The Nature of Competition in Electronic Markets:
An Empirical Investigation of Online Travel Agent Offerings

Eric K. Clemons
clemons@wharton.upenn.edu

Il-Horn Hann
hann36@wharton.upenn.edu

Lorin M. Hitt
lhitt@wharton.upenn.edu

Department of Operations and Information Management
The Wharton School, University of Pennsylvania
1300 Steinberg Hall-Dietrich Hall
Philadelphia, PA 19104

June, 1999

We would like to thank David Croson, Rachel Croson, Moti Levi, Eli Snir, Ho-Geun Lee, Rebecca Tsui, the participants of the 1997 Workshop on Information Systems and Economics, and several anonymous referees for helpful comments on earlier drafts of this paper. We would also like to thank Chad Ross for programming assistance. Funding for this work was provided by the Reginald Jones Center Project on Information: Industry Structure and Competitive Strategy.

The Nature of Competition in Electronic Markets:
An Empirical Investigation of Online Travel Agent Offerings

Abstract

Several authors have argued that because modern computing and communications technologies reduce buyer search costs and other market inefficiencies, there should be intense price competition between sellers in electronic markets. This paper examines this prediction using data on the airline ticket offerings of online travel agents (OTA). We find that different OTAs offer tickets with substantially different prices and characteristics when given the same customer request. Some of this variation appears to be due to product differentiation -- different OTAs specialize by systematically offering different tradeoffs between ticket price and ticket quality (minimizing connections, depart or return when requested). However, even after accounting for differences in ticket quality, ticket prices vary by as much as 18% across OTAs. In addition, OTAs return tickets that are strictly inferior to the ticket offered by another OTA for the same request between 2.2% and 28% of the time suggesting that this price dispersion is not a modeling artifact. We conclude that this market is not characterized by perfect competition, but find that OTAs engage in both horizontal product differentiation and price discrimination in addition to simply having some degree of random inefficiency. Our results further suggest that product differentiation is an important component of strategy for electronic markets, even in goods with relatively simple and unambiguous product descriptions.

1. Introduction

Numerous articles in the business press have declared electronic commerce to be a new paradigm for the sale and delivery of goods and services. The Internet provides a potentially low-cost channel for retail distribution that can reach customers 24 hours a day, anyplace in the world. In addition, as customer interaction shifts to electronic channels, detailed data can be collected to improve targeted advertising and sales efforts as well as to reduce other operational costs.

The Internet has increased the availability of information about prices and products, enabling customers to identify the best deal or at least to improve their bargaining position with vendors both online and in traditional channels. For example, Combes and Patel (1997) describe the new customer environment for travel services:

> "...a whole new level of convenience and ubiquity to the shopping experience. Consumers are empowered with the ability to price and compare features with ease. They can inquire about various aspects of a travel destination without having to speak to a travel agent… *or they can quickly and simply find the lowest fare to Las Vegas.*" (italics added)

While there is little debate whether electronic commerce has the potential to create value, it is unclear how this value is allocated between producers and consumers. One stream of argument suggests that as markets move from physical to electronic, they remove inefficiencies on the consumer side, principally search cost. Search cost represents the real and opportunity costs of identifying vendors, locating products, comparing offerings, and making a purchasing decision, These costs may be lower in electronic markets due to factors such as increased convenience, reduced waiting times, and elimination of travel time. The reduction in consumer search costs in electronic markets puts vendors under increased price competition, resulting in converging prices and ultimately eliminating any extraordinary profits (Malone, Yates and Benjamin 1987; Bakos 1991, 1997; Benjamin and Wigand 1995).[1]

[1] There is some anecdotal evidence that support this notion. For example, Chrysler's chairman, Bob Eaton, estimates that in 1998, 25% of all American car buyers will use the Internet to acquire information about the true costs for each model and subsequently bargain with dealers about the mark-up. Information about true costs are the sort of trade secrets that car dealers "…never would have revealed to their wives, let alone their customers" (The Economist, Feb. 14th 1998). With access to nearly unlimited access to product and price information, vendors lose the previously enjoyed information advantage that enabled them to charge different prices to different customers. The ability of charging different prices for the same good – a phenomenon termed price dispersion – has been

However, it is well known that products with complex characteristics enable producers to avoid the outcome of pure price competition through product differentiation (Porter 1985). By offering different products – either different price-quality combinations or different non-quality attributes (color, size) – firms can soften competition and increase their profits in markets where consumers have heterogeneous preferences (Hotelling 1929; Salop 1979; Gabszewicz and Thisse 1979, 1980; Shaked and Sutton 1982, 1983; Perloff and Salop 1985). Product differentiation applies to retailers as well. Retailers differentiate themselves in a variety of ways; such as the range of products that they offer, their geographic locations, and the choice of product or service quality (Chamberlin 1933). It is the ability to appeal to different tastes that enable vendors to segment their markets, to limit direct competition, and thus to charge a premium.

The objective of this paper is twofold: Our first goal is simply to examine whether prices offered by different vendors converge in electronic markets. While declines in search costs are very difficult to measure directly, we can observe whether firms behave as if search costs are zero. Second, we try to understand the extent to which strategic actions prevalent in traditional markets, such as product differentiation, are utilized in electronic markets.

In this study, we investigate the competition among online travel agents (OTA) in the market for air travel. This market is particularly relevant to study for several reasons. First, it has a long history of utilizing electronic distribution (computerized reservation services, electronic booking, electronic ticketing). Second, it is one of the first retail businesses on the Internet, and online air travel sales represent as much as 40% of e-commerce revenue historically, and nearly $4Bn of air travel is expected to be transacted on-line in 1999. (Combes and Patel 1997). Finally, air travel is ideal for studying product differentiation since the underlying good is both complex and unambiguously describable.

We selected five systems, which are available on the World Wide Web (WWW); four popular online travel agents (hereafter named OTA1, OTA2, OTA3, and OTA4), and one proprietary system (OTA5). We examine the prices and characteristics of the tickets to address three questions:

---

explicitly linked to the existence of search costs (see, Stigler 1961; Salop and Stiglitz 1976; Pratt, Wise, and Zeckhauser 1979).

1) Is the flight selection behavior of OTAs consistent with converging prices?

2) Do OTAs utilize strategies of product differentiation?

3) Other than product differentiation, what other economic models can explain the observed flight selection behavior of OTAs?

To summarize our main results, we find that on-line agents offer tickets with substantial price dispersion – using simple averages, we find that the highest price agent is, on average, 28% more expensive than the lowest price agent for the same set of customer requests. After accounting for quality and route differences in tickets using a hedonic price model, the variation in average price across OTAs drops to about 18%; this suggests that some of the difference is due to systematic differences in the quality of tickets offered (product differentiation) in addition to residual price dispersion. This is further supported by the observation that OTAs given the same ticket request return a ticket unambiguously inferior to one offered by another OTA between 2.2% and 28% of the time.

We offer various possible explanations for our empirical results. First, we can clearly identify some degree of product differentiation in OTA services – different OTAs demonstrably are tailored for different customer segments. Second, it is possible that search costs faced by customers are still sufficiently high to allow for equilibrium price dispersion; that is, even though access to online travel agents (OTAs) is readily available, few customers are likely to search all OTAs. This may enable strategic pricing behavior. Perhaps more likely, it may enable firms to continue to operate with less than perfect search engines and still retain customers. Finally, we find informal evidence that the design of the OTA interface can potentially be coupled with product differentiation to price discriminate among consumers. Overall, this suggests that a range of strategies are in use by OTAs and that this market has little in common with the Bertrand-style, perfectly competitive markets described in the literature on electronic markets.

The remainder of this paper is organized as follows: Section 2 reviews the previous literature on search and product differentiation; Section 3 introduces the OTA market and formulates the hypotheses we test; Section 4 describes the methods and data; Section 5 contains the statistical results, hypothesis tests and the discussion; and conclusions appear in Section 6.

2. Previous Literature

In order to understand the nature of competition in electronic markets, it is important to consider models of competition in traditional markets. When competing on price in undifferentiated product markets, Bertrand (1883) argued that firms have to charge the same competitive price; any firm that charged a price higher than the lowest marginal production cost would be undercut and have no buyers, any firm that charged lower than the competitive price would be unprofitable. However, this outcome relies on some fairly stringent assumptions: free entry; full information on prices and product characteristics; large numbers of buyers and sellers; a homogeneous good; and no switching costs. In the context of electronic markets, Bakos (1991) argues that the provision of price information leads to a reduction of buyers' search cost that reduces sellers' profits and increases consumer welfare. Benjamin and Wigand (1995) termed this effect the *market maker effect*, in which "…the consumer does very well, producers lose their profit margins, and the market makers gain the remaining profits. (p. 66)". While there is some anecdotal evidence for such an effect, this has not been confirmed in initial empirical explorations. Bailey, Brynjolfsson, and Smith (1997) found that for near-commodity products like books, CDs, and software, the price dispersion is actually higher among Internet retailers than among conventional retailers. However, at least in the book market, the entry of Barnes & Noble has forced Amazon, the leader in on-line book retailing at the time, to adjust its prices and prices of the two market leaders converged fairly quickly. Lee (1998) explored the Japanese auto-auction market and compared the prices of traditional auctions with the ones of an electronic auction. He found that prices in the electronic auction are higher, but attributed the mark-up to unique characteristics of the electronic auction: higher quality of the cars, buyer externalities, and the option of stating a reservation price.

There is a substantial body of work in economics that has relaxed the Bertrand assumptions for analyzing traditional product markets; these extensions may be relevant in electronic markets as well. The two most promising approaches are models where different goods are not perfect substitutes and customers have heterogeneous preferences (product differentiation) or where consumers do not have full information on prices (information asymmetry).

The first model of product differentiation was provided by Hotelling (1929), who represented consumer tastes as different points on a line (known as the "linear city" model). Salop (1979) considered a similar model (the "circular city" model) where preferences were distributed over a unit circle, avoiding some issues of continuity at the ends of the line. In these models, firms were able to "locate" at different points in the line or circle and capture sales from the closest consumers. Because consumers trade product "closeness" against price, firms are able to charge prices above marginal cost and profits are maximized when firms locate as far away from each other as possible. This price differential exists whenever there is a finite number of firms (a fixed cost of entry is a sufficient assumption for this to occur). Hotelling's and Salop's models belong to the category of *horizontal product differentiation*, which can be interpreted as models of product variety – all products are functionally identical and carry the same price, but consumers make different choices of products based on their idiosyncratic preferences over other characteristics (such as color).[2]

A second class of papers examines pricing and competition when consumers have differing information about prices. The classic result (Stigler 1961) is that when the cost of obtaining information on product price are non-zero, firms will price above marginal cost. Other models have considered different variations of information and search costs[3] with the same general type of result: as consumers are less well informed about prices, firms can charge higher prices. Bakos (1991, 1997) combines search cost models with horizontal product differentiation models to model the effects of reduced search costs on competition in electronic markets. He finds that as search costs for price information decline, prices converge to the competitive price, although the rate of convergence is slowed by the presence of product differentiation.

[2] Products can also be differentiated in quality, which is termed vertical product differentiation. However, this form of differentiation is not relevant in our context because OTAs do not have control over product price.

3 For example, Salop and Stiglitz (1976) examine the effects of consumers with different search costs; Burdett and Judd (1983) examine the effects of consumers who start off equally informed, but through random chance obtain different amounts of price information; Varian (1980) considers firm strategies to induce consumer uncertainty by varying price over time.

3. Competition Among Online Travel Agents (OTA)

3.1. Introduction to the Online Travel Market

Online travel agents (OTAs) provide a point of contact via the World Wide Web (WWW) to enable customers to search for appropriate flights and fares and make a selection, which is then booked and ticketed by the OTA. There are literally dozens of OTAs representing independent on-line travel agents, airlines, traditional travel agents, and reservation systems (e.g. Sabre Group's Travelocity), although traffic is fairly concentrated. Of the top 50 travel sites in June, 1999 tracked by 100hot.com, only 10 were full service OTAs and 8 were airlines. According to Forrester research, total on-line travel sales are growing 60% per year and will total 12% of the travel market by 2003.

The air travel industry is ideal for evaluating the effects of product differentiation. The characteristics of the underlying good, airline tickets, are fairly standardized and have well understood means of description, enabling clear identification and differentiation of offerings in electronic markets. Based on interviews with market participants, the key factors that differentiate airline tickets are:

- Timeliness: an inquiry for a specific flight includes the desired departure time for both legs of the flight. An OTA is not bound to report only flights that meet the time requirements specified by the traveler, but can also select other flights, which might be less expensive.

- Number of connections: connections often permit more options to be explained, which may make a flight cheaper and may sometimes be necessary to ensure the timeliness of the flight.

- Length of connections: an OTA has to make the decision about the duration of a connection that is acceptable for the customer, given his or her priorities for time and price.

Additional differences among online travel agents involve the design of their Web site (for example, the quality of the user interface), the speed of response to ticket inquiries, the extent to which they allow users to express preferences about tickets (e.g., search for low cost or search for no connections) and the choice of the computerized reservation system (CRS). The OTAs draw from a common set of products (available flights), except for some minor differences in airline participation in the CRSs. Therefore, if all systems apply the same decision criteria for selecting flights, they will generally return the same information to the consumer.

In this study, we take the structure of the upstream market (airlines) as a given and focus only on how OTAs select from a common set of alternative flights. It is well known that the airline industry does not exhibit perfect competition for a variety of reasons, such as capacity constraints and collusion (Borenstein, 1989, 1991; Borenstein and Rose, 1994). While this creates the opportunity for product differentiation among OTAs, it does not require that OTAs offer different products (with different prices) for a given customer query.

An OTA performs several services in sequential order: reservation information and recommendation service, reservation services, and ticketing service. The first service that OTAs provide is the compilation and delivery of flight reservation information and recommendations in a convenient form. The recommendations that are offered to the customer are sometimes based on the customer's criteria, such as whether price or timeliness is more important (three of the five OTAs we examine provide this facility). The reservation information and recommendation service is also the basis of competition among OTAs. After the customer has selected a flight, the OTA makes a reservation on behalf of the customer. These services are free of charge to the traveler, but they represent the OTA's "gateway" to the revenue stream. When the customer has decided on a flight, he or she authorizes the OTA to carry out the ticketing service. It is the ticketing service that is tied to the compensation for OTA.

The OTA's revenue is derived mostly through payments from the airlines[4]. At the time of data collection, some major domestic airlines compensate online travel agents through a fixed fee, while other airlines compensate OTAs with reduced commissions of 5% or $40, whichever is less. A few airlines pay traditional commissions of 8% of the ticket price, capped by $50. Such a variable commission arrangement may provide incentive for agents to sell more expensive tickets (assuming agents have some market power); the agent optimally trades the higher revenue of offering a higher price ticket against the reduced probability of sale at a higher price.

---

[4] OTAs also generate revenues through banner advertisement and commissions for car rental and hotel bookings.

However, it is important to note that OTAs have no pricing ability themselves, they can select tickets with different prices, but cannot change the price of a ticket.

3.2. Application of the Theory of Search and Product Differentiation to OTAs

A central issue in determining the structure of competition among OTAs is buyer search costs. Search costs in this market represent the time, effort or other costs of locating and obtaining prices from competing OTAs. In this context, it is plausible that these costs are quite low. For the novice OTA user, the ability to search is greatly enhanced through search engines and directory services such as Yahoo! or GoNetwork, which are among the most visited Web sites. For example, it takes less than 5 mouse clicks to get a selection of 12 OTAs through "Yahoo!" and 11 OTAs through "Infoseek". Moreover, some releases of Internet browsers such as Netscape Communicator 4.0 contain direct links to popular OTAs. With the ability to "bookmark" these OTA sites, a customer who repeatedly flies has virtually no search costs at all for locating vendors. Once a customer has input his or her itinerary, the OTA presents the user with potential flight routes within a few minutes. In addition, a customer can simultaneously open several windows and search on different OTAs at the same time. This enables the customer to compare the different flights in a very short time period.

One approach to investigating the decline in search costs is to measure them directly. However, because the idea of a search cost includes both actual costs (e.g. time) as well as costs which are almost impossible to measure (e.g. cognitive effort, disutility of waiting) this would be very difficult to do in the abstract. In addition, it would be very difficult, costly and imprecise to attempt to syncronize behavior between multiple on-line travel agents and multiple physical agents. However, with much greater precision we can observe if market participants behave as if search costs were zero by utilizing results from well-established economic models.

As Bakos (1991, 1997) has shown, even if products are horizontally differentiated, as search costs approach zero, we would expect all vendors to provide the product with the best combination of customer "fit" and price. In addition, it is not more costly to offer an additional flight in the list of suggestions. Therefore, as long as agents do not differ in their information about a particular consumer request (which is true by design in our study) and they have the same prior information about what customer segments exist (which seems reasonable to assume), then their response to a customer query should contain the same flights, including a flight optimal for each customer segment. This translates into:

> Hypothesis H1: For a given customer request (preferred times of departure, origin and destination cities, number of days between request and departure), all OTAs will return an identical set of recommendations.

However, tests of this hypothesis are likely to be too powerful (it will almost certainly be rejected) since a single deviation from the exact set of recommendations is evidence that the systems do not behave the same. This is particularly problematic because offering an additional flight is costless; an OTA could be indifferent between offering the list of optimal tickets and offering the optimal list plus some randomly chosen flights.[5] To avoid the results from being biased by extraneous flight choices, to provide some room for occasional error by the OTAs, and to simplify empirical testing and interpretation, we reformulate this hypothesis with two changes: 1) We restrict the analysis only to flights that would be chosen by two representative types of customers (business and leisure travelers) by selecting a single optimal ticket from the list of choices returned by the OTAs appropriate for each segment, and 2) we focus separately on price attributes and ticket quality attributes.

Thus, we test the following hypotheses:

> Hypothesis H1a: For a given customer request and decision rule, the selected recommendation for each OTA will have the same price

> Hypothesis H1b: For a given customer request and decision rule, the selected recommendation for each OTA will have the same non-price characteristics (specifically, meeting departure and return time windows, number of connections)

On the other hand, there are reasons to believe that not all OTAs will recommend the same flight for a given decision rule. If all OTAs provided exactly the same service and tickets, essentially a commodity service, then they would have little bargaining power against the airlines to retain their commissions. Moreover, since there are moderate to high fixed costs (such as advertising and site design) absent some form of differentiation, this leaves few profit levers except for

5 This depends, of course, on the extent to which an additional choice imposes additional costs on the consumer. In general, the number of choices (with the exception of OTA5) is consistent across different OTAs, the number of alternatives ranged from 1 to 4, with an average of 2 choices for a given request.

scale. Since there are often many alternatives available for travel from point A to point B, OTAs can choose to offer a different set of flights to the customer. In other words, the existence of alternatives may enable them to engage in product differentiation.

Testing for product differentiation among online travel agents is different from testing for product differentiation among producers, such as car manufacturers. Travel agents are middlemen between the airlines and the consumer; they cannot create different air travel products nor alter the price to the consumer, but rather are limited to a selection from available alternatives. Moreover, there is no difference in cost to the OTA for providing a flight of greater convenience. When competing for business, all agents essentially have access to all flights; they generally cannot compete by offering a flight or price not available elsewhere. However, an agent can offer travelers many possible routes from the departure to the destination city at different times and at different prices. Hence, travel agents compete to find the *best* flight for the customer, which represents a bundle of price and non-price characteristics.

In order to see if OTAs are making these choices, it is necessary to consider both price and convenience simultaneously. One method that enables this comparison is to construct a hedonic price model (Chow 1967; Griliches 1961) that relates ticket price to ticket characteristics. By adjusting the price of tickets for their variation in characteristics, we can then compare the prices of the ticket offerings among OTAs on an equivalent basis. If agents differ only because they offer flights with different characteristics and we are able to model the effect of characteristics on price accurately, then we would expect the following to be true:

> Hypothesis H2: For a given customer request (preferred time, preferred number of connections, origin and destination, number of days before departure) and decision rule, all OTAs will return identical ticket prices after controlling for price variation due to ticket characteristics.

If the variation among OTA responses is reduced in the characteristic-adjusted model when compared to the base model, then that suggests at least some of the price difference among the OTAs in the base case is due to differentiation. To the extent that quality variation does not remove all the price variation among OTAs we will have to consider alternative explanations, such as inefficiency of the search process by the OTAs and other market imperfections. These will be explored in the discussion section.

4. Data and Methodology

4.1. Data Sources and Collection Methods

Our analysis begins with a data set of actual tickets written by a major corporate travel agent in the U.S. for five corporate clients in the month of April 1997. We replicated these tickets as closely as possible on major OTA systems using intelligent software agents. By using a base of actual tickets for the analysis we are able to make comparisons among OTAs under a scenario that closely matches how they would be actually used by a traveler. To ensure that all systems could process these tickets and that tickets we were replicating represented the actual ticketing and flight patterns, we applied a number of screening rules to eliminate problematic tickets.[6]

The five online travel agents to be studied were selected based on recommendation of market participants due to a lack of hard market data to make such an evaluation. As mentioned before, three of the five OTAs provided users with the option to indicate a preference (for either price or time). We treated each OTA that offered the ability to choose between time preference and price preference as two separate OTAs, which resulted in our studying eight OTAs.

Using this screened sample of tickets, the reservations were then replicated on each of the OTAs using intelligent agents.[7] The use of automated data collection was critical, given the dynamic pricing behavior of airlines. As Hopper (1990) reports, prices are updated dynamically based on yield management algorithms and can change at any time - in a normal day, 200,000 fare changes per day are loaded into the CRS. To prevent inconsistencies due to spanning price changes during the inquiries, our system of intelligent agents made a specific reservation request on all online travel agents at the same time. This approach makes the probability that we find spurious price variation due to fare changes is vanishingly small.[8]

---

[6] Specifically, we required that all tickets have either 2 or 4 segments, had the same departure and final destination, and represented travel entirely within the U.S.

[7] Besides the guaranteed price stability at the moment of reservation request for all online travel agents, the use of intelligent agents had another major advantages: they prevented any manual mistakes that humans tend to make and therefore ensured high data quality.

[8] Assuming 200,000 fare changes per day over 24 hours represents 139 fare changes per minute. On average, our agents complete a request every 30 seconds (note that due to failed requests, this is consistent with the 300-500 reservations/day reported elsewhere in the paper). Even if our dataset represented 10% of all tickets that were available in the market (an assumption conservative by at least an order of magnitude), the chance of a mid-query

The set of flight alternatives offered by OTAs was stored in a database. Based on the input of market participants, we applied two different decision rules to select a flight from the set of alternative flights offered. In our first decision rule, we emulated the consumer preferences of business traveler by making timeliness our top priority. Of all flights offered, we selected those flights that departed in the time window of one hour before or after the specified departure time for the departure and return flight.[9] A ticket that met the time window constraint in one leg was preferred to a ticket that did not meet the time window constraint at all. If a ticket met the time window in only one leg, the tickets where the time window was met in the departure leg was given preference. In case of a tie, the cheaper ticket was given preference. The second decision rule reflects the consumer preferences of price sensitive leisure travelers. Hence, price was the top priority, with timeliness (meeting the time window) as the tiebreaker.[10] For discussion purposes, we label these the *time-priority* and *price-priority* data sets. We also collected other data, such as the difference of the desired and actual departure time as well as the length of the connections.

We ran our intelligent agents for four consecutive days for 24 hours. Each day, our set of intelligent agents made reservations for flights that were originally made on that weekday with the same number of days to departure, seeking to replicate travel attributes such as trip duration and departure times. On average, each of our agents was able to make between 300 and 500 reservations per day. The total number of reservations requests that the intelligent software agents made depended highly on the availability of the OTAs[11] and the response time of the slowest OTA. In order to obtain a consistent data set, we discarded all reservations where one or

---

fare change is <0.7% (~70 fare changes/ticket cycle/(939*10 tickets) . Borenstein (1994) found that the expected difference in prices paid by two passengers selected at random on a route is about 36 percent of the airline's mean ticket price on the route. Using this number, fare changes cannot account for more than 0.7% x .36 = 0.27% of price variation.

[9] The length of the time window was set after consultation with travel agents. The one-hour time period before and after the specified departure time represents common practice.

[10] In some respects, these decision rules appear arbitrary. However, they are necessary to enable comparisons of a single preferred ticket from each system and to prevent the results being skewed by tickets that are offered that would never be selected; moreover, they accord with actual traveler behavior (only one ticket is chosen). The results are consistent across the different preference choices, suggesting that this procedure is not biasing the results.

[11] It was not uncommon for one or several of the OTAs to be unavailable. In that case, the intelligent software agents stopped the reservation request after a specified waiting time and continued with the next reservation.

more OTA had made no recommendation. This reduced our data set to 939 unique tickets for which we had reservation recommendations for all eight OTAs.

4.2. Data Characteristics

Table 1 shows some basic statistics of our sample. Overall, we have a total of 7512 tickets with an average price of $557 in the time priority data set and $515 in the price priority data set. The tickets otherwise show similar characteristics between the two data sets except for the percentage of flights meeting the time window; the price priority data set contains more than twice as many tickets that violate the time window. There is large variation across the sample in prices, primarily due to route differences. The standard deviation of price is nearly $400 and the tickets range in price from $81 to $2118.

5. Empirical Analysis and Hypothesis Testing

In this section, we test the hypotheses outlined in Section 3.2 using regression analysis and non-parametric statistics. First, we describe the construction of the econometric model.   We then investigate a baseline case where the prices of the various OTAs' quotes are compared. This enables us to test the first hypothesis, that there should be no price differences across OTAs (Hypothesis 1a). We then use non-parametric tests to examine whether or not OTAs select tickets with different characteristics (Hypothesis 1b). Next, we construct a hedonic model of ticket prices as a function of convenience characteristics, and then use this model to repeat the comparison of different OTAs controlling for differences in ticket quality (testing Hypothesis 2).

5.1. Model Formulation

We propose a model in which prices are largely determined by the route (embodying characteristics such as distance, competition, and demand), and secondarily determined by ticket characteristics such as the number of connections, whether there is a Saturday night stay, and whether or not we impose a constraint that the flight leaves or returns at a specified time.   We also include a variable for which CRS is used, in order to pick up any residual variation due to fare differences in the various CRS. These characteristics were determined by interviews with market participants and span most of the relevant characteristics.[12] In general, connections are less convenient for the traveler but more efficient for the airline, suggesting that connections should lower prices. Imposing the constraint to select a ticket within a specified time window limits available options and should generally increase fares. Saturday night stay is a standard business-leisure segmentation device used by the airlines and should result in lower prices.

There is substantial debate in the literature on hedonic price models as to the functional form of the price-characteristic relationship. In our context, the most plausible model is the log-linear form. In this formulation, characteristics lead to a percentage increase in the base price rather than an absolute increment. This appears more consistent with actual pricing behavior in the market (e.g. permitting a connection to a $1000 flight could reduce the price by $200, while permitting a connection to a $150 flight may only save $30). In addition, the combination of city-pair fixed effects and the log-linear model virtually eliminates the heteroskedasticity that is present due to the fact that higher price tickets have more price variation (the remaining heteroskedasticity is not economically meaningful and is addressed by the use of White Robust standard errors).[13] Thus, we model pricing of airline tickets as:

$$\begin{aligned}\log\left(p_{OTA}\right) = {} & d_1 * TimeWindow_{departure} + d_2 * TimeWindow_{return} + d_3 * Connection_{departure} + \\ & d_4 * Connection_{return} + d_5 * Saturday + d_6 * CRS + \sum_i g_i * Citypair_i + e \end{aligned} \qquad (1)$$

For our main formulation, all variables are binary. $TimeWindow_{departure}$ and $TimeWindow_{return}$ are 1 if the time window is met on departure and return respectively, and zero otherwise. $Connction_{departure}$ and $Connection_{return}$ are 1 if there is a connection on departure and return respectively, and zero otherwise. *Saturday* is 1 if there is a Saturday night stay (zero otherwise), and the *CRS* is 1 if the ticket was priced from Apollo (zero otherwise). The variables $Citypair_i$ represent a dummy variable for each of the 436 combinations of departure and return cities in our

[12] The only omitted characteristic was airlines, since the choice of a specific flight from an airline is the essential functionality of the OTA, in which the OTA acts as an information filter and essentially allows the OTA to engage in product differentiation.

[13] The log-linear formulation reduced heteroskedasticity substantially compared to the linear formulation, resulting in a reduction of the R^2 of the White Test regression from 12% to ~0.9%. Due to our large sample size, this is still statistically significant but not economically significant. We therefore report White heteroskedasticity-consistent standard errors (White, 1980) in all the tables (which are essentially identical to the OLS errors).

dataset. We also explored models in which we examine deviation from desired time window and length of connection (both measured in minutes) with virtually identical results (available from authors). We thus conduct all the analysis using the simple, more easily interpreted model.

In interpreting the coefficients of Equation 1, it is important to consider the interaction between the quality characteristics and the city-pair characteristics. In this model, the quality characteristics do not represent the pure difference due to connections or time, but the differential when the number of connections is above the norm for that route. For example, if a route always has connections, then there will be no premium attributed to a connection on that route. Thus, while this accurately removes variation in ticket prices due to these characteristics (our primary objective) and is directionally correct, these coefficients should not be interpreted as the simple price premium or discount due to these quality characteristics in isolation.

To examine the pricing behavior of OTAs, we add them as additional binary independent variables to the estimating equation (OTA1 - OTA8 with the variable 1 if the ticket came from that OTA and zero otherwise) . The coefficients can then be interpreted as the average percentage difference in ticket price, after controlling for route and other characteristics we include.

5.2. Baseline Comparison

Our analysis begins with a relatively simple log-linear model that relates the log of the ticket price to the identity of the OTA and the city pair. The city pair variable is retained to remove most of the (uniniteresting) variance in price due to route-specific characteristics. However, for this analysis we drop the variables for ticket quality to test the initial hypotheses that all tickets are the same (unconditionally). Formally, we analyze the following model:

$$\begin{aligned}\log\left(p_{OTA}\right) = {} & b_1 * OTA2_{time} + b_2 * OTA2_{price} + b_3 * OTA3_{time} + b_4 * OTA3_{price} + \\ & b_5 * OTA4_{time} + b_6 * OTA4_{price} + b_7 * OTA5 + \sum_i g_i * Citypair_i + e \end{aligned} \quad (2)$$

Note that in the fixed effect model the intercept is subsumed by the citypair variables and the b s then represent the percentage change in ticket price over and above the ticket price reported by *OTA1*, which simple averages suggest is the lowest price OTA. In this model, Hypothesis 1a can be restated as $b_1 = b_2 = b_3 = b_4 = b_5 = b_6 = b_7 = 0$. In other words, on average the OTAs should yield the same prices. If the coefficients are not significantly different from zero, we cannot reject the hypothesis that all OTAs return the same prices as would be expected if there were perfect competition with complete price transparency.

The first column of Table 2 depicts the coefficients for the time-priority data set, while the second column depicts the coefficients for the price-priority data set. In both models we can clearly reject hypothesis H1a, that the prices are equal across OTAs (for time-priority data set $F_{443, 7069} = 49.34$, $p<.0001$; for price-priority data set $F_{443, 7069} = 56.25$, $p<.0001$).

OTA4 shows the greatest price variation from OTA1, with b_4= 27.8% for the time-priority data set and b_4= 25.9% for the price-priority data set. All coefficients are significant at a p<.001 level. Interestingly, OTA2 is the only OTA whose deviation from the average price changes noticeably when applying a different decision criterion. For the time-priority data set, OTA2 is about 17.2% more expensive than OTA1, while it is only 5.2% more expensive for the price-priority data set. This suggests that OTA is the only agent that returns substantially different offerings to appeal separately to time vs. price sensitive travelers.

To explore this result further, we also investigate the difference between time preference and price preference on systems that allow these preferences to be expressed (namely OTA2, OTA3, and OTA4). The point estimates suggest that these options have little effect on price and this is confirmed by several hypothesis tests. For each of the three systems, we cannot reject the null hypothesis that expressed time/price preference has no effect on quoted price (p-values are all above .85). This suggests that although different systems request information on time versus price preference from consumers, this has little effect on the tickets that these systems actually offered. We verify this by analyzing the rank of the chosen recommendation among all alternatives offered by the OTA and find that the ranks offered by the system with price preference do not differ significantly from the one with time preference.[14]

When coding the intelligent agents, we discovered that OTA2 and OTA3 appear to share a common search engine. This suggests that they could offer similar tickets. To verify this in our

---

[14] We perform a t-test to explore whether or not the mean ranks of the systems with the option of indicating different preferences (e.g., $OTA2_{time}$ and $OTA2_{price}$) are the same. For the time-priority (price-priority) data set we obtain the following results: OTA2 t=0.10 (t=0.15), OTA3 t=0.31 (t=0.15) and OTA4 t=0 (t=0).

data, we tested whether OTA2 and OTA3 quoted identical prices (the actual test is $OTA2_{time} = OTA2_{price} = OTA3_{time} = OTA3_{price}$). The results suggest that these two systems and their time/price options are virtually identical ($F_{3,7069}$=.051, p>.98 for the time priority data set; $F_{3,7504}$=.002, p>.99 for the price priority data set).

Because the time/price options and OTA2/OTA3 data appear to be redundant and including them in the analysis would only increase the power of our statistical results (perhaps artificially), we take a conservative approach and delete the redundant systems and from the analysis. Specifically, we remove OTA3 and the price options for OTA2 and OTA4 leaving four unique OTAs for analysis. We then repeat the baseline analysis (new results appear in Table 2, columns 3 and 4) and find similar results –a price dispersion of approximately 28% between the highest and lowest price OTA for both decision criteria.

Altogether, these findings are inconsistent with a unique price equilibrium. One plausible explanation for this result is that different OTAs systematically select tickets with different characteristics and, therefore, have systematically different prices. This is explored in the next section.

5.3. Variations in Characteristics of Selected Tickets

Tickets in our sample have four possible characteristics indicative of "quality": meeting time window requirements on departure, meeting time window requirement on return, having no connection on departure, and having no connections on return. This yields 16 possible types of tickets. In Table 3a and 3b we count the number of ticket recommendations from each OTA of each quality. For example, the first row of Table 3a indicates that OTA2 quoted 5 ticket recommendations that had no connections but failed to meet the time window requirements on both departure and return. OTA1 quoted 54 ticket recommendations, OTA4 quoted 14 ticket recommendations and OTA5 quoted only 1 ticket recommendation with these characteristics.

To examine whether there are systematic variations across OTAs in the type of tickets they issue, we employ the simple sign tests and the Wilcoxon signed rank tests to evaluate whether the distributions across ticket types are the same for all OTAs. This is a test of hypothesis H1b. Moreover, given our earlier results, we can check to see if the systems with higher prices also tend to yield tickets with higher quality, which would be indicative of differentiation.

Specifically, if price behavior and ticket quality choice were related, we would expect OTA4 and OTA5 to have similar quality tickets since their average price is similar. In addition, we may also expect OTA2 and OTA1 to be similar in the data set where tickets were selected on price.

The results are shown in Table 4a and 4b. Our results suggest that in both data sets OTA5 and OTA4 are quite similar ($p > .10$ for sign and Wilcoxon test for time and price priority data set) and that they are different from OTA1 (at least $p < .05$ for the sign test for both data sets and at least $p < .10$ for the Wilcoxon test for both data sets). In some cases we can also distinguish OTA5 and OTA4 from OTA2. We also find that OTA1 and OTA2 are similar in the price priority data set but not in the time priority data set ($p > .10$ for the price priority data set and $p < .10$ for the time priority data set). Overall this is consistent with the idea that variation in prices is matched to variation in ticket quality. Moreover, the results appear to be fairly robust to alternative priority criteria (price/time) and statistical tests (Wilcoxon/simple sign test).

5.4. Testing for Price Variation Across OTAs Accounting for Ticket Quality

The previous results suggest that different OTAs appear to be targeting different price-quality segments among consumers. In this section we estimate the full hedonic price model including the terms for time window, connections, CRS and Saturday night stay. The specific model is shown in Equation 1.

To examine the behavior of the hedonic model, we estimate it first without the OTA effects (Table 5, columns 1 and 2). Overall, we find strong support for the idea that quality characteristics affect ticket price. OLS estimates on both the time-priority and price-priority data sets result in coefficients with the predicted sign. Furthermore, all ticket quality coefficients are significant at the (p<0.01) level except for the coefficient of the variable $Connection_{departure}$. The various characteristics have a large effect on price: the OLS estimates indicate that based on an average ticket price of $549.98 ($521.39) for the time-priority (price-priority) data set, requiring flights that meet the time window makes travel between 11.3% and 18.8% more expensive than accepting those that don't; refusing to accept a connection can increases prices up to 20.7% on a route that normally has connections. Saturday night stays decrease the ticket price by up to 40%.

We now combine the hedonic price model and the dummy variable regression with OTAs to separate the price effects of quality variation from other OTA specific characteristics:

$$\begin{aligned}\log\left(p_{OTA}\right) = & d_1 * TimeWindow_{departure} + d_2 * TimeWindow_{return} + \\ & d_3 * Connection_{departure} + d_4 * Connection_{return} + d_5 * Saturday + \quad (3) \\ & b_1 * OTA2_{time} + b_5 * OTA4_{time} + b_7 * OTA5 + \sum_i g_i * Citypair_i + e\end{aligned}$$

Similar to the earlier models, the coefficient estimates on price are relative to OTA1 and estimates are made using the log-linear formulation with fixed city-pair effects. The hypothesis that quality differences account for all the variation in ticket prices (H2) can be expressed as: $b_6 = b_7 = b_8 = 0$. In other words, after controlling for ticket characteristics, we should observe no price differences in the price of the underlying good across OTAs.

In Table 5, columns 3 and 4, we show the results for both data sets (time priority and price priority). Overall, we are able to reject equality of the OTA effects in all models at (p<.0001) level,[15] suggesting that quality variation is not the only source of price variation in the sample. However, the estimates show that price variation across OTAs is reduced; for the time and price regression the price variation is 17.9% and 18.9% respectively as opposed to 27.8% for the regression of the base model. This suggests that product differentiation accounts for some of the previously observed price variation. Other coefficients generally have the predicted sign except for *Connection Departure* variable.

To further investigate the extent to which the price dispersion is driven by OTAs as opposed to the specific characteristics of the route, we specified the following model:

[15] $F_{3,3312}$=35.36 for the time-priority data set and $F_{3,3312}$=49.34 for the price-priority data.

$$
\begin{aligned}
\log\left(p_{OTA}\right) = {} & b_1 * OTA2_{time} + b_5 * OTA4_{time} + b_7 * OTA5 + \\
& d_1 * TimeWindow_{departure} + d_2 * TimeWindow_{return} + \\
& d_3 * Connection_{departure} + d_4 * Connection_{return} + d_5 * Saturday + d_6 * CRS \\
& \sum_i w_i * TimeWindow_{departure} * Citypair_i + \qquad (4) \\
& \sum_j x_j * TimeWindow_{return} * Citypair_j + \\
& \sum_k y_k * Connection_{departure} * Citypair_k + \\
& \sum_l z_l * Connection_{return} * Citypair_l + e
\end{aligned}
$$

(Note: 1 citypair variable is dropped from each sum to avoid perfect multicollinearity)

While this formulation consumes an enormous number of degrees of freedom, it provides a reasonable lower bound for the amount of residual price variation. Interestingly, the price dispersion across OTAs is only slightly lower compared to the previous model formulation. For the time-priority data set, $OTA2_{time}$ was 10.3% more expensive than OTA1, $OTA4_{time}$ and OTA5 were 16.4% and 13.4% more expensive. All estimates for the OTAs were significant at the $p<.001$ level. For the price-priority data set, $OTA2_{time}$ was 0.2% cheaper than OTA1 (not significant), while $OTA4_{time}$ and OTA5 were 18% and 12% more expensive. Both estimators were significant at the $p<.001$ level.

Overall, this suggests that product differentiation indeed accounts for at least 10% of the variation in actual ticket price between OTAs. This represents a variation of about $50 in the price of an average ticket, or 35% of the overall price dispersion across OTAs.

6. Discussion

Overall, we find that different OTAs offer different types of tickets at substantially different prices. In addition, the variation in prices across OTAs is reduced substantially, but not entirely, when we control for variations in ticket characteristics. These results collectively suggest that product differentiation is indeed occurring. However, there still remains a substantial portion of unexplained price variation across OTAs. In this section, we reconcile these results first with the economics of product differentiation, and then explore other possible explanations for the remaining price dispersion (search engine inefficiency, principal-agent problems, and price discrimination).

6.1. Product Differentiation

Our results clearly indicate that OTAs are targeting different customer segments with different tradeoffs between price and ticket quality. OTA5 and OTA4 consistently recommend high price/high convenience tickets, whereas OTA1 offers low price/low convenience tickets. OTA2's offerings are between these extremes. This suggests that OTAs compete on two dimensions: price and quality. A straightforward way to see this is to plot the average position of each of the OTAs in terms of price versus quality to sketch out an efficient frontier. The plot for price satisfaction (ratio of the prices of OTAX to OTA1) vs. connection satisfaction (percentage of flights without connections) is shown in Figure 1a. All systems except OTA2 are on the frontier; that is, they offer a price-quality tradeoff that (on average) is not dominated by any other.[16] OTA1 is clearly the price leader, but it does so by having connections on over 35% of tickets while OTA4, the most expensive, only has connections on about 15% of tickets. A plot of price satisfaction vs. timeliness satisfaction (percentage of flights meeting both time windows) (Figure 1b) shows a similar result. Three of the four systems are on the frontier, only OTA4 is dominated; it is more expensive than OTA5, yet it fails to meet the time requirement 13% more of the time. Overall, while this chart ignores the variance around these averages it clearly suggests that OTA1 is pursuing a low cost/low quality strategy while OTA5 is a balanced high-cost/high-quality, and OTA4 concentrates on high-quality, but emphasizes minimizing connections.

OTAs achieve this differentiation through parameterization of their systems. Since the booking process is completely automated after the customer's preferences, departure and return information have been entered, every OTA must have decision rules in place that select the best flight recommendations based on the customer's information. Interviews with designers of these

[16] The difficulty of this graphical analysis is that it ignores the variance around these mean points - the dominance tabulations (table 8a and 8b) avoid this difficulty. However, this provides a simple graphical representation of the key insights of this analysis.

systems suggested that parameters they use to capture their own search strategies, and to differentiate their products, include:

- permitted deviation from requested departure or return time
- minimum savings required to justify a connection
- maximum duration permitted for a connection
- minimum distance before a connection flight is considered

These dimensions directly affect the set of potential flight routes selected. Generally, and not surprisingly, the more strictly these variables are set, the more expensive the flights that the OTA selects.

In order to test specifically for the differences among the parameter settings, we repeated our earlier analysis of differences among OTAs allowing the relationship between ticket characteristics and price to vary by OTA. Specifically, we tested the following model:

$$
\begin{aligned}
\log\left(p_{OTA}\right) = {} & b_1 * OTA2_{time} + b_4 * OTA4_{time} + b_7 * OTA5 + \\
& t_1 * OTA1 * Connection + t_2 * OTA2_{time} * Connection + \\
& t_3 * OTA4_{time} * Connection + t_4 * OTA5 * Connection + \\
& t_5 * OTA1 * TimeWindow + t_6 * OTA2_{time} * TimeWindow + \\
& t_7 * OTA4_{time} * TimeWindow + t_8 * OTA5 * TimeWindow + \\
& \sum_i g_i * Citypair_i + e
\end{aligned}
\qquad (5)
$$

Intuitively, this is a crude measure of the "shadow cost" of deviations from ideal ticket characteristics imposed by the parameter settings chosen by the OTA.[17] To calculate these effects, we simply interact each OTA variable with ticket characteristic variables. To limit the number of coefficients in the model, we aggregate time window satisfaction and connections into two dummy variables. For time window, the variable *TimeWindow* is 1 if both departure and

[17] This measure is crude because it is only identified if the ticket set returned by an agent has substantial variation on these dimensions. For example, to estimate the shadow cost of missing connections, we need to observe the same agent returning tickets with and without connections for a similar request. For agents which aggressively optimize on a single dimension, there will be little variation in the sample set leading to potentially anomolous results. However, for agents which tend to trade off multiple dimensions, this analysis will more accurately reveal this tendency.

return time windows are met and zero otherwise. For connections, the variable *Connection* is 1 if there are any connections and zero otherwise. The results are shown in Table 6.

Overall, there is a wide variation in the effects of connections and satisfaction of time window requirements across OTAs. For the time-priority data set, $OTA4_{time}$'s ticket prices decrease by only 4.7% when a connection is involved (estimate is not significantly different from zero)[18], while all other OTAs reduce the prices by 9.6% to 22% (all estimates are significant at the $p<.001$ level). OTA5's ticket prices are actually 12.8% cheaper when they leave on time than when they leave outside of the specified time window. $OTA4_{time}$'s price increase by 7.4% (significant at $p<.01$ level) for flights that leave on time, while OTA2 time and OTA1 have an increase of 11.5% and 24.2% respectively (all significant at $p<0.0001$ level). Similar results are found for the price-priority data set. The high variation in the coefficients for the interaction terms of OTA and ticket characteristics clearly reflects the differences in the set of parameters that an OTA has chosen. For example, OTA5 apparently requires the highest saving to justify a connection and OTA1 allows for the highest deviation from the specified time of departure.

6.2. Price Dispersion

We observe significant price dispersion in our data, above that created by differentiation; even after accounting for ticket characteristics, average price still varies by as much as 18.9% across OTAs. On average, OTA4 and OTA5 recommend more expensive flights than OTA2 and OTA1. This result runs counter to the widespread belief that reduced search costs on the WWW has led to better informed consumers, has increased price transparency, and has eliminated the ability of vendors to retain customers when they offer inferior prices.

One explanation is that even though consumer search costs have come down significantly, they may have been replaced by other costs. These costs include: signing up for another OTA, learning its user interface, entering flight preferences into multiple OTAs, and evaluating flight recommendations from multiple OTAs. Of these costs, the first two are one-time costs, while the last two are incurred with each additional search. Hence the consumer faces a trade-off between

[18] This result is probably due to the fact that $OTA4_{time}$ offered the fewest connections of all OTAs. After accounting for all 'necessary' connections through the city-pair variables, there is not enough data variation left to achieve statistical significance.

a potential savings and a known and certain evaluation cost. This tradeoff has been discussed in the literature (see e.g., Salop and Stiglitz 1977) and has been shown to enable producers to charge consumers different prices for the same good.

In addition to search, the inefficiencies in the search process of the OTAs may also explain some of the variation. One possible explanation is that the currently available search parameterization is too inflexible. The specific parameter setting might work well for some city pairs but not for others. For instance, the parameter that defines the minimum distance before a connection is considered is set as an absolute number, independent of the actual distance between departure and destination airport. Further, as an industry participant pointed out to us, the number of carriers competing in one city-pair market or the number of possible connections affects the flight recommendation significantly, yet these factors are not considered in the parameterization of these systems or, more importantly, in our analyses.

If this lack of flexibility does indeed hold, we would expect that differently parameterized systems would yield some flights that were optimal for some requests, while others with a different parameterization could find better flights for other requests. To investigate this possibility, we examined the relationship between each pair of OTAs for each customer request in our data set. We categorized the relationships as either: 1) identical – the OTAs gave the same ticket, 2) one OTA strictly dominates another – an OTA provided a ticket which was at least as good on all characteristics (time window, connection, price) and strictly better on at least one, or 3) non-comparable – one OTA gave a ticket that was superior in one dimension, while the other returned a ticket superior on another dimension. We focus particularly on dominated tickets since this is a clear example of inefficiency and is independent of the decision rule (if a ticket is dominated, it is necessarily dominated for both decision rules).

An analysis of dominated tickets for pair of OTAs is presented in Table 7a and 7b. To interpret the results, consider the relationship between OTA2 and OTA1. $OTA2_{time}$ strictly dominates OTA1 82 times, but OTA1 strictly dominates $OTA2_{time}$ 122 times. When comparing the strict dominance of all OTA pairs, we observe that OTA1 is least often strictly dominated; this may be attributable to their strategy of finding low ticket prices making it unlikely that they are dominated on price. OTA5 is most often strictly dominated, perhaps because they have the most balanced strategy of optimizing across price as well as quality and thus have more opportunity to offer a suboptimal ticket. However, we cannot find any OTA pair for which we can clearly state that one OTA clearly dominates the other.[19]

This suggests that even though we have evidence for inefficiencies, we cannot point out to one single OTA that is clearly inferior. The large number of dominated tickets for all systems suggests that parameterizations are indeed too inflexible; moreover, systems that aggressively pursue high quality tickets may have the highest ticket cost penalty associated with these "mistakes."

To summarize, we can clearly observe price dispersion. Our analysis suggests that the inflexibility of the search criteria may be largely responsible for this effect. In addition, this inefficiency can persist in competitive markets if customers' search cost of examining another OTA is non-trivial or if there are other barriers to search such as switching costs. This might also suggest that OTAs are very well aware that they fall short on being able to provide a system that strictly dominates all other OTAs and that they rely on consumers' search cost to sell the strictly dominated tickets.

6.3. Agency Problems

Given that agents are representatives of the airlines and not the consumer, it is possible that agents offer higher price tickets in an effort to increase their own revenue (or the revenue of their owners for those OTAs affliated with airlines). In a fully efficient market any agent that deliberately offered unnecessarily expensive tickets would receive no business, making agency problems unlikely. However, we have already shown that there are substantial inefficiencies in pricing that suggest a variety of factors limiting perfect competition. When OTAs enjoy some, perhaps small, amount of local monopoly power, it is possible that this can be exploited. Of course, this factor itself will only affect observed price dispersion if incentives indeed differ among OTAs.

To investigate this possibility we examined the various commission structures in place for OTAs at the time of our study. While there is a substantial variation across different airlines in

[19] We extended this analysis for all 16 possible subgroups of ticket categories to see whether or not an OTA is strictly dominant for specific ticket characteristics. Our results did not indicate any strict dominance of an OTA for any subgroup.

commission structure, making such a comparison difficult, one OTA (OTA5) performs their services entirely on a flat fee basis for their customers and rebates all commissions. Thus, of all the systems, they should have no incentive to offer tickets with inflated prices. If agency problems are a concern, then we would expect OTA5 to have lower ticket prices on average. However, our data suggest that OTA5 offers tickets with the second highest average price, even after controlling for quality. While this doesn't completely rule out possible agency effects, this provides evidence against the hypothesis that agency problems lead to the price dispersion we observe.

6.4. Price Discrimination

The strategy of the two OTAs that are operated by the same company, namely OTA4 and OTA1, is also interesting. In this case, one company offers two different on-line services with different interface characteristics and very different prices. OTA1 offers the cheapest tickets with lowest convenience and OTA4 offers the most expensive tickets with highest convenience. Simultaneously, the front-end for OTA1 can be best described as archaic, while the front-end for OTA4 is state-of-the-art.

At the first glance, it may seem foolish to burden a consumer with using an unforgiving user interface that requires cryptic inputs, when an ergonomic user interface is readily available. Presumably, a good user interface should only increases the consumers' value for the service. However, this argument is only true if travelers have homogeneous preferences. For heterogeneous tastes, a bad or good interface can be used as a screen for the traveler's willingness to pay if the traveler with the highest willingness to pay also has the highest distaste for a bad user interface. For time-sensitive travelers it can certainly be argued that they do not have the patience to work through the unintuitive mainframe menu of OTA1, while price sensitive travelers may endure the procedure. In other words, the difficulty in using OTA1's user interface serves as a screen to prevent the time sensitive-travelers from exercising personal arbitrage.[20] Further, a traveler who is willing to learn OTA1's user interface can be assumed to

[20] See Tirole (1988) for a general argument.

be computer-savvy and therefore more likely to be knowledgeable about other OTAs and their interfaces and hence, also well informed about other alternatives.

Such a behavior is fairly widespread. The most famous example of such a segmentation technique is discussed by Dupuit's (1849), who reported the following about the railroad tariffs for passenger traffic:

It is not because of the few thousand francs which would have to be spent to put a roof over the third-class seats that some company or other has open carriages with wooden benches… What the company is trying to do is prevent the passengers who can pay the second-class fare from traveling third-class; it hits the poor, not because it wants to hurt them, but to frighten the rich… And it is again for the same reason that the companies, having proved almost cruel to third-class passengers and mean to second-class ones, become lavish in dealing with first-class passengers. Having refused the poor what is necessary, they give the rich what is superfluous (p. 23).

Given that OTAs are differentiated, it should not surprise that the owner of OTA1 and OTA4 tries to price-discriminate users on its residual demand curve. This strategy might have been helped by the historic development of OTAs. OTA1 was originally developed as proprietary software and was designed to connect to a mainframe. Hence, its user interface was cumbersome and error prone. OTA4 was built with the interactivity in mind that the World Wide Web provides. With this accidental sagacity, the owner of the two OTAs might have coupled the ticket selection behavior that corresponds with the customer segment and interface design.

6.5. Conclusion

Competition in electronic markets appears to be more complex than can be explained with the simple models of pure price competition, zero search costs, or complete price transparency. Even in the face of declining search costs, market imperfections persist and sellers do not immediately jump to perfect price competition. It is arguable that search costs prevalent in traditional markets are replaced by new types of costs. For example, OTAs may create switching costs by requiring the customer to "sign-up" by entering personal information that later reduces the time to find and book flights, or frequent user programs that reward repeat customers. They may also require customer log-ins or other time consuming steps that increase the cost of searching multiple agents. When non-zero switching and search costs are combined with uncertainty about which OTAs will truly provide the best flight for the consumers preference, it may be optimal for consumers to limit their search to a single or to a few OTAs. If this is the case, then consumers would be unambiguously better off using a "Super OTA" that searches all available OTAs and returns their flight recommendations. One system indeed saw this opportunity and offered to search 4 different OTAs for the best fare. Of those 4 OTAs, 2 OTAs – OTA2 and OTA3 – were in our sample. The "Super OTA" itself was operated by the same company that operated OTA2 and OTA3. But even before this system was upgraded from a limited beta version to a full version, the two other systems were withdrawn, claiming they already offered the best prices. The two remaining systems offer, according to our analysis, the same flight recommendations.

Our results also show that service differentiation is a key strategic component of online sellers that offer access to heterogeneous goods. This result should not be surprising, since it mirrors behavior in non-electronic markets: by exploiting consumers' heterogeneity in tastes and uncertainty of vendor quality, vendors can ease price competition by segmenting the market. Moreover, with a product with little cost of differentiation, new strategies may emerge; for example, creating "Sister OTAs" strategically to capture different customer segments and utilizing user interface quality to segment the market.

These results may to apply to other settings, which involve the distribution of differentiated goods in electronic markets or information products, which are easily differentiated at low cost. For these markets, the range of possible producer behaviors is much richer than would be suggested by the tendency of the business press to equate electronic markets with low cost strategies and pure price competition.

Tables

Table 1: Ticket Characteristics

Characteristic	Time Priority	Price Priority
Total Tickets	7,512	7,512
Average Price	$557	$515
Standard Deviation	$395	$376
Tickets with connections	38.0%	40.8%
Tickets not w/in time window	17.6%	38.8%

Table 2: Price Differences Among OTAs (Baseline Regression)

VARIABLE	Log-linear TIME Full Data Set	Log-linear PRICE Full Data Set	Log-linear TIME Reduced Data Set	Log-linear PRICE Reduced Data Set
OTA1				
$OTA2_{time}$	0.172^{***} (0.016)	0.052^{***} (0.014)	0.172^{***} (0.016)	0.052^{***} (0.015)
$OTA2_{price}$	0.172^{***} (0.016)	0.052^{***} (0.014)		
$OTA3_{time}$	0.175^{***} (0.016)	0.052^{***} (0.014)		
$OTA3_{price}$	0.178^{***} (0.016)	0.053^{***} (0.014)		
$OTA4_{time}$	0.278^{***} (0.016)	0.259^{***} (0.015)	0.278^{***} (0.016)	0.259^{***} (0.015)
$OTA4_{price}$	0.278^{***} (0.016)	0.259^{***} (0.010)		
OTA5	0.252^{***} (0.018)	0.208^{***} (0.010)	0.252^{***} (0.017)	0.208^{***} (0.016)
N	7512	7512	3756	3756
F (Prob>F)	49.34 p<0.0001	56.25 p<0.0001	22.88 p<0.0001	25.43 p<0.0001
R^2	75.52%	77.86%	75.13%	77.05%

White standard errors in parenthesis; ***-p<.001, ** - p<.01, * - p<.05

Table 3a: Ticket Characteristic Combinations by OTAs ($_{\text{Time}}$ Preference)

TWD	TWR	CND	CNR	OTA1	OTA2 time	OTA4 time	OTA5	Total
				54	5	14	1	74
				13	1	2	0	16
				6	0	0	0	6
				52	6	1	1	60
				56	28	49	13	146
				20	3	27	0	50
				8	1	0	0	9
				60	25	9	7	101
				87	14	32	25	158
				13	3	0	0	16
				19	2	25	0	46
				69	28	4	11	112
				279	467	634	638	2018
				15	24	6	0	45
				19	21	9	0	49
				169	311	127	243	850
				939	939	939	939	3756

✓ - Means that this criteria was met

Table 3b: Ticket Characteristic Combinations by OTAs (Price Preferences)

TWD	TWR	CND	CNR	OTA1	OTA2 time	OTA4 time	OTA5	Total
				79	44	32	18	173
				13	3	2	0	18
				6	3	1	0	10
				68	69	1	15	153
				54	40	62	67	223
				22	3	22	0	47
				6	4	0	0	10
				72	66	12	36	186
				84	76	61	68	289
				13	5	1	0	19
				9	10	15	0	34
				74	116	5	27	222
				267	287	593	520	1667
				15	16	3	0	34
				16	11	6	0	33
				141	186	123	188	638
			Total	939	939	939	939	3756

✓ - Means that this criteria was met

TWD (Time Window Departure) TWD is checked if the departure flight left within the time window (one hour before and after the desiredtime).

TWR (Time Window Return) TWR is checked if the return flight left within the time window (one hour before and after the desired$_{\text{time}}$).

CND (Connection Departure) CND is checked if the departure flight involved at least one connection.

CNR (Connection Return CNR is checked if the return flight involved at least one connection.

Table 4a: Distribution Tests for Similarity of Characteristics Across Systems (Time Preference)

Sign Test TIME	OTA1	OTA2	$OTA4_{time}$	OTA5
OTA1	-	n = 16 M=4*	n = 16 M = 13**	n = 16 M=14***
OTA2	-	-	n = 15 M = 8	n = 15 M = 13***
$OTA4_{time}$	-	-	-	n = 12 M = 9
OTA5	-	-	-	-

Wilcoxon TIME	OTA1	OTA2	$OTA4_{time}$	OTA5
OTA1	-	n = 16 T = 35*	n = 16 T = 21**	n = 16 T = 31*
OTA2	-	-	n = 15 T = 6.5***	n = 15 T = 23**
$OTA4_{time}$	-	-	-	n = 12 T = 20.5
OTA5	-	-	-	-

***-p<.01, ** - p<.05, * - p<.1

Table 4b: Distribution Tests for Similarity of Characteristics Across Systems (Price Preference)

Sign Test PRICE	OTA1	OTA2	OTA4time	OTA5
OTA1	-	n = 16 M=5	n = 15 M = 12**	n = 16 M=13**
OTA2	-	-	n = 16 M = 12*	n = 16 M = 13**
OTA4time	-	-	-	n = 15 M = 9
OTA5	-	-	-	-

Wilcoxon PRICE	OTA1	OTA2	$OTA4_{time}$	OTA5
OTA1	-	n = 16 T = 50	n = 15 T = 21.5**	n = 16 T = 33.5*
OTA2	-	-	n = 16 T = 42.5	n = 16 T = 29**
$OTA4_{time}$	-	-	-	n = 15 T = 59
OTA5	-	-	-	-

***-p<.01, ** - p<.05, * - p<.1

Table 5: Hedonic price models and comparison across OTA controlling for ticket characteristics

VARIABLE	Log-linear TIME	Log-linear PRICE	Log-linear TIME	Log-linear PRICE
$OTA2_{time}$			0.114*** (0.016)	0.048*** (0.014)
$OTA4_{time}$			0.179*** (0.016)	0.189*** (0.016)
OTA5			0.156*** (0.017)	0.152*** (0.015)
Time Window Departure	0.147*** (0.023)	0.137*** (0.017)	0.107*** (0.023)	0.105*** (0.017)
Time Window Return	0.188*** (0.022)	0.113*** (0.016)	0.130*** (0.022)	0.069*** (0.016)
Connection Departure	-0.055* (0.032)	-0.041 (0.032)	-0.019 (0.031)	-0.005 (0.031)
Connection Return	-0.207*** (0.035)	-0.169*** (0.034)	-0.169*** (0.034)	-0.116*** (0.033)
Saturday Night Stay	-0.378*** (0.025)	-0.407*** (0.025)	-0.380*** (0.025)	-0.407*** (0.025)
CRS	-0.034*** (0.011)	-0.012 (0.011)		
N	3756	3756	3756	3756
F (Prob>F)	26.05 p<0.0001	28.68 p<0.0001	26.91 p<0.0001	30.13 p<0.0001
R^2	77.61%	79.24%	78.26%	80.12%

White standard errors in parenthesis; ***-p<.001, ** - p<.01, * - p<.05

Table 7: The Effect of Interactions between Ticket Characteristics and OTAs on Ticket Price

VARIABLE	Log-linear TIME	Log-linear PRICE
$OTA2_{time}$	0.242^{***} (0.043)[21]	0.129^{***} (0.029)
$OTA4_{time}$	0.303^{***} (0.041)	0.272^{***} (0.037)
OTA5	0.510^{***} (0.061)	0.253^{***} (0.032)
OTA1 * CN	-0.096^{***} (0.027)	-0.053^{*} (0.026)
$OTA2_{time}$ * CN	-0.211^{***} (0.029)	-0.110^{***} (0.027)
$OTA4_{time}$ * CN	-0.047 (0.037)	-0.013 (0.037)
OTA5 * CN	-0.220^{***} (0.034)	-0.159^{***} (0.032)
OTA1 * TW	0.242^{***} (0.025)	0.209^{***} (0.024)
$OTA2_{time}$ * TW	0.115^{***} (0.039)	0.097^{***} (0.026)
$OTA4_{time}$ * TW	0.074^{**} (0.037)	0.080^{**} (0.034)
OTA5 * TW	-0.128^{**} (0.058)	0.096^{***} (0.029)
N	3756	3756
F (Prob>F)	2314.92 p<.0001	2416.00 p<.0001
R^2	99.68%	99.69%

White standard errors in parenthesis; ***-p<.001, ** - p<.01, * - p<.05

Table 8a: Comparison of OTAs for Similar and Dominated Tickets (Time Preference)

		Y			
	X Strictly Dominates Y	OTA1	$OTA2_{time}$	$OTA4_{time}$	OTA5
X	OTA1	0	122	74	207
	$OTA2_{time}$	82	0	112	241
	$OTA4_{time}$	20	110	0	268
	OTA5	103	193	199	0

[21] White Standard Error

	X is equal to Y	Y: OTA1	OTA2$_{time}$	OTA4$_{time}$	OTA5
X	OTA1	939	297	337	161
	OTA2$_{time}$	297	939	430	266
	OTA4$_{time}$	357	430	939	316
	OTA5	161	266	316	939

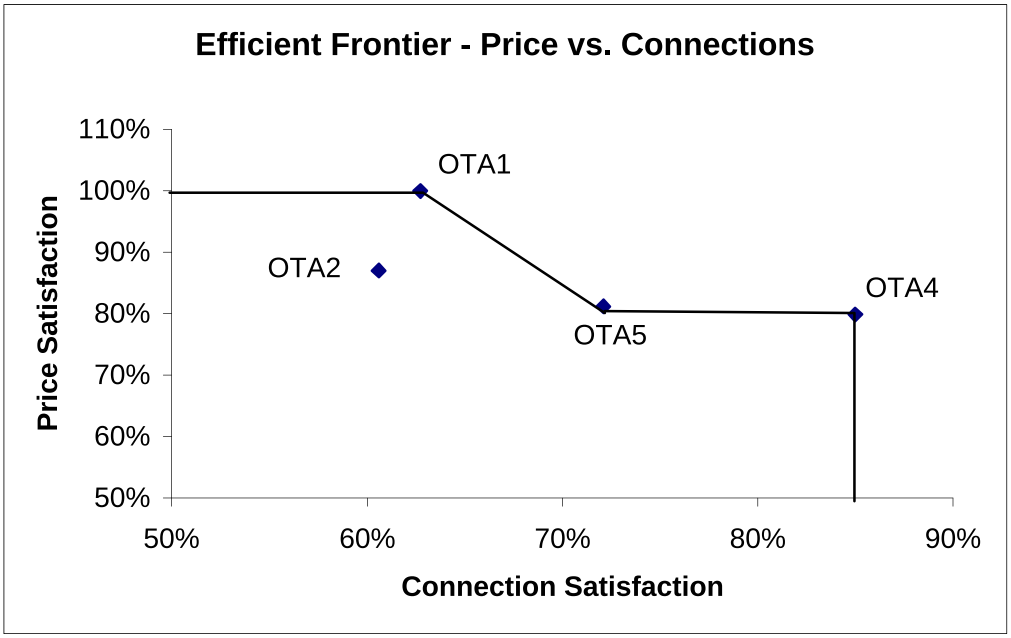
Table 8b: Comparison of OTAs for Similar and Dominated Tickets (Price Preference)

	X Strictly Dominates Y	Y: OTA1	OTA2$_{time}$	OTA4$_{time}$	OTA5
X	OTA1	0	163	82	216
	OTA2$_{time}$	88	0	80	185
	OTA4$_{time}$	21	67	0	258
	OTA5	102	122	177	0

	X is equal to Y	Y: OTA1	OTA2$_{time}$	OTA4$_{time}$	OTA5
X	OTA1	939	350	337	161
	OTA2$_{time}$	350	939	308	192
	OTA4$_{time}$	337	308	939	293
	OTA5	161	192	293	939

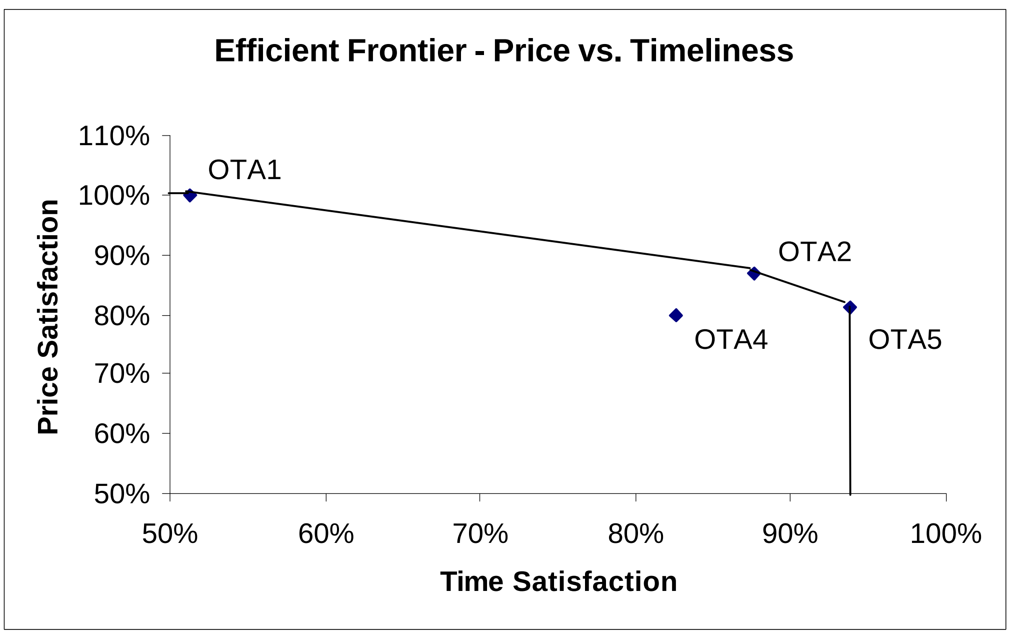
Figures

Figure 1a: Efficient Frontier - Price Satisfaction vs. Connections



Price Satisfaction = Price Price(OTA1)/Price(OTAX); Connection Satisfaction is the percentage of flights offered without connections.

Figure 1b: Efficient Frontier: Price Satisfaction vs. Timeliness



References

Bailey, J., Brynjolfsson, E. and Smith M.D. "In Search of "Friction-Free Markets": An Exploratory Analysis of Prices for Books, CDs and Software Sold on the Internet," working paper, Technology, Management and Policy, Massachusetts Institute of Technology, 1997.

Bakos, J. Y. "A Strategic Analysis of Electronic Marketplaces," *MIS Quarterly*, Vol. 15, No. 3, September 1991, pp. 295-310.

Bakos, J. Y. "Reducing Buyer Search Costs: Implications for Electronic Marketplaces," *Management Science*, Vol. 43, No. 12, 1997, pp. 1676-1692.

Beath, J. and Katsoulacos, Y. *The Economic Theory of Product Differentiation.* Cambridge University Press, Cambridge, 1991.

Benjamin, R. and Wigand, R. "Electronic Markets and Virtual Value Chains on the Information Superhighway," *Sloan Management Review*, Winter 1995, pp. 62-72.

Bertrand, J. "Théorie Mathématique de la Richesse Sociale," *Journal des Savants*, 1883, pp. 499-508.

Borenstein, S. "Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry," *Rand Journal of Economics*, Vol. 20, Autumn 1989, pp. 344-65.

Borenstein, S. "The Dominant-firm Advantage in Multiproduct Industries: Evidence from the U.S. Airlines," *Quarterly Journal of Economics*, Vol. 106, November 1991, pp. 1237-66.

Borenstein, S. and Rose, N. L. "Competition and Price Dispersion in the U.S. Airline Industry," *Journal of Political Economy*, Vol. 102, No. 4, 1994, pp. 653-683.

Burdett, K. and Judd, K. L. "Equilibrium Price Dispersion," Vol. 51, No. 4, July 1983, *Econometrica*, pp. 955-969.

Chamberlin, E. *The Theory of Monopolistic Competition.* Cambridge, Harvard University Press, 1933.

Chow, G. C. "Technological Change and the Demand for Computers," *American Economic Review*, Vol. 57, No. 5, December 1967, pp. 1117-1130.

Clemons, E. K. and Row, M. C. "Information Technology at Rosenbluth Travel: Competitive Advantage in a Rapidly Growing Global Service Company," *Journal of Management Information Systems*, Vol. 8, No. 2, Fall 1991, pp. 53-79.

Combes, G. C. and Patel, J. J. "Creating Lifelong Customer Relationships: Why the Race for Customer Acquisition on the Internet is so Strategically Important," *iword*, Vol. 2, Issue 4, Hambrecht&Quist 1997.

Copeland D. G. and McKenney, J. L. "Airline Reservation Systems: Lessons From History," *MIS Quarterly*, Vol. 12, No. 3, September 1988, pp. 353-370.

Dupuit, J. “On Tolls and Transport Charges,” in *International Economic Papers*, London, Macmillan, 1962, pp. 7-31; original version: “De l’influence des péages sur l’utilité des voies de communication,” in *Annales des Ponts et Chaussées*, 2e série, 1849, pp. 207-248.

Gabszewicz, J. and Thisse J.-F. “Price Competition, Quality and Income Disparities,” *Journal of Economic Theory*, Vol. 20, 1979, pp. 340-359.

Gabszewicz, J. and Thisse J.-F. “Entry (and Exit) in a Differentiated Industry,” *Journal of Economic Theory*, Vol. 22, 1980, pp. 327-338.

Griliches, Z. “Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change,” in *The Price Statistics of the Federal Government*, General Series No. 73, New York, Columbia University Press for the National Bureau of Economic Research, 1961, pp. 137-196.

Gurbaxani, V. and Whang, S. “The Impact of Information Systems on Organizations and Markets,” *Communications of the ACM*, Vol. 34, No. 1, January 1991, pp. 59-73.

Hardie, M. E., Bluestein, W. M., McKnight J., and Davis K. *The Forrester Report: Entertainment & Technology Strategies.* Vol. 1, No. 2, May 1, 1997.

Hopper, M. D. “Rattling SABRE - New Ways to Compete on Information,” *Harvard Business Review*, Vol. 90, No. 3, May-June 1990, pp. 118-125.

Hotelling, H. “Stability in Competition,” *Economic Journal*, Vol. 39, 1929, pp. 41-57.

Lee, H. G. “Do Electronic Marketplaces Lower the Price of Goods?” *Communications of the ACM*, January 1998, Vol. 41, No. 1, pp. 73-80.

Malone, T. W., Yates, J., and Benjamin, R. I. “Electronic Market and Electronic Hierarchies,” *Communications of the ACM*, Vol. 30, No. 6, 1987, pp. 484-497.

Perloff, J. M., Salop, S. C. “Equilibrium with Product Differentiation,” *Review of Economic Studies*, Vol. 52, 1985, pp. 107-120.

Porter, M. E. “*Competitive Advantage: Creating and Sustaining Superior Performance*,” New York, Free Press, 1985.

Pratt, J. W., Wise, D. A., Zeckhauser, R. “Price Differences in Almost Competitive Markets,” *The Quarterly Journal of Economics*, May 1979, pp. 189-211.

Salop, S.C. “Monopolistic Competition with Outside Goods,” *Bell Journal of Economics*, Vol. 10, 1979, pp. 141-156.

Salop, S., Stiglitz, J. “Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion,” *Review of Economic Studies*, Vol. 44, 1976, pp. 493-510.

Shaked, A. and Sutton, J. “Relaxing Price Competition through Product Differentiation,” *Review of Economic Studies*, Vol. 49, 1982, pp. 3-13.

Shaked, A. and Sutton, J. “Natural Oligopolies,” *Econometrica*, Vol. 51, 1983, pp. 1469-1484.

Stigler, G. J. “The Economics of Information,” *The Journal of Political Economy*, Vol. , No. 3, June 1961, pp. 213-225.

The Economist “Who will Deal in Dealerships?” *The Economist*, February 14th 1998, pp. 61-63.

Tirole, J. *The Theory of Industrial Organization.* Cambridge, Massachusetts: The MIT Press, 1988.

Varian, H. R. “A Model of Sales,” *The American Economic Review*, Vol. 70, No. 4, pp.651-659.

White, H. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity,” *Econometrica*, Vol. 48, pp. 817-838, May 198