategic consumers is higher at the beginning of the booking horizon, that is, when risk and waiting costs are lower, or close to departure, likely in search of last-minute deals. This pattern is less prominent in popular markets such as Las Vegas (LAS) and Orlando (MCO), which is in line with

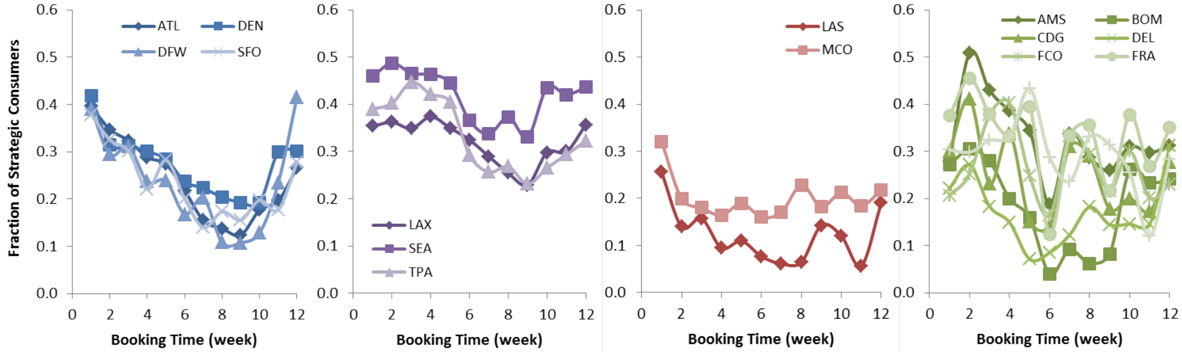


Figure 3.1: Fraction of Strategic Consumers Varying with Booking Time (grouped by destination)

the observation that last-minute deals are often offered in less-popular markets where there is distressed inventory. Those consumers who arrive for the last-minute deals are also flexible in their travel schedules, so they can decide to travel without much advance planning if the price is low, or decide not to travel if the price is high.

3.8 Counterfactual Analysis

Now that we have robust evidence of strategic consumers, what are the revenue implications? What strategies should airlines take in response? We answer these questions using counterfactual analysis. We focus on 1) the revenue impact of the presence of strategic consumers, and 2) alternative pricing schemes to eliminate strategic waiting.

3.8.1 Revenue Impact of Strategic Consumers

The commonly held belief regarding strategic consumers is that they pose a threat to firms' revenues (see Anderson and Wilson 2003, Aviv and Pazgal 2008, Levin et al. 2009).

Surprisingly, we find from our counterfactual analysis that this is not always true. As a matter of fact, strategic waiting may sometimes benefit the seller, an effect that is in line with the predictions by Su (2007) and Cho et al. (2008). Table 3.10 illustrates this through two examples, based on 100 simulations using the most complex model specification with heterogeneity in price sensitivity and time-variant fraction of strategic consumers. The results hold across other model specifications as well. Under rational expectations, each simulation involves computing a new equilibrium of consumer beliefs and actual future prices. We ask the following question: how much will revenues change if all consumers turn from myopic to strategic? To isolate the effect of price sensitivities and the effect of being strategic, we let myopic consumers become strategic without changing their original price sensitivities. As shown in Column 5, for business markets such as Market B, revenue decreases by 0.40% to 1.01%, while in leisure markets such as Market L, revenue increases by 0.22% to 1.39% (with small standard deviations in both markets). This pattern is not specific to these two example markets but holds in general, as we demonstrate shortly.

To explain the phenomenon, consider the two effects of strategic consumers: the price-reduction effect and the demand-increasing effect. Strategic consumers drive down prices, and as a result firms cannot charge high prices even to high-value customers. Yet strategic consumers drive up demand in two ways. While the first is obvious — demand is higher since prices are lower— the second is more subtle. Some low-value customers will not buy at the price when they first arrive. The flexibility of being able to come back in later periods offers them more purchasing chances, and they may end up buying.

Which effect dominates depends on the heterogeneity in the composition of high-value and low-value customers in the market. In business markets with proportionally more high-value customers, the price-reduction effect dominates, whereas in leisure markets

Table 3.10: Counterfactual Analysis

	daily	(1) all myopic	(2) current	(3) all strategic	(4) monotone price	(5)= (3)-(1) strategic - myopic	(6)=(4)-(2) gains from monotone price
				Market L (leisure)			
perfect foresight	**revenue ($1,000)**	**102.2**	**102.45**	**103.62**	**98.71**	**1.39% (0.16%)**	**-3.66% (0.43%)**
	demand (# of pax)	405.66	407.09	423.36	356.39	4.36%	-12.45%
	weighted price ($)	252.02	251.75	244.84	277.04	-2.85%	10.06%
	price ($)	242.97	243.03	242.19	261.1	-0.32%	7.44%
strong-form rational expectation	**revenue ($1,000)**	**104.99**	**104.88**	**105.22**	**102.34**	**0.22% (0.08%)**	**-2.42% (0.36%)**
	demand (# of pax)	414.25	414.11	417.88	368.61	0.88%	-10.98%
	weighted price ($)	253.51	253.34	251.87	277.71	-0.65%	9.63%
	price ($)	243.79	243.59	241.77	261.46	-0.83%	7.34%
				Market B (business)			
perfect foresight	**revenue ($1,000)**	**55.359**	**55.377**	**54.801**	**56.954**	**-1.01%(0.19%)**	**2.85%(0.34%)**
	demand (# of pax)	197.61	197.77	200.75	190.71	1.59%	-3.58%
	weighted price ($)	280.14	280	273	298.65	-2.55%	6.67%
	price ($)	239.69	239.69	238.92	260.97	-0.32%	8.88%
strong-form rational expectation	**revenue ($1,000)**	**55.601**	**55.59**	**55.377**	**57.585**	**-0.40%(0.11%)**	**3.59%(0.39%)**
	demand (# of pax)	198.41	198.48	199.03	193.18	0.31%	-2.68%
	weighted price ($)	280.2	280.05	278.2	298.09	-0.71%	6.45%
	price ($)	239.85	239.76	238.71	260.86	-0.48%	8.80%

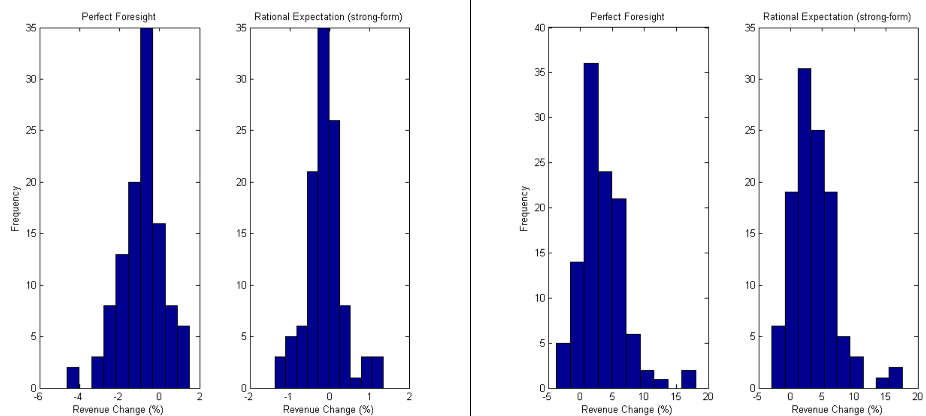
Note: Standard deviations are shown in parentheses.

with proportionally more low-value customers, the demand-increasing effect dominates. Table 3.10 illustrates the dominant effect for each market. The average price decreases by 2.55% in business market B, similar to that (2.85%) in leisure market L assuming perfect foresight (or 0.71% compared to 0.65% assuming rational expectation); however, demand increases by only 1.59% in business market B, significantly smaller than the 4.36% increase in leisure market L (or 0.31% compared to 0.88% assuming rational expectation).

We present summaries and histograms of the impacts of strategic consumers on revenues across all markets in the left panel of Table 3.11. The ranges of revenue impacts are [-2.7%, 1.0%] under perfect foresight and [-1.0%, 0.8%] under rational expectations, as measured by the 5th and 95th percentiles. Since our data includes inventory classes, we are able to compute the percentage of consumers who purchased full fares (inventory class Y and B) as a measure of the extent to which the route is a business route. This measure is negatively correlated with the revenue impact. The coefficient of correlation is -0.366 and -0.216 under perfect foresight and rational expectations, respectively, suggesting that the more "business" the route is, the more likely it is that the presence of strategic consumers will have a negative effect on the revenues. Plus, we find no significant effect of competition when regressing the revenue changes on various market characteristics. Table 3.10 also shows that the effects of strategic consumers tend to be more negative when they are more sophisticated in their prediction of future prices.

Table 3.11: Counterfactual Analysis Summarized across All Markets

	revenue change (strategic - myopic)		**revenue gains from monotone price**	
	perfect foresight	**rational Exp. (strong-form)**	**perfect foresight**	**rational exp. (strong-form)**
percentile				
5%	-2.7%	-1.0%	-1.4%	-1.1%
50%	-0.8%	-0.1%	2.8%	3.1%
95%	1.0%	0.8%	9.2%	10.6%
mean	-0.9%	-0.1%	3.4%	3.6%
std	1.1%	0.5%	3.6%	3.5%



3.8.2 Non-Decreasing Price Commitment

One common approach to eliminate strategic waiting is to commit to fixed or non-decreasing price paths, which is the practice often used by low-cost carriers. For example, Southwest has much more transparent and simpler fare structures than the legacy airlines, and their prices only go up as the departure date approaches, except for some occasional temporary sales. The question is: would it be better for airlines to commit to non-decreasing prices? Computing the optimal non-decreasing price path is itself a question requiring separate analytical efforts beyond the scope of this paper. Instead, we examine the impact of a heuristic non-decreasing price scheme which adds a small twist to the current pricing strategy. We take the current pricing strategy and use it to predict a candidate price point for the next period. We set the future price as the maximum of this candidate price and the current price to guarantee that it is non-decreasing for all airlines.

Column 6 in Table 3.10 shows that for business markets where strategic consumers are undesirable, commitment to non-decreasing price ensures higher revenues — 2.9% to 3.6% higher in this example. However, on the leisure markets, such commitment would not be beneficial to the firms: revenues decrease by 2.4% to 3.7% in the example. The impacts of committing to non-decreasing price schemes across different markets are presented in the right panel of Table 3.11. The ranges of the revenue gains are [-1.4%, 9.2%] under perfect foresight and [-1.1%, 10.6%] under rational expectations, as measured by the 5th and 95th percentiles. The coefficient of correlation between the measure of how business-like the market is and the revenue gains is positive, i.e., 0.249 and 0.220 under perfect foresight and rational expectations, respectively. This suggests that the more business-like the route is, the more likely it is that airlines will benefit from a non-decreasing price commitment.

3.9 Conclusions

We provide evidence of strategic consumers in the air-travel industry, and this evidence is robust to various modeling assumptions. We obtain an estimate of 4.9% to 44.9% proportion of strategic consumers on average across markets under rational expectation assumption. Contrary to the predominant belief, our counterfactual analysis shows that strategic consumers do not always hurt revenues; the effect differs by market type. In business markets, where a large proportion of consumers are relatively price-inelastic, high value consumers, the presence of strategic consumers tends to drive down the total revenue through lower prices. However, in leisure markets, the presence of strategic consumers may boost market revenue by inducing higher demand. Therefore, commitment to non-decreasing prices is more likely to be beneficial in business markets than in leisure markets. The median revenue improvement on those markets that benefit from a non-decreasing pricing strategy is 3.5%, with quartiles 1.8% and 5.6%.

Our results have important implications for both theory and practice. In many industries with significant price fluctuations over time, it is crucial to model demand with consumers' inter-temporal choices, and we propose a structural model for this purpose. Failing to do so will result in suboptimal pricing or inventory decisions. Practically, our counterfactual analysis provides important guidance to airline managers: improve demand forecasts accounting for inter-temporal demand substitution, assess the impact of strategic consumers, and decide in what circumstances it would be desirable to eliminate strategic waiting and what is the potential benefit. Our approach can also be applied to many other industries to help firms figure out whether their business is subject to the presence of strategic consumers, and what action they should take in response.

Naturally, our study is not free of limitations, which offer many avenues for future research. As we discussed earlier, we do not model details of inter-firm competition, which merits a separate paper, although our simple controls for competition through

Herfindahl Index were insignificant. Other most obvious restrictions and assumptions are a single waiting time period (rather than multiple); a discrete, two-point distribution of strategic consumer behavior (rather than a continuous distribution); and weekly aggregation of the data. We also conduct only limited counterfactual analysis: a separate study might attempt to devise an entirely new optimization algorithm to maximize revenues. These and other extensions will inevitably run into computational difficulties, and resolving them will be part of the challenge.

Chapter 4

Who are My Competitors? Let the Customer Tell

4.1 Introduction

With whom you compete? Many competitive and fragmented industries, such as the hotel industry, are challenged by this question. The US hotel and motel industry consists of about 40,000 companies that operate about 50,000 properties. 50 largest companies generate only 45 percent of total revenue. Take New York City for example, where more than 500 hotels compete for business in the city and its vicinity. Besides the sheer number of potential competitors, hotels also compete on multiple dimensions, both spatially (i.e., locations) and vertically (i.e., quality). Increasing adoption of internet as a convenient and reliable searching and transaction channel has made the problem even more complex

- it used to be the case that hotels competed mostly with other hotels in close proximity, but now they can be competing with hotels located far away but offering appealing services and rates.

Understanding competition structure in such a market has important theoretical and practical implications. Theoretically, we do not understand quite well how firms in a market like described above compete – researchers analyzed and measured competition in markets with few competitors, e.g., duopoly and oligopoly markets, in which every player competes with each other. However, there is currently no sufficient understanding of competitive relationships in a networked market which includes both spacial and vertical differentiation. Practically, monitoring competitors' performance is a critical part of hotels' strategy development as well as daily operations. However, industry professionals still mostly depend on rules of thumb to define competition set (Comp Set), often with little support from data-driven analytical tools. Current prevalent practice to define Comp Set in the hotel industry varies from looking across the street and identifying properties that charge the same basic rates (appealing to the same price customer), to weighing and scoring property attributes. For example, Smith Travel Research(STR) advices hotels to look at the following attributes in determine competitive set: price, location, restaurant and room service in hotel, meeting space, complimentary breakfast, loyalty program, full-service amenities, and brand.

In this paper we propose a new customer-centric approach to study competition structure. At the core of our methodology is the idea that hotels should see themselves as potential customers see them, because ultimately they are competing for customers' demand. Thus, instead of asking themselves who they think they are competing with, a better question to ask is "which hotels do customers consider as competitors?" The ability to track consumer behavior online is making such approach possible because, as noted by HotelNewsNow (a professional industry website launched by STR) "[According to Forrester Research,] 90 percent of all travel-related purchasing decisions are made online. So, despite our best efforts to define who we think our competition is, it's really customer perception that drives purchasing decisions[22]." Analyzing consumer footprints has been made possible by emerging availability of online search data. Which hotels did consumers seriously consider before reaching the final decision? Which hotels did they click to see details? Whose websites did they visit? Answers to these questions can inform us about which hotels customers perceive as competitors. Note that clickstream data has actually been around for about ten years or more (e.g., Fader and Moe 2004) but both industry professionals and academic researchers are still finding ways to mine gold out of this data. While most studies focus on using click streams to make better prediction of conversion rate, we unveil in this paper another potential of clickstream data — to analyze market competition structure.

In this paper, we develop a simple and intuitive approach to identify market competition structure and conduct an empirical study using a combination of click-stream data containing customer page views, search data containing displayed search results even if not clicked, and hotel data from a major online travel agency (OTA). There

[22] "Who are my 'true' hotel competitors?" Trevor Stuart-Hill. HotelNewsNow.com

are three key components in this paper. We first measure competition from customers' point of view – which pairs of hotels are most likely to be compared by customers? We identify one important caveat of measuring competition using online data. One must be careful about the effect of travel agent's ranking systems on measures of competition: it is a well-documented fact that higher ranks are associated with higher click-through rate (see Hoque and Loshe 1999 for example). To adjust for this possibility, we consider several measures of co-location commonly used in word analysis in linguistic literature. These measures control baseline probability of one hotel being clicked when calculating the level of co-location (i.e., competition). Furthermore, we show that there can potentially be an aggregation bias in these measures if we ignore heterogeneity in hotel ranks and consumer tastes. To account for this effect, we imbed a Random Coefficient Choice Model into the construction of co-location measures. This can only be done by combining search data (which provides information on all displayed hotels) and clickstream data (which provides information on clicked hotels).

Second, we calculate a proxy of how competition is perceived by hoteliers, where we propose to measure the strength of the competitive relationship using price responsiveness (price matching). With these two measures, we are able to visualize local competition structure in the form of competition networks, build upon customers' and hoteliers' perspectives, respectively. Finally, we measure the extent of mismatch between the two competition networks and we examine when mismatches are likely to occur. We find that independent hotels and distant hotels are likely to be left out from competition sets while branded hotels and hotels in close proximity are more likely to be included.

Moreover, hotels tend to benchmark themselves with hotels that have lower star ratings, lower price levels, lower ranks and higher customer reviews. By showing evidence of Comp Set misperception among hoteliers, we illustrate how data-driven analytical approach can bring hotel managers closer to truly understand their competition and their competitive positions.

We note that, although the paper uses hotel industry data, the methodology proposed here can be easily applied to other products and services and the data required for this type of co-location or co-occurrence analysis prevails in various online and even offline settings. Many merchandisers' websites such as Verizon and Bestbuy allow consumers to choose a couple of products (usually up to 4 or 5) and click "compare" to compare them in detail. Many websites such as TripAdvisor and Amazon provide recommendations to consumers by showing "customers who viewed this product also viewed...". Consumer generated contents on social media also allows to analyze which products tend to be mentioned together by customers (Netzer et al. 2012). Furthermore, emerging technologies such as Shopper Tracker and Euclid are even able to track consumers' physical actions in retail stores or nearby. One commonality of all this data is that, unlike before when we relied almost solely on transaction data to understand consumer choice and resulting competition, now we are able to know not only consumers' final choice but also which options they have considered – in other words, which products/services competed for each customer. Thus, the customer-centric approach and the view of networked competition structure that we propose in this paper are likely to benefit many other business and future research as well. More questions can be examined under this framework, such as how competition network evolves dynamically over time, and how small local perturbation may travel across a network through competitive links.

4.2 Literature

First and foremost, our customer-centric approach is part of the current trend of incorporating consumer-driven models into analysis of revenue management problems and associated empirical applications. Talluri and van Ryzin (2004) introduces a general discrete choice model to a single-leg Revenue Management problem in the airline industry. A stream of studies have followed up on this line of research including Zhang and Cooper (2006), Gallego et al. (2009), Farias et al. (2012), to name a few. Customer-driven demand models have also been studied empirically in several revenue management contexts. Vulcano et al. (2010) shows that, by accounting for customer choice behavior in inventory control optimization, airline can improve their average revenue by 1–5% from the current ESMR-b (Expected Marginal Seat Revenue, version b) policy. Li et al. (2011) show that accounting for dynamic consumer choices can improve revenue in certain types of markets in the airline industry by 3-5%. In the realm of hotel revenue management, Bodea et al. (2009) collect data from five U.S. properties of a major hotel chain and illustrate how choice-based RM can be used with real data. Anderson and Xie (2011) estimate a nested logit model on data from firms selling hotel rooms through an opaque channel, and they optimize firms' dynamic pricing decisions with the knowledge of customer choice behavior. Using hotel transaction data supplemented with user-generated content from social media, Ghose et al. (2012) optimize hotel ranking system on estimated customer valuation of hotel stays. Our approach is similar in adopting a customer-centric perspective. Yet, we differ from the literature in three significant aspects: 1) we adopt a different yet simple and intuitive approach to use customer data; 2) the data we use is different from transaction data that is commonly used in this literature. We use search and clickstream data – which allows us to track actions leading to final decisions; 3) the research goal is to identify competition set for firms to match pricing and benchmark performances on. This is a different type of revenue management problem than the most commonly studied inventory control and dynamic pricing problems.

What we propose is also an innovative approach towards understanding market competition structure. While competition is an old and widely studied question in economics and operations, most existing approaches fall into two categories in terms of methodology. The first approach uses reduced form analysis to study the impact of competition on firms' performance using measures such as Herfindahl-Hirschman Index (Borenstein and Rose. 1994), number of competitors (Olivares and Cachon 2009) or competitor entry and exit (Buell et al. 2011). The second approach applies structural estimation to competitive models. See Berry et al. (1995) for an example of static competition model, and Aguirregabiria and Ho (2012) for an example of a dynamic competition model. In this approach, estimated cross-price elasticities can be understood as a measure of competitiveness between firms. The majority of the literature assumes that every player in the market competes with everyone else, and customers choose from all available products. The relevant papers usually utilize firm-level performance data or customer-level transaction data – neither stream attempts to understand which options were among the real competing choices for each customer. In studying localized competition, there are two general applications similar to our setting. One is fast food industry (Allon et al. 2011), and the other is gasoline industry. For example, Pinkse et al. (2002) show that price competition among gasoline market is highly localized.

In adopting this customer-centric approach, we are able to visualize a networked competition structure. This attempt is a part of the recent burgeoning efforts towards modeling economic markets as networks. Similar to the idea raised in (Kranton and Minehart 2001), we posit that firms do not compete with every other firm in a market, or at least not as close competitors. Our work is also associated with a few other streams of network-related literature. Customer associative networks have been used to understand brand preferences and associations. A recent work by Netzer et al. (2012) analyze how often consumers mention two products from online user-generated contents to map out the competitive market structure. Our work is similar in its research goals and methodology, but it is derived from a different type of data – data that measures real actions, i.e., customer views of a product page, rather than data that measures human memory and perceptions. In addition, we not only describe the customer-based competition network, we also compare it against hotelier-based competition network to examine degree of network mismatch. Network misalignment has also been studied in other settings by Sosa et al. (2004), Gokpinar et al. (2010).

In addition, our work also contributes to the empirical literature stemming from the increasing availability of online clickstream data (e.g., Fader and Moe 2004, Fader and Park 2004). Most papers in this stream focus on predicting conversion rate or understanding customer browsing behavior and no studies we are aware of use the data to understand market competition structure. We contribute by pointing out one other important direction of how such data can be utilized.

4.3 Data and Industry Overview

4.3.1 Data

We combine three data sets associated with online search for hotels sponsored by a major Online Travel Agency (OTA). The first data set, i.e., search and transaction data, contains complete histories of all product searches at the sponsoring OTA's website conducted by approximately 4,000 cookie-based users in Manhattan, New York during the first two weeks of October 2009. The travel dates actually span time period from October 2009 to September 2010. 90.4% requests are for dates within 3 months, and 96.2% are for dates within 6 months. This data set includes information on what users searched for, what searching criteria were specified, which hotels were returned in response (together with room availability, price and promotions, and reviews), which hotels visitors looked at in details and which hotels were booked, if any. Search and transaction data is tracked internally by the firm. The second data set, clickstream data, contains the complete clickstream history of all actions by the users identified in the search and transactions data. The clickstream dataset contains usage of all web-properties belonging to the sponsoring OTA for the same users included in the transaction data from January 2009 to October 2010. Each record describes a clickstream event such as viewing a particular web page or submitting a form. This data set is tracked by a third-party web analytics firm. The third dataset, hotel data, contains information associated with hotels appearing on the OTA's website, such as brand and chain name, ownership type, room capacity, star level, etc.

In both search/transaction data and clickstream data, a unique ID is used to identify each web user (cookie-based). The two data sets can be matched based on user ID and event time. Note that certain information in these two data sets are overlapping – such as a click of a hotel displayed in a search result. Meanwhile, each data set contains its unique information. Since clickstream data only records clicks, hotels displayed but not clicked are only recorded by search/transaction data, but not clickstream data. In search/transaction data, we see everything that a user sees in a search result – hotels displayed and their associated prices, promotions and customer reviews. This hotel-specific information is not recorded by clickstream data. However, since clickstream data records all user *actions*, hotel page visits that are not resulting from standard search requests are also recorded. A visit of a specific hotel's webpage does not necessarily originate from a direct search on the OTA's website, but can also originate from external sources and other internal pages of the OTA's website. Examples of external sources include search engines such as Google or Yahoo, other online booking sites, email promotions.

Examples of internal pages include a page promoting recent deals, flight reservation or car-rental pages, another hotel page linking to similar hotels in the same destination. In fact, clicks which do not originate from direct search request on the OTA's website take up 69.8% of all page visits. In summary, access to clickstream data gives us a more comprehensive view of the real consideration set of a customer. We could have used the 20-month clickstream data to obtain Comp Sets. However, in order to match it with the search/transaction data to control for the effect of the ranking system and for heterogeneity in choice probabilities, we restrict our attention to the period in which two data sets overlap, i.e., the first two weeks of October 2009.

In all, our data contains 3,514 users who searched for hotel stays in New York and its vicinity during the first two weeks of October, 2009. This covers 309 hotels in Manhattan[23]. In total, these hotels received 22,901 page views in our data, an equivalent to an average of 76.8 page views per hotel. Naturally, hotels may be viewed multiple times by the same user. 37.5% of users did not view any specific hotel page at all. The rest viewed 10.4 pages on average per user, or 3.5 distinct hotels. We consolidate multiple visits of the same hotel within a short period of time: we define restarting of a new search session using an inaction period of 24 hours[24]. Prices, promotions, customer reviews and hence ranks of the same hotel may change the next day – the same hotel would essentially be a different hotel for a new search after 24 hours. Thus, for our

[23]New York City and its vicinity include 255 hotels outside Manhattan (e.g., New Jersey or Long Island). We restrict our attention to hotels in Manhattan which receive, on average, six times more page views than hotels outside of Manhattan. The unique geography of Manhattan makes it convenient for our purposes or constructing competition networks without making them too sparse.

[24]We also use 8 hours and 48 hours as robustness checks and the results are very similar.

purposes, within a search session multiple visits of a hotel page we count as only one visit. Thus, we consolidate 22,901 page views to 7,764 distinct page visits.

In our data, we mainly observe price variations along two dimensions: 1) day-of-travel and 2) day-of-observation. Variation along day-of-travel is mostly based on segmentation of customer types. For example, business travelers tend to stay on weekdays, while leisure travelers tend to stay on weekends. Hotels targeting different segments would demonstrate different pricing strategies based on different days of the week. Variation along day-of-observation is mostly du to inventory-based Revenue Management models – how to price based on days in advance and how to dynamically change rates given remaining room inventory. An important input for price changes along both dimensions is how competitors adjust pries along the same two dimensions. Later, we will estimate a pricing model to examine this aspect of the pricing behavior.

4.3.2 Hotel Industry Background

Hotel industry is characterized by intense competition with various types of branded and non-branded properties competing for business. Hoteliers make continuous efforts to differentiate their services from others, but they are still faced with fierce price competition. As prices becoming more and more transparent with Internet search engines, competition among hoteliers only increases. In our data, 309 hotels compete for business in Manhattan, New York, offering 68,584 rooms. Among them, 59.2% are independent properties, or 41.4% in terms of room capacity. Among the branded properties, a few major chains (i.e., Hilton Worldwide, Choice, Marriott, IHG, Starwood) operate 77 properties under 32 brand names ranging from two to five stars. Competition exists not only between chains but also within a chain – hotels in the same chain are typically operated by different individuals or management firms.

Price Patterns. Hotel industry exhibits a great amount of price variation across different properties, room types, customer types, days of stay and days of booking. Our sample of 309 hotels in Manhattan offers an average room rate of \$295 with a standard deviation of \$200. Part of the variation is manifested through frequent promotions – a hotel is on promotion 41% of time in this period. Hotels sometimes reach full capacity – 2.60% on average. The main drivers of the price variations are two-folds. First, the industry has successfully adopted revenue management tools for the past 30 years. Further, rooms of a hotel are usually classified into several room types, which are further grouped into a few (typically 10 to 12) rate bands together with customer types for revenue management purposes. Rates typically start off from rack rates and decrease from there. Discounted rates are characterized by percentage discounts off rack rates. As room inventory depletes when the travel date approaches, the revenue management system suggests opening or closing certain rate bands. At the same time, revenue managers also retain rights to override the system when needs arise.

Such needs usually arise due to observing changes in competitors' prices – an important input for pricing decisions in this industry. As a matter of fact, monitoring competitors' prices is part of everyday operations in the industry. "Call-Around" is a common practice in hotel industry, whereby "Hotels engage in regular communications, typically by telephone with the hotels on their Call-Around Lists, two or three times daily, to exchange with such hotels: (i) each hotel's non-public current occupancy rate (generally expressed as a percentage of hotel rooms occupied) and (ii) the standard rate currently being charged for hotel rooms to be occupied that same day (generally expressed as the "BAR rate" or the "rack rate", which would not include any available discount rates)... Revenue Managers have the ability to manually override the preset grid andor computer reservation system to adjust the applicable room rate.... In determining whether to manually override the reservation system, the Regional Revenue Manager may periodically consult various sources of information concerning competitor rates, including publicly listed rates through internet sites or other market information[25]. In addition to the "Call-Around" practice, hoteliers subscribe to automated tools to monitor competitors' rates closely, such as MarketVision, PriceTrack, and RateVIEW. These tools help hoteliers shop rates and availabilities of hotels in their pre-specified competition set (Cross et al. 2009). In addition, development of online search engines exposes a wealth of readily-accessible price information not only to customers but also to competitors.

Online booking channels. Transient business and leisure markets are captured increasingly through online reservations. 81.5% consumers perform travel-related search, and 70% visit a travel-related site prior to booking at the supplier's website (Withiam

[25] An Agreement By and Among the Attorney General of the State of Connecticut, LQ Management L.L.C., and La Quinta Franchising, L.L.C. March, 2010

2011). OTAs currently capture a large market share of online hotel reservations. Based on our analysis of a random sample of ComScore data in 2009, major OTAs — Expedia.com, Hotel.com, Hotwire, Priceline, Travelocity and Orbitz in descending order of market share— contribute 43.7% of all online observations. This percentage is even higher when we consider eye-visit market share. For example, 10.5% of customers who booked hotels online made their reservations through Expedia.com, meanwhile, 29.8% customers actually have visited Expedia.com before purchasing.

In an era of great price transparency enabled by online search engines, price matching with wrong competitors has a potential to generate sizable revenue losses because any pricing error is much more visible. Customers are quicker to punish misaligned prices by booking away from hotels that have rates too high or pouncing on rates that are below market. This new reality calls for an accurate definition of competition set and a properly designed data-driven analytical approach, which will allow hotel managers to benchmark prices with the right set of competitors and take proper actions to observed price changes in the marketplace.

4.4 Methodology

In this section, we discuss the methodology used to measure competition intensity between pairs of hotels which further leads to construction of two competition networks based on customers' perspective and hoteliers' perspective, respectively.

4.4.1 Customer-Based Measure of Competition

To understand which hotel competes with which other hotels from customers' perspective, we use click-stream data to analyze which hotels customers have compared before making the final choice, or, in other words, which hotels have competed for each customer's demand. The general idea is simple and intuitive: those hotels which are compared frequently tend to be competing hotels. However, there are several caveats associated with how we should define "frequent comparison". If the number of comparisons between hotel A and B is higher than that of A and C, should B be considered more of a competitor to A than C?

Illustrative example. Suppose there are three hotels A, B and C. $1,000$ customers arrive and choose whether or not to click each hotel. Suppose that we observe hotel A is clicked 200 times, B 500 times, and C 200 times. We also observe that A and B are clicked together (i.e., compared) 100 times, while A and C are only compared 50 times. Do we conclude that competition between A and B is more intense than that between A and C? A simple 2 by 2 contingency table answers this question.

Table 4.1: Example: 2 by 2 Contingency Table

	B	$\bar{B}$	**total**		C	$\bar{C}$	**total**
A	100	100	200	A	50	150	200
$\bar{A}$	400	400	800	$\bar{A}$	150	650	800
total	500	500	1000	**total**	200	800	1000
Chi-square	0			**Chi-square**	3.9063		
P-value	1			**P-value**	0.0481		

Note: A represents A is clicked; $\bar{A}$ represents A is not clicked.

We observe that hotel A is compared with hotel B twice as often as it is compared to Hotel C. However, the chi-square test shows that A and B are two independent events, while A and C have a significantly positive dependence. If customers click on hotels randomly, we expect to observe a frequency of 100 comparisons of A and B. However, we expect to see only 40 comparisons of A and C, which is smaller than what we actually observe, i.e., 50 comparisons.

From this simple example it is clear that we need to account for probability of each hotel being clicked when considering the extent to which two hotels compete for demand. That is, we need to control for popularity of each hotel. Even if clicking a particular hotel is completely independent from clicking other hotels, a popular hotel will be compared more often with others simply because it receives more clicks. This is especially important when we consider potential effects of online ranking systems: it is a well-known fact that higher-ranked options receive more clicks in general. In our data, probability of clicking a hotel descends almost monotonically as one moves down along the OTA suggested ranks displayed on the page, and the top 10 hotels receive about half (47.7%) of all click-throughs (see Figure 4.1). Since higher-ranked hotels tend to receive more clicks, if we do not adjust for probability of hotel being clicked, we would come to a biased conclusion that there is more competition among top-ranked hotels than among lower-ranked hotels.

Measures of co-location. To test whether including hotel A and B into the same consideration set is a random event, we adopt four co-location measures that are widely

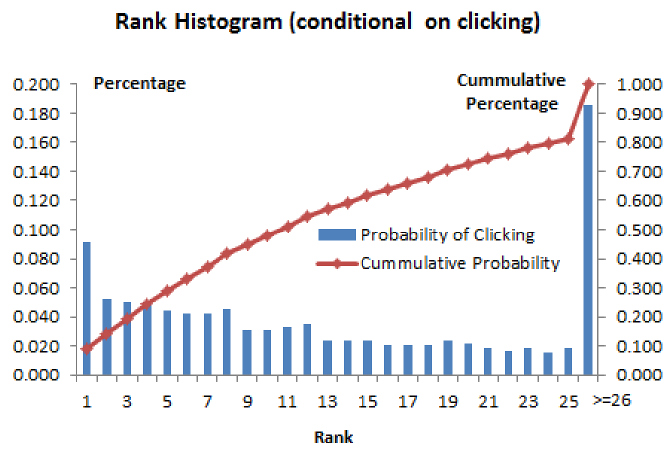


Figure 4.1: Histogram of Rank conditional on Clicking

used in both in the linguistic literature (see Manning and Schutze 1999 for a general discussion) and in the data-mining literature. The null hypothesis is that including hotel A in the consideration set is independent from including hotel B in the consideration set, i.e., $P(AB) = P(A)P(B)$. The four measures are 1) t statistics, 2) Chi-Square statistics , 3) Pointwise Mutual Information (PMI, also called "lift" in data-ming literature), and 4) and an adjusted version of PMI — normalized PMI.

T-test uses normal approximation for the number of successes of a series of Bernoulli trials with probability $p = P(A)P(B)$ under the null hypothesis. Chi-Square test implements the chi-square test for independence for the 2 by 2 contingency table as in Table 4.1. Unlike t-test, Chi-Square test does not rely on the normal approximation to the Binomial distribution.

Pointwise Mutual Information (proposed by Church and Hanks 1990) is defined as follows,

$$log\Big(\frac{P(AB)}{P(A)P(B)}\Big). \tag{4.1}$$

PMI is simply the ratio of A and B's actual joint probability over the expected joint probability if they were independent events. PMI being 0 means that the two events are independent. A positive PMI means that occurrence of the two events is positively correlated. One known limitation of this measure, however, is that PMI is particularly sensitive to low frequency events (even though all co-location measures are sensitive to low frequency data to some extent). To see why, assume perfect dependence between clicking two hotels A and B so we have

$$log\Big(\frac{P(AB)}{P(A)P(B)}\Big) = log\Big(\frac{1}{P(B)}\Big). \tag{4.2}$$

A low probability of $P(B)$ indicates a large PMI value in this case. To mitigate its impact, we introduce another measure – normalized PMI defined as follows (Bouma 2007),

$$log\Big(\frac{P(AB)}{P(A)P(B)}\Big)/ - log(P(AB)). \tag{4.3}$$

This measure is proposed to overcome two shortcomings of PMI: 1) sensitivity to low-frequency data; 2) lack of a fixed upper-bound. Normalized PMI is essentially a PMI scaled by joint probability, which gives lower weights to low frequency events. When two events are completely dependent, the above measure equals its upper bound 1.

Alternative measures of co-location.

Consider the following example of clicking probabilities which are dependent on ranks. Suppose clicking hotel A, B, C are independent events. The probability of one hotel being clicked is completely determined by its rank – 0.7 if ranked first, 0.5 if ranked second, and 0.3 if ranked last. Also suppose all six orders appear with equal probabilities.

Table 4.2: Example: Aggregation Bias

rank	order 1	order 2	order 3	order 4	order 5	order 6	Pr(click)
1	A	A	B	B	C	C	0.7
2	B	C	A	C	A	B	0.5
3	C	B	C	A	B	A	0.3

In this example, we can calculate the following probabilities:

$$\begin{aligned} P(AB) &= \frac{1}{6}(0.35 + 0.21 + 0.35 + 0.21 + 0.15 + 0.15) = 0.237, \\ P(A) &= P(B) = \frac{1}{3}(0.7 + 0.5 + 0.3) = 0.5, \\ P(AB) &\neq P(A)P(B). \end{aligned}$$

If the probability of an event occurring varies with different contexts (such as different orders of results), aggregating probabilities may induce bias. The following equation shows why it is the case in general. Under null hypothesis of independence,

$$\begin{aligned} P(AB) &= \frac{1}{N}\sum_{i=1}^{N} P_i(AB) = \frac{1}{N}\sum_{i=1}^{N} P_i(A)P_i(B) \\ &\neq (\frac{1}{N}\sum_{i=1}^{N} P_i(A))(\frac{1}{N}\sum_{i=1}^{N} P_i(B)) = P(A)P(B). \end{aligned}$$

In different types of searches and on a different observation day, the same hotel may appear at different positions in the search result. If we do not account for this possibility, we will have a biased measure of co-location. Note that, this aggregation bias can also be induced through other channels. For instance, the choice of viewing a hotel is also dependent on prices, promotions and customer reviews, which are also context-dependent.

Suppose we know the clicking probability $P_i(A), P_i(B), i = 1, ..., n$, it is then straightforward to adjust for this potential aggregation bias. Simply substitute $P(A)P(B)$ by $\frac{1}{N}\sum_{i=1}^{N} P_i(A)P_i(B)$. A special note to validity of using t-statistics in this case. Under null hypothesis of independence, we expect to observe a sequence of independent but non-identical Bernoulli trials with succuss probabilities: $p_i = P_i(A)P_i(B), i = 1, ..., N$. Total number of successes then follows a Poisson Binomial Distribution, which can also be approximated by a Normal distribution with mean $\sum_i p_i$ and variance $\sum_i p_i(1-p_i)$. Thus, we can still use t-test of independence. Consequently, $P(AB)$ can be approximated by a normal distribution with mean $\frac{1}{N}\sum_i p_i$ and variance $\frac{1}{N^2}\sum_i p_i(1-p_i)$. The remaining question is: in t-statistics, can we still use $\frac{P(AB)(1-P(AB))}{N}$ as estimates of the sample variance? Even if null hypothesis holds, in general, $\frac{P(AB)(1-P(AB)}{N} \neq \frac{1}{N^2}\sum_i p_i(1-p_i)$. However, when p_i's are small, we have $1 - p_i \approx 1$. Then, we have $\frac{P(AB)(1-P(AB))}{N} \approx \frac{P(AB)}{N} = \frac{1}{N^2}\sum_i p_i \approx \frac{1}{N^2}\sum_i p_i(1-p_i)$.

How to obtain estimates of $P_i(A), P_i(B), i = 1, ..., N$? These probabilities can be estimated through a logistic regression. For each displayed hotel, a consumer decides to either click or not click[26], and this action is a function of both time-variant and time-invariant characteristics of the displayed hotel: rank, room rate, promotion status, customer review, star rating and location. Other information such as amenities, transportation, complementary services do not affect the decision of clicking as they are not part of customers' information set yet.

One shortcoming using logistic regression is its underlying assumption of independent error terms and the Independence of Irrelevant Alternatives (IIA) property. In reality, a consumer's clicking decision can be correlated across hotels. To account for this heterogeneity in consumers' tastes and to overcome the IIA problem, we use the Random Coefficient Logit (or Mixed Logit) model (Train 2003). We predict clicking probability $P_i(A), P_i(B), i = 1, ..., N$ based on this model while allowing for random coefficients.

$$\begin{aligned} u_{ijk} &= \beta_i X_{jk} + \epsilon_{ijk}, \beta_i \sim N(\beta_0, \Sigma), \\ u_{ijk} &= \beta_0 X_{jk} + \xi_{ijk}, \text{where } \xi_{ijk} = (\beta_i - \beta_0) X_{jk} + \epsilon_{ijk}, \end{aligned}$$

where i represents an individual, j represents an alternative, i.e., a hotel, and k represents a search occasion. An example of a search occasion is a search request made on October 1, 2010 for a two-night stay in New York City starting from November 1, 2010. X_{jk} represents characteristics of hotel j at search occasion k. For example, \$295 per night were quoted for the above request. β_i is a vector of random coefficients. ϵ_{ijk} are i.i.d.

[26]Suppose that the total number of hotels displayed is exogenously given. In reality, it is chosen endogenously, as consumers do choose to wether or not to continue on to the next page or result. Should they choose to continue, more hotels will be displayed to them. However, in our data, almost all customers (96.4%) stayed on the first page of results and did not scroll down.

shocks across consumers, hotels and search occasions. This means that even though, at the person level, clicking of hotels A and B are completely independent as ϵ_{iAk} and ϵ_{iBk} are independent, the aggregate choice of the two hotels can be dependent due to heterogeneity in customers' tastes, that is, ξ_{iAk} and ξ_{iBk} are correlated within the same person. We want to adjust for this possibility, and verify if the data and the proposed methodology capture additional patterns of competition in addition to what is captured by the conventional model.

4.4.2 Hotelier-Based Measure of Competition

Competitor's price is a key component when a hotel determines its own price. Each property has a list of hotels which they consider as competitors. Traditionally, they use such a list in Call-Around practice to solicit information on standard rates and occupancy rate from competitors. Now with proliferation of online booking tools, prices become even more transparent. Revenue managers no longer need to call several times a day to learn about competitors' prices and availability, it is just a click away. Any change in a hotel's price such as newly launched promotions, or changes in room availability, can be immediately spotted and may invoke price changes from competitors should they decide to react. As a result of price matching practice, prices of competing hotels will be highly correlated, and we will use this as basis to identify competitors from hoteliers' point-of-view.

Based on theories of product differentiation, if firms are able to differentiate their services or products through dimensions such as quality or location, they do not need to engage in fierce price competition. Thus, a lack of price differentiation among firms indicates a lack of differentiation along other aspects, or in other words, firms perceive themselves offering very similar products. In the hotel industry in particular, 70% of variations in prices, measured usually by Average Daily Rate (ADR), can be explained by a hedonic model of product characteristics (e.g., Thrane 2007), similar to housing markets. This is to say, price levels in this industry are a symbol of how similar hoteliers think their services are. Kim and Canina (2009) exploit ADR clustering patterns to identify hotel competition set.

We use correlation of two price series $\{p_{jk}, j = A, B, k = 1, 2, ..., K\}$ to capture competitiveness between hotel A and B from hoteliers' perspectives. j corresponds to hotels, and k corresponds to search occasions. For example, Hotel A may charge \$299 per night for a two-night stay request made on a particular date for a particular travel starting date, with or without specifications on the number of adults, number of children and type of room. Hotel B may charge \$289 for the exact same request. Note that the availability of search data allows us to obtain price correlation down to the level of search requests, while most other analysis using hotel prices are based on aggregated ADR. This benefit also comes with a cost. Not all hotels are seen by a customer, due to his particular specifications of the travel request or his inaction beyond the first page of results. As a consequence, we have only a few observations for some hotel pairs if they are seldom shown in the same search. However, the pairs for which we do not have much data are also those pairs that charge quite different rates and hence are not likely to be close competitors. For hotels that charge competitive rates, we would have thousands of price pairs to calculate price correlation. Further, as price quotes are search-specific and not customer-specific, we can group the same search made by multiple customers with adjacent time, such as within the same searching day, to obtain more observations of price pairs.

Alternative Measures of Price Matching.

To better understand how hotels match prices, we decompose price correlation into two components according to the following pricing model:

$$p_{jk} = \alpha_j + \beta_j X_{jk} + \epsilon_{jk}, \tag{4.4}$$

α_j is a hotel-specific intercept which represents the average level of price of hotel j. It includes how a hotel sets prices based on its static characteristics such as location, star rating, and brand. β_j estimates a hotel-specific factor of how hotels price characteristics of each travel request, such as length of stay, day-of-week of travel dates, days of advanced purchase, number of adults and children etc. We suspect that price matching mostly happens along two dimensions – travel dates and booking dates, as they are the main variations of different travel requests and two most important identifiers of travel types (i.e., business vs. leisure). Comparison of how similar β_j's are will inform us about the level of price matching between hotel pairs. The error component ϵ_{jk} represents idiosyncratic shocks to prices, such as unpredicted changes in demand and unpredicted adjustments in competitors' prices. Applying the above model to each hotel, we are able to obtain a predicted price series $\{\hat{p}_{jk}, k = 1, 2, ..., K\}$ and residual price series $\{p_{jk} - \hat{p}_{jk}, k = 1, 2, ..., K\}$. Similar to correlation of original price series, we obtain correlations of predicted price series and residual price series for each hotel pair. Correlation of predicted prices is mainly based on price matching along travel dates and booking dates. Correlation of price residuals is mostly based on price matching of idiosyncratic price changes and common demand shocks.

One concern of using price correlation as a measure of price matching is that two price series can go in parallel with each other, but with a large discrepancy (for example, one around \$100 and the other around \$500). This means that it is probably the overall demand shock that is driving the correlation rather than the hotel trying to match prices. To address this concern, we define another measure — Average Price Difference between two price series $\frac{1}{K}\sum_{k=1}^{K}|p_{1k} - p_{2k}|$. Under perfect price matching, the measure should be zero. A larger value means less price matching and less competition.

4.5 Empirical Results

4.5.1 Consumer-based Competition Measure

We compare four measures of consumer-based competition — t-statistics, Chi-Square statistics, PMI, NPMI, and then select the most appropriate for our application. As shown in many other cases of co-location analysis, PMI is particularly sensitive to low-frequency data. Normalized PMI is proposed to address this issue. We observe that Normalized PMI indeed adjusts this sensitivity, but only to a certain extent, as shown in Table 4.3 listed in descending order of NPMI. Take first and forth rows of Table 4.3, for example, both PMI and NPMI favor the low-frequency event of hotel “John Street Suites” which has been clicked only once during our period of study but was compared with “Hotel Belleclaire” at that time. Chi-Square statistic further adjusts sensitivity to low-frequency data. Within the above example, Chi-Square favors “Grand Union” with 36 clicks and 10 comparisons over the small-probability event of clicking “John Street Suites”. However, comparing the first row and the twelfth row, Chi-Square favors the low frequency event in this case. T-statistic is the least sensitive to low-frequency data among all four measures. It is also one of the three measures (PMI, NPMI, and t-statistics) that can be consistently implemented when we account for heterogeneity and observe independent but non-identical Bernoulli trials. Due to these reasons, we focus on t-statistics as our main measure of competition[27]. The shortcoming of t-statistics is its normal approximation for Binomial distribution. A commonly applied rule of thumb is to restrict attention to cases in which $Np >= 5$. We note that, when we only choose t-statistics greater than 1.96 (i.e., $p - value = 0.05$), this number is greater than or equal to 4 for all selected pairs of hotels and 97.0% of these pairs have greater than 5 comparisons.

[27]We also applied NPMI with a cut-off value where the event has to occur at least 3 times, suggested by (Manning and Schutze 1999). Results are consistent.

Table 4.3: Sensitivity to Low Frequency Data

hotel1	**hotel2**	**# click1**	**# click2**	**# compare**	**PMI**	**NPMI**	**chi square**	**t**
Hotel Belleclaire	John Street Suites	204	1	1	4.813	0.385	27.117	0.965
Hotel Belleclaire	Brownstone Bed and Breakfast	204	3	2	4.228	0.368	34.848	1.339
Hotel Belleclaire	Oxbridge Carnegie Hill Apartments	204	7	3	3.591	0.329	31.553	1.589
Hotel Belleclaire	Grand Union	204	36	10	2.965	0.324	61.950	2.760
Hotel Belleclaire	Holiday Inn New York City - Wall Street	204	12	4	3.228	0.308	31.079	1.787
Hotel Belleclaire	Hotel 373 Fifth Avenue	204	21	6	3.006	0.304	38.444	2.146
Hotel Belleclaire	The Ritz-Carlton New York, Battery Park	204	14	4	3.006	0.287	25.598	1.752
Hotel Belleclaire	Comfort Inn Manhattan Bridge	204	10	3	3.076	0.282	20.418	1.527
Hotel Belleclaire	Thirty Thirty Hotel New York	204	75	14	2.392	0.276	50.573	3.032
Hotel Belleclaire	Hampton Inn Manhattan Soho	204	11	3	2.939	0.270	18.069	1.507
Hotel Belleclaire	Belnord Hotel	204	59	11	2.390	0.265	39.553	2.686
Hotel Belleclaire	254 East Vacation	204	7	2	3.006	0.262	12.783	1.238
Hotel Belleclaire	Washington Jefferson Hotel	204	54	10	2.380	0.260	35.570	2.557
Hotel Belleclaire	Sutton Place Suites	204	3	1	3.228	0.259	7.757	0.893
Hotel Belleclaire	Best Western Hospitality House	204	3	1	3.228	0.259	7.757	0.893

Table 4.4: Intensity of Competition between Hotels – sorted by t statistics

hotel1	**hotel2**	**# click1**	**# click2**	**# compare**	**PMI**	**NPMI**	**chi square**	**t**
top 10								
Park Central New York Hotel	Paramount Hotel Times Square New York	164	93	21	2.981	0.368	132.352	4.010
Sheraton New York Hotel And Towers	The Manhattan at Times Square Hotel	141	89	19	3.118	0.379	134.511	3.863
Hilton Garden Inn Times Square	Hampton Inn Times Square North	146	68	18	3.378	0.406	158.769	3.841
Hilton New York	Roosevelt Hotel New York	133	79	18	3.296	0.396	148.108	3.817
Hilton Garden Inn Times Square	The New Yorker Hotel	146	150	21	2.459	0.304	81.454	3.756
The Edison Hotel	Paramount Hotel Times Square New York	180	93	19	2.702	0.328	92.973	3.695
Sheraton New York Hotel And Towers	Hilton Garden Inn Times Square	141	146	20	2.478	0.304	78.925	3.676
Hilton Garden Inn Times Square	Roosevelt Hotel New York	146	79	17	3.079	0.367	116.229	3.641
The Edison Hotel	Salisbury Hotel	180	127	20	2.327	0.285	67.925	3.587
Hilton Garden Inn Times Square	The Belvedere Hotel	146	114	18	2.633	0.317	82.229	3.564
bottom 10								
Hotel Pennsylvania	Doubletree Guest Suites Times Square NYC	46	191	1	-0.615	-0.049	0.193	-0.532
Doubletree Guest Suites Times Square NYC	Ink48 Hotel, a Kimpton Hotel	191	47	1	-0.646	-0.052	0.213	-0.565
St. Giles - The Court New York	Hotel Belleclaire	45	204	1	-0.679	-0.054	0.236	-0.601
Hilton New York	Latham Hotel	133	73	1	-0.760	-0.061	0.294	-0.693
Doubletree Guest Suites Times Square NYC	The Pierre, A Taj Hotel	191	51	1	-0.764	-0.061	0.300	-0.699
Sheraton New York Hotel And Towers	Latham Hotel	141	73	1	-0.844	-0.068	0.365	-0.795
The Edison Hotel	Sofitel New York	180	65	1	-1.029	-0.082	0.554	-1.040
Doubletree Guest Suites Times Square NYC	The Waldorf Astoria	191	105	2	-0.806	-0.070	0.675	-1.059
The Plaza	Hotel Belleclaire	59	204	1	-1.069	-0.086	0.603	-1.099
Empire Hotel	Doubletree Guest Suites Times Square NYC	70	191	1	-1.221	-0.098	0.796	-1.331

Based on t-statistics, the top 10 and bottom 10 pairs of competitors are listed in Table 4.4. On the top of this list is Park Central New York Hotel which has been viewed 164 times, and Paramount Hotel Times Square New York which has been viewed 93 times, and the pair have been compared 21 times. At the bottom of the list is Empire Hotel and Doubletree Guest Suites Times Square NYC. Each of them has been viewed multiple times (70 and 191, respectively). However, they are compared only once[28].

Using a cut-off value of 1.96 of t-statistic, we build an (undirected) competition network of Manhattan hotels in Figure 4.2. As we expected, there is a strong clustering pattern within the same star level. Occasionally, hotels also compete with hotels of adjacent star levels. The degree distribution of the network has a mean of 8.42, median of 6.00, and minimum of 1 and maximum of 36. That is on average, a hotel has 6 to 8 competitors, consistent with the average size of Benchmark Comp Set provided directly by 2,833 hotels to Hotel Compete[29]. The average geodesic distance in this graph is 2.93, with a maximum (diameter) of 8. That is, on average, a hotel can be linked to another hotel through three steps of competitive relationship, or two competitors in between. This small degree of separation can have interesting implications – it means that some kind of local shock to a particular hotel (such as promotions resulted from demand or supply shocks) may spread quickly to other parts of the network through price matching.

One caveat of using a universal cut-off is that some hotels with only weak competition

[28] For the purpose of illustration, we only list those pairs of hotels which were compared at least once. There are many other pairs which have no comparisons at all.

[29] "Hotel Comp Set Analysis – Untapped Opportunity #1: Market Dynamics." Hotel Compete. May 16, 2012.

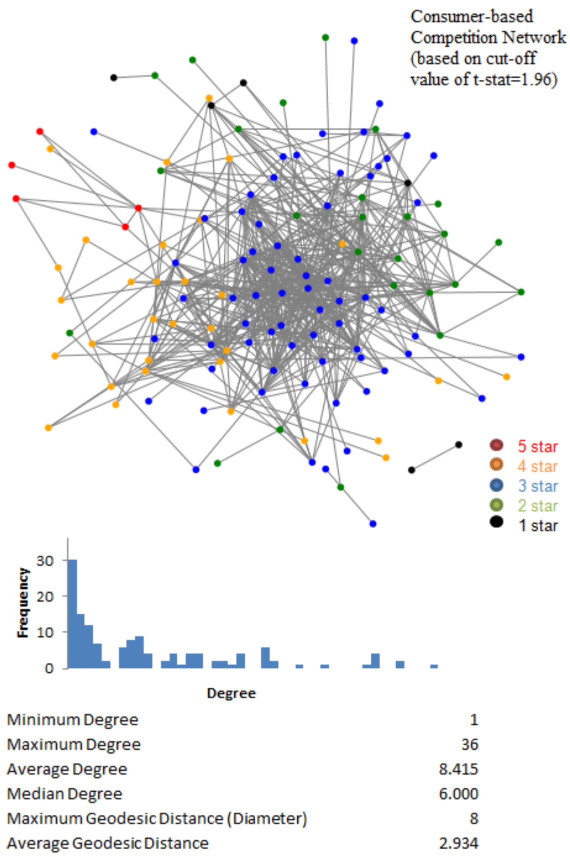


Figure 4.2: Visualize Customer-based Competition Network (t-stat> 1.96)

links (usually those which do not appear often and also these that are not compared often with others) may be left out from this graph. Of course, as we move down the ladder of t-statistics and choose a lower cut-off value we would visualize more competitive links.

4.5.2 Evidence of Price Matching

To illustrate price matching among competitive hotels, we compare pricing strategies among hotels. We take the top two pairs of competitive hotels from Table 4.4, for example. The basic information for each hotel is listed in the table below:

	Hotel Name	**Star**	**Brand**	**Location**	**# Room**
Pair1	**Park Central New York Hotel**	3.5	Independent	Midtown	935
	Paramount Hotel Times Square NY	3	Independent	Times Square	567
Pair2	**Sheraton New York Hotel And Towers**	3.5	Starwood Sheraton	Midtown	1750
	The Manhattan at Times Square Hotel	3	Starwood other brand	Midtown	665

Following the pricing model described in Equation 4.4, we analyze pricing strategy of each hotel separately, the results of which are shown in Table 4.5. Note that the estimated coefficients are very similar within each pair, but quite different across pairs. For example, days in advance have positive coefficients for both hotels in Pair 1 and negative coefficients for both hotels in Pair 2. The two hotels from Pair 1 increase prices linearly from Sunday to Saturday, with the highest room rate on Saturday-night stay. Hotels of Pair 2 tend to charge the highest rate on Friday night, with the second highest being Wednesday night. These patterns can be seen clearly in Figures 4.3 and 4.4. It is obvious that Park Central and Paramount adopt similar pricing strategies, and Sheraton and Manhattan adopt similar pricing strategies. This evidence suggests that hoteliers

Table 4.5: Price Matching Patterns

	Park Central New York Hotel	Paramount Hotel Times Square New York
days in advance	0.59*** (0.14)	1.89*** (0.14)
days in advance squared	-0.02*** (0.00)	-0.02*** (0.00)
length of stay	-9.42*** (0.64)	-10.88*** (0.63)
Mon	7.82 (10.68)	-19.23 (9.60)
Tue	62.86*** (9.35)	50.82*** (8.43)
Wed	91.00*** (8.73)	68.20*** (7.72)
Thu	106.12*** (8.36)	79.57*** (7.51)
Fri	115.73*** (7.91)	84.98*** (7.04)
Sat	118.52*** (9.56)	100.94*** (10.16)
Christmas/New Year	95.22*** (5.18)	43.67*** (4.89)
# adults per room	15.40*** (1.96)	49.16*** (1.86)
# children per room	29.43*** (3.27)	87.59*** (3.01)
const	225.61*** (7.99)	161.23*** (7.06)
Adj R_sq	0.3267	0.4781
# obs	3945	3690
	Sheraton New York Hotel And Towers	**The Manhattan at Times Square Hotel**
days in advance	-0.45*** (0.13)	-0.34*** (0.13)
days in advance squared	-0.01*** (0.00)	-0.01*** (0.00)
length of stay	-4.85*** (0.69)	-6.30*** (0.62)
Mon	5.27 (9.11)	-1.94 (10.13)
Tue	30.75*** (8.04)	17.16** (8.64)
Wed	63.09*** (7.53)	53.59*** (8.06)
Thu	55.50*** (7.60)	46.85*** (8.07)
Fri	88.32*** (6.75)	83.51*** (7.45)
Sat	51.53*** (7.01)	45.65*** (7.68)
Christmas/New Year	3.82 (6.39)	-11.06* (6.17)
# adults per room	4.98*** (1.88)	1.63 (2.05)
# children per room	8.82*** (3.40)	2.95 (5.00)
const	308.76*** (6.93)	313.28 (7.69)
Adj R_sq	0.3888	0.3691
# obs	3231	3306

match prices with their competitors, which supports our approach of using similarity in prices to represent who hotel managers identify as their close competitors.

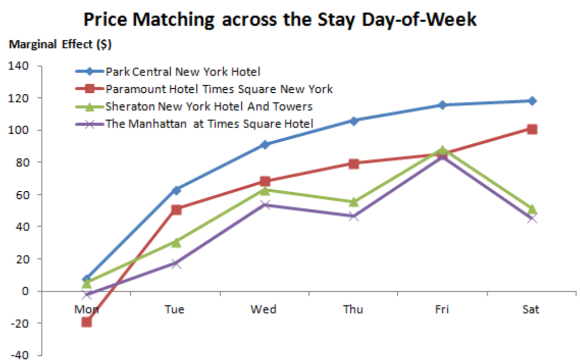


Figure 4.3: Price Matching across Day of Week

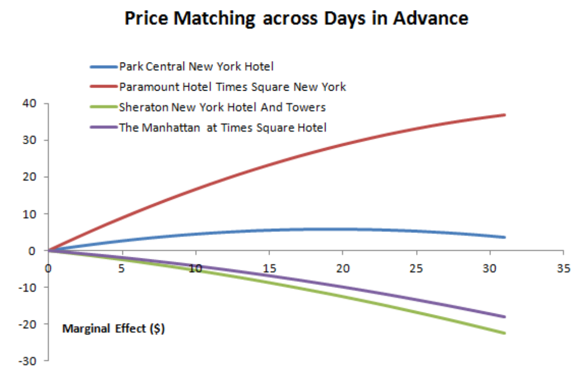


Figure 4.4: Price Matching across Days in Advance

Table 4.6 lists 4 measures of price matching for top 10 and bottom 10 pairs displayed in Table 4.4. Generally, higher t-statistics are associated with higher correlation scores and lower average price difference, even though the order might not be exactly the same. The association is even stronger once we combine multiple measures to determine competitiveness.

Table 4.6: Measures of Price Competitiveness

				Price Correlaiton		Price
			total	**predicted**	**unpredicted**	**Difference**
hotel1	**hotel2**	**t**	$\rho(p)$	$\rho(\hat{p})$	$\rho(p_{res})$	**$**
top 10						
Park Central	Paramount Hotel Times Sq	4.010	0.777	0.814	0.739	42.3
Sheraton NY Hotel And Towers	The Manhattan at Times Sq	3.863	0.985	0.993	0.973	9.9
Hilton Garden Inn Times Sq	Hampton Inn Times Sq North	3.841	0.885	0.842	0.784	33.2
Hilton New York	Roosevelt Hotel New York	3.817	0.773	0.878	0.726	43.1
Hilton Garden Inn Times Sq	The New Yorker Hotel	3.756	0.671	0.811	0.559	54.7
The Edison Hotel	Paramount Hotel Times Sq	3.695	0.602	0.664	0.518	90.1
Sheraton NY Hotel And Towers	Hilton Garden Inn Times Sq	3.676	0.842	0.916	0.774	37.4
Hilton Garden Inn Times Sq	Roosevelt Hotel New York	3.641	0.827	0.872	0.803	33.1
The Edison Hotel	Salisbury Hotel	3.587	0.724	0.955	0.522	54.9
Hilton Garden Inn Times Sq	The Belvedere Hotel	3.564	0.851	0.971	0.794	39.8
bottom 10						
Hotel Pennsylvania	Doubletree Guest Suites Times Sq	-0.532	0.538	0.504	0.641	509.2
Doubletree Guest Suites Times Sq	Ink48 Hotel, a Kimpton Hotel	-0.565	0.475	0.410	0.497	66.6
St. Giles - The Court New York	Hotel Belleclaire	-0.601	0.481	0.353	0.270	77.4
Hilton New York	Latham Hotel	-0.693	0.551	0.686	0.572	210.4
Doubletree Guest Suites Times Sq	The Pierre, A Taj Hotel	-0.699	0.122	-0.001	0.484	262.4
Sheraton New York Hotel And Towers	Latham Hotel	-0.795	0.641	0.844	0.551	179.5
The Edison Hotel	Sofitel New York	-1.040	0.735	0.902	0.559	153.2
Doubletree Guest Suites Times Square	The Waldorf Astoria	-1.059	0.170	-0.060	0.238	278.7
The Plaza	Hotel Belleclaire	-1.099	0.457	0.821	0.259	527.2
Empire Hotel	Doubletree Guest Suites Times Sq	-1.331	0.569	0.474	0.575	502.5

4.5.3 Network Mismatch

After obtaining both consumer-based and hotelier-based competition measures, we are ready to evaluate the mismatch of the two networks. As some hotels, especially star 1 and 2 hotels, do not appear quite often to have a robust measure of price correlation, we restrict our attention to the sub-network of hotels with 3 stars or higher, which gives us around 193 hotels. There is not a sufficient number of comparisons or price pairs for some pairs of hotels (which most likely are not strong competitors in any way). We further restrict our attention to those pairs where we have sufficient comparisons for from the consumer data (i.e., at least 3 comparisons) and sufficient price pairs (i.e, at least 30 price pairs). In this way, our networks are finally constructed on 89 hotels, which actually represent 64.1% of the total number of rooms offered collectively by hotels with a minimum star level of three.

We note that, when it comes to price matching, different hotels match prices on different levels. Some hotels tend to have higher correlation with most other hotels, while some other hotels have a lower price correlation with others in general. For example, 5-star hotels usually do not engage in much price competition. Due to this reason, using a universal cut-off of price correlation may seem unfair to different hotels. Instead, we decide to identify up to top 5 competitors for each hotel based on price correlations of this hotel with all other hotels. Similar logic also applies to consumer-based network. We choose up to top 5 hotels[30] based on t-statistics as competitors of the focal hotel.

[30] Top 5 is a conservative choice as the average number of competitors of a hotel is usually 6 to 8. We also tried top 10 competitors, and the results are similar.

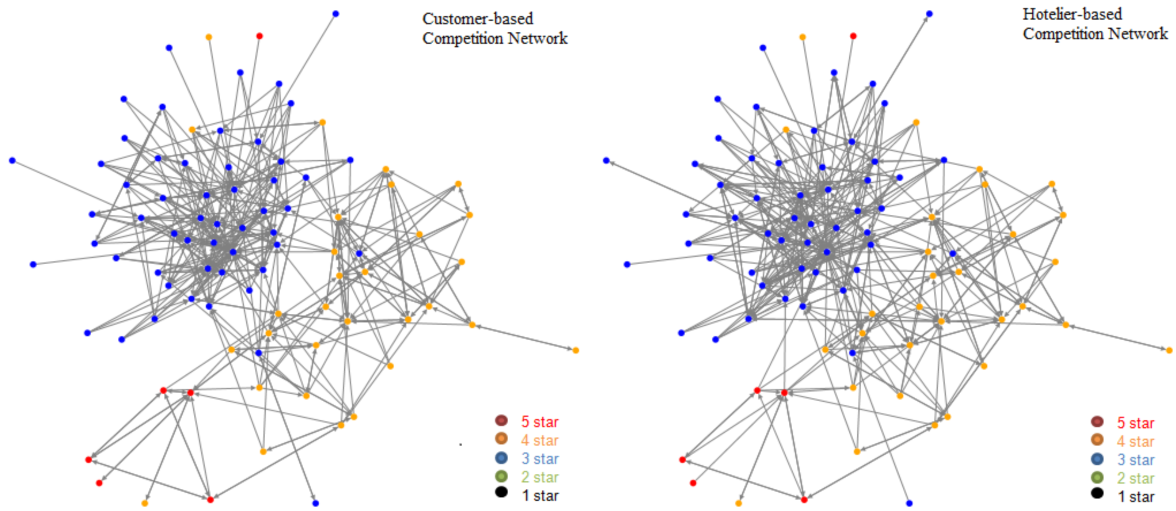


Figure 4.5: Top 5 Competitors Networks of Hotels with 3 Stars and Up

In this way, we derive two directed networks – top 5 competitors' networks as shown in Figure 4.5. Directed links in this figure point from a focal hotel to its competitors.

First, we make a note on the average amount of overlap between the two networks. The amount of mismatch is visualized in Figure 4.6, and the numbers are shown in Table 4.7. Among 386 competitors identified by customer-based measures, 191 or 49.5% are also competitors recognized by hoteliers. 160 of the 386 competitive links in hotelier-based competition network are also reciprocal – that is, both hotels consider each other among top 5 competitors. A similar number of links, i.e., 172, are reciprocal in customer-based competition network. To further examine when mismatch is likely to occur, we run regressions of true competitors being leftout or wrong competitors being included on characteristics of hotels relative to the hotel in consideration.

Results of when mismatch is likely to occur are shown in Table 4.8. Coefficients in Column 1 are marginal effects obtained from a logit model for predicting a competitor

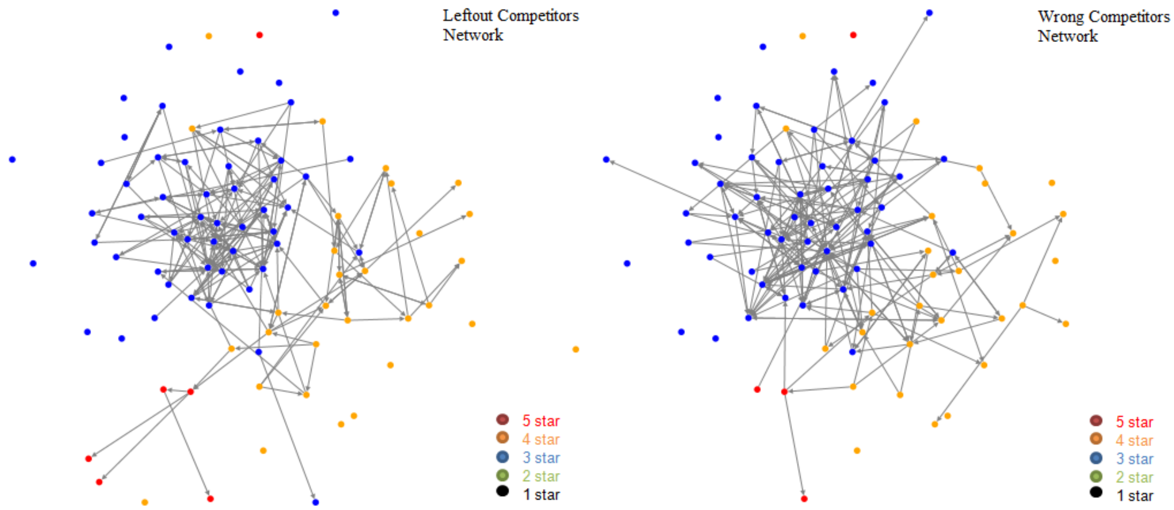


Figure 4.6: Network Mismatch

Table 4.7: Network Overlap and Mismatch among Hotels with 3 Stars and Up

	hotelier-based competitors		
customer-based competitors	0	1	Total
0	859	195	1,054
1	195	191	386
Total	1,054	386	1,440

being left out from competition set by hoteliers conditional on customers consider it as a competitor. Coefficients in Column 2 are marginal effects obtained from a logit model for predicting a competitor being included in competition set by hoteliers conditional on it is not a competitor from customer perspective. We find that independent hotels are likely to be leftout when hoteliers define their competition set, and branded hotels are often included in the competition set while they should not be. Reading from Column 1 and 2, an independent hotel has a 20% higher chance being left out of competition set than a branded hotel, and a branded hotel has a 10.5% higher chance of being included in the competition set while it should not. In forming their competition set, hoteliers also have a bias for hotels in the same district[31]. A hotel in a different district has a 10.9% higher chance of being left out from competition set[32]. Hotels with lower traveler reviews are likely to be left out from competition set as well. It can be that hotels set up a high benchmark for themselves or just think they are as good as those better rated ones. Finally, we also notice that hotels tend to benchmark themselves with other hotels with lower star ratings, lower price levels and lower ranks.

[31] Districts are defined according to the OTA's definition as follows: Broadway – Times Square, Upper East Side, Upper West Side, Financial District / Downtown, Grand Central, Lower, Lower East Side, Lower West Side, Mid-town East, Mid-town West, North Manhattan, Uptown East.

[32] To test the location effect, we also used distance rather than the same district indicator. The signs are exactly the same but sometimes not statistically significant.

Table 4.8: When Network Mismatch Is Likely to Occur

	Basic measures		**Alternative measure of t-stat**		**Alternative measure of price matching**	
	Leftout	**Wrong**	**Leftout**	**Wrong**	**Leftout**	**Wrong**
Higher Starrating	0.012 (0.008)	-0.010*** (0.002)	0.011 (0.007)	-0.011*** (0.003)	-0.007 (0.008)	-0.007** (0.003)
Larger Hotel	0.008 (0.032)	0.019 (0.013)	-0.006 (0.029)	0.008 (0.013)	0.003 (0.030)	0.006 (0.013)
Independent Hotel	0.203*** (0.053)	-0.105*** (0.025)	0.065 (0.055)	-0.146*** (0.025)	0.102* (0.053)	-0.078*** (0.025)
More Expensive Hotel	0.001 (0.000)	0.000** (0.000)	0.002 (0.004)	-0.005*** (0.001)	0.005 (0.004)	-0.001 (0.001)
Higher Customer Review	-0.076** (0.036)	-0.022 (0.019)	-0.074** (0.035)	-0.014 (0.020)	-0.067** (0.034)	-0.012 (0.021)
Higher Avg. Rank	0.008*** (0.002)	0.002* (0.001)	0.009*** (0.002)	0.002 (0.001)	0.004* (0.002)	0.000 (0.001)
Different sub-market	0.109* (0.060)	-0.032 (0.028)	0.117* (0.061)	-0.044 (0.029)	0.066 (0.058)	-0.061** (0.030)
Log Likelihood	-244.8	-472.3	-248.9	-486.8	-256.4	-496.2
# Obs	386	1054	386	1054	386	1054

Note: Coefficients are marginal effects predicted from logit models.

To check the robustness of mismatch patterns discovered above, we use alternative measures of customer-based and hotelier-based competition. For customer-based competition, we adjust aggregation bias using Logit and Mix Logit model. The estimation results of the two logit models are displayed in Table 4.9. Mismatch patterns based on t-statistics resulting from Mix Logit model are shown in the third and forth columns of Table 4.7. For hotelier-based competition measure, we present results based on the average price difference in columns 5 and 6. We first note that the two alternative measures yield similar levels of overall overlapping between two networks. The amounts of overlap are 45.1% and 48.4% respectively, as compared to 49.5% previously. Second, these results confirm the findings with regard to when mismatch is likely to occur, though with minor changes in certain significance levels.

Table 4.9: Choice Models Predicting Probability of Clicking

	Logit	**Mixed Logit**	
		Mean	**S.D.**
rank	-0.045*** (0.004)	-0.063*** (0.004)	0.045*** (0.003)
total # of displayed hotels	-0.073*** (0.004)	-0.086*** (0.005)	0.016*** (0.002)
no available room	-0.180*** (0.022)	-0.549*** (0.130)	0.039 (0.245)
price	-0.122 (0.106)	-0.244*** (0.028)	0.051** (0.023)
promotion	0.530*** (0.049)	0.392*** (0.065)	0.892*** (0.065)
star rating	0.066** (0.028)	-0.019 (0.043)	0.006 (0.013)
customer review	-0.013 (0.039)	0.061** (0.029)	0.006 (0.011)
sub-market dummies			
financial	-0.092 (0.592)	-0.084 (0.621)	0.347 (0.332)
central	-0.709 (0.609)	-1.505 (0.957)	1.610** (0.699)
lower	-0.349 (0.608)	-0.330 (0.651)	0.290 (0.623)
lowereast	-0.061 (0.621)	-0.117 (0.673)	0.440 (0.675)
lowerwest	-0.210 (0.575)	-0.160 (0.600)	0.411* (0.234)
mideast	-0.003 (0.573)	0.056 (0.595)	0.302* (0.163)
midwest	-0.019 (0.573)	0.062 (0.594)	0.139 (0.121)
timessquare	0.021 (0.573)	0.082 (0.596)	0.312** (0.159)
uppereast	-0.301 (0.584)	-0.354 (0.621)	0.584* (0.351)
upperwest	0.206 (0.578)	0.346 (0.600)	0.191 (0.352)
uptowneast	1.357 (0.986)	1.221 (1.196)	0.886 (1.559)
const	-0.197 (0.588)	0.317 (0.610)	0.011 (0.046)
Log likelihood	-7800.5209	-7542.7694	
# obs	32223	32223	

4.6 Conclusions

In this paper, we propose methodology to identify competition sets from customers' perspective. Our main contributions are three-folds: 1) We develop a customer-centric, data-driven approach to tackle the question that challenges many businesses – who are we really competing with? 2) We provide a network view to understand market competition structure; 3) We point out one important potential of using online data – to study competition among firms. We apply this methodology to hotel industry, using a combination of search, clickstream and hotel data from a major OTA. Using this data, we also identify price matching behavior among hoteliers, and use it as a basis to develop competition sets from hoteliers' perspective. We contrast the two competition networks on a subset of hotels in Manhattan. We find 50% mismatch of the two competition sets. We find that usually independent and distant hotels which are left out from competition sets. Meanwhile, hotels have a tendency to compare themselves to other hotels with lower star ratings, lower price levels, lower ranks and higher customers reviews.

A shortcoming with our application is that the data is from a single OTA. That is, we do not observe customers actions on other websites, such as other OTAs or hotels' own websites. Considering the amount of overlap of listed hotels among multiple OTA sites (especially in the major destination that we study in this paper – Manhattan), and the fact that 70% of the customers visit a travel-related site before booking on supplier's own website, the concern is mitigated. Certainly, a more comprehensive dataset would allow for more precise estimation.

Our approach can easily be applied to many other industries. We note that there exist various formats of data which can support the type of analysis that we conduct in this paper. Some examples of this type of data are: "customers who viewed this product also viewed ...", product comparison data tracked by sites providing "compare" options for online shoppers, and even offline technologies tracking customer movements in supermarkets. We would like to highlight that the business analytics opportunity of using this type of data is vast and promising.

We see this study as opening up many future research avenues on network competition as well. In the competition network of Manhattan hotels, we find that, on average, a hotel has 6 to 8 close competitors, and the average number of steps needed to reach from one hotel to another is 3. The small degree of separation potentially means that a small perturbation, such as a price promotion, may quickly spread to other parts of the market through competitive links. This would be an interesting phenomenon to study. Additionally, competition network is an endogenous network. How does it form in the market-place, and how do firms choose to position themselves in such a network is also an interesting question to answer. Other topics like dynamic evolution of competition network and overlap of competition network with other networks are also worth addressing.

Chapter 5

Appendices

5.1 Partnering with Competitors — An Empirical Analysis of Airline Alliances and Multimarket Competition

Table 5.1: Replacing Regional Carriers by Major Carriers

Regional Carrier Code	Regional Carrier Name	Ticketing Carrier	Major Carrier or LCC	Regional Carrier Code	Regional Carrier Name	Ticketing Carrier	Major Carrier or LCC
16	PSA Airlines Inc.		US	F8	Freedom Airlines d/b/a HP Expr	DL	DL
CP	Compass Airlines		NW	F8		HP	HP
CS	Continental Micronesia		CO	NA	North American Airlines	AA	AA
EV	Atlantic Southeast Airlines		DL	NA		NA	NA
G7	GoJet Airlines, LLC d/b/a United Express		UA	OO	SkyWest Airlines Inc.	OO	OO
HQ	Harmony Airways		DL	OO		CO	DL
J7	Valujet Airlines Inc.		FL	OO		DL	DL
L4	Lynx Aviation d/b/a Frontier Airlines		F9	OO		NW	DL
MQ	American Eagle Airlines Inc.		AA	OO		UA	UA
OH	Comair Inc.		DL	OO		US	UA
OW	Executive Airlines		AA	QX	Horizon Air	AS	AS
RD	Ryan International Airlines		FL	QX		F9	F9
RU	Expressjet Airlines Inc.		CO	RW	Republic Airlines	F9	F9
TB	USAir Shuttle		US	RW		HP	HP
U2	UFS Inc.		UA	RW		UA	UA
XE	Expressjet Airlines Inc.		CO	RW		US	US
XJ	Mesaba Airlines		NW	RW		YX	YX
9E	Pinnacle Airlines Inc.	CO	NW	S5	Shuttle America Corp.	S5	S5
9E		DL	DL	S5		CO	DL
9E		NW	NW	S5		DL	DL
9E		YX	YX	S5		NW	DL
9L	Colgan Air	CO	CO	S5		UA	UA
9L		UA	UA	S5		US	UA
9L		US	US	YV	Mesa Airlines Inc.	HP	HP
AX	Trans States Airlines	AA	AA	YV		LH	
AX		DL	DL	YV		UA	UA
AX		UA	UA	YV		US	US
AX		US	US	YV		YV	YV
DH	Independence Air	CO	DL	ZW	Air Wisconsin Airlines Corp	UA	UA
DH		DH	DH	ZW		US	US
DH		DL	DL				
DH		NW	DL				
DH		UA	UA				

Table 5.2: Airports to MSAs

Airport	MSA	Airport	MSA
ALB	Albany-Schenectady-Troy, NY (MSA)	MKE	Milwaukee-Waukesha-West Allis, WI (MSA)
ABQ	Albuquerque, NM (MSA)	MSP	Minneapolis-St. Paul-Bloomington, MN-WI (MSA)
ATL	Atlanta-Sandy Springs-Marietta, GA (MSA)	BNA	Nashville-Davidson-Murfreesboro-Franklin, TN (MSA)
AUS	Austin-Round Rock, TX (MSA)	MSY	New Orleans-Metairie-Kenner, LA (MSA)
BWI	Baltimore-Towson, MD (MSA)	EWR	New York-Newark-Bridgeport, NY-NJ-CT-PA (CSA)
BHM	Birmingham-Hoover, AL (MSA)	ISP	New York-Newark-Bridgeport, NY-NJ-CT-PA (CSA)
BOI	Boise City-Nampa, ID (MSA)	JFK	New York-Newark-Bridgeport, NY-NJ-CT-PA (CSA)
BOS	Boston-Cambridge-Quincy, MA-NH (MSA)	LGA	New York-Newark-Bridgeport, NY-NJ-CT-PA (CSA)
BUF	Buffalo-Niagara Falls, NY (MSA)	OKC	Oklahoma City, OK (MSA)
RSW	Cape Coral-Fort Myers, FL (MSA)	OMA	Omaha-Council Bluffs, NE-IA (MSA)
CLT	Charlotte-Gastonia-Concord, NC-SC (MSA)	MCO	Orlando-Kissimmee, FL (MSA)
MDW	Chicago-Naperville-Joliet, IL-IN-WI (MSA)	PHL	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD (MSA)
ORD	Chicago-Naperville-Joliet, IL-IN-WI (MSA)	PHX	Phoenix-Mesa-Scottsdale, AZ (MSA)
CVG	Cincinnati-Middletown, OH-KY-IN (MSA)	PIT	Pittsburgh, PA (MSA)
CLE	Cleveland, TN (MSA)	PDX	Portland-Vancouver-Beaverton, OR-WA (MSA)
CMH	Columbus, OH (MSA)	PVD	Providence-New Bedford-Fall River, RI-MA (MSA)
DAL	Dallas-Fort Worth-Arlington, TX (MSA)	RDU	Raleigh-Cary, NC (MSA)
DFW	Dallas-Fort Worth-Arlington, TX (MSA)	RNO	Reno-Sparks, NV (MSA)
DEN	Denver-Aurora, CO (MSA)	RIC	Richmond, VA (MSA)
DTW	Detroit-Warren-Livonia, MI (MSA)	ROC	Rochester, NY (MSA)
ELP	El Paso, TX (MSA)	SMF	Sacramento-Arden-Arcade-Roseville, CA (MSA)
BDL	Hartford-West Hartford-East Hartford, CT (MSA)	SLC	Salt Lake City, UT (MSA)
HOU	Houston-Sugar Land-Baytown, TX (MSA)	SAT	San Antonio, TX (MSA)
IAH	Houston-Sugar Land-Baytown, TX (MSA)	SAN	San Diego-Carlsbad-San Marcos, CA (MSA)
IND	Indianapolis-Carmel, IN (MSA)	OAK	San Jose-San Francisco-Oakland, CA (CSA)
JAX	Jacksonville, FL (MSA)	SFO	San Jose-San Francisco-Oakland, CA (CSA)
MCI	Kansas City, MO-KS (MSA)	SJC	San Jose-San Francisco-Oakland, CA (CSA)
LAS	Las Vegas-Paradise, NV (MSA)	SEA	Seattle-Tacoma-Bellevue, WA (MSA)
BUR	Los Angeles-Long Beach-Riverside, CA (CSA)	GEG	Spokane, WA (MSA)
LAX	Los Angeles-Long Beach-Riverside, CA (CSA)	STL	St. Louis, MO-IL (MSA)
ONT	Los Angeles-Long Beach-Riverside, CA (CSA)	TPA	Tampa-St. Petersburg-Clearwater, FL (MSA)
SNA	Los Angeles-Long Beach-Riverside, CA (CSA)	TUS	Tucson, AZ (MSA)
LGB	Los Angeles-Long Beach-Santa Ana, CA (MSA)	TUL	Tulsa, OK (MSA)
SDF	Louisville-Jefferson County, KY-IN (MSA)	ORF	Virginia Beach-Norfolk-Newport News, VA-NC (MSA)
MHT	Manchester-Nashua, NH (MSA)	DCA	Washington-Arlington-Alexandria, DC-VA-MD-WV (MSA)
MEM	Memphis, TN-MS-AR (MSA)	IAD	Washington-Arlington-Alexandria, DC-VA-MD-WV (MSA)
FLL	Miami-Fort Lauderdale-Pompano Beach, FL (MSA)	PBI	West Palm Beach-Boca Raton-Boynton Beach,
MIA	Miami-Fort Lauderdale-Pompano Beach, FL (MSA)		FL Metropolitan Division

5.2 Are Consumers Strategic? Structural Estimation from the Air-Travel Industry

Estimation under Fixed-Effect and Serial Correlation

To estimate the model with fixed effects and serial correlation in the baseline demand, we apply the following transformation. For simplicity of illustration, in the following discussion we assume a constant fraction of strategic consumers and the same price sensitivity for both non-strategic and strategic consumers. Similar logic can easily be applied to more complicated models. In the demand model, let X_{mt} represent market-departure date characteristics X_m, polynomials of booking time t and the constant

$$\begin{aligned} d_{mt} &= (1-\theta z_{mt})\big(-\beta p_{mt} + X_{mt}\delta + \varepsilon_{mt}\big) + \theta z_{m,t-1}\big(-\beta p_{mt} + X_{m,t-1}\delta + \varepsilon_{m,t-1}\big) \\ \varepsilon_{mt} &= \mu_m + \epsilon_{mt}, \\ \epsilon_{mt} &= \rho\epsilon_{m,t-1} + \nu_{mt}. \end{aligned}$$

The demand model can also be written as

$$\begin{aligned} d_{mt} &= \hat{d}_{mt} + (1-\theta z_{mt})\nu_{mt}, \\ \hat{d}_{mt} &= (1-\theta z_{mt})\big(-\beta p_{mt} + X_{mt}\delta + \rho\varepsilon_{m,t-1}\big) + \theta z_{m,t-1}\big(-\beta p_{mt} + X_{m,t-1}\delta + \varepsilon_{m,t-1}\big) \end{aligned}$$

Note that previously realized demand shock $\varepsilon_{m,t-1}$ can be used to predict demand $\hat{d}_{mt}$ once it is estimated from the previously realized demand $d_{m,t-1}$.

To obtain consistent estimates through non-linear least square, all we need is the following assumption,

$$E\big[(1-\theta z_{mt})\nu_{mt}|X,p\big]=0,$$

which is satisfied by our assumption that $E\big[\nu_{mt}|X,p\big]=0$.

Our goal is to find the parameter set $(\theta,\beta,\rho,\delta)$ that minimizes the sum of squares $\sum(d_{mt}-\hat{d}_{mt})^2=\sum(1-\theta z_{mt})^2\hat{\nu}_{mt}^2$, while allowing for possible price endogeneity. Price p_{mt} can be correlated with μ_m. Meanwhile, price p_{mt} can also be correlated with ε_{mt} through correlation with $\varepsilon_{m,t-1}$.

To estimate the model, the general idea is to transform the nonlinear problem to a linear problem — the traditional panel data problem with fixed effect and serial correlation. First, move price to the left-hand side in the demand model, and the right-hand side can be written as the the multiplication of a transformation matrix $Z(\theta)$ and a column vector

$$\begin{aligned}
d_{mt}+(1-\theta z_{mt}+\theta z_{m,t-1})\beta p_{mt} &= (1-\theta z_{mt})\big(X_{mt}\delta+\mu_m+\epsilon_{mt}\big)\\
&\quad+\theta z_{m,t-1}\big(X_{m,t-1}\delta+\mu_m+\epsilon_{m,t-1}\big),\\
d+\beta(1-\theta z+\theta lagz)p &= Z(\theta)\big(X\delta+\mu+\epsilon\big),\\
Z^{-1}(\theta)\big(d+\beta(1-\theta z+\theta lagz)p\big) &= X\delta+\mu+\epsilon.
\end{aligned}$$

where $lagz$ is z lagged by one period of time. Let $\tilde{d}_{mt}(\theta, \beta)$ denote the left-hand side. Since

$$\begin{aligned} \tilde{d}_{mt}(\theta, \beta) &= X_{mt}\delta + \mu_m + \epsilon_{mt} = X_{mt}\delta + \mu_m + \rho\epsilon_{m,t-1} + \nu_{mt}, \\ \tilde{d}_{m,t-1}(\theta, \beta) &= X_{m,t-1}\delta + \mu_m + \epsilon_{m,t-1}. \end{aligned}$$

Take partial difference to remove serial correlation,

$$\tilde{d}_{mt}(\theta, \beta) - \rho\tilde{d}_{m,t-1}(\theta, \beta) = (X_{mt} - \rho X_{m,t-1})\delta + (1-\rho)\mu_m + \nu_{mt}, \tag{5.1}$$

$$\tilde{\tilde{d}}_{mt}(\theta, \beta, \rho) = \tilde{X}_{mt}(\rho)\delta + \tilde{\mu}_m(\rho) + \nu_{mt}. \tag{5.2}$$

Now, the problem is transformed to a traditional fixed-effect problem with transformed fixed-effect $\tilde{\mu}_m(\rho)$.

$$\tilde{\tilde{d}}_{mt}(\theta, \beta, \rho) - \bar{\tilde{\tilde{d}}}_{m\cdot}(\theta, \beta, \rho) = (\tilde{X}_{mt}(\rho) - \bar{\tilde{X}}_{m\cdot}(\rho))\delta + \nu_{mt} - \bar{\nu}_{m\cdot}, \tag{5.3}$$

$$\dot{d}_{mt}(\theta, \beta, \rho) = \dot{X}_{mt}(\rho)\delta + \dot{\nu}_{mt}, \tag{5.4}$$

where $\bar{\tilde{\tilde{d}}}_{m\cdot}$, $\bar{\tilde{X}}_{m\cdot}$, and $\bar{\nu}$ represent the average of corresponding variables across time t.

Now, given any (θ, β, ρ), we are be able to obtain an unbiased estimate of δ using OLS estimator. The sum of squares, i.e. $\sum(1-\theta z_{mt})^2\hat{\nu}_{mt}^2$, can be calculated as follows,

1. For any given (θ, β, ρ), compute $\dot{d}_{mt}, \dot{X}_{mt}$.

2. Compute the OLS estimator of δ in Equation 5.4; denote as $\hat{\delta}$.

3. Compute estimates of residuals in Equation 5.2, $\hat{\tilde{\mu}}_m + \hat{\nu}_{mt} = \tilde{\tilde{d}}_{mt}(\theta, \beta, \rho) - \tilde{X}_{mt}\hat{\delta}$.

4. Estimate transformed fixed effect $\hat{\tilde{\mu}}_m$ using the average of estimated residuals obtained in step 3. Calculate $\hat{\nu}_{mt}$ subsequently.

5. Compute *sum of squares*$(\theta, \beta, \rho) = \sum(1 - \theta z_{mt})^2 \hat{\nu}^2_{mt}$.

The nonlinear least square procedure is applied to find the optimal (θ, β, ρ) that minimizes *sum of squares*(θ, β, ρ). One additional comment about the initial observation, since we take partial difference in the model, the degree of freedom decreases by one within each group, i.e., market-departure-date. Simply omitting the first observation is likely to cause inefficiency when the number of groups is small, i.e., 45 departure dates in each market. We apply Prais-Winsten transformation to the first observation in each group, i.e. multiply by the first error term by $\sqrt{1 - \rho^2}$.

Additional Table

Table 5.3: Prediction Models for Weak- and Strong-Form Rational Expectations: Market L

Weak-Form Rational Exp. Probit Model $Pr(P_t + 1 < p_t)$		**Strong-Form Rational Exp. Supply Model**	
current_to_last	-0.035	**Price**$_{t-1}$	0.924
	(0.732)		(0.033)
current_to_inital	10.818	**cumulative demand**$_{t-1}$	0.046
	(2.819)		(0.059)
current_to_mktavg	-8.257		
	(2.854)		
weeklyfare_cv	-4.167		
	(1.756)		
initial price	0.041	**initial price**	0.043
	(0.013)		(0.011)
$\boldsymbol{t}$	-0.580	$\boldsymbol{t}$	6.186
	(0.216)		(0.055)
$\boldsymbol{t^2}$	0.127	$\boldsymbol{t^2}$	-1.382
	(0.039)		(0.946)
$\boldsymbol{t^3}$	-0.008	$\boldsymbol{t^3}$	0.076
	(0.002)		(0.045)
		final week	31.463
			(7.177)
high season	-1.070	**high season**	10.699
	(0.325)		(5.450)
day-of-week dummies	yes	**day-of-week dummies**	yes
const	-12.190	**const**	-3.255
	(3.381)		(16.176)
pseudo R-square	0.1729	**R-square**	0.8751

Bibliography

Aguirregabiria, V., C. Ho. 2012. A dynamic oligopoly game of the US airline industry: Estimation and policy experiments. *Journal of Econometrics* Forthcoming.

Allon, G., A. Federgruen. 2009. Competition in service industries with segmented markets. *Management Science* **55**(4) 619–634.

Allon, G., A. Federgruen, M. Pierson. 2011. How much is a reduction of your customers' wait worth? An empirical study of the fast-food dirve-thru industry based on structural estimation methods. *Manufaturing and Service Operations Management* **13**(4) 489–507.

Anderson, C.K., J.G. Wilson. 2003. Wait or buy? The strategic consumer: pricing and profit implications. *Journal of the Operational Research Society* **54**(3) 299–306.

Anderson, C.K., X.Q. Xie. 2011. A Choice-Based Dynamic Programming Approach for Setting Opaque Prices. *Production and Operations Management* **21**(3) 590–605.

Anderson, S., R.G. Newell. 2003. Simplified marginal effects in discrete choice models. *Economics Letters* **81**(3) 321–326.

Arellano, M., S. Bond. 1991. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies* **58**(2) 277–297.

Arikan, M., V. Deshpande. 2010. The impact of airline flight schedules on flight delays. Working paper, Purdue University, West Lafayette, IN.

Arikan, M., V. Deshpande, M. Sohoni. 2010. Building reliable air-travel infrastructure using empirical data and stochastic models of airline networks. Working paper, Purdue University, West Lafayette, IN.

Armantier, O., O. Richard. 2008. Domestic airline alliances and consumer welfare. *RAND Journal of Economics* **39**(3) 875–904.

Aviv, Y., A. Pazgal. 2008. Optimal pricing of seasonal products in the presence of forward-looking consumers. *Manufacturing & Service Operations Management* **10**(3) 339–359.

Bain, J. 1956. *Barriers to New Competition*. Harvard University Press, Cambridge, MA.

Bajari, P., L. Benkard, J. Levin. 2007. Estimating dynamic models of imperfect competition. *Econometrica* **75**(5) 1331–1370.

Baum, J.A.C., H.J. Korn. 1996. Comeptitive dynamics of interfirm rivalry. *Academy of Management Journal* **39**(2) 255–291.

Baum, J.A.C., H.J. Korn. 1999. Dynamics of dyadic competitive interaction. *Strategic Management Journal* **20** 251–278.

Belobaba, P.P. 1989. Application of a probabilitics decision model to airline seat inventory control. *Operations Research* **37**(2) 183–197.

Benkard, C.L., A. Bodoh-Creed, J. Lazarev. 2010. The long run effects of us airline mergers. Working paper, Yale University, New Haven, CT.

Berry, S. 1992. Estimation of a model of entry in the airline industry. *Econometrica* **60**(4) 889–917.

Berry, S., J. Levinsohn, A. Parkes. 1995. Automobile prices in market equilibrium. *Econometrica* **63**(4) 841–890.

Besley, T., R. Burgess. 2004. Can Labor Regulation Hinder Economic Performance? Evidence From India. *Quarterly Journal of Economics* **119**(1) 94–134.

Bodea, T., M. Ferguson, L. Garrow. 2009. Choice-Based Revenue Management: Data from a Major Hotel Chain. *Manufacturing & Service Operations Management* **11**(2) 356–361.

Borenstein, S., N.L. Rose. 1994. Competition and Price Dispersion in the U.S. Airline Industry. *Journal of Political Economy* **102**(4) 653–683.

Bouma, G. 2007. Normalized (Pointwise) Mutual Information in Collocation Extraction. *Proceedings of GSCL* 31–40.

Boyd, E.A. 2007. *The Future of Pricing: How Airline Ticket Pricing Has Inspired a Revolution*. Palgrave-MacMillan, New York.

Brueckner, J.K. 2003. International airfares in the age of alliances: The effect of codesharing and antitrust immunity. *The Review of Economics and Statistics* **85**(1) 105–118.

Buell, R.W., D. Campbel, F.X. Frei. 2011. How do incumbents fare in the face of increased service competition? Working Paper, Harvard University, Cambridge, MA.

Busse, M. 2002. Firm financial condition and airline price wars. *RAND Journal of Economics* **33**(2) 298–318.

Cachon, G. 2001. Stock wars: Inventory competition in a two echelon supply chain. *Operations Research* **49**(5) 658–674.

Cachon, G., M. Olivares. 2009. Competing retailers and inventory: an empirical investigation of General Motors' dealerships in isolated U.S. markets. *Managment Science* **55**(9) 1586–

1604.

Cachon, G.P., R. Swinney. 2009. Purchasing, pricing, and quick response in the presence of strategic consumers. *Management Science* **55**(3) 497–511.

Chamberlain, G. 1980. Analysis of covariance with qualitative data. *Review of Economics Studies* **47**(1) 225–238.

Chellappa, R.K., R.G. Sin, S. Siddarth. 2011. Price Formats as a Source of Price Dispersion: A Study of Online and Offine Prices in the Domestic U.S. Airline Markets. *Information Systems Research* **22**(1) 83–98.

Chevalier, J., A. Goolsbee. 2009. Are durable goods consumers forward-looking? Evidence from the college textbooks. *The Quarterly Journal of Economics* **124**(4) 1853–1884.

Cho, M., M. Fan, Y.P. Zhou. 2007. An empirical study of revenue management practices in the airline industry. Working Paper, University of Washington, Seattle, WA.

Cho, M., M. Fan, Y.P. Zhou. 2008. Strategic consumer response to dynamic pricing of perishable products. Working paper, University of Washington, Seattle, WA.

Church, K.W., P. Hanks. 1990. Word Association Norms, Mutual Information and Lexicography. *Computational Linguistics* **16**(1) 22–29.

Clemons, E.K., L. Hitt, I.H. Hann. 2002. Price dispersion and differentiation in online travel: An empirical investigation. *Management Science* **48**(4) 534–550.

Cross, R.G., J.A. Higbie, D.Q. Cross. 2009. Revenue Management's Renaissance: A Rebirth of the Art and Science of Profitable Revenue Generation. *Cornell Hospitality Quarterly* **50**(1) 56–81.

Dana, J., E. Orlov. 2009. Internet penetration and capacity utilization in the us airline industry.

Working Paper, Northeastern University, Boston, MA.

Dillon, W.R., A. Kumar. 1994. Latent structure and other mixture models in marketing: An integrative survey and overview. R.P. Bagozz, ed., *Advanced ethods of Marketing Research*. Blackwell Publisher, Cambridge, 352–388.

Edwards, C.D. 1955. *Conglomerate Bigness as a Source of Power*. Princeton University Press.

Erdem, T., M.P. Keane. 1996. Decision-making under uncertainty: Capturing dynamic choice processes in turbulent consumer goods markets. *Marketing Science* **15**(1) 1–20.

Etzioni, O., R. Tuchinda, C.A. Knoblock, A. Yates. 2003. To buy or not to buy: Mining airfare data to minimize ticket purchase price. *In Proceedings of KDD03*. ACM Press, 119–128.

Fader, P., W. Moe. 2004. Dynamic Conversion Behavior at e-Commerce Sites. *Management Science* **50**(3) 326–335.

Fader, P., Y.H. Park. 2004. Modeling Browsing Behavior at Multiple Web Sites. *Marketing Science* **23** 280–303.

Farias, V. F., S. Jagabathula, D. Shah. 2012. A New Approach to Modeling Choice with Limited Data. *Management Science* Forthcoming.

Gallego, G., L. Lin, R. Ratliff. 2009. Choice-based emsr methods for single-leg revenue management with demand dependencie. *Journal of Revenue & Pricing Management* **8** 207–240.

Gallego, G.R., G. van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Marketing Science* **40**(8) 999–1020.

Gayle, P.G. 2008. An empirical analysis of the competitive effects of the delta/continental/northwest code-share alliance. *Journal of Law and Economics* **40**(4) 743–766.

General Accounting Office. 1998. Aviation competition- proposed domestic airline alliances raise serious issues.

Ghose, A., P. Ipeirotis, B. Li. 2012. Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowd-Sourced Content. *Marketing Science* Forthcoming.

Gimeno, J., C.Y. Woo. 1996. Hypercompetitionin a Multimarket Environment: The Role of Strategic Similarity and Multimarket Contact on Competitive De-Escalation. *Organization Science* **7** 322–341.

Gokpinar, B., W.J. Hopp, S.M.R. Iravani. 2010. The Impact of Misalignment of Organizational Structure and Product Architecture on Quality in Complex Product Development. *Management Science* **56**(3) 468–484.

Goyal, M., S. Netessine. 2007. Strategic technology choice and capacity investment under demand uncertainty. *Management Science* **53**(2) 192–207.

Granados, N.F., A. Gupta, R.J. Kauffman. 2012. Online and offline demand and price elasticities: Evidence from the air travel industry. *Information Systems Research* Forthcoming.

Greve, H.R. 2006. The intent and extent of multimarket contact. *Strategic Organization* **4**(3) 249–274.

Greve, H.R., J.A.C. Baum. 2001. A multiunit, multimarket world. J.A.C. Baum, H.R. Greve, eds., *Multiunit Organization and Multimarket Strategy: Advances in Strategic Management*, vol. 18. JAI Press, Oxford, 1–28.

Guajardo, J. A., M.A. Cohen, S. Netessine. 2011. Service competition and product quality in the U.S. automobile industry. Working paper, University of Pennsylvania, Philadelphia, PA.

Hendel, I., A. Nevo. 2006. Measuring the implications of sales and consumer inventory behavior. *Econometrica* **74**(6) 1637–1673.

Hendel, I., A. Nevo. 2011. Intertemporal price discrimination in storable good markets. Working paper, Northwestern University, Evanston, IL.

Hoque, A.Y., G.L. Loshe. 1999. An Information Search Cost Perspective for Designing Interfaces for Electronic Commerce. *Journal of Marketing Research* **36**(3) 387–394.

Hu, X., R. Caldentey, G. Vulcano. 2010. Revenue sharing in airline alliances. Working Paper, New York University, New York City, NY.

Ito, H., D. Lee. 2007. Domestic codesharing, alliances and airfares in the u.s. airline industry. *Journal of Law and Economics* **50**(2) 355–380.

Jayachandran, S., J. Gimeno, P.R. Varadarajan. 1999. The Theory of Multimarket Competition: A Synthesis and Implications for Marketing Strategy. *The Journal of Marketing* **63**(3) 49–66.

Jerath, K., S. Netessine, S.K. Veeraraghavan. 2010. Revenue management with strategic customers: Last-minute selling and opaque selling. *Management Science* **56**(3) 430–448.

Kim, J.Y., L. Canina. 2009. Product Tiers and ADR Clusters: Integrating Two Methods for Determining Hotel Competitive Sets. *Cornell Hospitality Report* 207–240.

Kranton, R.E., D.F. Minehart. 2001. A theory of buyer-seller network. *The American Economic Review* **91**(3) 485–508.

Levin, Y., J. McGill, M. Nediak. 2009. Dynamic pricing in the presence of strategic consumers and oligopolistic competition. *Management Science* **55**(1) 32–46.

Li, J., N. Granados, S. Netessine. 2011. Are Consumers Strategic? Structural Estimation from the Air Travel Industry. Working paper, University of Pennsylvania, Philadelphia, PA.

Li, W., M. Lipson, K. Ramdas, J. Williams. 2010. Can stock price movements inform operational improvement efforts? Evidence from the airline industry. Working Paper, University of Virginia, Charlottesville, VA.

Lippman, S.A., K.F. McCardle. 1997. The competitive newsboy. *Operations Research* **45**(1) 54–65.

Liu, Q., G.J. van Ryzin. 2008. Strategic capacity rationing to induce early purchases. *Management Science* **54**(6) 1115–1131.

Magnac, T., D. Thesmar. 2002. Identifying dynamic discrete decision processes. *Econometrica* **70**(2) 801–816.

Mahajan, S., G. van Ryzin. 2001. Inventory competition under dynamic consumer choice. *Operations Research* **49**(5) 646–657.

Manning, C., H. Schutze. 1999. *Foundations of Statistical Natural Language Processing.*. MIT Press, Cambridge, MA.

Mantin, B., D. Gillen. 2011. The hidden information content of price movements. *European Journal of Operational Research* Forthcoming.

Morrison, S.A., C. Winston. 1996. Causes and consequences of airline farewars. *Brookings Papers: Microeconomics* 85–123.

Mundlak, Y. 1978. On the pooling of time series and cross section data. *Econometica* **46**(1) 69–85.

Nair, H. 2007. Intertemporal price discrimination with forward-looking consumers: Application to the us market for console video-games. *Quantitative Marketing and Economics* **5**(3)

239–292.

Netessine, S., N. Rudi. 2003. Centralized and competitive inventory models with demand substitution. *Operations Research* **51**(2) 329–335.

Netessine, S., R.A Shumsky. 2005. Revenue management games: Horizontal and vertical competition. *Management Science* **51**(5) 813–831.

Netessine, S., C. Tang, eds. 2009. *Consumer-driven demand and operations management models*. Springer Publisher, New York, NY.

Netzer, O., R. Feldman, J. Goldenberg, M. Fresko. 2012. Mine Your Own Business: Market-Structure Surveillance Through Text Mining. *Marketing Science* Forthcoming.

Newman, J.P., M.E. Ferguson, L.A. Garrow. 2010. Estimation of choice-based models using sales data from a single firm. Working paper, Georgia Institute of Technology, Atlanta, GA.

Olivares, M., G.P. Cachon. 2009. Competing Retailers and Inventory: An Empirical Investigation of General Motors Dealerships in Isolated U.S. Markets. *Management Science* **55**(9) 1586–1604.

Osadchiy, N., E. Bendoly. 2010. Are consumers really strategic? Implications from an experimental study. Working paper, Emory University, Atlanta, GA.

Perakis, G., A. Sood. 2006. Competitive multi-period pricing for perishable products: A robust optimization approach. *Mathematical Programming* **107**(1–2) 295–335.

Pinkse, J., M. Slade, C. Brett. 2002. Spatial Price Competition: A Semiparametric Approach. *Econometrica* **70** 1111–1153.

Ramdas, K., J. Williams. 2009. An empirical investigation into the tradeoffs that impact on-time performance in the airline industry. Working Paper, London Business School, London, UK.

Rust, J. 1994. Structural estimation of markov decision processes. R. Engle, D. McFadden, eds., *Handbook of Econometrics*, vol. 4. Elsevier, Amsterdam: North Holland, 3081–3143.

Scott, J.T. 1988. Diversification versus co-operation in R&D investment. *Managerial and Decision Economics* **9** 173–186.

Shen, Z.M., X. Su. 2007. Customer behavior modeling in revenue management and auctions: A review and new research opportunities. *Production and Operations Management* **16**(6) 713–728.

Sosa, M.E., S.D. Eppinger, C.M. Rowles. 2004. The misalignment of product architecture and organizational structure in complex product development. *Management Science* **50**(12) 1674–1689.

Stephan, J., J.P. Murmann, W. Boeker, J. Goodstein. 2003. Bringing Managers into Theories of Multimarket Competition: CEOs and the Determinants of Market Entry. *Organization Science* **14**(4) 403–421.

Su, X. 2007. Inter-temporal pricing with strategic consumer behavior. *Management Science* **53**(5) 726–741.

Talluri, K.T., G. J. van Ryzin. 2004. *The Theory and Practice of Revenue Management*. Kluwer Academic Publishers, Boston, MA.

Talluri, K.T., G.J. van Ryzin. 2005. *The Theory and Practice of Revenue Management*. Springer, New York.

Thrane, C. 2007. Examining the determinants of room rates for hotels in capital cities: The Oslo experience. *Journal of Revenue & Pricing Management* **5**(4) 315–323.

Train, K. 2003. *Discrete choice methods with simulation*. Cambridge University Press, Cambridge, UK.

Vulcano, G., G. van Ryzin, W. Chaar. 2010. Choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing and Service Operations Management* **12**(3) 371–392.

Withiam, G. 2011. Brave New World: Online Hotel Distribution. *Cornell Hospitality Roundtable & Conference Proceedings* **3**(4).

Wooldridge, J.M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. The MIT Press, Cambridge, Massachusetts, London, England.

Wright, C.P., H. Groenevelt, R.A. Shumsky. 2010. Dynamic revenue management in airline alliances. *Transportation Science* **44**(1) 15–37.

Zhang, D., W.L. Cooper. 2006. Revenue management for parallel flights with customer-choice behavior. *Operations Research* **53**(3) 415–431.

Zhao, X., D.R. Atkins. 2008. Newsvendors under simultaneous price and inventory competition. *Manufacturing & Service Operations Management* **10**(2) 539–546.