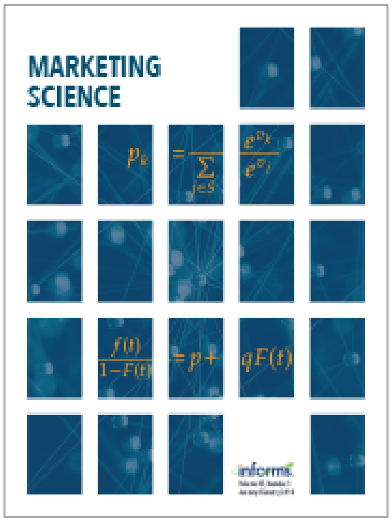


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Fare Prediction Websites and Transaction Prices: Empirical Evidence from the Airline Industry

Benny Mantin, Eran Rubin

MARKETING SCIENCE
Vol. 35, No. 4, July–August 2016, pp. 640–655
ISSN 0732-2399 (print) | ISSN 1526-548X (online)
http://dx.doi.org/10.1287/mksc.2015.0965

Fare Prediction Websites and Transaction Prices: Empirical Evidence from the Airline Industry

Benny Mantin
Department of Management Sciences, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada, bmantin@uwaterloo.ca

Eran Rubin
Faculty of Technology Management, Holon Institute of Technology, Holon 5810201, Israel, eranr@hit.ac.il

The marketing and operations disciplines have increasingly accounted for the presence of strategic consumer behavior. Theory suggests that such behavior exists when consumers are able to consider future distribution of prices, and that this behavior exposes firms to intertemporal competition that results with a downward pressure on prices. However, deriving future distribution of prices is not a trivial task. Online decision support tools that provide consumers with information about future distributions of prices can facilitate strategic consumer behavior. This paper studies whether the availability of such information affects transacted prices by conducting an empirical analysis in the context of the airline industry. Studying the effect at the route level, we find significant price reduction effects as such information becomes available for a route, both in fixed-effects and difference-in-differences estimation models. This effect is consistent across the different fare percentiles and amounts to a reduction of approximately 4%–6% in transactions' prices. Our results lend ample support to the notion that price prediction decision tools make a statistically significant economic impact. Presumably, consumers are able to exploit the information available online and exhibit strategic behavior.

Data, as supplemental material, are available at http://dx.doi.org/10.1287/mksc.2015.0965.

Keywords: strategic consumers; decision support; revenue management; information availability; airline industry
History: Received: September 4, 2014; accepted: August 9, 2015; K. Sudhir served as the senior editor and Vishal Singh served as associate editor for this article. Published online in *Articles in Advance* March 24, 2016.

1. Introduction

In recent years, researchers have come to recognize the importance of strategic consumer behavior. A growing number of contributions explicitly model consumers—or a fraction of them—as intertemporal utility maximizers. These consumers, referred to as strategic consumers, consider future realizations of prices when timing their purchases (e.g., see the editorial by Shugan 2006 and reviews by Aviv et al. 2009, Shen and Su 2007). Even though multiperiod dynamic pricing increasingly relies on the assumption that some of the consumers behave strategically, there is limited evidence to indicate the extent to which such behavior impacts pricing. This paper contributes to the literature by estimating the relationship between the presence of online tools that support strategic consumer behavior and transacted prices in the context of the airline industry.

One of the most prominent effects of the Internet is its change of the retail landscape. On the consumers' side, new and innovative decision tools aimed at optimizing the product selection process are constantly being introduced: shopbots, recommendation agents, and price comparison websites are a few examples. On the retailer's side, the Internet facilitates revenue increasing through new strategies (e.g., Jiang et al. 2011) such as obfuscation[1] (Ellison and Ellison 2009), differentiation of product offering (Clemons et al. 2002), and the offering of online decision tools to effect consumer choice (Rubin and Mantin 2012). Whereas earlier empirical work associated with the effect of the Internet on pricing was primarily concerned with price dispersion and reduction of prices due to decreased search costs, in this work we focus on the pricing effects stemming from the availability of online intertemporal decision support tools. Specifically, we examine the effect of the introduction of a unique recommendation system for strategic consumers.

Recommendation systems help consumers make more accurate choices and draw consumers' attention for possible offerings that they did not consider (e.g., Brynjolfsson et al. 2011, Chen et al. 2013). Such systems have been shown to affect consumer choices, especially when faced with multiple choice options and when considerable difficulty is associated with making the choice. These systems serve to reduce the effort required to make the choice, as well as reduce the uncertainty surrounding the decision, effectively reducing the overall choice-making effort and increasing

[1] Retailers that use obfuscation embrace the effect of reduced search costs to attract customers to their website through one product, only to lure a transaction of another product.

choice confidence (Fitzsimons and Lehmann 2004). We consider the effect of a special type of recommendation system that extends the choice set to account for predicted *future* realizations of the alternatives and hence may support strategic consumer behavior.

The setting of this research is in the context of the airline industry. Strategic consumer behavior in the airline industry exhibits some unique attributes that naturally control for factors that may hinder strategic behavior analysis in other contexts. Unlike many other products, for which strategizing to wait might impose the risk of a reduced future valuation for the good (e.g., fashion goods), effective strategic consumer behavior for purchasing airline tickets is purely manifested in buying at lower prices. Since a flight ticket cannot be used until the flight date has arrived, the primary risk borne when strategizing to wait for lower airfares is that the price could, in fact, increase. By examining the effect of a purchase timing decision support tool in this industry, we are able to examine a tool whose core effect is on prices. Furthermore, timing purchases in this industry is particularly valuable for consumers, as the revenue management mechanisms adopted by airlines often prompts high variations in the observed prices across time.

In this paper, we analyze the transacted price behavior exhibited on routes for which price prediction information is publicly available, and we compare it against a control group, consisting of routes that are not supported with such information. The routes with price prediction information were extracted from Farecast—which was part of Microsoft's online travel website—a unique site that provides statistical information regarding future fare distributions for flights. Consumers interested in timing their purchase for a low priced airfare could consult this website and gain knowledge about the predicted movements of the lowest available fare in the ensuing week, the magnitude of these movements, and their likelihood. As consumers are provided with inferential statistics about the realization of future prices, they can develop informed expectations about their own surplus realizations and act accordingly. Thus, this site enabled consumers to be more informed prior to their decision making, and thus exhibit more effective intertemporal choices.

We first estimate the overall effect associated with the availability of fare prediction information and find that routes with this information exhibit transactions that are 4%–6% lower compared to other routes. Assessing the durability of the effect and possible differences in impact over the years, we employ difference-in-differences estimations for different time periods. These estimations reveal that when fare prediction information was first introduced (between 2006 and 2008), it resulted in a reduction in transacted fares across all fare percentiles. Successive introduction of prediction information (between 2008 and 2010), on other routes, is associated with fare reductions concentrated at the lower percentiles. This suggests that consumers who behave strategically are distributed across all fare percentiles when prediction information is not available. Alternatively, the reduction of fares at the lower percentiles may have pulled the entire fare distribution downward—an effect that airlines may have managed to control with later introduction of fare prediction information. As for the durability of the impact on fares, we find that markets that have had prediction information for two years or more exhibit no additional reduction in prices.

Because of the possibility of a route inclusion bias, we control for endogeneity concerns by utilizing a natural experiment event, in which the provision of fare prediction information was discontinued. This analysis reveals that fares on routes where price prediction information was discontinued experienced a significant increase compared to other routes. We also employ a two-stage instrumental variable (IV-2SLS) model, yielding consistent results. We conclude by examining additional effects that may be induced by information availability and lead to lower transacted prices, such as evidence for increased delayed purchases (manifested through changes in the percentage of restricted fare purchases), balancing of load factors across carriers in the market, and increased price sensitivity. These estimations show evidence that purchases are indeed delayed, but that the purchase pattern is not utterly different; i.e., purchases are not made from different carriers or from alternative venues, and the change is concentrated in purchase timing.

The remainder of this paper is organized as follows. In §2 we explain the fundamentals of revenue management and strategic consumer behavior in the context of the airline industry. Section 3 elaborates on the role of the Internet in providing decision support in purchasing airline tickets. Section 4 describes the data, whereas §5 presents the empirical models and estimation results. In §6 we summarize and conclude.

2. Strategic Consumer Behavior and Its Implications

Consumers who behave strategically recognize that the price of a desired product may drop at some point in time and time their purchase accordingly. To effectively behave strategically, an underlying premise is that consumers have access to sufficient information and are capable of solving the optimization problem faced by the firms, and thus are able to predict markdown patterns to tap on. Such assumptions may be plausible in many industries, such as fashion, where the intertemporal price discrimination generally takes the shape of dropping prices. However, in the travel industry, price movements are harder to predict. Although, in general, prices in this industry tend to increase toward the departure day due to arrival of business consumers (Belobaba 1989),[2] prices may also exhibit a large degree of fluctuation, mainly due to the employment of complex revenue management mechanisms.

Indeed, Li et al. (2014) find that the majority of the consumers in the travel industry exhibit myopic behavior; that is, rather than consider the possible realization of future prices, most consumers in the travel industry appear to base their purchase on the prices they observe on their arrival (i.e., shortly after they realize their need for travel). Similarly, Grubb (2015b) shows that in complex pricing environments, consumers tend to stick with initial choices or default options with excessive inertia. Such myopic behavior may also be supported by the expectation for a general price increase as time progresses. As a result, myopic consumers may miss out on opportunities for possible future markdowns. Furthermore, when consumers myopically purchase a product intended for future consumption, behavioral biases such as overconfidence, impatience, and loss aversion may yield further suboptimality in choices (DellaVigna and Malmendier 2006, Ho et al. 2006, Grubb 2015a). Other consumers may, of course, choose to delay their purchase and wait for a future lower price, but often fail to do so effectively because, for example, they may search too little or exhibit confusion when comparing prices (Grubb 2015a). Thus, faced with dynamically changing complex pricing environments coupled with a lack of information, consumers may miss out on the lower advance selling price as well as the timing of price markdowns.

However, consumers increasingly gain access to information and decision support systems that support strategic behavior (see §3). Such systems may help consumers weigh the probabilities of price changes (price drops versus price increases) and recommend a wait-or-buy decision that minimizes expected purchase price. In this research, we empirically analyze the degree to which such systems support strategic consumer behavior, because failing to account for strategic consumers can significantly reduce expected revenue (Talluri and van Ryzin 2004).

Starting with the seminal paper by Coase (1972), it has been argued, and later demonstrated by many (Stokey 1979, 1981; Besanko and Winston 1990; Aviv and Pazgal 2008; Levin et al. 2009; the recent review by Chen and Chen 2015), that the presence of strategic consumers exposes firms to intense intertemporal competition, which forces firms to drop prices to induce early purchase by consumers. Despite the growing interest in modeling and understanding strategic consumer behavior, there is, thus far, limited empirical evidence of strategic consumer behavior and its implication on firms. The problem stems from identifying these types of consumers. Nair (2007) developed a two-step process to estimate demand parameters in the presence of strategic consumers which are then used to numerically solve the firm's dynamic pricing problem. Soysal and Krishnamurthi (2012) estimated a structural model in the context of fashion goods, where waiting is associated with reduced utility and greater risk of stockouts. Osadchiy and Bendoly (2010) studied strategic consumer behavior in a laboratory setting. They found that providing subjects with information about the probability of getting an item induced a larger portion of the subjects to exhibit strategic behavior. Using a structural model, Li et al. (2014) estimated the extent to which consumers exhibited strategic consumer behavior. In this paper, we provide empirical support to these findings and further highlight the link between fare prediction websites and transacted prices.

2.1. Revenue Management and Strategic Consumer Behavior in the Airline Industry

At the core of revenue management mechanisms used by airlines is the expected marginal seat revenue (EMSR) of Belobaba (1989).[3] The intuition behind the EMSR mechanism is simple: protect seats for later arrival of consumers who might purchase at higher prices. Once the protection levels are established, the airline will open and close fare levels depending on the realization of demand.

To illustrate the concept of the protection levels, assume there are n fare classes, where F_i denotes fare of class i, with $F_i > F_{i+1}$. Traditionally, those fare classes are determined well in advance. The number of seats to be protected for fare classes 1 to i from class $i+1$ is denoted y_i, and thus the booking limit on classes $i+1, \ldots, n$ is $b_{i+1} = Capacity - y_i$; that is, b_i seats will be available for passengers at the fare of class i or lower.[4]

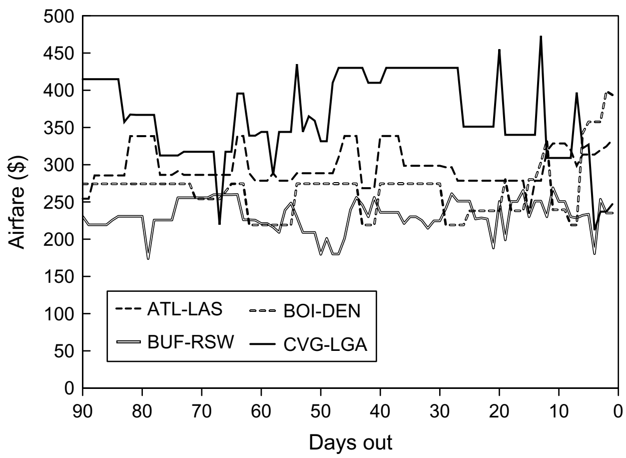
In a dynamic model, the estimations of seat protection levels and booking limits for the remaining capacity are revised as time, t, progresses from $t = T$ toward the departure day, $t = 0$, and fare classes may open and close—an important aspect when consumers consider exhibiting strategic behavior. Thus, an outcome of the use of revenue management as described above is that airfares, as perceived by consumers, may fluctuate—sometimes quite dramatically—as is illustrated in

[2] A price increase is also expected in this industry since early commitment to purchase a ticket is expected to be compensated for by a lower transaction price (Xie and Shugan 2001).

[3] Essentially, revenue management systems combine three economic pricing principles: cost-based, demand-based, and service-based pricing (Belobaba et al. 2009). For comprehensive reviews of the theory and practice of pricing and revenue management, see, e.g., Talluri and van Ryzin (2004), Phillips (2005), and Boyd (2007).

[4] Technically, y_i equals to the fractile $1 - F_{i+1}/F_{1,i}$, where $F_{1,i} = \sum_{j=1}^{i} F_j \mu_j / \sum_{j=1}^{i} \mu_j$, and μ_j denotes the mean demand for class j.

Figure 1 Sample of Lowest Available Fare Histories for Airport Pairs That Were Covered by Forecast for Seven-Day Return Flights Departing on March 4, 2008



Note. The airport codes refer to the airports in Atlanta (ATL), Las Vegas (LAS), Boise (BOI), Denver (DEN), Buffalo (BUF), Southwest Florida (RSW), Cincinnati/Northern Kentucky (CVG), and New York City's LaGuardia (LGA).

Figure 1. This phenomenon has been well documented in the literature (see, e.g., Etzioni et al. 2003, McAfee and te Velde 2007, Escobari 2012). Therefore, timing the purchase can lead to substantial savings to consumers.

Essentially, strategic consumers wishing to time their purchase need to estimate the chances that lower fare classes will open in the future. This may be done on the basis of intuition or some heuristic, but most accurately this can be done by estimating the probabilities of classes opening at the end of each period. Using the estimated probabilities, strategic consumers can weigh the benefit of waiting (savings due to opening of a lower fare class) versus the risk of waiting (loss due to closing of a fare class).

To demonstrate, let $y_{i,t}$ denote the protection limit for fare classes 1 to i at time t, let $d_{i,t}$ denote the realized demand for class i during period t, and let c_t denote the seating capacity of the aircraft at the beginning of period t. The actual sales, $s_{i,t}$, is a function of the class's booking limit, $b_{i,t}$, as well as the sales to higher fare classes during period t. If at time t the lowest available fare is of class j, F_j, then a strategic consumer would be interested in the probability that class $j+1$ will reopen in the next period. Since class j is the lowest class open, the available capacity exceeds the protection level for the higher fare class, $j-1$, at time t. Class $j+1$ will reopen if at the beginning of the next period, the capacity will exceed that period's protection level for higher fare classes (i.e., j and higher); that is, if $c_{t-1} > y_{j,t-1}$.[5]

Table 1 Fare Classes' Demand Estimations and Protection Levels by Period

		Period		
		3	2	1
Class 1	$N(\mu_{i,t}, \sigma_{i,t})$	$N(1,1)$	$N(7.5, 4.69)$	$N(9, 3.38)$
	$y_{i,t}$	19	18	10
Class 2	$N(\mu_{i,t}, \sigma_{i,t})$	$N(5,5)$	$N(5,5)$	$N(5,5)$
	$y_{i,t}$	39	32	18
Class 3	$N(\mu_{i,t}, \sigma_{i,t})$	$N(9, 3.38)$	$N(7.5, 4.69)$	$N(1,1)$

Note. $\mu_{i,t}$ and $\sigma_{i,t}$ denote the mean and standard deviation of demand, respectively, for class i at period t.

We illustrate using the example (Table 1) from Anderson and Wilson (2003), where the aircraft has a seating capacity of 50, and the fares for classes 1, 2, and 3 are \$500, \$200, and \$100, respectively.[6] The behavior of strategic consumers can be influential when, for example, class 2 consumers who arrive at period 2 consider the probability that the fares in class 3 will reopen.[7] Considering the realization of sales in period 3 $s_{1,3}=0$, $s_{2,3}=8$, and $s_{3,3}=11$, the capacity at the end of period 3 is $c_3 = 50-19 = 31$. Thus, $c_3 < y_{2,2}$, and class 3 closes at the end of period 3. If period 2 demand, $d_{1,2}+d_{2,2}$, is less than $c_1 - y_{2,1}$, then class 3 reopens at the beginning of period 1. This occurs with a probability of $\Pr(d_{1,2}+d_{2,2} < 31-18) = 0.53$. If period 2 sales, $s_{1,2}+s_{2,2}$, exceed $c_1 - y_{1,1}$, then class 2 closes. This occurs with a probability of $\Pr(s_{1,2}+s_{2,2} > 31-10) = 0.11$.[8] Factoring the difference between the fares (\$200–\$100 and \$500–\$200, respectively), the expected savings is \$53, which exceeds the expected loss, \$32, and hence in such a situation, a strategic consumer might choose to wait. Notably, considerable information and skill sets are required for predicting the reopening of fare classes. The complexity of this task amplifies when alternative carriers and flights are considered.

This process suggests several predictions: (i) as consumers start exhibiting strategic behavior and time

[5] Technically, the sales can be expressed as $s_{i,t} = \min\{\min\{b_{i,t}, d_{i,t}\}, c_t - \sum_{j=1}^{i-1} s_{j,t}\}$, where the first term captures the minimum between the realized demand and the booking limit, and the second term further reduces the available seats due to sales to higher fare classes during period t. The capacity updating can be written as $c_{t-1} = \max\{c_t - \sum_{i=1}^{n} s_{i,t}, 0\}$, the probability that class $j+1$ will reopen next period is given by $\Pr(\sum_{i=1}^{j} s_{i,t} < c_t - y_{j,t-1})$, and the probability that class j will close by next period is given by $\Pr(\sum_{i=1}^{j} s_{i,t} > c_t - y_{j-1,t-1})$.

[6] In Table 1, $y_{1,3}=19$ is derived by generating the sum of class 1 demand distributions over the three periods, $N(17.5, 5.86)$, and locating the fractile $1-200/500=0.6$. Similarly, $y_{2,3}=39$ is derived by locating the fractile $1-100/361.54=0.72$ in the sum of class 1 and 2 demand distributions over the three periods, $N(32.5, 10.46)$, and so forth.

[7] In their paper, Anderson and Wilson (2003) assumed that strategic consumers wait if the probability of a price drop exceeds a certain threshold, ignoring the potential savings and abstracting away from the risk the fare will actually rise.

[8] The probability $\Pr(s_{1,2}+s_{2,2} > 31-10)$ can be expressed as $\Pr(d_{1,2}+d_{2,2} > 31-10 \cap d_{1,2} > 18-10)$; since the protection level for class 1 is 18, at least eight of the seat sales need to come from class 1. To estimate this probability, one can discretize the normal distribution.

their purchase to take advantage of future fare fluctuations, the transacted price is expected to decrease; (ii) increased consumers' strategic waiting may result with an increased proportion of transactions being delayed;[9] and (iii) since low fill rates are associated with the reopening of low fare classes, improved strategic behavior can lead to balancing of load factors across carriers on a route, as demand stemming from strategic consumers may gravitate to carriers with lower load factors. In what follows, we empirically test those predictions.[10]

3. Information and Decision Support in the Airline Industry

Derivation of the probabilities of reopening and closing of fare classes, as demonstrated in the previous section, is not an easy task. Farecast, which was founded in 2003 and later sold to Microsoft in 2008, was the first and only website to provide support for deciding on the flight ticket purchase timing.[11] The mechanism provides users with statistically inferred airfare predictions. To build the database of historical market airfares, the site bought extensive data from ITA Software, an airfare and pricing provider, and employed inferential techniques to predict the direction and magnitude of movement of the lowest available airfare. Although the algorithm is proprietary,[12] a three-month audit reported that Farecast correctly predicts the direction of fares 75% of the time and saves passengers $27 on average.[13] This tool enables effective strategic consumer

Figure 2 (Color online) Search on August 29, 2011, for JFK–LAX for a Seven-Day Return Trip Departing on November 10, 2011

behavior for purchasing flight tickets because, by using this tool, consumers can knowledgably anticipate the probabilities of fare classes closing and reopening as well as the magnitude of change in fares, and then purchase accordingly. This provides a unique opportunity to isolate the effect of strategic consumer behavior on transacted prices.

Figure 2 gives an example of the type of information this tool provides to its users. The confidence levels and price ranges of the predicted price are broken down into three groups: a price increase within a certain range ($18–$43 in the example), a price decrease within a certain range ($22–$72 in the example), or a stable price between these two ranges.[14] With this information on the predicted fare distribution, a strategic consumer can assess the expected utility from waiting, without any possession of knowledge about demand distributions, fare classes, capacities, or booking limits. Considering the example from Figure 2, factoring in the probabilities

[9] Although strategic consumer behavior does not necessarily imply strategic waiting, it is often assumed that myopic consumers simply consider a purchase as soon as they arrive (see, e.g., empirical support by Li et al. 2014). Thus, even if only a fraction of strategic behavior results with strategic waiting, overall the expected outcome is an overall delay in the transactions.

[10] Additional, longer run effects could be suggested based on this theory. For example, a process similar to the spiral-down effect (Cooper et al. 2006) could emerge: seats are protected based on past realizations of demand, oftentimes ignoring the effect of strategic waiting for low-fare tickets. If demand gravitates from higher fare classes to cheaper tickets, predictions are updated resulting with lower protection levels and hence greater availability of low-fare tickets (as they might remain open for longer periods of time) resulting with an even greater number of lower transacted fares. Also, if airlines realize the impact of strategic consumer waiting, they might revise their pricing strategy in such a way that reopening of lower fare classes will be minimized or even eliminated altogether.

[11] Farecast has captured wide media attention. For example, it was recognized as one of PC World's 20 Most Innovative Products of 2007, "Best of What's New for 2006" by *Popular Science*, one of *Time* magazine's 50 Coolest Websites for 2006, and one of the "Best Trip Planning Tools" by *BusinessWeek* readers as cited by PR Newswire (2007).

[12] The data mining technique that lies at the core of the prediction process is described by Etzioni et al. (2003).

[13] See http://www.minitime.com/trip-tips/Snag-Low-Airfares-Like-A-Pro-article.

[14] Additionally, the user is provided with information about the airfare history starting from 90 days prior to the flight date, complemented with statistical indicators (lowest/highest airfares, and some measure of historical volatility). In the online appendix (available as supplemental material at http://dx.doi.org/10.1287/mksc.2015.0965), we demonstrate that the prediction information provided by the tool has not changed between 2008 and 2010.

and using the middle fare in each of the ranges, the expected savings of about \$29.14 ($= 0.62 \cdot \$47$) exceeds the expected loss of about \$2.44 ($= 0.08 \cdot \$30.5$). Hence, strategic consumers might prefer to wait, exactly as suggested by Farecast, which provided consumers with a recommendation (Tip): wait or buy now.

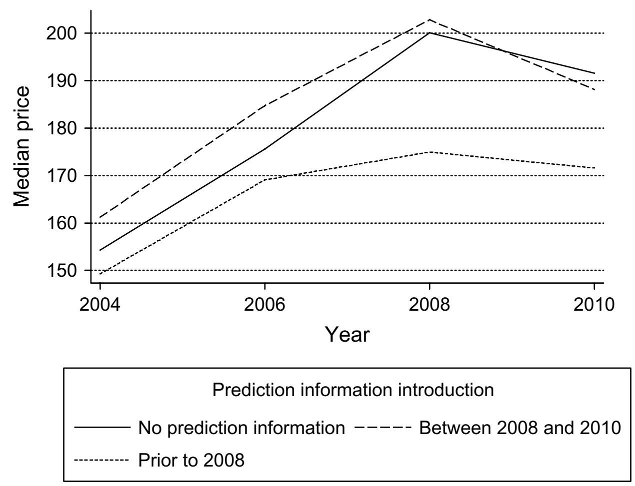
Hence, Farecast is a type of a recommendation system that supports consumers in behaving strategically. Recommendation systems and the information shown to users can have a large impact on the outcome of their decision processes (Chen et al. 2013), possibly leading customers to purchase products that they did not plan on purchasing (Schonfeld 2007, Flynn 2006). For example, Adomavicius et al. (2011) have shown a strong anchoring effect on consumer preferences following the observation of recommendations from a system. This implies that a recommendation by Farecast to wait for a lower airfare may induce consumers to act accordingly. Farecast may also lead consumers to purchase cheaper itineraries they were not even aware of, simply because they are expected to become cheaper.

Another aspect of recommendation systems is their potential contribution to enhance decision accuracy. A decision process is commonly interpreted as involving a trade-off between the decision-making effort and the accuracy of the decision outcome. The complexity of decision tasks, the potentially limited cognitive resources and knowledge of the users, and the tendency to reduce the overall decision effort potentially lead to suboptimal outcomes (e.g., Mandl et al. 2011). Recommendation systems, however, reduce the required effort, which can increase decision accuracy. This means that even if consumers are able to decipher airlines' revenue management mechanisms, a recommendation system such as Farecast can support their strategic behavior simply because information is more easily accessible.

4. Data

Our analysis of the effects of strategic consumer behavior on fares paid in the airline industry is based on data from two main sources: Farecast's website, for data on fare prediction availability, and the U.S. Department of Transportation's (DOT) databases, for data on transacted fares and other competitive measures. From Farecast's website, we extracted the list of airport pairs for which the website provides fare prediction information. Farecast was launched in July 2006, and we collected this list twice: during the first quarter of 2008 and during the first quarter of 2010. In 2008, fare prediction was provided for a total of 1,639 U.S. domestic airport pairs, which increased to a total of 3,885 airport pairs in 2010.[15] We constructed a binary variable, *PredictionInfo*, that takes a value of 1 if airfare prediction is available for an airport pair on the website and 0 otherwise. In conjunction with Farecast's data, we analyzed transacted fares on nonstop flights from the immediately following quarter (i.e., second quarter of 2006, 2008, and 2010, respectively). The DOT provides a quarterly report of a 10% random sample of transacted airline tickets from reporting carriers (the Airline Origin and Destination Survey, also known as the DB1B).[16] To have reliable percentile estimates, we consider only airport pairs that have at least 50 quarterly observations, and we included only tickets that were purchased for at least US\$50.

Figure 3 **Average Median Transacted Fare (in US\$) During the Second Quarter of Each Year Grouped by the Time of Prediction Information Introduction**



In Figure 3, we plot the relationship between prediction information availability and transacted fares in quarter 2 of the years 2004, 2006, 2008, and 2010. The plot provides the average median transacted fares of three groups of routes (defined according to the timing of the introduction of fare prediction information on the route). Evidently, prior to the availability of fare prediction information (i.e., between 2004 and 2006, as the site was launched later in 2006), no apparent difference in price behavior emerges between the three groups. However, once prediction information is introduced, a lower increase (or higher decrease) appears in the transacted prices of the relevant group. In our analysis we empirically test for this and related outcomes after controlling for the necessary variables.

Our control variables are consistent with those typically used in the air-transport literature. To capture

[15] Between 2008 and 2010, four city pairs were removed from the list of airport pairs covered. Because of the small sample of pairs that were removed, we ignore these four airport pairs. The 2010 coverage is close to 80% of the U.S. airport pairs providing domestic commercial flights.

[16] Certain aspects of the itinerary are not provided by the DB1B such as the exact transaction date, the exact flight date, timing of the flights, and duration of layovers. Such information, which is not publicly available, could have allowed us to explore additional aspects pertaining to consumers' strategic behavior.

route characteristics, we collected data on the great circle distance for each airport pair (*Distance*), and derived the average population (*AvgPop*) of the metropolitan areas served by the origin and destination airports. Using the data from the Bureau of Economic Analysis, we further derived the average income per capita (*AvgIncPop*) of the origin and destination metropolitan areas. The competitive environment is captured via three market structure measures: (1) the Herfindahl–Hirschman index (*HHI*), which quantifies the degree of market concentration (based on airlines' seats market shares derived from the DOT's T-100 database, which provides records on market aggregate data such as frequency, seats, and enplanement); (2) the seats market share of low-cost carriers (*LCC*);[17] and (3) the market size using the total number of passenger seats available in the market (*MarketSeats*). We also account for the average magnitude of activity at the origin and destination airports using the number of passengers served by each airport (*AvgAirportPax*). This measure may reflect airports' location-based rents and level of connectivity (e.g., from serving as a hub for certain carriers), which would ultimately allow airports to charge the carriers and passengers a premium for using the facility.

5. Model and Empirical Analysis

In our empirical analysis, we estimate the effect imposed by the availability of fare prediction information. We carry out the analysis using fixed effects as well as difference-in-difference estimations (§5.1). Following these estimations, we provide estimations that account for the concern of a possible route selection bias. We address this issue by employing a natural experiment event (§5.2) and by applying a two-stage IV model endogenizing market inclusion choice (§5.3).

5.1. The Effect of Fare Prediction Information

We first analyze whether there is evidence that the presence of fare prediction information is significantly related to the prices paid. For this purpose, we apply a fixed-effects regression with control variables for both time- and route-specific effects. Fixed-effects regressions are commonly used for analysis of airfares with panel data (e.g., Gerardi and Shapiro 2009, Brueckner et al. 2013; among others). The results of fixed-effects regressions reveal the magnitude effect of variables as changes occur over time.

Because the effect of fare prediction information may vary across the different fare percentiles, we consider the effect at the primary three quartiles: the 25th, 50th, and 75th, as well as the lower and upper deciles—the 10th and 90th percentiles.[18] Similar to other researchers who studied airfares utilizing panel data (e.g., Lederman 2007, Kwoka and Shumilkina 2010, Brueckner et al. 2011), we estimate the following log-linear reduced-form inverse demand fixed-effects equations:

$$\begin{aligned}\ln(Fare_t^{ij}) = {}& \alpha_0 + \alpha_1(PredictionInfo_t^i) + \alpha_2(LCC_t^i)\\ &+ \alpha_3(HHI_t^i) + \alpha_4(MarketSeats_t^i)\\ &+ \alpha_5(AvgAirportPax_t^i) + \beta_t^j + \gamma^{ij} + \varepsilon_t^{ij}. \quad (1)\end{aligned}$$

Hence we estimate five separate models, one for each percentile, where $Fare_t^{ij}$ is the fare paid for a ticket at the jth percentile of ticket prices in market i for an itinerary during quarter t. Market refers to any of the airport pairs covered in the U.S. DOT origin and destination database. In each estimated equation, $t \in \{2006, 2008, 2010\}$ and $j \in \{10\text{th}, 25\text{th}, 50\text{th}, 75\text{th}, 90\text{th}\}$ percentiles, β_t^j is the annual fixed effects, and γ^{ij} is the market fixed effects. The regressions are estimated separately for each of the percentiles. The market fixed effects control for systematic differences in demand across the different markets. The year fixed effects control for industrywide changes, such as changes in economic conditions, weather, and technology, that may have taken place in the airline industry over the sampled period.[19] The estimation results are reported in Table 2.

Our interest is in the effect of the availability of prediction information. The coefficient of the *PredictionInfo* variable measures the magnitude effect of introducing such information on transacted airfares in the corresponding markets, with respect to the change experienced in other markets, where change in prediction information availability has not taken place (i.e., have had prediction information available both in 2008 and 2010, or have had no prediction in either year).

Observation 1. Introduction of fare prediction information is associated with a significant negative effect on transacted airfares.

[17] The presence of LCCs have been shown to impose pressure on airfares (Goolsbee and Syverson 2008). The list of LCCs include AirTran Airways, Allegiant Air, JetBlue Airways, Southwest Airlines, Spirit Airlines, Sun Country Airlines, USA3000 Airlines, Virgin America, and Frontier Airlines.

[18] Other contributions that account for fare percentiles include those of, e.g., Lederman (2007), who considers the 20th and 80th percentiles; Brueckner et al. (2013), who considers the 25th and 75th percentiles; and Gerardi and Shapiro (2009), who focus on the 10th and 90th percentiles.

[19] Since debundling markets affects all markets, the change will be experienced in a systematic way and not expected to influence markets where prediction information is introduced between 2008 and 2010 more than other markets. Similarly, if other changes took place in markets (such as departure time), they are not expected to correlate with the prediction information provision. In any case, we provide endogeneity analysis to account for the effect of potential unobserved variables.

Table 2 **Impact of *PredictionInfo* on Transacted Airfares at Different Percentiles: Fixed-Effects Regressions**

Ln(*Fare*)	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile
PredictionInfo	−0.0389***	−0.0582***	−0.0589***	−0.0634***	−0.0559***
	(0.00648)	(0.00656)	(0.00663)	(0.00681)	(0.00682)
AvgAirportPax (M)	−0.0154***	−0.0168***	−0.0146***	−0.00300	0.00364
	(0.00370)	(0.00375)	(0.00379)	(0.00389)	(0.00390)
LCC	−0.173***	−0.205***	−0.227***	−0.255***	−0.326***
	(0.0199)	(0.0201)	(0.0203)	(0.0209)	(0.0209)
MarketSeats (K)	−0.00181***	−0.00196***	−0.00229***	−0.00264***	−0.00275***
	(0.000196)	(0.000198)	(0.000200)	(0.000205)	(0.000206)
HHI	0.118***	0.109***	0.107***	0.127***	0.148***
	(0.0150)	(0.0152)	(0.0154)	(0.0158)	(0.0158)
N	9,774	9,774	9,774	9,774	9,774
Adj. R^2	0.722	0.744	0.771	0.804	0.837

Notes. Standard errors are in parentheses. All regressions include annual and market fixed-effects dummies. (K) and (M) indicate the variables are measured in thousands and in millions of units, respectively.
$^{***}p < 0.01$.

The coefficient of the *PredictionInfo* variable is negative and significant at all percentiles. This is an important result. It suggests that fare prediction information has a negative impact on transacted airfares. Possibly, as discussed earlier, such information is utilized by consumers to time their purchase more wisely. This lends support to the idea that consumers may desire to behave strategically, but without the availability of relevant information fail to do so.

The magnitude of the effect on transacted airfares is about 4%–6%.[20] The effect is not limited to the median fare or to the lower transacted fare percentiles, where one might argue that price sensitive consumers are more likely to concentrate. In fact, the effect shifts the entire transacted fare distribution downward. If indeed strategic consumers are more likely to be concentrated at the lower percentiles, then the downward shift effect these consumers impose on prices, even those at higher percentiles, echoes the observation made by Gerardi and Shapiro (2009) on the impact induced by the entry of LCCs; namely, as fares in the lower percentiles drop, the rest of the distribution shifts as well. Importantly, this could be induced by the spiral-down effect mentioned in §2.1.

Another reason for this shift could be that when fare prediction information is not available, price-sensitive consumers are not necessarily concentrated only at the lower percentiles and end up purchasing at a high fare (e.g., if consumers fail to effectively time their purchase or if they wrongly estimate future fare distributions). As fare prediction information becomes available, those consumers can time their purchase more effectively. Consequently, the distribution of transacted airfares shifts, and we observe a strong effect at the higher percentiles as well. Note that since these estimates are in the context of a route fixed-effects regression, this effect relates to markets in which fare prediction is available, without controlling for possible differences in magnitude of the effect between 2006 and 2008. In §5.1.1 we further explore how the effect differs over the years and as the prediction information continues to be available.

5.1.1. Standard Difference-in-Differences Estimation. In this section we conduct a difference-in-differences estimation to capture the net effect experienced by routes for which fare prediction information was made available online (the treatment group) compared with the markets without fare prediction information (control group), while taking into account different levels of maturity of information and a possible difference in effect between 2006 and 2008. Hence, we provide the estimates of the effects of prediction information inclusion in separate estimations for two time periods: (i) between the years 2006 and 2008, the time period that prediction information was made available for the first time, and (ii) between 2008 and 2010, the time period when additional routes were incorporated. For the panel consisting of 2006 and 2008, we estimate the following:

$$\begin{aligned}\ln(Fare_t^{ij}) &= \alpha_0+\alpha_1(PredictionInfo_t^i)+\alpha_2(LCC_t^i)+\alpha_3(HHI_t^i)\\ &\quad+\alpha(MarketSeats_t^i)+\alpha_5(AvgAirportPax_t^i)+\alpha_6(2008_t^i)\\ &\quad+\alpha_7(2008_t^i\times PredictionInfo_t^i)+\alpha_8(AvgPop^i)\\ &\quad+\alpha_9(AvgIncPop^i)+\alpha_{10}(Distance^i)+\varepsilon_t^{ij},\end{aligned} \quad (2)$$

where the 2008 dummy variable captures industry-wide changes between 2006 and 2008, and the 2008 × *PredictionInfo* interaction variable captures the incremental effect between 2006 and 2008 in the treatment markets as compared with the control markets. The relationship between the variables and controls is provided in Table 3.

[20] In log-linear regressions, the coefficients reflect a relation of e^{α}. For example, in the 10th percentile the coefficient is −0.0389, implying $e^{-0.0389} = 0.962$. Since this is a multiplicative relationship, the fares dropped by a factor of $1 - 0.962 = 0.038$, or 3.8%.

Table 3 Difference-in-Differences Estimations: 2006–2008

	2006	2008	Difference
Prediction info: Markets with prediction info in 2008	$\alpha_0+\alpha_1$	$\alpha_0+\alpha_1+\alpha_6+\alpha_7$	$\alpha_6+\alpha_7$
No prediction info: Control group	α_0	$\alpha_0+\alpha_6$	α_6
Difference	Prediction − Control $=\alpha_1$	Prediction − Control $=\alpha_1+\alpha_7$	Prediction − Control $=\alpha_7$

Table 4 Difference-in-Differences Estimations: 2008–2010

	2008	2010	Difference
Mature: Markets with prediction info prior to 2008	$\alpha_0+\alpha_1$	$\alpha_0+\alpha_1+\alpha_7+\alpha_8$	$\alpha_7+\alpha_8$
New: Markets with prediction info added between 2008 and 2010	$\alpha_0+\alpha_2$	$\alpha_0+\alpha_2+\alpha_7+\alpha_9$	$\alpha_7+\alpha_9$
No prediction info: Control group	α_0	$\alpha_0+\alpha_7$	α_7
Difference	Mature − Control $=\alpha_1$ New − Control $=\alpha_2$	Mature − Control $=\alpha_1+\alpha_8$ New − Control $=\alpha_2+\alpha_9$	Mature − Control $=\alpha_8$ New − Control $=\alpha_9$

Similarly, for the panel consisting of 2008 and 2010, we estimate

$$\begin{aligned}\ln(Fare_t^{ij}) &= \alpha_0+\alpha_1(PredictionInfoMature_t^i)\\ &\quad+\alpha_2(PredictionInfoNew_t^i)+\alpha_3(LCC_t^i)\\ &\quad+\alpha_4(HHI_t^i)+\alpha_5(MarketSeats_t^i)+\alpha_6(AvgAirportPax_t^i)\\ &\quad+\alpha_7(2010_t^i)+\alpha_8(2010_t^i\times PredictionInfoMature_t^i)\\ &\quad+\alpha_9(2010_t^i\times PredictionInfoNew_t^i)+\alpha_{10}(AvgPop^i)\\ &\quad+\alpha_{11}(AvgIncPop^i)+\alpha_{12}(Distance^i)+\varepsilon_t^{ij},\end{aligned} \tag{3}$$

where *PredictionInfoMature* is a dummy variable that has a value of 1 if a market has had prediction information prior to 2008, and *PredictionInfoNew* is a dummy variable which equals 1 if the prediction information was added between 2008 and 2010. The interaction variables $2010_t^i \times PredictionInfoMature$ and $2010 \times PredictionInfoNew$ capture the incremental effect between 2008 and 2010 in those market types compared with the control markets. The relationship between the variables and the controls is illustrated in Table 4.

The estimation results of both difference-in-differences are provided in Table 5. The first set of estimations show that for markets where fare prediction was introduced between 2006 and 2008, a fare decrease in the range of 3%–5% was experienced at the different percentiles (except for the 10th percentile) compared

Table 5 Impact of *PredictionInfo* on Transacted Airfares: Difference-in-Differences Regressions

Ln(*Fare*)	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile
			2006–2008		
PredictionInfo	−0.0764***	−0.113***	−0.132***	−0.122***	−0.0988***
	(0.00886)	(0.00952)	(0.0110)	(0.0123)	(0.0126)
2008 × *PredictionInfo*	−0.00878	−0.0290**	−0.0489***	−0.0472***	−0.0326**
	(0.0110)	(0.0118)	(0.0136)	(0.0152)	(0.0156)
N	7,242	7,242	7,242	7,242	7,242
Adj. R^2	0.452	0.439	0.376	0.371	0.438
			2008–2010		
PredictionInfoNew	−0.0984***	−0.157***	−0.241***	−0.226***	−0.185***
	(0.0118)	(0.0127)	(0.0142)	(0.0157)	(0.0161)
PredictionInfoMature	−0.164***	−0.262***	−0.370***	−0.348***	−0.289***
	(0.0126)	(0.0136)	(0.0151)	(0.0167)	(0.0171)
2010 × *PredictionInfoNew*	−0.0542***	−0.0590***	−0.0162	−0.0370*	−0.0347
	(0.0168)	(0.0180)	(0.0201)	(0.0222)	(0.0228)
2010 × *PredictionInfoMature*	−0.00775	−0.00744	0.0357*	0.0223	0.0184
	(0.0168)	(0.0180)	(0.0202)	(0.0222)	(0.0228)
N	7,214	7,214	7,214	7,214	7,214
Adj. R^2	0.469	0.453	0.422	0.407	0.453

Notes. Standard errors are in parentheses. All regressions include *AvgPop*, *AvgIncPop*, *Distance*, *LCC*, *MarketSeats*, *HHI*, and *AvgAirportPax*.
$^{*}p<0.10$; $^{**}p<0.05$; $^{***}p<0.01$.

to markets without fare prediction. Considering the second set of estimations, we can observe that markets with mature fare prediction—namely, those that were introduced prior to 2008—show no significant fare change compared to markets without fare prediction between 2008 and 2010. This may suggest that the reduction in prices exhausted the possible surplus that can be shifted from the airlines to the consumers, and therefore no more reduction can be obtained. It is important to note that although an additional reduction in fare for those markets is not observed, these markets did not experience a relative increase in fare. Thus, the surplus that shifted to consumers in these markets in 2008 was sustained in 2010. We note the following observation.

Observation 2. The effect of fare prediction information is significant upon introduction. As the information matures, no further incremental change in fares is observed.

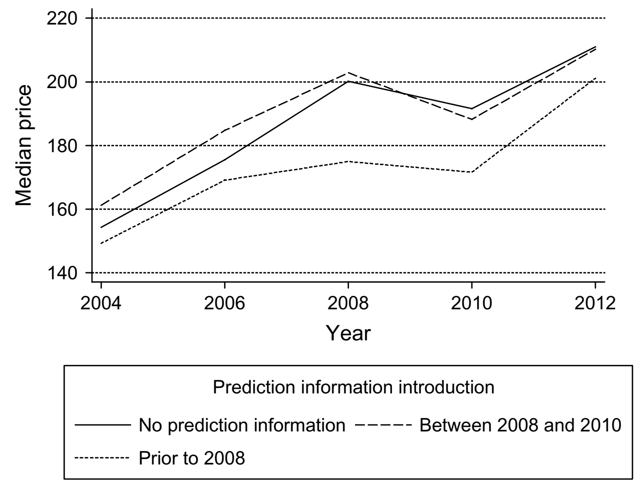
We note that markets with *PredictionInfoNew* (i.e., those that were introduced after 2008) show a significant difference of about 5%–6% only at the lower fare percentiles. A possible explanation for the concentration at the lower percentiles could be that as the spiral-down effect, which is induced by strategic consumer behavior, was recognized by airlines and they were able to avoid it; i.e., the availability of low-fare seats was not increased. Thus, surplus shifted only to consumers at the lower fare percentiles.

Hence, the estimations in this section show a consistent picture on the range of surplus that may be shifted to consumers. The overall shift in surplus appears to have been highest after the first set of routes was introduced, as surplus shift was distributed across the different transacted fare percentiles. Consequently, carriers have possibly learned how to cope with this changing consumer behavior, and therefore the surplus shift was limited to the lower percentiles.

5.2. The Effect of Discontinuing Fare Prediction Service

Farecast relied on data purchased from ITA Software to generate its fare prediction for the different markets. However, in 2011, ITA was purchased by Google, a move that had detrimental effects on Farecast's ability to continue providing prediction data on routes and eventually led to the termination of the fare prediction service. The imperiling effects of the ITA purchase were well documented in the media and led to the formation of an antideal FairSearch coalition, which resulted in the U.S. Department of Justice (DOJ) filing an antitrust lawsuit (Goldman 2011). Yet, the purchase was approved in 2011 following a settlement with the DOJ. Critically, after this acquisition, Microsoft decided to terminate its purchase of airfare data from ITA (since "Bing didn't want to pay Google for the data required to power Farecast," according to Oren Etzioni, the founder of Farecast[21]). In addition, at around the same time, the Farecast executive and information mining groups left the company.[22] These issues severely affected Farecast's ability to provide useful prediction information and led Microsoft to abandon the fare prediction service and terminate it completely. Hence, this event provides us with the opportunity for an additional experiment examining the effects of changes in fare prediction information availability on transacted prices. Under this experiment, endogeneity concerns are naturally controlled for since the event is purely external and was imposed on the market.

Figure 4 **Average Median Transacted Price Grouped According to the Time of Prediction Information Introduction: 2004–2012**



An expected outcome of the ITA purchase event would be that routes for which fare prediction information was provided in 2010 would experience a significant price increase, when compared to prices which had no fare prediction information. Figure 4, which complements Figure 3 with 2012 data (i.e., after the acquisition of ITA by Google), depicts an apparent increase in median prices between 2010 and 2012 in all market groups. This price increase may be a result of increased fuel prices or other marketwide effects. Most importantly, a careful inspection reveals that the relative price increase for markets that had fare prediction data (the dashed lines) appears to be higher than the increase experienced by the control group (the solid line).

Similar to the analysis carried out in §5.1.1, we consider the three market groups consisting of *PredictionInfoNew, PredicitonInfoMature*, and the control group of markets with no fare prediction information and perform a difference-in-differences analysis

[21] See Cook (2014).

[22] See Cook (2014).

Table 6 Impact of *PredictionInfo* on Transacted Airfares: Difference-in-Differences Regressions, 2010–2012

Ln(*Fare*)	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile
PredictionInfoNew	−0.143***	−0.200***	−0.227***	−0.227***	−0.177***
	(0.0137)	(0.0135)	(0.0138)	(0.0150)	(0.0156)
PredictionInfoMature	−0.159***	−0.254***	−0.317***	−0.302***	−0.236***
	(0.0143)	(0.0141)	(0.0145)	(0.0157)	(0.0163)
2012 × *PredictionInfoNew*	0.0652***	0.0369**	0.0258	0.0363*	0.0144
	(0.0187)	(0.0185)	(0.0190)	(0.0206)	(0.0213)
2012_t^i × *PredictionInfoMature*	0.111***	0.103***	0.0932***	0.0835***	0.0713***
	(0.0186)	(0.0184)	(0.0189)	(0.0205)	(0.0213)
N	6,999	6,999	6,999	6,999	6,999
Adj. R^2	0.459	0.463	0.436	0.412	0.452

Notes. Standard errors are in parentheses. All regressions include *AvgPop*, *AvgIncPop*, *Distance*, *LCC*, *MarketSeats*, *HHI*, and *AvgAirportPax*.
$^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

over the period 2010 to 2012. The estimation results (Table 6) show compelling evidence for a significant price increase after the loss of prediction information. One can see that markets that had fare prediction information prior to 2008 experienced a fare increase of about 7%–11% (compared to markets with no predictive information) between 2010 and 2012, and markets that had this information added between 2008 and 2010 saw an increase of approximately 3%–6% (compared to markets with no predictive information) in the lower percentiles. It is interesting to see that the impact of the change is more profound in the markets that were initially introduced to Farecast. This is in line with the idea that Farecast initially included routes with large volumes of price sensitive consumers. Therefore, fare prediction information loss may have had a more significant impact in those routes.

5.3. Endogenizing Market Choice

Our next estimation, which also addresses the endogeneity concern, contemplates a possible route inclusion bias by employing an IV-2SLS estimation model. In this estimation, our endogenous variable (i.e., treatment) is the choice to include a market on Farecast and was modeled in our first-stage estimation after consulting with Farecast executives regarding the company's expansion strategy; that is, in the first stage, we model the endogenous binary variable *PredictionInfo*, and in the second stage we model the percentage fare change, $\Delta FarePct_{2008}^{ij} = (Fare_{2008}^{ij} - Fare_{2006}^{ij})/Fare_{2006}^{ij}$, experienced on the routes.

According to the company's executives, a main factor that influenced Farecast's choice of markets as it launched its fare predication tool was market and passenger characteristics; that is, when Farecast launched, it quickly expanded to include markets that are geographically spread across the continental United States and in which it assumed patient and price-sensitive consumers are concentrated. Many variables related to this aspect of Farecast's choice are controlled for via our earlier regressors (*AvgPop*, *AvgIncPop*, *Distance*, *HHI*, *LCC*, *AvgAirportPax*, and *MarketSeats*).

For our two-stage model, we have developed a new instrument that captures the company's preference to spread geographically rather than concentrate on specific regions. Specifically, utilizing the Federal Aviation Administration's classification of airports into nine geographical regions, we use an ordinal ranking for airports within each region, with 1 being the busiest airport, and then normalize for region size by dividing this ranking by the number of airports in the region. Averaging the normalized ranking of the origin and destination airports in each route yields our instrument, *AvgRegionAPRank*.[23]

The estimation results, for both a linear and probit first-stage estimation, are provided in Table 7.[24, 25] We learn that Farecast focused its entry into thicker markets with larger populations and higher income levels. It is also highly notable that the results support our prediction that geographical coverage is a strong instrument for fare prediction introduction. Most importantly, when examining the second-stage coefficients, we see that the coefficient of *PredictionInfoFit* is negative and significant.[26] Hence, consistent with

[23] Finding an instrument when estimating prices in the aviation markets is considered especially challenging, and therefore often different mechanisms are used to reduce endogeneity concerns. For example, Dresner and Tretheway (1992) use a two-stage indirect least squares model to capture the endogeneity of passenger demand in the first stage and price in the second stage. Alderighi et al. (2015) study the effect of code sharing on posted prices. To control for the endogeneity of the existence of a code-sharing agreement, they separately estimate a probit model in the first stage and then use the inverse Mills ratio to correct for a selection bias. Another paper that discusses the issues of endogeneity is Brueckner and Luo (2014).

[24] Technically, these relate to the ivreg2 and treatreg packages in Stata to estimate these IV-2SLS regressions.

[25] We shall note that the *C* test—which is equivalent to the Hausman test comparing IV and ordinary least squares (OLS) estimates—rejects the OLS in favor of the IV. Also, the Cragg–Donald Wald *F* statistic for testing the null that the equation is weakly identified is 304.06, which is rejected based on the Stock–Yogo weak ID test critical values.

[26] We present the second stage estimations only for the 50th percentile. Complete estimations are provided in the online appendix and show

Table 7 Estimations Results of the IV-2SLS (Second Stage Reports Only 50th Percentile)

	Linear first stage		Probit first stage	
	1st stage: *PredictionInfo*	2nd stage: Δ*FarePct*	1st stage: *PredictionInfo*	2nd stage: Δ*FarePct*
PredictionInfoFit		−0.188***		−0.163***
		(0.0387)		(0.027)
AvgPop (M)	0.0065**	0.00358**	0.0353***	0.00470***
	(0.00259)	(0.00179)	(0.01078)	(0.00169)
MarketSeats (K)	0.0013***	−0.000323**	0.0067***	−0.000329***
	(0.00014)	(0.000128)	(0.00073)	(9.2E−05)
LCC	−0.0273	−0.107***	−0.0608	−0.117***
	(0.02060)	(0.0139)	(0.08004)	(0.0111)
Distance (K)	0.1708***	0.0132	0.6380***	0.0133
	(0.01264)	(0.0129)	(0.05377)	(0.00957)
HHI	−0.2144***	−0.0162	−0.9172***	−0.0466**
	(0.02837)	(0.0222)	(0.11610)	(0.0199)
AvgAirportPax (M)	0.0065***	0.00125	0.0145**	−0.00083
	(0.00149)	(0.00106)	(0.00669)	(0.00087)
AvgIncPop (K)	0.0102***	0.00817***	0.0381***	0.00494***
	(0.00118)	(0.000909)	(0.00493)	(0.00049)
AvgRegionAPRank	−2.0977***		−10.151***	
	(0.12030)		(0.67071)	
Constant	0.0974	−0.178***	−1.146***	
	(0.06506)	(0.0427)	(0.27831)	
N	3,399	3,399	3,399	3,399
RMSE		0.258		0.257
Centered R^2/Pseudo R^2	0.4135		0.3966	
F Statistic/Chi^2	298.7		1,855.81	

Notes. Standard errors are in parentheses. RMSE, Root mean square error. (K) and (M) indicate the variables are measured in thousands and in millions of units, respectively.
$^{**}p < 0.05$; $^{***}p < 0.01$.

our previous results, after controlling for a possible selection bias, the introduction of route prediction information exhibits a significant negative effect on transacted fares. The coefficients of *PredictionInfoFit* are larger than those reported in the earlier estimations (i.e., local treatment effects are larger than the average treatment effects). A possible explanation is that Farecast may have favored entering markets where prediction information affects prices the most.[27]

5.4. Evidence for Enhanced Strategic Consumer Behavior

Below we explore how strategic consumer behavior is manifested in different purchase patterns such as increased acquisition of fare classes associated with later purchases (§5.4.1), purchase balancing across carriers (§5.4.2), and increased consumer price sensitivity (§5.4.3).

5.4.1. Evidence for Delayed Purchases. As noted in §2, because of the general price increase of tickets as the travel date approaches, when consumers lack relevant price fluctuation information, they may purchase shortly after the need for travel is finalized. Such behavior may be supported by consumer myopia that leads to suboptimal purchase timing. Fare prediction information may result with more consumers who are willing to delay their purchase in expectation of monetary savings. In this section, we test whether the introduction of fare prediction information leads to an increased portion of delayed purchases. For this purpose we exploit the DB1B data set classification of the transacted coach tickets into restricted and unrestricted fare classes. Eligibility for restricted fare classes is associated with different requirements, one of which is advance purchase. Hence, earlier purchases are more likely to be associated with restricted fare classes. Accordingly, because fare prediction information may encourage consumers to delay their purchase, we hypothesize that its introduction will result with an increase in the portion of transacted unrestricted tickets.

We define the portion of transacted unrestricted fare classes as $URRatio_t^i = T_t^{i,U}/(T_t^{i,U} + T_t^{i,R})$, where $T_t^{i,j}$ captures the number of restricted ($j = R$) or unrestricted ($j = U$) seats sold in market i during period t, and we let $\Delta URRatio_t^i = URRatio_t^i - URRatio_{t-2}^i$, where $t \in \{2010, 2008\}$.[28]

that the coefficient of *PredictionInfoFit* is negative and significant to different extents at all percentiles except for the 10th percentile under the linear first stage.

[27] We are grateful to the anonymous referee who proposed this explanation and for other useful suggestions on IV-2SLS.

[28] Since the notion of fare classes is generally limited to legacy carriers, our analysis includes transactions made with these carriers only.

Table 8 Fare Class Purchase Ratio: Difference of Means ($\Delta URRatio_t^i$)

Year	Group	N	Mean	SE	t-statistic	p-value
2006–2008	Prediction introduced	948	−0.010	0.001	2.89***	0.0040
	Control	296	−0.021	0.004		
2008–2010	Prediction introduced	121	−0.002	0.003	3.48***	0.0005
	Control	566	−0.013	0.001		

***$p < 0.01$.

Table 8 shows the distribution of $\Delta URRatio$ experienced on the routes with new prediction information (first row) and the control group (second row) for markets for which prediction information was not introduced between $t-2$ and t. Apparently the average proportion of unrestricted fare classes in all markets marginally decreases between 2006 and 2010, with a significantly lower decrease in routes for which fare prediction information was introduced compared to the control group. The results support the notion that with fare prediction information, some consumers switch from being myopic to being strategic and thus increasingly delay their purchase.

5.4.2. Balanced Fill Rates. Reopening of low fare classes is generally an outcome of low fill rates. This suggests that prediction information may guide strategic consumers to purchase from carriers that exhibit, or are expected to exhibit, lower fill rates. Because consumer may gravitate to flights with lower fill rates, the fill rates may eventually balance across carriers. We examine whether the dispersion of the final fill rates across competing carriers decreases as a result of the introduction of prediction information. Using difference-in-differences, we estimate the effect of prediction information on the standard deviation of legacy carriers' fill rates on a route, $\mathrm{SD}(FillRate_t^i)$.

Although the results (Table 9) show there may be a reduction in the standard deviation of fill rates following the introduction of fare prediction introduction between 2006 and 2008, we cannot reject the null hypothesis (of no change in fill rate balancing across competing carriers on a route) with sufficient confidence. A possible explanation for this result is that a reduction in fares by one carrier triggers similar fare reductions by competitors, thus exhibiting minor temporal price differences across carriers.

5.4.3. Evidence for Increased Price Sensitivity. Since provision of prediction information reduces cognitive effort in considering purchase timing, strategic consumers may enhance their search for lower fares also in other aspects. Accordingly, in this section we examine whether there is evidence for consumers' improved selection processes beyond timing. Specifically, we examine whether there are increased purchases

Table 9 Impact of *PredictionInfo* on Fill Rate Dispersion ($\mathrm{SD}(FillRate_t^i)$): Difference-in-Differences Regressions

SD(*FillRate*)	2006–2008	2008–2010
PredictionInfoNew	0.013	0.013
	(0.008)	(0.012)
PredictionInfoMature		0.007
		(0.012)
Year dummy	0.008	0.021
	(0.007)	(0.014)
Year × *PredictionInfoNew*	−0.0163*	−0.007
	(0.009)	(0.017)
Year × *PredictionInfoMature*		−0.019
		(0.016)
N	2,197	2,190
Adj. R^2	0.078	0.062

Notes. Standard errors are in parentheses. All regressions include *AvgPop*, *AvgIncPop*, *Distance*, *LCC*, *MarketSeats*, *HHI*, and *AvgAirportPax*.
*$p < 0.10$.

of tickets that are typically marketed at lower prices—the tickets that are marketed by the operating carrier (Chen and Gayle 2007, Gilo and Simonelli 2015).[29]

The estimation results (Table 10) provide no support to the notion that the reduction of cognitive effort in decision timing results in increased purchases of tickets that are typically marketed at lower fares.[30]

6. Concluding Remarks

This research provides empirical evidence for the significant economic impact stemming from the availability of airfare prediction information that can support strategic consumer behavior. Strategic, or forward looking,

[29] This information is available in DB1B through code-sharing indications. Code sharing is a common practice in the airline industry, under which marketing airlines sell tickets for flights operated by other carriers. Although code sharing is primarily used in interlining flights to improve a carriers' connectivity, it is commonly used for nonstop flights as well for marketing considerations only. In those flights the operating carrier typically offers lower fares due to double marginalization when purchasing from other carriers (Gilo and Simonelli 2015).

[30] In our regression, we include routes for which there was a mixture of tickets sold by the operating carrier and tickets sold by other marketing carriers. The predicted variable is the portion of nonstop tickets sold by legacy carriers for flights they operate on a route out of all nonstop tickets sold by legacy carriers on the route.

Table 10 Impact of *PredictionInfo* on Proportion of Seats Marketed by Operating Carrier: Difference-in-Differences Regressions

Number of direct transactions	2006–2008	2008–2010
PredictionInfoNew	0.1759***	−0.1720***
	(0.0145)	(0.0218)
PredictionInfoMature		0.0225
		(0.0227)
Year dummy	−0.0723***	−0.0413
	(0.0126)	(0.0298)
Year × *PredictionInfoNew*	0.0069	0.0243
	(0.0178)	(0.0325)
Year × *PredictionInfoMature*		−0.0265
		(0.0321)
N	4,129	4,198
Adj. R^2	0.47	0.51

Notes. Standard errors are in parentheses. All regressions include *AvgPop*, *AvgIncPop*, *Distance*, *LCC*, *MarketSeats*, *HHI*, and *AvgAirportPax*.
$^{***}p < 0.01$.

behavior is a characteristic of consumers who consider future possible price markdowns when timing their purchase. Such consumer behavior has increasingly been incorporated into dynamic pricing and revenue management models. Yet, empirical evidence for the effects of this behavior has thus far been limited.

An underlying assumption of models that incorporate strategic consumer behavior is that these consumers are able to solve the same model faced by the firms and are able to decipher the path of prices that will materialize in the future. In the context of the airline industry, this assumption raises concerns. Airfares are subject to complex algorithms that frequently change prices, and the amount of information available to consumers is rather limited. In this respect, much of the literature has ignored the fact that strategic consumer behavior can be facilitated, rather than assumed; namely, in the absence of information and a limited capacity to solve the pricing problem, consumers' decisions on whether to wait or buy are essentially a shot in the dark. However, with the availability of proper decision support, consumers can make a conscious and knowledgeable decision on whether to delay their purchase. Thus, the presence of online decision tools has a critical role in enabling strategic consumer behavior.

Our empirical analysis shows that enabling consumers to behave strategically significantly affects transacted prices. The results support the idea that although consumers may desire to behave strategically they fail to do so without the supporting information. The effect appears to shift surplus from the carriers to the consumers, and we have estimated the magnitude of this effect to be about 4%–6%. Evidence provided in this research shows that not long after enabling strategic consumption, the price drop takes place, and that surplus does not seem to shift back to the seller.

How does this impact airlines' performance? Consider, for example, Alaska Airlines, which generated about 80% of its revenue from passengers in 2008. If fare prediction information were available in all of Alaska's markets, one would expect an impact of up to 3.2% in total revenue. In 2008, Alaska Airlines had an operating loss of about \$130 million, which could have been almost entirely (roughly \$100 million) recouped had the fares been higher.

Our research could be useful for airline managers as the work provides figures that help understand the costs and benefits of strategic consumer behavior. The literature has proposed different approaches to help counter and mitigate strategic consumer behavior including, for example, rationing capacity (Liu and van Ryzin 2008), making price commitments (Su and Zhang 2008), employing opaque distribution channels (Jerath et al. 2010), offering price-matching guarantees, or limiting inventory related information (see review by Aviv et al. 2009).[31] Implementing any of these practices entails some costs to the company; this research helps better assess the potential benefits.

Our work contributes to the growing literature on strategic consumers by quantifying the importance of decision support information in facilitating such a behavior. One venue for future work could be associated with such decision tools as new ones emerge.[32] It would be interesting to study the evolution of different tools and the effects of their unique attributes. Other venues for future research could explore dynamic pricing patterns and tools in other industries (e.g., fashion, technology) or even in different segments of the travel industry (e.g., car rental and hotel booking).

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mksc.2015.0965.

Acknowledgments

The authors thank the review team for constructive feedback that improved this paper significantly. The authors also thank Brian Cozzarin, Amnon Gonen, Martin Dresner, Arik Sadeh, and Tuba Toru for their comments on earlier versions of this manuscript. This research was partially supported by

[31] Some other advanced methods are becoming available to firms in the context of the airline industry. For example, Farelogix provides airlines with the distribution of airfares to facilitate improved intertemporal airfare planning.

[32] For example, in 2013, a prediction tool similar to the one discussed in this research was adopted by Kayak.com. There seems to be, however, one major difference between the tool provided by Farecast and the one provided by Kayak. Whereas Farecast relied on active queries of airfare data, Kayak relies on consumer-generated queries. Hence, Kayak may not be able to provide predictions if consumers do not query certain markets. Thus, Kayak's tool may be effective only for thick markets.

a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant.

References

Adomavicius G, Bockstedt J, Curley S, Zhang J (2011) Recommender systems, consumer preferences, and anchoring effects. *RecSys 2011 Workshop on Human Decision Making in Recommender Systems, Chicago*, 35–42.

Alderighi M, Gaggero AA, Piga CA (2015) The effect of code-share agreements on the temporal profile of airline fares. *Transportation Res. Part A: Policy Practice* 79:42–54.

Anderson CK, Wilson JG (2003) Wait or buy? The strategic consumer: Pricing and profit implications. *J. Oper. Res. Soc.* 54(3):299–306.

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

Aviv Y, Levin Y, Nediak M (2009) Counteracting strategic consumer behavior in dynamic pricing systems. Netessine S, Tang CS, eds. *Consumer-Driven Demand and Operations Management Models* (Springer, New York), 323–352.

Belobaba PP (1989) Application of a probabilistic decision model to airline seat inventory control. *Oper. Res.* 37(2):183–197.

Belobaba PP, Odoni A, Barnhart C (2009) Fundamentals of pricing and revenue management. Belobaba PP, Odoni A, Barnhart C, eds. *The Global Airline Industry* (John Wiley and Sons, Chichester, UK), 73–112.

Besanko D, Winston WL (1990) Optimal price skimming by a monopolist facing rational consumers. *Management Sci.* 36(5):555–567.

Boyd AE (2007) *The Future of Pricing* (Palgrave Macmillan, New York).

Brueckner JK, Luo D (2014) Measuring strategic firm interaction in product-quality choices: The case of airline flight frequency. *Econom. Transportation* 3(1):102–115.

Brueckner JK, Lee DN, Singer ES (2011) Alliances, codesharing, antitrust immunity, and international airfares: Do previous patterns persist? *J. Competition Law Econom.* 7(3):573–602.

Brueckner J, Lee DN, Singer ES (2013) Airline competition and domestic U.S. airfares: A comprehensive reappraisal. *Econom. Transportation* 2(1):1–17.

Brynjolfsson E, Hu Y, Simester D (2011) Goodbye Pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Sci.* 57(8):1373–1386.

Chen L, de Gemmis M, Felfernig A, Lops P, Ricci F, Semeraro G (2013) Human decision making and recommender systems. *ACM Trans. Interactive Intelligent Systems* 3(3):1–17.

Chen M, Chen ZL (2015) Recent developments in dynamic pricing research: Multiple products, competition, and limited demand information. *Production Oper. Management* 24(5):704–731.

Chen Y, Gayle PG (2007) Vertical contracting between airlines: An equilibrium analysis of codeshare alliances. *Internat. J. Indust. Organ.* 25(5):1046–1060.

Clemons EK, Hann IH, Hitt L (2002) Price dispersion and differentiation in online travel: An empirical investigation. *Management Sci.* 48(4):534–549.

Coase RH (1972) Durability and monopoly. *J. Law Econom.* 15(1):143–149.

Cook J (2014) Farewell, Farecast: Microsoft kills airfare price predictor, to the dismay of its creator. *GeekWire* (April 8), http://www.geekwire.com/2014/farewell-farecast-microsoft-kills-airfare-price-predictor-dismay-creator/.

Cooper WL, Homem-de-Mello T, Kleywegt AJ (2006) Models of the spiral-down effect in revenue management. *Oper. Res.* 54(5):968–987.

DellaVigna S, Malmendier U (2006) Paying not to go to the gym. *Amer. Econom. Rev.* 96(3):694–719.

Dresner M, Tretheway MW (1992) Modelling and testing the effect of market structure on price: The case of international air transport. *J. Transport Econom. Policy* 26(2):171–184.

Ellison G, Ellison F (2009) Search obfuscation, and price elasticities on the Internet. *Econometrica* 77(2):427–452.

Escobari D (2012) Dynamic pricing, advance sales and aggregate demand learning in airlines. *J. Indust. Econom.* 60(4):697–724.

Etzioni O, Tuchinda R, Knoblock CA, Yates A (2003) To buy or not to buy: Mining airfare data to minimize ticket purchase price. *Proc. Knowledge Discovery Data Mining* (ACM, New York), 119–128.

Fitzsimons GJ, Lehmann DR (2004) Reactance to recommendations: When unsolicited advice yields contrary responses. *Marketing Sci.* 23(1):82–94.

Flynn LJ (2006) Like this? You'll hate that. (Not all Web recommendations are welcome.) *New York Times* (January 23), http://www.nytimes.com/2006/01/23/technology/23recommend.html.

Gerardi KS, Shapiro AH (2009) Does competition reduce price discrimination? New evidence from the airline industry. *J. Political Econom.* 117(1):1–37.

Gilo D, Simonelli F (2015) The price-increasing effects of domestic code-sharing agreements for non-stop airline routes. *J. Competition Law Econom.* 11(1):69–83.

Goldman D (2011) DOJ approves Google's ITA deal . . . with caveats. *CNN Money* (April 8), http://money.cnn.com/2011/04/08/technology/google_ita/index.htm.

Goolsbee A, Syverson C (2008) How do incumbents respond to the threat of entry? Evidence from the major airlines. *Quart. J. Econom.* 123(4):1611–1633.

Grubb MD (2015a) Behavior consumers in industrial organization. *Rev. Indust. Organ.* 47(3):247–258.

Grubb MD (2015b) Failing to choose the best price: Theory, evidence, and policy. *Rev. Indust. Organ.* 47(3):303–340.

Ho TH, Lim N, Camerer CF (2006) Modeling the psychology of consumer and firm behavior with behavioral economics. *J. Marketing Res.* 43(3):307–331.

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

Jiang Y, Shang J, Kemerer CF, Liu Y (2011) Optimizing e-tailer profits and customer savings: Pricing multistage customized online bundles. *Marketing Sci.* 30(4):737–752.

Kwoka J, Shumilkina E (2010) The price effect of eliminating potential competition: Evidence from an airline merger. *J. Indust. Econom.* 58(4):767–793.

Lederman M (2007) Do enhancements to loyalty programs affect demand? The impact of international frequent flyer partnerships on domestic airline demand. *RAND J. Econom.* 38(4):1134–1158.

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

Li J, Granados N, Netessine S (2014) Are consumers strategic? Structural estimation from the air-travel industry. *Management Sci.* 60(9):2114–2137.

Liu Q, van Ryzin GJ (2008) Strategic capacity rationing to induce early purchases. *Management Sci.* 54(6):1115–1131.

Mandl M, Felfernig A, Teppan E, Schubert M (2011) Consumer decision making in knowledge-based recommendation. *J. Intelligent Inform. Systems* 37(1):1–22.

McAfee RP, te Velde V (2007) Dynamic pricing in the airline industry. Hendershott TJ, ed. *Handbook of Economics and Information Systems*, Vol. 1 (Elsevier, New York), 527–561.

Nair H (2007) Intertemporal price discrimination with forward-looking consumers: Application to the US market for console video-games. *Quant. Marketing Econom.* 5(3):239–292.

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

Phillips RL (2005) *Pricing and Revenue Optimization* (Stanford University Press, Stanford, CA).

PR Newswire (2007) Farecast launches innovative new fare guard service. (January 22), http://www.prnewswire.com/news-releases/farecastlaunches-innovative-new-fare-guard-service-53660022.html.

Rubin E, Mantin B (2012) Tabulated decision aids and airfare pricing. *Electronic Commerce Res. Appl.* 11(2):159–170.

Schonfeld E (2007) Click here for the upsell. *CNNMoney.com* (July 1), http://money.cnn.com/magazines/business2/business2_archive/2007/07/01/100117056/.

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

Shugan SM (2006) Editorial–Are consumers rational? Experimental evidence? *Marketing Sci.* 25(1):1–7.

Soysal GP, Krishnamurthi L (2012) Demand dynamics in the seasonal goods industry: An empirical analysis. *Marketing Sci.* 31(2):293–316.

Stokey NL (1979) Intertemporal price discrimination. *Quart. J. Econom.* 93(3):355–371.

Stokey NL (1981) Rational expectations and durable goods pricing. *Bell J. Econom.* 12(1):112–128.

Su X, Zhang F (2008) Strategic consumer behavior, commitment, and supply chain performance. *Management Sci.* 54(10):1759–1713.

Talluri KT, van Ryzin GJ (2004) *The Theory and Practice of Revenue Management*, Internat. Series Oper. Res. Management Sci., Vol. 68 (Springer Science+Business Media, New York).

Xie J, Shugan SM (2001) Electronic tickets, smart cards, and online prepayments: When and how to advance sell. *Marketing Sci.* 20(3):219–243.