

# Stochastic peak-load pricing with real-time demand learning in the U.S. airline industry\*

Diego Escobari<sup>†</sup>

October 4, 2007

## Abstract

Dynamic pricing and demand learning play an important role in airlines because (1) in low demand flights unsold tickets are of little value after departure, and (2) in high demand flights carriers may forego important profits if the flight sells out and some relatively high willingness-to-pay consumers have to be rationed. Under a price sensitive demand, *stochastic* peak-load pricing suggests that at any point prior departure airlines should set higher fares in expected *peak* flights, where demand is more likely to exceed capacity. Moreover, in order to promote sales, lower fares should be set in expected *off-peak* flights. Using a unique panel of U.S. airlines fares and seat inventories, this paper shows that airlines learn about the demand as sales progress and the flight date approaches. Forecasted values of occupancy rates and sold out probabilities are employed to calibrate an *ex-ante* –before sales begin– distribution of demand uncertainty. Nonparametric techniques are then used to construct a latent variable to identify different expected demand states at different points prior departure. This latent variable is utilized to dictate the regime shift in a panel endogenous threshold model. Consistent with the *stochastic* peak-load pricing predictions, the results show that higher fares are set in the *peak* regime, while lower fares in the *off-peak* regime. The results proved to be robust to an alternative specification of a GMM dynamic panel, were the assumption of strict exogeneity is relaxed. This is the first paper to provide formal evidence of *stochastic* peak-load pricing in airlines or to show that airlines learn about the demand and respond to early sales.

*Keywords:* Peak-load pricing, Demand Learning, Airlines

*JEL Classifications:* C23, L93

---

\*The author thanks Li Gan for helpful discussions. The paper also benefited from comments by Eric French, Manuel Hernandez, and Thomas R. Saving. Stephanie Reynolds provided excellent research assistance. Financial support from the Private Enterprise Research Center at Texas A&M and the Bradley Foundation is gratefully appreciated.

<sup>†</sup>Department of Economics, Texas A&M University, College Station, TX 77843, Phone: (979) 845-4344, Fax: (979) 845-6636, Email: [escobari@tamu.edu](mailto:escobari@tamu.edu), URL: <http://econweb.tamu.edu/descobari>

# 1 Introduction

The term dynamic pricing, most commonly known as ‘yield management,’ is used to describe pricing and seat-inventory control decisions. It is important in industries that deal with perishable products such as airlines, where unsold seats perish when the flight leaves the gate. Similar examples involve hotel rooms, fashion apparel, cabins on cruise liners, car rentals, entertainment and sporting events, and restaurants. In all these cases the seller can improve its revenues by dynamically adjusting the price of the product rather than committing to a price schedule or a unique price throughout the selling period. Demand uncertainty plays an important role in airline’s dynamic pricing because tickets are sold in advance with prices being set when carriers have limited information about the total number of potential consumers. Moreover, capacity is set in advance and can only be modified at a relatively high marginal cost. When demand is relatively low, unsold tickets are of little value for the carrier. Likewise, when demand is relatively high the airline may give up important profits if some consumers with a relatively high willingness-to-pay have to be rationed out. This may be the case of a business traveler who has a very high willingness-to-pay for a ticket but arrives when there are no tickets left.

Learning about a price sensitive demand as sales progress and the flight date nears is crucial for airlines to price accordingly. The shadow cost of capacity for the seats on a flight will be different at different points in time prior departure depending on the expected demand. When the probability that demand will exceed capacity is large, the shadow cost of capacity is large. Peak-load or congestion pricing, defined as the practice of charging higher prices during peak periods when capacity constraints cause marginal costs to be high, is the pricing strategy that takes into account this shadow cost. Borenstein and Rose (1994) provide a clear distinction between two types of peak-load pricing in airlines. *Systematic* peak-load pricing reflects variations in the expected shadow cost of capacity at the time the flight is scheduled and before any ticket is sold, while *stochastic* peak-load pricing that reflects uncertainty about individual flights that is resolved as the flight date nears and tickets are sold. In this paper we control for *systematic* peak-load pricing and analyze the impact of demand learning and *stochastic* peak-load pricing on fares. If at the moment the ticket is sold carriers expect to have a *peak* flight, they will charge higher fares. Moreover, expected *off-peak* flights will be associated with lower fares.

Despite the large theoretical literature on airlines’ pricing in economics, marketing and operational research journals, there is few empirical under-

standing on how carriers are actually setting their fares and the dynamics that govern their evolution as the flight date nears. This is the first empirical paper that evaluates the very intuitive predictions of *stochastic* peak-load pricing in airlines and to test whether airlines can reduce the cost of demand uncertainty by responding to early sales. One of the reasons why this hasn't been done before is the lack of adequate data.<sup>1</sup> While most of the empirical research in airlines uses the *Bureau of Transportation Statistics' DB1B*, which is a 10% random sample of tickets, recent research has begun analyzing more detailed data that allows tracking day by day decisions by airlines. Stavins (2001), and more recently Chen (2006), and Bachis and Piga (2007) among others look at offered fares by airlines. However, none of these papers has information on inventories of seats for each offered fare. This paper takes advantage of a unique U.S. airline's panel disaggregated at the ticket level that contains the evolution of offered fares and seat inventories over a period of 103 days for 228 domestic flights that departed on June 22<sup>nd</sup>, 2006. The data collection resembles experimental data which perfectly controls for product heterogeneities and 'fences' that segment consumers. This is key, since many price discrimination tools that define ticket characteristics (e.g. saturday-night-stayover) are also used for *stochastic* peak-load pricing to reduce demand uncertainty.<sup>2</sup>

To test whether airlines learn about the demand and implement a *stochastic* peak-load pricing strategy, the empirical section initially obtains the optimal price schedule under *no* demand learning by calibrating the *ex-ante* distribution of demand states. This is done using information on sold-out probabilities and forecasted values of occupancy rates. The sold-out probabilities are calculated using a second dataset from *Expedia.com*, and the forecasted occupancy rates are obtained using time-series data on occupancy rates from the *T-100* of the *Bureau of Transportation Statistics*. Under the Prescott (1975)'s type of models, this distribution should give us the optimal price schedule, which holds through the entire selling horizon as long as airlines do not learn about the demand or if price commitments hold. The basic idea in the testing is to analyze whether airlines deviate from this *ex-ante* optimal price schedule as information about the demand is revealed. To capture the information about the demand that is revealed as sales progress, the paper uses the techniques described in Racine and Li (2004) and estimates

---

<sup>1</sup>Another could be the challenge of coming up with an adequate empirical test.

<sup>2</sup>A round-trip ticket that involves *peak* flights may not benefit from a saturday-night-stayover discount even if the stay has a Saturday night. Moreover, if it involves *off-peak* flights, then the discount will have an *stochastic* peak-load pricing component and a price discrimination component.

a nonparametric model using both categorical and continuous data based on flight- and route-level information to explain the evolution of sales. The nonparametric results are used to construct a latent variable that can distinguish between different expected demand states at different points prior departure for every flight in the sample. This latent variable is later utilized to control the regime shift in the estimation of a panel endogenous threshold model as described in Hansen (1999) and Hansen (2000).

The results are consistent with the *stochastic* peak-load pricing predictions. The panel endogenous threshold estimates found the existence of two pricing regimes. In expected *peak* flights, where demand is more likely to exceed capacity and the shadow cost of a seat is high, airlines set higher fares. For expected *off-peak* flights where demand is expected to be low, the shadow cost of capacity will also be low. Hence, airlines will set lower fares to incentivate sales. To control for the potential interaction between previous price levels and cumulative sales, the paper also estimates a dynamic panel, where the assumption of strict exogeneity of the regressors is relaxed. The GMM dynamic panel results from the *difference* estimator, as explained in Holtz-Eakin et al. (1988) and Arellano and Bond (1991), and the *system* estimator, as described Arellano and Bover (1995) and Blundell and Bond (1998), were found to be consistent with the two pricing regimes and the *stochastic* peak-load pricing predictions found before. Based on the *system* GMM estimates, evaluated at the sample average fare of 291.09 dollars and for a 100 seats airplane, selling one more seat increases fares by 38.1 cents in an expected *off-peak* flight while increases fares by 58.5 cents in a expected *peak* flight.

By testing for *stochastic* peak-load pricing and demand learning, this paper explains an important source of price dispersion as well. Borenstein and Rose (1994) calculated that the expected absolute difference in fares between two passengers on a route is 36% of the airline's average ticket price. One cost based source of this price dispersion is *stochastic* peak-load pricing. Even though the figures found in this paper are not directly comparable to this 36%, we find that an increase of one standard deviation in capacity utilization increases *peak* fares by 6.5% *within* flight standard deviations more than *off-peak* fares. This estimate is after controlling for *systematic* peak load pricing, unobserved flight and route characteristics, and 'fences' that restrict consumers that are commonly used as price discrimination tools.

By focusing on the role of the evolution of inventories on dynamic pricing, this paper was able to identify three components in the evolution of fares as the flight date nears. First, the *stochastic* peak load pricing component as the difference in fares between *peak* and *off-peak* flights. Second, the

effect of demand uncertainty and costly capacity on fares as explained in the theoretical works by Prescott (1975), Eden (1990), Dana (1999b), and more recently by Deneckere and Peck (2005). The importance of this effect to explain price dispersion has been previously documented empirically in Escobari and Gan (2007). Finally, the third component, advance purchase discounts. This last component is consistent with the price discrimination argument in Dana (1998), where the existence of second degree price discrimination takes the form of advance purchase discounts. Moreover, it is also consistent with the existence of advance purchase discounts under an uncertain peak demand period in Gale and Holmes (1992) and advance purchase discounts with perfectly predictable peak demand times in Gale and Holmes (1993). It is important to mention that both of the works by Gale and Holmes do not consider the shadow cost of capacity and there is no cost-based price variation. The predicted price dispersion suggests price discrimination.

The organization of the paper is as follows. Section 2 presents a model of pricing under demand uncertainty that extends to the empirical testing. The data is explained in section 3. The empirical model is presented in section 4. Finally, section 5 concludes the paper.

## 2 Airline Pricing under Demand Uncertainty

Airline pricing has three basic characteristics that make its study fascinating. First, capacity is fixed and can only be augmented at a relatively high marginal cost. It is unlikely for carriers to change the size of the aircraft once they have already started selling tickets. Doing so would involve a large rescheduling of the fleet and airport slots. Second, air tickets expire at a point in time; once the plane departs carriers can no longer sell tickets. Tickets that haven't been sold by then have little value to the carrier. On the other hand, the carrier may still want to reserve a certain number of seats if it expects to have last-minute travelers who are willing to pay substantially higher fares (see Lin (2006)). Finally, the third characteristic, there is uncertainty in the demand. This becomes crucial because airlines sell in advance and fares have to be set when carriers have limited information about the total number of potential passengers that will show up to get a ticket. Under this basic scenario, it is key for carriers to learn about the final state of demand as tickets are sold and the departure date nears. If information about a price sensitive demand becomes available, airlines will want to adjust their prices accordingly to maximize profits. These

characteristics, common to various industries originated a large amount of theoretical literature on optimal pricing of a perishable non-renewable asset with stochastic demand. However, there is still few empirical understanding about how actual prices are set. This is the first paper that provides evidence of *stochastic* peak-load pricing in airlines. Moreover, it is also the first to presents empirical evidence that shows that airlines learn about the demand as the departure date nears.

Airline pricing, nevertheless, is much more sophisticated than dealing with an an inventory of seats that expire at a point in time. Usually air tickets involve complex itineraries and carriers exploit ‘fences’ such as saturday-night-stayover requirement, minimum- and maximum-stay, nonrefundable purchases, frequent flyer miles, blackouts, days in advance requirements, or volume discounts to segment consumers. The nature of the dataset used in this paper, that resembles a quasi-experiment, and the econometric techniques employed, control for all these fencing devices and complex itineraries to allow us focusing on the pricing of an inventory of seats that expire at departure. Therefore, this overview of airline pricing under demand uncertainty discusses just this case. We begin with the case where there is *no* demand learning and capacity is costly. We then explain the implications for pricing decisions when airlines learn about the demand through the information contained in early sales. At the end of the section we discuss about the implications of two cost-based sources of price dispersion for airlines: *systematic* and *stochastic* peak-load pricing.

## 2.1 Pricing *without* Demand Learning

The simplest model that explains dispersed prices for a homogeneous good under costly capacity and demand uncertainty is Prescott (1975). He considers a model of hotel rooms where prices are set *ex-ante* –before the total number of buyers is known–. Motivated with the airlines’ problem, Prescott (1975)’s model assumes that there is a stochastic demand  $n$  for homogeneous airline seats with a probability distribution function  $F(n)$ . Consumers are identical and purchase only one seat if the price is lower than a reservation value  $\bar{p}$ . The equilibrium prices will be dispersed with  $H(p)$  being the equilibrium number of seats priced at  $p$  or below. Travelers observe all prices and buy the less expensive unit available. In equilibrium, a seat priced  $p$  will be vacant with probability  $F[H(p)]$ . Let  $\lambda$  be the unit cost of capacity incurred on *all* units, whether these units are sold or not. In a perfectly competitive market, the zero expected profit condition implies that expected revenue should be equal to the unit cost of capacity,  $[1 - F[H(p)]] \cdot p = \lambda$ . This last

equation can also be written as

$$p = \frac{\lambda}{1 - F[H(p)]} \equiv ECC \quad (1)$$

for all  $p \in [\lambda, \bar{p}]$ . Any price offered in equilibrium must be equal to the unit cost of capacity divided by the probability that a unit offered at that price will be sold. Dana (1999b) interprets this last term as the *effective cost of capacity* (*ECC*), which is the revenue the carrier must earn if the seat is sold in order to cover the unit capacity cost it incurs whether or not the seat is sold. The intuition from this result is simple. Consider the case where there are two equally likely demand states and the cost of holding a seat in the aircraft is \$100. If a given seat in the aircraft is only sold during the high demand state, the carrier has to charge a fare equal to  $\$100/0.5 = \$200$ , to compensate the times the seat is not sold during the low demand state.

The key implication from the Prescott (1975) model is that lower-priced units will be sold with higher probability and higher-priced unit with lower probability. Therefore, sellers face the trade-off between price and the probability of making a sell. Even though Equation 1 is constructed for a perfectly competitive market, Dana (1999b) derives an analog of Equation 1 for perfect competition, monopoly, and oligopoly. In all cases the key implication is the same, however, in noncompetitive markets, the effect of *ECC* on fares has to be adjusted by the size of the markup.

Prescott's model was later formalized by Eden (1990) in a setting where consumers arrive sequentially, observe all offers and after buying the cheapest available offer they leave the market. Eden derives an equilibrium that exhibits price dispersion even when sellers are allowed to change their prices during trade and have no monopoly power. It is important to realize that the absence of price commitments alone is not enough to generate *stochastic* peak-load pricing. Information about the final state of the demand has to be revealed to price accordingly. Prescott's "hotels" model, as pointed out by Eden (1990) and Lucas and Woodford (1993), has an interesting time-consistency property. Deneckere and Peck (2005) explain that even if firms could sequentially compete by choosing a price after each market transaction fares will still follow Equation 1. The requirement for fares to depart from the original price schedule predicted by Equation 1 is that some additional information about the final state of the demand is revealed as the flight date approaches.

## 2.2 Pricing *with* Demand Learning

If fares can adjust as information about the demand is revealed over time the predictions from the dynamic pricing literature are very intuitive. As explained in Lin (2006), when an airline sells seats for the same class, the fares offered will be different depending on the time to departure and the current seat inventory. The airline has incentives to promote sale when departure time is approaching and inventories are high. On the other hand, the airline may still want to reserve some seats if it expects to have some last-minute travelers willing to pay substantially higher prices. Lin (2006) presents a theoretical dynamic pricing model where customers arrive in accordance with a conditional Poisson process. Sellers learn about the final state of the demand as sales move forward and the optimal price adjusts dynamically to maximize expected total revenue. The results indicate that higher prices should be set when demand is expected to be larger. A similar result is found in Gallego and van Ryzin (1994) and Kincaid and Darling (1963), where at a given point in time prior departure optimal prices will be higher if the inventory is lower, signaling a higher demand state.

There is a key difference between the models presented in Kincaid and Darling (1963), Gallego and van Ryzin (1994), and Lin (2006), and the Prescott (1975)'s type of models. In the first ones all costs related to the production of the seat are sunk, so the value for the seller for an unsold item is zero. If demand is expected to be low, fares are allowed to drop, result that explains the 'last minute deals' or cheap fares that airlines offer in some flights in order to promote sales. On the other hand, Prescott (1975)'s type of models assume that capacity is costly; airlines have to be able to cover the unit cost of capacity adjusted by the probability of sale for each of the seats. Moreover, since there is no demand learning, fares will always increase as sales progress, prediction that can be easily seen from Equation 1. The simplification of *no* costly capacity in the first type of models is not realistic, at least for airlines pricing, while the existence of *no* demand learning in the second type seems also restrictive. However, as information about the demand becomes available, pricing accordingly gains significance while dealing with costly capacity loses attractiveness. This difference in predicted outcomes as information about the demand becomes available will let us identify the existence of demand learning.

The basic information carriers use to learn about the final state of the demand is *realized* demand up to a given point prior departure. This is basically how many seats have been sold up to a given point in time, which contains some information about the speed of selling tickets and can be



used to predict whether final demand may exceed capacity. Models that take into account *realized* demand in its pricing policies include Gallego and van Ryzin (1994), Chatwin (2000), and Lin (2006). The exact nature of how information about current sales is taken into account to forecast the final state of the demand is not necessarily important in this paper. Airlines may have very different ways to use information about early sales to adjust price later on, however, all these models have the same testable prediction. At a given point in time, lower inventory levels, signaling an expected higher demand, results in higher prices. Likewise, if time passes by and no seats have been sold, this is evidence of approaching a low demand state, hence, lower fares should be set. As will be pointed out below, this prediction from operational research literature is equivalent to the *stochastic* peak-load pricing prediction found in economics journals.

### 2.3 Peak-load Pricing in Airlines

Typical peak-load pricing models under certainty (*systematic* peak-load pricing, e.g. Boiteaux (1949), Steiner (1957), Hirshleifer (1958), Williamson (1966)) and peak-load pricing models under uncertainty (*stochastic* peak-load pricing, e.g Brown and Johnson (1969), Visscher (1973), Carlton (1977)) they all suggest charging higher prices during peak times and lower prices during off-peak times. However, as explained in McAfee and te Velde (2006) these models poorly suit airline pricing. One important reason is that they assume the existence of a spot market, where all consumers in the *peak* demand pay a higher price and all consumers in the *off-peak* demand pay a lower price. However, advance purchases and different expectations about the demand prior departure are an important ingredient in the pricing problem.

For the airlines, fluctuations in the demand can be broken down into two parts. The *deterministic* component that refers to fluctuations in the demand which are known to carriers before selling starts, and the *stochastic* component of the demand for a flight, that is orthogonal to all information carriers have at the time of scheduling. As explain in Borenstein and Rose (1994), this two components of demand give rise to two different types of peak-load pricing in airlines.

#### 2.3.1 *Systematic* Peak-load Pricing

Fluctuations in capacity utilization across flights and across

In order for carriers to follow a *systematic* peak-load pricing strategy,

they require prior knowledge of peak flights (or peak periods) when they create their flight schedule. Hence, they can restrict the number of lower priced seats in peak flights to divert demand from peak to off-peak flight in order to expand output. In their empirical study, Borenstein and Rose (1994) control for *systematic* peak-load pricing under the assumption that this one is correlated with the variability in airlines' fleet utilization rates and airports' operation rates. However, they are not able to measure the effect of this peak-load pricing of fares. Escobari (2006) provides empirical evidence of the existence of *systematic* peak-load pricing and shows that airlines set higher fares in *ex-ante* known congested periods. Moreover he estimates a congestion premia and provides support for the main empirical prediction in Gale and Holmes (1993), less discount seats on peak periods.

### 2.3.2 *Stochastic* Peak-load Pricing

Demand learning can exist for *systematic* peak-load pricing. Learning models such as Burnetas and Smith (2000) have the seller learning from repetition of identical experiments (same flight number through different days). *Stochastic* peak-load pricing requires learning throughout the sales horizon of a single event.

However, the typical dynamic pricing literature that comes from operational research journals somehow overlooks some concepts economists consider important. Explain the shadow cost of a seat This is how

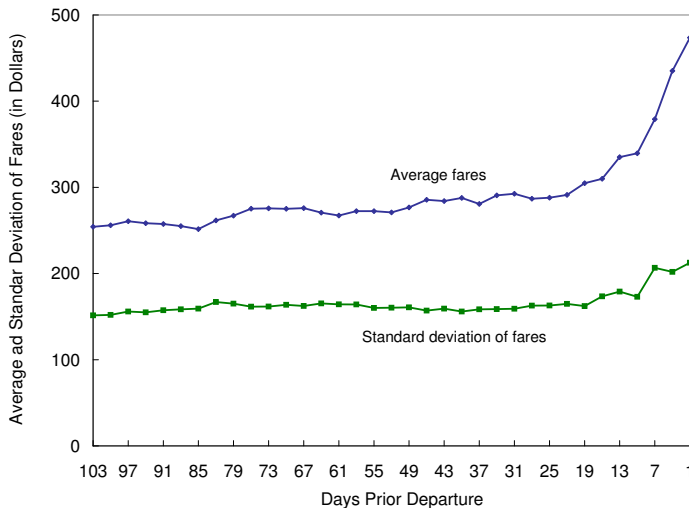
If carriers set their prices as demand is revealed over time, Crew and Kleindorfer (1986) explain that the optimal *stochastic* peak-load pricing strategy will depend on the probability at the time the ticket is sold that demand will exceed capacity and the expected shadow cost if this happens.

## 3 Data

The paper has two main sources of data, the Online Travel Agency (OTA) *Expedia.com*, used to build two datasets, and the *T-100* from the *Bureau of Transportation Statistics* used to construct a third dataset. The first dataset from *Expedia.com* is a panel with 228 cross-sectional observations over 35 periods making a total of 7980 observations. Each cross sectional observation is a specific carrier's non-stop flight in one of the 81 routes considered, where a route is a pair of departing and destination cities. The observations in time start 103 days prior departure and were gathered every three days up until one day prior departure, making the 35 observations in time. All flights depart the same date, Thursday June 22<sup>nd</sup>, 2006. The carriers considered

are American, Alaska, Continental, Delta, United, and US Airways. The number of flights per carrier was chosen to make sure that the share of each of these carriers is close to its share in the US airlines' market. This dataset has similar characteristics the one used in Stavins (2001) with two important differences. The data here is a panel and it has information about seat availability at each fare, where fare is the cheapest available economy class fare. The only two previous papers that work with such a detail information on prices and inventories are Escobari (2006) and Escobari and Gan (2007). The second dataset from *Expedia.com* was collected to obtain an estimate of the sold out probabilities for each of the 81 routes.

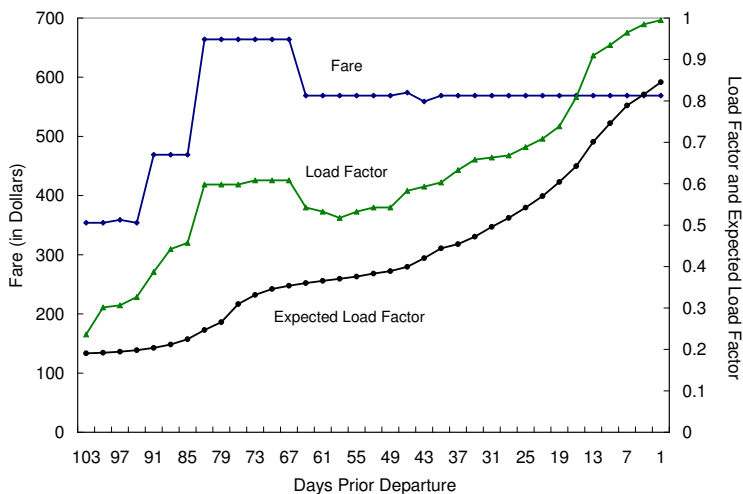
Figure 1: Average and Standard Deviation of Fares



The third dataset comes from the *T-100* obtained from the *Bureau of Transportation Statistics*. This is a panel containing average load factors at departure for the same 81 routes over the period 1990 to 2005. This dataset will be useful to estimate the expected number of tickets sold in each route, used to derive the *ex-ante* demand uncertainty.

A flight that illustrates the *stochastic* peak-load pricing we are testing in this paper is the one presented in Figure 2. This is flight Delta 1588, covering the 2111 miles between Atlanta, GA (ATL) and San Jose, CA (SJC) with a Boeing 737-800 that has a total capacity of 199 economy class seats, departs at 7:54 p.m. and arrives at 10:00 p.m. Figure 2 shows the evolution of fares, inventories of seats and the expected evolution of seat inventories for

Figure 2: Fares, Loads and Expected Loads (Delta 1588 ATL-SJC)



a period of 103 days prior departure. As required by Prescott (1975), Eden (1990), and Dana (1999b), fares represent the cheapest available fare for a given flight at each point in time prior departure. A detailed explanation of why the fares from *expedia.com* used in this paper are representative for the industry is presented in Escobari and Gan (2007) section 2.2. The evolution of inventories is best viewed as the ratio of available seats to total seats in the aircraft. Along this paper we refer to this ratio as the load factor, which is a ticket level load factor. The airline literature defines load factor only once the plane departed as the percentage of seats filled with paying passengers. Our load factor will go from zero when the plane is empty to one when it is full. In this Delta flight 1588, the load factor went from 0.235, 103 days prior departure to 0.995 one day prior departure. An interesting feature on this particular flight is that load factor is not necessarily increasing monotonically. The decrease between 67 and 61 days prior departure may be because some tickets have been reserved and never bought or maybe bought but cancelled later. The exact calculation of the expected load factor will be explained below. For now, it is just important to know that it is a measure of how the carrier, Delta, expects sales to evolve over time for this particular flight under normal conditions and price commitments. If at a given point in time the actual load factor is significantly above expected load factor, it is reasonable for the carrier to believe that demand will exceed

capacity and a *stochastic* peak-load pricing strategy would suggest charging higher fares. This is exactly what happened during the period between 94 and 67 days prior departure. Load factor was relatively high as compared to the expected evolution of load factor. At this consumers' arrival rate, demand would exceed capacity. Even if the arrival rate of future consumers is independent of this high arrival rate, demand would still exceed capacity if Delta keeps the same price schedule. The optimal *stochastic* peak-load pricing would be to set higher fares, which is exactly what they do. When load factor decreases between 67 and 61 days prior departure, fares also decrease. Moreover, notice that during the last month of sales, load factor increased even to higher levels, but this increase is just explained by the expected load factor, so there is no reason to charge higher fares at this point.

The evolution of sales in each flight in the sample is the result of tickets being bought across a huge number of potential alternatives, where the observed leg may be just part of a larger trip. What is important to realize is that the fare charged by the carrier is the carrier's response to the level of inventories and this one has its own dynamics. Here we are just making explicit what previous studies that work with non-transactions data implicitly assumed, e.g. Stavins (2001), Chen (2006), and McAfee and te Velde (2006). It is reasonable to believe that fares for more complicated itineraries vary accordingly with the one way fare. Bachis and Piga (2007) explain how some European carriers price all its legs independently, so there is no extra charge for one-way tickets. Actually, observing higher fares on one-way tickets is perfectly consistent with the predictions in Prescott (1975), Eden (1990), and Dana (1999b), where earlier purchasers benefit from lower fares. The idea is simple, a round-trip fare is the combination of two parts. Because the return date is further away from the purchase date, the second part is being bought with more days in advance than the first, with the presumably lower load factor. Hence, a round-trip ticket, measured as the summation of these two parts will be less expensive than just multiplying the first part of the ticket by two.

## 4 Empirical Model

The empirical model developed in this section is based on two testing procedures for *stochastic* peak-load pricing. The first one is just based on analyzing how the pricing of the next available ticket differs depending on the expectations of demand. The testing will see whether an expected peak

or an expected off-peak flight follow different pricing strategies. The second testing procedure is build on the models developed by Prescott (1975), Eden (1990), and Dana (1999b), where price dispersion exists in a setting with capacity constraints and demand uncertainty. The key feature in these models is that there is *no* demand learning as sales progress. Therefore, we start building the price schedule based on these models and then we test how pricing differs as carries learn about the demand. For the second testing procedure we derive the *ex-ante* distribution of demand uncertainty, which is before any information about actual sales is revealed. This will let us calculate the *effective cost of capacity*, which should have a positive impact on fares.

Common to both testing procedures, using nonparametric techniques we then develop a measure of the evolution of the expected number of seats sold. This measure is exogenous to the actual evolution of sales, so by comparing actual sales with expected sales at any point prior departure we can obtain information on the likelihood that demand will exceed capacity. An endogenous panel threshold model is then estimated to separate between expected peak and expected off-peak flight, allowing for different pricing strategies in different regimes. To control for potential enogeneities, the empirical section closes with the estimation of dynamic panels with a exogenous selection of the threshold.

#### 4.1 *Ex-ante* Distribution of Demand Uncertainty

The *ex-ante* distribution of demand uncertainty reffers to the distribution of arriving consumers known to the carrier before any ticket is sold. Based on this distribution, Prescott (1975) showed that the equilibrium prices will be dispersed. In this subsection we calibrate the *ex-ante* distribution of demand uncertainty. Under price commitments or if no information about the final state of the demand is revealed as tickets get sold, this *ex-ante* distribution of demand uncertainty should explain the observed price dispersion.

There exist uncertainty in the demand because carriers do not know *ex-ante* the total number of passangers that will buy tickets. Consider the case of having an infinite number of demand states. Let  $N_h$  be the number of consumers who arrive at demand state  $h$ , where  $h = 0, \dots, \infty$  and  $N_h \leq N_{h+1}$ . This last inequality imply that consumers who arrive at demand state  $h$  will also arrive at a higher-numbered demand state  $h + 1$ . Define a batch as the additional number of travelers who arrive at demand state  $h$  when compared to the inmediate lower demand state  $h - 1$ , therefore batch  $h$  is given by  $N_h - N_{h-1}$  with the first batch given by  $N_0$ .

Each demand state  $h$  occurs with probability  $\rho_h$ . Because all demand states have at least  $N_0$  travelers, the probability that  $N_0$  travelers arrive is  $Pr_0 = \int_0^\infty \rho_\kappa d\kappa = 1$ . In general, the probability that  $N_h$  travelers arrive is given by  $Pr_h = \int_h^\infty \rho_\kappa d\kappa$ , the summation of all demand states that have at least  $N_h$  consumers. Assume that each batch has one consumer buying a ticket, hence the probability of selling seat  $h$  is the summation all demand states that have at least  $h$  travelers buying a seat. Additionally, when demand states are normally distributed  $\rho_h = \phi_h$ , with  $\phi$  being the *pdf* of a normal distribution, the probability of selling seat  $h$  is given by:

$$Pr_h = \int_h^\infty \phi_\kappa d\kappa | q(p) = 1 - \Phi_h | q(p) \quad (2)$$

with  $q(p)$  being the distribution of prices and  $\Phi$  the *cdf* of a normal distribution.

This  $Pr_h$  corresponds to the  $F[h(p)]$  in Equation 1. To derive a measure of the *effective cost of capacity* and its impact on fares, we will calibrate the this distribution of demand uncertainty at a route level. To do this we follow Escobari and Gan (2007) and assume normally distributed demand states. The key feature that allows the calibration process is that demand states are censored when transformed to tickets sold. Once the aircraft is sold out, higher demand states are no longer observed. To get the values of the mean  $\mu$  and the standard deviation  $\sigma$ , at the route level, for the normally distributed demand states we first need two pieces of information, the sold-out probabilities and the expected number of tickets sold for each of the routes.

#### 4.1.1 Sold-out probabilities

The sold out probabilities for each of the 81 routes are obtained using the second dataset from *Expedia.com*. The fact that allows calculating these sold-out probabilities is that airlines and online travel agencies do not display their sold-out flights on their websites.<sup>3</sup> First, a couple of weeks in advance when no flight was expected to be sold-out yet, we made a census of all the available non-stop flights in each of the 81 routes during seven days between February 2<sup>nd</sup> and February 8<sup>th</sup>, 2007. The total number of flights

---

<sup>3</sup>The reason, according to Roman Blahoski, spokesman of Northwestern, is that they do not want to disappoint the travelers. Keeping the online display simple may also be a motive, and according to Dan Toporek, spokesman of *Travelocity.com*, “showing sold-out flights alongside available flights could be confusing.” Both of these quotes are from David Grossman, “Gone today, here tomorrow,” *USA Today*, August 2006.

were 5,881. Then, late the night before each of those seven days, we counted the number of flights still available at each route. If a flight was no longer there, it was assumed to be sold-out. The calculated sold out probability is just the ratio of sold out flights to total number of flights for each route.

#### 4.1.2 Expected number of seats sold

The expected number of seat sold are calculated using the *T-100* from the *Bureau of Transportation Statistics*. From the *T-100*, we obtain the average load factors at departure time for the 81 routes over the period 1990 to 2005. Each of these 81 series is used to estimate an ARMA model. Then, using a one-step forecast we obtain the expected number of seats sold for 2006. For routes where the expected number of seats sold is high, meaning that most of the seats are expected to be sold, the calibration procedure will assign higher probabilities to higher demand states. The details of the estimation are available upon request.

#### 4.1.3 Calibration

Let the underlying demand state  $h^*$  be distributed  $N(\mu, \sigma^2)$  and let  $m$  be the total number of seats in the aircraft. The number of seats sold  $h$  is equal to demand state  $h^*$  before the plane sells out,  $h = h^*$  if  $h < m$ , and equal to total number of seats in the aircraft,  $h = m$ , otherwise. The expected number of tickets sold is given by the first moment of the censored normal:

$$E(h) = Prob(h = m) \cdot E(h|h = m) + Prob(h < m) \cdot E(h|h < m) \\ \left(1 - \Phi\left(\frac{m - \mu}{\sigma}\right)\right) \cdot m + \Phi\left(\frac{m - \mu}{\sigma}\right) \cdot \left[1 - \sigma \frac{\phi\left(\frac{m - \mu}{\sigma}\right)}{\Phi\left(\frac{m - \mu}{\sigma}\right)}\right] \quad (3)$$

$E(h|h < m)$  comes from the mean of a truncated density and the *pdf* and *cdf* are evaluated at the moment the flight sells out. Therefore,  $\Phi\left(\frac{m - \mu}{\sigma}\right)$  is interpreted as the sold out probability. With information on the sold-out probabilities obtained in subsection 4.1.1 and the information on the expected number of tickets sold obtained in subsection 4.1.3, we use Equation 3 to obtain the values of  $\mu$  and  $\sigma$  at the route level.

## 4.2 Learning the Stochastic Demand

As carriers learn about the state of the demand they may want to depart from any price commitments to increase their profits. The way carriers use actual bookings to infer about the state of the demand can be complex



and may differ across carriers, but once some information is revealed, the outcome predicted by the *stochastic* peak-load pricing is simple. *Stochastic* peak-load pricing suggests charging higher fares in expected peak flights, while charging lower fares in expected off-peak flights. To test if this is true, the first step is to separate between expected peak and expected off-peak flights.

Under ‘normal’ conditions, let’s say, when a flight is not expected to be peak nor off-peak, sales should have a natural evolution over time as the flight date approaches. The rate at which tickets are sold need not be constant in time and may differ from route to route or across carries. If tickets are sold faster than the ‘normal’ rate and at a given point prior departure there are less seats left unsold than under ‘normal’ conditions, it would be reasonable for the carrier to believe that this is a peak flight. Clearly, this expected peak flight was not known to the carrier *ex-ante*, before the flight was opened for booking.

To test for the existence of demand learning with the corresponding *stochastic* peak-load pricing as the response to information about the final state of the demand we take the following steps. First, using nonparametric techniques we came with a measure of the evolution of sales under ‘normal’ or average conditions. Then we estimate a panel endogenous threshold model to see whether there are different pricing regimes when the expectation of demand differs. The latent variable that dictates the regime switch is the ratio of actual sales to expected sales at a point in time prior departure. Higher sales relative to normal sales would be evidence of a peak-demand flight. Finally, to control for potential endogeneity in the regressors we estimate a dynamic panel with an exogenous distinction of expected peak and expected off-peak flights.

#### **4.2.1 Nonparametric Estimation of Expected Sales**

In this section we came with a measure of the evolution of sales under average or normal conditions. That is, we estimate an exogenous measure of expected sales as the departure date nears for each of the flights in the sample. This measure of the evolution of sales for each flight is expected to be captured by the flight, carrier and the route’s characteristics. Consider the following nonparametric model of cumulative sales on various flight, carrier, and route characteristics.

$$\begin{aligned}
LOAD_{ijt} = & g(DAYADV_{ijt}, DEPTIME_{ij}, DIST_j, ROUSHARE_{ij}, \\
& HHI_j, HUB_{ij}, SLOT_j, DIFRAIN_j, DIFSUN_j \\
& DIFTEMP_j, AVEHHINC_j, AVEPOP_j, AA_j, AL_j \\
& CO_j, DE_j, UN_j, US_j) + \eta_{ijt}
\end{aligned} \tag{4}$$

The subscript  $i$  refers to flight,  $j$  to route, and  $t$  is time. Equation 4 is a panel estimated using kernel methods for mixed datatypes as explained in Li and Racine (2007). The dependent variable is  $LOAD$ , defined as the total number of seats sold divided by the total number of seats in the aircraft. The explanatory variables include the number of days in advance  $DAYADV$ , and various flight and route characteristics denoted by  $\mathbf{X}$ . Table 1 provides the summary statistics of these variables.<sup>4</sup> The evolution of the expected cumulative sales for flight  $i$ ,  $E(LOAD_{ijt}|DAYADV, \mathbf{X})$ , is obtained first by estimating Equation 4 using the observations from all other routes except the route from flight  $i$  as train data. Then flight  $i$ 's characteristics are used as evaluation data. This means that Equation 4 is estimated 81 times, once for each route, and evaluated 228 times at the corresponding flight's characteristics. To illustrate part of the results, the estimated nonparametric expected sales at different points prior departure and for different trip distances is shown in Figure 3. This was done using all datapoints as train data and with the remaining variables held constant at their median values for the evaluation points.<sup>5</sup>

---

<sup>4</sup>A detailed description of the explanatory variables is included in Appendix A.

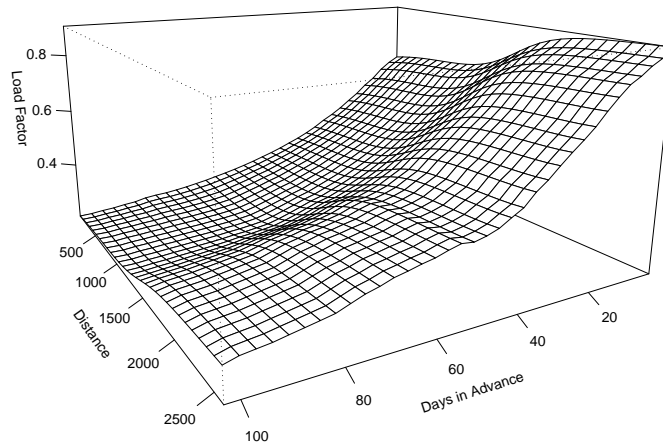
<sup>5</sup>Because of the large number of observations and for computational tractability, the bandwidths were obtained via rule-of-thumb (see Li and Racine (2007)).

Table 1: Summary Statistics

Variables	Mean	Std. Dev.	Min.	Max.	Obs.
<i>For the nonparametric estimation</i>					
<i>LOAD</i>	.509	.252	.012	1.000	7933
<i>DAYADV</i>	52.289	30.154	1.000	103.000	7933
<i>DEPTIME</i>	.451	.176	.229	.910	7933
<i>DIST</i>	1104.380	620.720	91.000	2604.000	7933
<i>ROUSHARE</i>	.665	.314	.119	1.000	7933
<i>HHI</i>	.684	.287	.259	1.000	7933
<i>HUB</i>	.737	.440	.000	1.000	7933
<i>SLOT</i>	.298	.458	.000	1.000	7933
<i>DIFRAIN</i>	2.010	1.484	.000	4.900	7933
<i>DIFSUN</i>	7.911	8.461	.000	45.000	7933
<i>DIFTEMP</i>	6.210	4.137	.000	19.000	7933
<i>AVEHHINC</i>	35580	4620	25198	53430	7933
<i>AVEPOP</i>	1044072	631862	187704	2897818	7933
<i>For the calibration of demand uncertainty</i>					
Forecasted LF	.739	.083	.469	.890	81
Sold-out probability	.227	.104	.037	.571	81
<i>For the endogenous panel threshold estimation</i>					
<i>FARE</i> <sup>a</sup>	291.087	171.879	54.000	1224.000	7933
<i>LOAD</i>	.509	.252	.012	1.000	7933
<i>ECC</i>	1.557	.940	1.000	11.668	7933
<i>E(LOAD DAYADV, X)</i>	.504	.208	.000	.980	7933
<i>S</i> (latent variable)	1.026	.417	.024	3.977	7933
<i>For the dynamic panel</i>					
<i>PEAK = I(S &gt; 1.18)</i>	.284	.451	.000	1.000	7933

Notes: <sup>a</sup> The standard deviation for *FARE* between flights is 152.933, and within is 78.751.

Figure 3: Estimated Nonparametric Expected Sales



### 4.2.2 Endogenous Threshold Estimation

Under the existence of menu costs, the benefits from switching pricing strategies may still be lower than the costs. Therefore even if carriers get to learn about the state of the demand, some degree of price flexibility is necessary in order to have *stochastic* peak-load pricing. Moreover, demand is never fully learned; as sales take place and some information is revealed about the demand, there will always be some uncertainty remaining about its final state of the demand. Under this scenario, carriers will want to wait until they have enough evidence toward having an expected peak or an expected off-peak flight before deciding to switch its pricing strategies. This suggests the existence of different pricing regimes for different expected demand states rather than a continuum of fully adjustable fares sensitive to every new piece of information about the expected final state of the demand. In this subsection we estimate an endogenous panel threshold model to test for the existence of different pricing regimes. The different regimes are given by the different expectations of the final state of the demand.

The  $E(LOAD_{ijt}|DAY ADV_{ijt}, \mathbf{X}_{ijt})$  estimated in the previous section is a measure of the expected evolution of sales for flight  $i$  under average conditions and it is independent to the actual evolution of sales given by  $LOAD_{ijt}$ . This is because the observations of the load factor of flight  $i$  were never included in the nonparametric estimation of  $E(LOAD_{ijt}|DAY ADV_{ijt}, \mathbf{X}_{ijt})$  for the same flight  $i$ . By independent we mean that if flight  $i$  is expected to be a peak-flight, is independent from the average of the other  $-i$  flights from being peak-flights. Therefore, the ratio

$$S_{ijt} = \frac{LOAD_{ijt}}{E(LOAD_{ijt}|DAY ADV_{ijt}, \mathbf{X}_{ijt})} \quad (5)$$

contains the necessary information to know whether at time  $t$  prior departure actual sales are high, low or about the same as compared to sales under average conditions. If at a given point prior departure this ratio is relatively large, it would be reasonable for carriers to think they are in a peak period and that expected demand will be greater than the allocated capacity. On the other hand, low values indicate that sales are low relative to average or normal sales and it would be reasonable for airlines to think they are in an off-peak period and some seats may be left unsold.

The ratio  $S_{ijt}$  has some interesting properties. Recall that the dataset was constructed in a way that all flights share the same departure date, hence they also share the same dates prior departure. If, for example, sales are higher/lower during weekends, this should affect all flights and will change

both,  $LOAD_{ijt}$  and  $E(LOAD_{ijt}|DAYADV_{ijt}, \mathbf{X}_{ijt})$ . Therefore the ratio  $S_{ijt}$  should remain unchanged. We assume carriers already know whether specific dates affect sales (e.g. weekends) and take this into account in their calculations of expected demand. Higher or lower sales on a given point in time common to all flight will have no impact on the definition of expected peak and expected off-peak flight. What is even more important, the construction of this ratio allows us to control for *systematic* peak-load pricing. During *ex-ante* know congested periods, *stochastic* peak-load pricing suggests that carriers will charge higher fares. As explained in Borenstein and Rose (1994), this type of peak-load pricing arises at an airport or fleet level. Here, the most likely capacity constraint is given by the total number of aircrafts. As a result *systematic* peak-load pricing should affect *all* flights keeping the ratio  $S_{ijt}$  unchanged. The drawback in this approach is that we will not be able to measure the effect of *systematic* peak-load pricing on fares. For an estimation of the congestion premia on fares due to *systematic* peak-load pricing, see Escobari (2006).

This section estimates a threshold model to test whether carriers have different pricing strategies for different expected states of the demand. *Stochastic* peak-load pricing suggests that when demand is expected to be greater than fixed capacity, carriers will set higher fares. The latent variable that will control the shift between expected peak and expected off-peak flights is  $S_{ijt}$ . To avoid an arbitrary selection of the number of pricing regimes and selection of the threshold(s), we estimate the model using the panel threshold regression methods with individual-specific fixed effects of Hansen (1999). The equation to be estimated has the form

$$\begin{aligned} \ln(FARE)_{ijt} = & \delta_0 DAYADV_{ijt} + \delta_1 \ln(ECC)_{ijt} \cdot I(S_{ij,t-1} \leq \gamma) \\ & + \delta_2 \ln(ECC)_{ijt} \cdot I(\gamma < S_{ij,t-1}) + \nu_{ij} + \varepsilon_{ijt} \end{aligned} \quad (6)$$

where  $I(\cdot)$  is the indicator function,  $S_{ijt}$  is the threshold variable and  $\gamma$  is the threshold. Moreover,  $\nu_{ij}$  is the unobserved carrier- and flight-specific effect,  $\varepsilon_{ijt}$  is error term, and as before the subscripts  $i$  denotes flight,  $j$  is route and  $t$  is time. Another way of writing Equation 6 is

$$\begin{aligned} \ln(FARE)_{ijt} = & \delta_0 DAYADV_{ijt} \\ & + \begin{cases} \delta_1 \ln(ECC)_{ijt} + \nu_{ij} + \varepsilon_{ijt} & \text{if } S_{ij,t-1} \leq \gamma \text{ (off-peak)} \\ \delta_2 \ln(ECC)_{ijt} + \nu_{ij} + \varepsilon_{ijt} & \text{if } \gamma < S_{ij,t-1} \text{ (peak)}. \end{cases} \end{aligned}$$

For the case of Equation 6, the observations are divided into two pricing regimes depending on whether the threshold variable  $S_{ij,t-1}$  is smaller or

larger than the threshold. The regime-independent variables  $DAYADV$ , is included to control for a time trend. Even though Equation 6 is illustrated for only one threshold, the actual estimation process test for the existence of up to three thresholds, allowing for up to four different pricing regimes. In the absence of regime changes, Equation 6 follows the form suggested by the theory under *no* demand learning in Equation 1.

Given the construction of the dataset we perfectly control for important sources of price dispersion observed in the industry (e.g. saturday-night stay-over, minimum and maximum stay, different connections/legs, fare class, refundability). Moreover, estimating the model using flight fixed effects allows controlling for unobservable time invariant characteristic, which include all the time invariant control variables included in Stavins (2001) (e.g. flight, carrier, and route characteristics). The main coefficient of interest is the Effective Cost of Capacity  $ECC$ . Prescott (1975)'s type of models predict a positive effect of  $ECC$  on fares. However, this is true under no demand learning or under price commitments. Under the specification of Equation 6, the coefficient on  $ECC$  is allowed to be different across flights and at different points prior departure, depending on the expectations of the demand. As predicted by *stochastic* peak-load pricing, higher expected demand states will be associated with a greater impact of  $ECC$  on fares, while lower expected demand states will be associated with lower or even a negative coefficient on  $ECC$ . The empirical specification is estimated as a constant elasticity model in  $\ln - \ln$  form. This is because both variables  $FARE$  and  $ECC$  are measured in dollars. Moreover, recall that  $ECC = \lambda/Pr$ , then estimating the equation using the logarithm of  $ECC$  allows separating its components in two.  $\ln(\lambda)$  goes as part of the regression intercept while the coefficient on  $\ln(Pr)^{-1}$  remains the same as the coefficient on  $\ln(ECC)$ . We can then interpret this coefficient as the impact of a percentage increase in  $ECC$  or a percentage decrease in the selling probability,  $Pr$ , on fares. The interpretation of this elasticity measure does not need to know the value of  $\lambda$ . An alternative specification replaces  $\ln ECC$  with  $LOAD$  in Equation 6. The *stochastic* peak-load pricing analysis follows the same logic as with  $\ln ECC$ , however, the interpretation is somehow different. Here a change in  $LOAD$  represents an increase capacity utilization.

In order to estimate the nonlinear specification in Equation 6 we follow the procedure proposed in Hansen (1999). First, to eliminate the unobserved carrier- and flight-specific effects, for a given  $\gamma$  and for each flight we obtain the deviations from the time averages. Stacking the data over all flights we obtain  $Y = \mathbf{V}(\gamma)\delta + \varepsilon$ , where  $Y$  and  $\mathbf{V}(\gamma)$  are just the stacked fixed-effects transformation just explained on  $\ln(FARE)$  and the set of explana-

tory variables respectively. Notice the values of the explanatory variables are a function of the value of the threshold. For any given  $\gamma$ , the vector of slope coefficients  $\delta$  can be estimated by ordinary least squares to obtain  $\hat{\delta}(\gamma)$ . Chan (1993) and Hansen (2000) recommend the estimation of  $\gamma$  by least squares, hence its estimator is

$$\hat{\gamma} = \arg \min_{\gamma} Y'(I - \mathbf{V}(\gamma)'(\mathbf{V}(\gamma)\mathbf{V}(\gamma))^{-1}\mathbf{V}(\gamma)')Y. \quad (7)$$

After  $\hat{\gamma}$  is found, the estimate for the slope coefficients is  $\hat{\delta}(\hat{\gamma})$ . Then the next step is to find out if the threshold is statistically significant. The null hypothesis of no threshold in Equation 6 can be characterized by  $H_0: \delta_1 = \delta_2$ . As explained in Hansen (1999), classical tests have non-standard distributions because under the null  $\gamma$  is not identified. Therefore, we follow Hansen (1996) and simulate the asymptotic distribution of the likelihood test by bootstrapping. The likelihood ratio to test  $H_0$  is

$$F_1 = (SSE_0 - SSE_1(\hat{\gamma}))/\hat{\sigma}^2 \quad (8)$$

where  $SSE_0$  is the sum of squared errors under the null after the the fixed-effects transformation is made. Similarly,  $SSE_1$  is the sum of squared errors of the fixed-effects transformation made on Equation 6. For a larger number of thresholds the idea is similar, with the important characteristic that sequential estimation is consistent. Therefore, in order to test for the number of thresholds, we allow for sequentially zero, one, two, and three thresholds. As in Hansen (1999), the observations are first sorted on the threshold variable and the search of the threshold is restricted to specific quantiles. The more quantiles the finer the grid to which the search is limited. Bootstrapping simulates the asymptotic distribution of the likelihood ratio test. This likelihood ratio is used to test whether the threshold is statistically significant under the null of no threshold. When rejecting the null, one more threshold is included.<sup>6</sup>

Table 2 provides the results that test for the number of thresholds: the test statistics  $F_1$  and  $F_2$ , along with the bootstrap p-values and critical values. From the bootstrap p-values, the null of no threshold for the one threshold model is rejected at a 1% level in all specifications. However, no evidence of further thresholds is found. The results for the three thresholds model are not reported since non of the second thresholds was found to be significant. Because the original dataset is unbalanced and the testing

---

<sup>6</sup>The estimation used 400 quantiles and 300 bootstrap replications for each of the bootstrap tests.

procedure implemented in this section only allows for balanced panels, we work with two subsets of the data. The first one has 198 flights over 35 time periods (covering a period of 100 days prior departure) is reported in Columns (1) and (3). The second has 216 flights over 34 time periods (103 days prior departure) and is reported in Columns (2) and (4).

Table 2: Tests for threshold effects

	<i>LOAD</i>		$\ln(ECC)$	
	(1)	(2)	(3)	(4)
<i>Test for a single threshold</i>				
$F_1$	77.735	113.161	109.968	161.568
p-value	0.010	0.000	0.010	0.003
Bootstrap critical values				
10%	47.578	45.580	62.897	57.310
5%	54.875	52.449	82.918	68.455
1%	69.513	63.036	103.242	130.663
<i>Test for a double threshold</i>				
$F_2$	19.136	23.838	16.713	33.609
p-value	0.473	0.360	0.720	0.243
Bootstrap critical values				
10%	33.728	36.242	45.522	46.067
5%	38.969	43.942	52.422	58.614
1%	54.101	63.603	75.602	76.973

Notes: Because non of the second thresholds was found significant, the tests for triple thresholds are not reported. Columns (1) and (3) have 6732 observations across 198 flights and columns (2) and (4) have 7128 across 216 flights.

The point estimates for the thresholds in all four specifications, along with the asymptotic 95% confidence intervals are presented in Table 3. All point estimates lie around 1.171 and 1.184 and the confidence intervals are very tight, indicating little uncertainty about the nature of this division. The results indicate the existance of two pricing regimes. The first pricing regime occurs when  $\gamma < S_{ij,t-1}$ . Notice that in this regime actual sales are relatively larger than sales under average conditions, hence we call this the peak period pricing regime. The second regime is characterized by  $\gamma \geq S_{ij,t-1}$ . This will be referred as the off-peak period pricing regime since actual sales are relatively lower that sales under average conditions.

The confidence interval construction shown in Figure 4, tabulated for specification in Column (1) of Tables 2 and 3, provides further insights for the threshold results. The point estimate is the value of  $\gamma$  at which the likelihood ratio is equal to zero. The confidence interval  $[\bar{\gamma}, \underline{\gamma}]$ , are the values for  $\gamma$  for which the likelihood ratio lies beneath the straight line. Moreover, there are no other major dips in the likelihood ratio, which would be evidence of a third pricing regime.

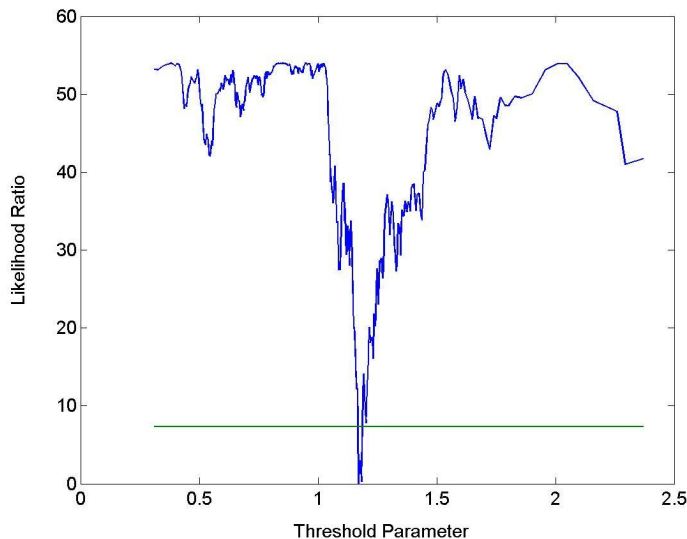


Table 3: Threshold estimates

	LOAD		ln(ECC)	
	(1)	(2)	(3)	(4)
$\gamma$	1.171	1.184	1.184	1.184
Asymptotic 95% confidence interval				
$\bar{\gamma}$	1.183	1.184	1.184	1.184
$\underline{\gamma}$	1.171	1.171	1.171	1.170

Notes: The test for a triple threshold not reported, since non of the second thresholds was found significant. Columns (1) and (3) have 6732 observations across 198 flights and columns (2) and (4) have 7128 across 216 flights.

Figure 4: Confidence interval construction in single threshold model



The regression estimates for the single threshold model are presented in Table 4. The first noticeable result is that Columns (1) and (2) are very similar, while (3) and (4) also look alike. Thus, the two balanced subsamples yield very similar results. The figures in parentheses are White-robust t-statistics. The regime-independent coefficient *DAYADV*, included as a control for a time trend is highly significant in all four specifications. The coefficient on *DAYADV* in Column (4) means that after controlling for capacity constraints and demand uncertainty, route, carrier and flight characteristics, ticket characteristics that segment consumers and *systematic* and *stochastic* peak-load pricing, buying a ticket one day in advance reduces the

ticket price by 56.3 cents.<sup>7</sup> This is a measure of second degree price discrimination in the form of advance-purchase requirements. As pointed out in Dana (1998), for advance-purchase discounts to be classified as discriminatory, it is necessary to define an appropriate measure of costs. Prices are considered discriminatory when the price markups over costs are different for different consumers. Hence the importance of having an appropriate measure of costs. In this analysis, the costs for different seats in the same aircraft may be different due to the existence of uncertain demand and costly capacity. These different costs are captured by *ECC*. Finally, notice that we are fully controlling for other sources of second degree price discrimination such as Saturday night stayover.

Table 4: Regression estimates: single threshold model

Regressor	<i>LOAD</i>		$\ln(\textit{ECC})$	
	(1)	(2)	(3)	(4)
<i>Regime-independent coefficients</i>				
<i>DAYADV</i> /10 <sup>3</sup>	-1.671 (-9.164)	-1.520 (-8.989)	-2.235 (-17.230)	-1.969 (-15.759)
<i>Regime-dependent coefficients</i>				
<i>LOAD</i> <sub>ijt</sub> · $I(\gamma \geq S_{ij,t-1})$	0.300 (9.033)	0.236 (7.752)		
<i>LOAD</i> <sub>ijt</sub> · $I(\gamma < S_{ij,t-1})$	0.428 (12.617)	0.386 (11.704)		
$\ln(\textit{ECC})_{ijt}$ · $I(\gamma \geq S_{ij,t-1})$			0.112 (7.959)	0.079 (5.564)
$\ln(\textit{ECC})_{ijt}$ · $I(\gamma < S_{ij,t-1})$			0.254 (11.821)	0.251 (11.460)
R-squared	0.893	0.895	0.897	0.899

Notes: The independent variable is  $\ln(\textit{FARE})$ . t-statistics in parentheses based on White-robust standard errors. All regressions are estimated with flight fixed effects, not reported. Columns (1) and (3) have 6732 observations across 198 flights and columns (2) and (4) have 7128 across 216 flights.  $I(\gamma < S_{ijt})$  is referred as the peak period, while  $I(\gamma \geq S_{ijt})$  is the off-peak period.

The variables we are mostly interested in are the regime-dependent. From Table 4 we observe that *LOAD* in Columns (1) and (2) and  $\ln(\textit{ECC})$  in Columns (3) and (4) are all highly significant and have a positive effect on fares on both regimes. In all four specifications the off-peak period regime,  $\gamma \geq S_{ij,t-1}$ , has a lower coefficient than the peak period regime,  $\gamma < S_{ij,t-1}$ . We know from the results in Table 2 that the coefficients in both regimes are significantly different. The results from Column (2), evaluated at the subsample average fare of 285.85 dollars indicate that in a 100 seat aircraft, having one seat less available increases fares by 67.5 cents in an expected

<sup>7</sup>This one is calculated using the average fare for the subsample used in the estimation of Column (4). This is  $285.85 \times -1.969/10^3 = -0.563$  dollars.

off-peak flight while increases fares by 110.3 cents in an expected peak flight. Columns (3) and (4) require some additional care. The effect of *ECC* on fares as predicted by Prescott (1975)'s type of models is positive. However, as sales progress and carriers learn about the state of the demand, the coefficient on *ECC* will be the outcome of two different type of models. The Prescott (1975)'s type and *stochastic* peak-load pricing. The later only predicts that fares will be larger during expected peak flights. Thus the only requirement on our regime-dependent coefficients is that the expected peak regime should have a larger coefficient than the expected off-peak regime. When capacity is not costly and expected demand is smaller than allocated capacity, carriers will be willing to sell the last seats in the aircraft for any price above the operating marginal cost (e.g. baggage transportation, soft drink and pretzels). Consequently the last seats could be priced very low and the coefficient on *ECC* could be negative indicating lower fares for later seats. However, the results are consistent with having costly capacity and provide important evidence supporting Prescott (1975)'s type of models, already documented in Escobari and Gan (2007). Columns (3) and (4) show that fares respond positively to *ECC* in both peak and off-peak regimes. Furthermore, there is also an important evidence supporting the existence of *stochastic* peak-load pricing with the peak regime coefficient being greater than in the off-peak regime.<sup>8</sup>

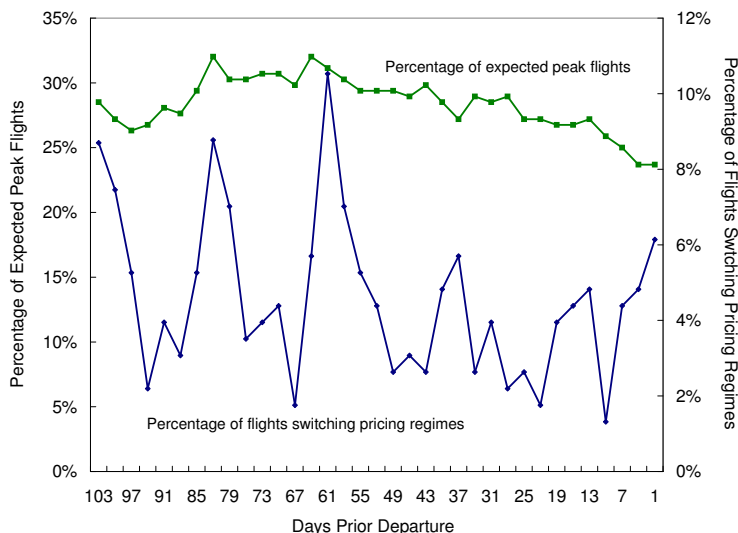
Fares will be increasing at a higher rate during expected peak regimes. Carriers forecasting that demand will be greater than allocated capacity will set higher fares to increase their profits and sell the remaining available capacity to travelers with higher valuations. If price commitments were to prevail or if carriers do not learn about the state of the demand, the flight will still sell-out in a high demand period. However in the absence of *stochastic* peak-load pricing existing capacity will be allocated to travelers that arrive first and not necessarily to travelers with higher valuations sorted by higher prices. On the other hand, when a low demand flight is expected, fares will increase at a lower rate. This is consistent with cheap fares offered close to departure and 'last minute deals'. Airlines offer this kind of tickets when demand falls short and allocated capacity is likely to remain underutilized.

Figure 5 shows the percentage of expected peak flights and the percentage of flights switching pricing regimes as the flight date approaches. We see that the percentage of flights expected to be peak is fairly constant and

---

<sup>8</sup>In this case a direct interpretation of the coefficient on *ECC*, as we did with *LOAD*, would not be entirely correct since *ECC* is constructed based on an *ex-ante* distribution of demand uncertainty.

Figure 5: Percentage of Expected Peak Flights



ranges from 24% to 32% of the sample over time. Moreover, the percentage of flights switching pricing regimes is very volatile. On average 5% of the flights switches regimes every three days. That is equivalent to say that each flight switches regime 1.5 times over the period studied.<sup>9</sup> It is interesting to see that the sharp increase in fares close to departure commonly observed in the industry is the result of second degree price discrimination and the combination of costly capacity and demand uncertainty, not the result of peak-load pricing. It is true that those higher fares close to departure are associated higher demand states, but they are higher because the effective cost of capacity *ECC* is larger (see Dana (1999b)).

### 4.2.3 Dynamic Panel with Exogenous Threshold

The previous endogenous threshold estimation used the methods described in Hansen (1999) and Hansen (2000) to identify two pricing regimes. This procedure developed for non-dynamic balanced panels required us to assume strict exogeneity of the regressors and to work with two balanced panels, subsets of the original unbalanced dataset. In this subsection we take care of these two issues. We will reestimate the model as suggested in Equation 1 to

<sup>9</sup>This should be viewed as a lower bound since increasing the frequency at which the data is observed would increase the number of times a flight switches regimes.

test for the existence of demand learning, but this time using dynamic panel techniques as developed in Holtz-Eakin et al. (1988), Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998). This will let us work with the entire unbalanced panel and relax the assumption of strict exogenous regressors. We will assume that the switch between the two pricing regimes is the same as the one estimated in the previous part. Specifically, the equation to be estimated is

$$\begin{aligned} \ln(FARE)_{ijt} = & \alpha \ln(FARE)_{ij,t-1} + \beta_1 DAYADV_{ijt} + \beta_2 PEAK_{ij,t-1} \\ & + (\delta_0 + \delta_1 PEAK_{ij,t-1}) \cdot \ln(ECC)_{ijt} + \nu_{ij} + \varepsilon_{ijt} \end{aligned} \quad (9)$$

The idea is the same as in the estimation of Equation 6. This means analyzing the effect of the effective cost of capacity on fares under no demand learning as suggested by Equation 1, while allowing for the existence of different pricing regimes when the expectation of future demand differs. As found in section 4.2.2, here we allow for the coefficient of  $ECC$  on fares to have two possible values that represent the expected peak and expected off-peak regimes. The division between these two regimes is assumed to be the same as before. Then the variable that dictates the shift is  $PEAK_{ijt} = I(S_{ijt} > 1.18)$ , with the 1.18 taken from estimates reported in Table 3.  $PEAK$  takes the value of one when the flight is expected to be a peak flight and is zero if it is expected to be an off-peak. These two pricing regimes will be significantly different if the interaction coefficient  $\delta_1$  is statistically significant. Then, during an expected off-peak flight the effect of  $ECC$  on fares will be  $\delta_0$ , while in an expected peak-flight it will be  $\delta_0 + \delta_1$ . As before  $DAYADV$  controls for any time trend. The coefficient on the lagged dependent variable,  $\ln(FARE)_{ij,t-1}$ , is not of direct interest, but allowing for dynamics in the underlying process may be crucial for recovering consistent estimates of the other parameters. As in the previous section, for a second specification  $\ln(ECC)$  will be replaced with  $LOAD$ .

The reason why a dynamic estimation is important is because both the *effective cost of capacity*,  $ECC$ , and the load factor,  $LOAD$ , are functions of cumulative sales. But the number of tickets that have already been sold—cumulative sales—depend on previous price levels. So there is reason to believe that the assumption of strict exogeneity of the regressors may be violated. The way the panel estimator presented in this section controls for endogeneity is by using ‘internal instruments’. We assume that the explanatory variables are only ‘weakly exogenous’, which means that the cumulative sales can be affected by current and past realization of fares, but must be uncorrelated with future realizations of the error term. Weak

exogeneity does not mean that consumers do not take into account expected future changes in fares in their decisions to buy or not a ticket; it just means that future (unanticipated) shocks in fares do not influence current cumulative sales or the decision to buy a ticket. We will assess the validity of this weak exogeneity assumption below.

To estimate Equation 9, we first take first-differences to eliminate carrier- and flight-specific effects. Then the resulting equation requires instruments to deal with the potential endogeneity of the explanatory variables and with the problem that the construction of the new error term,  $\varepsilon_{ijt} - \varepsilon_{ij,t-1}$ , is correlated with the lagged dependent variable,  $\ln(FARE_{ij,t-1}) - \ln(FARE_{ij,t-2})$ . The GMM *difference* panel estimator that we will report constructs its moment conditions under the assumptions that the error term,  $\varepsilon$ , is not serially correlated, and that the explanatory variables are weakly exogenous. Then the moment conditions used for the *difference* estimator are:

$$E[y_{ij,t-s}(\varepsilon_{ijt} - \varepsilon_{ij,t-1})] = 0 \quad \text{for } s \geq 2; t = 3, \dots, T, \quad (10)$$

$$E[\mathbf{W}_{ij,t-s}(\varepsilon_{ijt} - \varepsilon_{ij,t-1})] = 0 \quad \text{for } s \geq 2; t = 3, \dots, T. \quad (11)$$

where  $y_{ijt}$  is the natural logarithm of fare and  $\mathbf{W}_{ijt}$  is the set of explanatory variables other than the lagged logarithm of fare.

Blundell and Bond (1998) point out a statistical shortcoming with this *difference* estimator. When the explanatory variables are persistent over time, lagged levels of these variables are weak instruments for the regression equation in differences. To reduce the potential biases and imprecision associated with the usual difference estimator we employ the *system* estimator suggested in Blundell and Bond (1998). This *system* estimator combines the regression in differences with the regression in levels. The instruments for the regression in differences are the same as above. The instruments for the regression in levels are the lagged differences of the corresponding variables. The validity of these instruments relies on the following additional assumption: There is no correlation between the differences of the right-hand side variables in Equation 9 and the flight-specific effects, but there may be correlation between the levels of the right-hand side variables and the flight-specific effects. Then, for the regression in levels included as a second part of the system the additional moment conditions are:

$$E[(y_{ij,t-s} - y_{ij,t-1})(\nu_{ij} + \varepsilon_{ijt})] = 0 \quad \text{for } s = 1, \quad (12)$$

$$E[(\mathbf{W}_{ij,t-s} - \mathbf{W}_{ij,t-s-1})(\nu_{ij} + \varepsilon_{ijt})] = 0 \quad \text{for } s = 1. \quad (13)$$

To address the validity of the instruments we consider two specification tests suggested in Arellano and Bond (1991), Arellano and Bover (1995), and

Blundell and Bond (1998). To test the overall validity of the instruments we provide a Sargan test of over-identifying restrictions, which analyzes the sample analogs of the moment conditions used in the GMM estimation. To test the hypothesis that the error term,  $\varepsilon_{ijt}$ , is not serially correlated, we test whether the differenced error term is second-order serially correlated.

The dynamic panel estimates show that the load factor,  $LOAD$ , and the effective cost of capacity,  $\ln(ECC)$ , both have a significant impact on fares under both pricing regimes. Table 5 reports the results using the *differences* and the *system* estimators described above. Additionally, for comparative purposes we report the panel estimates when the estimation is done in *levels* using flight fixed effects. Table 5 also presents the p-values for the Sargan test and the serial correlation test. The null hypothesis for the Sargan test is that the instrumental variables are uncorrelated with the residuals (valid specification). Whereas in the null hypothesis for the serial correlation test is that errors in the differenced equation exhibit no second-order serial correlation (valid specification).

Table 5: Regression estimates: GMM Dynamic panel

	Levels (1) 1.18	First diff. (2)	System (3)
<i>Load Factor</i>			
$\ln(FARE)_{ij,t-1}$	0.691 (31.640)	0.742 (47.421)	0.992 (420.941)
$DAYADV_{ijt}/10^3$	-0.240 (-1.550)	1.704 (6.216)	-0.083 (-0.749)
$PEAK_{ij,t-1}$	-0.115 (-5.780)	-0.164 (-3.753)	-0.045 (-2.798)
$LOAD_{ijt}$	0.235 (8.950)	0.519 (11.543)	0.131 (7.128)
$LOAD_{ijt} \cdot PEAK_{ij,t-1}$	0.215 (6.200)	0.272 (4.032)	0.070 (2.493)
Sargan test <sup>a</sup> (p-value)		0.285	0.379
Serial correlation test <sup>b</sup> (p-value)		0.872	0.897
<i>Effective cost of capacity</i>			
$\ln(FARE)_{ij,t-1}$	0.695 (31.840)	0.693 (44.551)	0.973 (200.427)
$DAYADV_{ijt}/10^3$	-0.849 (-7.840)	0.839 (5.269)	0.972 (6.365)
$PEAK_{ij,t-1}$	0.004 (0.480)	-0.029 (-1.134)	-0.097 (-2.395)
$\ln(ECC)_{ijt}$	0.098 (8.620)	0.259 (15.016)	0.167 (8.188)
$\ln(ECC)_{ijt} \cdot PEAK_{ij,t-1}$	0.068 (3.980)	0.166 (4.974)	0.087 (1.650)
Sargan test <sup>a</sup> (p-value)		0.302	0.314
Serial correlation test <sup>b</sup> (p-value)		0.838	0.952

Notes: The dependent variable is  $\ln(FARE)$ . t-statistics in parentheses based on White robust standard errors.  $PEAK = I(S > 1.18)$ . <sup>a</sup> The null hypothesis is that the instruments are not correlated with the residuals (valid specification). <sup>b</sup> The null hypothesis is that the errors in the first-difference regression exhibit no second-order serial correlation (valid specification).

The results for the load factor in the upper part of Table 5, show that for the *levels*, *difference*, and *system* dynamic panel regressions, the effect of  $LOAD$  on fares is positive and highly significant. Moreover, it is greater in an expected peak flight than in an expected off-peak flight, with the difference being also statistically significant with all three estimators. The significance of this difference can be appreciated by looking at the t-statistics of the coefficient on the interaction variable  $LOAD_{ijt} \cdot PEAK_{ij,t-1}$ . In particular, based on the *system* GMM estimates, evaluated at the sample average fare of 291.09 dollars and for a 100 seats airplane, having one less seat available increases fares by 38.1 cents in an expected *off-peak* flight while increases fares by 58.5 cents in a expected *peak* flight. The *ECC* specification reported in the lower part of Table 5 supports the previous



findings. As suggested by the theory in the Prescott (1975)’s type of models, *ECC* has a positive impact on fares. Moreover, as information about the final state of the demand becomes available, the results are consistent with *stochastic* peak-load pricing with higher fares being set in expected peak demand periods and lower fares set in expected off-peak periods.

The regressions satisfy the specification tests. There is no evidence of second order serial correlation and the regressions pass the Sargan specification test. Regarding the sign on *DAYADV*, this one is no longer comparable with the one reported in Table 4 because of the existence of the lagged dependent variable in the dynamic panel regressions. This also explain why the sign on *DAYADV* is so volatile.

## 5 Conclusions

One important source of uncertainty for airlines is that they have limited information about the demand at the moment of scheduling a flight. Because tickets are sold in advance, prices should be set in an environment of uncertainty about the total number of arriving consumers. Having a good approximation of the expected demand is key because (1) seats left unsold have little value after departure, and (2) carriers may forgone important profits if the flight sells out and some consumers that would have paid even higher prices have to be rationed.

In this paper we initially calibrate the *ex-ante* –before any ticket is sold– distribution of demand uncertainty using information on sold-out probabilities and forecasted values of occupancy rates. Under the Prescott (1975)’s type of models of costly capacity and demand uncertainty, fares will be dispersed will lower priced units being sold before higher priced units. After controlling for restrictions that segment consumers (e.g. saturday-night-stayover, minimum and maximum stay, different connections/legs, fare class, refundability), the calibrated *ex-ante* demand uncertainty should be enough to explain the observed price dispersion if (1) there exist price commitments or (2) carriers do not learn about the final state of the demand as the flight date approaches.

Using nonparametric techniques we then construct a latent variable that is used as a proxy to identify different expected final demand states at different points prior departure. Using this latent variable we estimate a panel endogenous threshold model to test whether carriers abandon price commitments as they learn about the final state of the demand. The result identified two different pricing regimes. Consistent with the predictions of

*stochastic* peak-load pricing, in the expected peak flight regime fares will be higher, while in the expected off-peak flights fares will be lower. To control for potential endogeneity of the regressors and the interaction between cumulative sales and previous level of prices, we also estimate a dynamic panel model. The results also supported the existence of *stochastic* peak-load pricing in airlines.

The findings in this paper, as well as the

## A Appendix: Variable Description

*FARE<sub>ijt</sub>*: Price in US\$ paid for the one-way airfare.

*LOAD<sub>ijt</sub>*: Load factor, defined as total number of seats sold at time  $t$  divided by total number of seats in the aircraft.

*ECC<sub>ijt</sub>*: *Effective cost of capacity*, calculated by dividing costly capacity,  $\lambda$  (initially normalized to one), by the probability that this seat will be sold. For the *censored normal* case this one is given by

$$ECC_{ijt} = \frac{\lambda}{Pr_{h_{ijt}}} = \lambda \cdot \left[ \int_{h_{ijt}/m_{ij}}^{\infty} \sqrt{2\pi\sigma_j^2} \cdot \exp(-(\kappa - \mu_j)^2/2\sigma^2) d\kappa \right]^{-1}$$

$m_{ij}$  is the total number of seats in the aircraft and  $h_{ijt}$  is the number of seats that have already been sold. The values for  $\mu_j$  and  $\sigma_j$  are obtained from the calibration procedure in section 4.1.3.

*DAYADV<sub>ijt</sub>*: Number of days in advance the ticket was purchased.

*S<sub>ijt</sub>*: Threshold variable, defined as the ratio of actual seats sold to expected number of seats sold.  $S_{ijt} = LOAD_{ijt}/E(LOAD_{ijt}|DAYADV_{ijt}, \mathbf{X}_{ijt})$ .

*PEAK<sub>ijt</sub>*: Variable equal to one if flight  $i$  is expected to be a peak flight,  $PEAK_{ijt} = I(\gamma < S_{ijt})$ .

*DEPTIME<sub>j</sub>*: Time of the day the flight departed.

*DIST<sub>j</sub>*: Nonstop mileage between the two endpoint airports on a route.

*ROUSHARE<sub>ij</sub>*: Carrier's share on the route based on total number of seats in direct flights for the day of the flight.

*HHI<sub>j</sub>*: *Herfindahl-Hirshman Index* of concentration on the observed route, with *ROUSHARE* used as the measure of market share of each carrier.

$$HHI_j = \sum_{i=1}^N ROUSHARE_{ij}^2$$

*HUB<sub>ij</sub>*: Variable equal to one if the carrier has a hub in the origin or destination airports.

*SLOT<sub>j</sub>*: In some airports like Chicago O'Hare (ORD), Kennedy (JFK), La Guardia (LGA), and Reagan National (DCA), the U.S. government has imposed limits on the number of takeoffs and landings that may take place each hour. To take into account the scarcity value of acquiring a slot, the variable *SLOT* equals to one if either endpoint of route *j* is one of these airports and zero otherwise.

*DIFTEMP<sub>j</sub>*: Absolute difference in average end of October temperatures, measured in Fahrenheit degrees, between the departure and destination cities.

*DIFRAIN<sub>j</sub>*: Absolute difference in average end of October precipitation, measured in inches, between the departure and destination cities.

*DIFSUN<sub>j</sub>*: Absolute difference in average end of October sunshine, measured in percentage, between the departure and destination cities.

*AVEHHINC<sub>j</sub>*: Average of the median household income in the two cities.

*AVEPOP<sub>j</sub>*: Average population in the two cities. For cities with more than one airport, the population is apportioned to each airport according to each airport's share of total enplanements. Source: Table 3, Bureau of Transportation Statistics, Airport Activity Statistics of Certified Air Carriers: Summary Tables 2000.

*AA<sub>j</sub>, AL<sub>j</sub>, CO<sub>j</sub>, DE<sub>j</sub>, UN<sub>j</sub>, US<sub>j</sub>*: Variables equal to one if the carrier on route *j* is American, Alaska, Continental, Delta, United, or US Airways respectively.

## References

ARELLANO, M., BOND, S. (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *Review of Economic Studies* **58**, 277-297.

- ARELLANO, M., BOVER, O. (1995): "Another look at the instrumental variable estimation of error-components models," *Journal of Econometrics* **68**, 29-51.
- BACHIS, E., PIGA, C. (2007): "On-line price discrimination with and without arbitrage conditions," Loughborough University, Department of Economics Working Paper. Presented at the International Industrial Organization Conference, Savannah, GA, April, 2007.
- BLUNDELL, R.W., BOND, S. (1998): "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics* **87**, 115-143.
- BORENSTEIN, S., ROSE, N.L. (1994): "Competition and price dispersion in the U.S. airline industry," *Journal of Political Economy* **102**, 653-683.
- BOITEAUX, M. (1949): "La tarification des demandes en point: application de la theorie de la vente au cout marginal," *Revue Generale de l'electricite* **58**, 321-340; translated as "Peak load pricing," *Journal of Business* **33**, 157-179.
- BROWN, G., JOHNSON, B. (1969): "Public utility pricing and output under risk," *American Economic Review* **59**, 119-128.
- BURNETAS, A.N., SMITH C.E. (2000): "Adaptive ordering and pricing for perishable products," *Operational Research* **48** 436-443.
- CARLTON, D.W. (1977): "Peak load pricing with stochastic demand," *American Economic Review* **67**, 1006-1010.
- CHAN, K.S. (1993): "Consistency and limiting distributions of the least squares estimator of a threshold autoregressive model," *The Annals of Statistics* **21**, 520-533.
- CHATWIN, R.E. (2000): "Optimal dynamic pricing of perishable products with stochastic demand and a finite set of prices," *European Journal of Operational Research* **125**, 149-174.
- CHEN, J. (2006): "Differences in average prices on the internet: Evidence from the online market for air travel," *Economic Inquiry* **44**, 656-670.
- CREW, M.A., KLEINDORFER, P.R. (1986): "The economics of public utility regulation," MIT Press, Cambridge, MA.

- CREW, M.A., FERNANDO, C.S., KLEINDORFER, P.R. (1995): "The theory of peak-load pricing: A survey," *Journal of Regulatory Economics* **8**, 215-248.
- DANA, JR. J. (1998): "Advance-purchase discounts and price discrimination in competitive markets," *Journal of Political Economy* **106**, 395-422.
- DANA, JR. J. (1999a): "Using yield management to shift demand when the peak time is unknown," *Rand Journal of Economics* **30**, 456-474.
- DANA, JR. J. (1999b): "Equilibrium price dispersion under demand uncertainty: The roles of costly capacity and market structure," *Rand Journal of Economics* **30**, 632-660.
- DANIEL, J.I. (1995): "Congestion pricing and capacity of large hub airports: A bottleneck model with stochastic queues," *Econometrica* **63**, 327-370.
- DENECKERE, R., PECK, J. (2005): "Dynamic competition with random demand and costless search: A theory of price posting," University of Wisconsin, Department of Economics Working Paper.
- EDEN, B. (1990): "Marginal cost pricing when spot markets are complete," *Journal of Political Economy* **98**, 1293-1306.
- ESCOBARI, D. (2006): "Systematic peak-load pricing, congestion premia and demand diverting: Empirical evidence," Texas A&M University, Department of Economics Working Paper.
- ESCOBARI, D., GAN, L. (2007): "Price dispersion under costly capacity and demand uncertainty," NBER working paper 13075, Cambridge, MA.
- GALE, I., HOLMES, T. (1992): "The efficiency of advance-purchase discounts in the presence of aggregate demand uncertainty," *International Journal of Industrial Organization* **10**, 413-437.
- GALE, I., HOLMES, T. (1993): "Advance-purchase discounts and monopoly allocation of capacity," *American Economic Review* **83**, 135-146.
- GALLEGO, G., VAN RYZIN, G. (1994): "Optimal dynamic pricing of inventories with stochastic demand over finite horizons," *Management Science* **40**, 999-1020.
- HANSEN, B.E. (1996): "Inference when the nuisance parameter is not identified under the null hypothesis," *Econometrica* **64**, 413-430.

- HANSEN, B.E. (1999): "Threshold effects in non-dynamic panels: Estimation, testing, and inference," *Journal of Econometrics* **93**, 345-368.
- HANSEN, B.E. (2000): "Sample splitting and threshold estimation," *Econometrica* **68**, 575-603.
- HIRSHLEIFER, J. (1958): "Peak loads and efficient pricing: Comment," *Quarterly Journal of Economics* **72**, 451-462.
- HOLTZ-EAKIN, D., NEWEY, W., ROSEN, H. (1988): "Estimating vector autoregressions with panel data," *Econometrica* **56**, 1371-1395.
- KINCAID, W.M., DARLING, D.A. (1963): "An inventory pricing problem," *Journal of Mathematical Analysis and Applications* **7**, 183-208.
- LI, Q., RACINE, S. (2007): "Nonparametric econometrics," Princeton University Press, Princeton, NJ.
- LIN, K.Y. (2006): "Dynamic pricing with real-time demand learning," *European Journal of Operational Research* **174**, 522-538.
- LUCAS, R.E., WOODFORD, M. (1993): "Real effects of monetary shocks in an economy with sequential purchases," NBER working paper 4250, Cambridge, MA.
- MCAFEE, R.P., TE VELDE, V. (2006): "Dynamic pricing in the airline industry," In *Handbook of Economics and Information Systems*, vol. 1, edited by T.J. Hendershott. New York: Elsevier Sci.
- PRESCOTT, E. (1975): "Efficiency of the natural rate," *Journal of Political Economy* **83**, 1229-1236.
- RACINE, J.S., LI, Q. (2004): "Nonparametric estimation of regression functions with both categorical and continuous data," *Journal of Econometrics* **119**, 99-130.
- SAVING T.R., DE VANY, A.S. (1981): "Uncertain markets, reliability and peak-load pricing," *Southern Economic Journal* **47**, 908-923.
- STAVINS, J. (2001): "Price discrimination in the airline market: The effect of market concentration," *The Review of Economics and Statistics* **83**, 200-202.
- STEINER, P.O. (1957): "Peak loads and efficient pricing," *Quarterly Journal of Economics* **71**, 585-610.

VISSCHER, M.L. (1973): "Welfare-maximizing price and output with stochastic demand: Comment," *American Economic Review* **63**, 224-229.

WILLIAMSON, O.E. (1966): "Peak load pricing and optimal capacity under constraints," *American Economic Review* **56**, 810-827.