MPRA
Munich Personal RePEc Archive

The hidden side of dynamic pricing in airline markets

Alderighi, Marco and Gaggero, Alberto A and Piga, Claudio A

Universitá della Valle d'Aosta, Italy, University of Pavia, Italy, Keele University, UK

June 2016

Online at https://mpra.ub.uni-muenchen.de/71674/
MPRA Paper No. 71674, posted 01 Jun 2016 13:41 UTC

The hidden side of dynamic pricing in airline markets*

Marco Alderighi
Università della Valle d'Aosta
Università Bocconi

Alberto A. Gaggero
University of Pavia

Claudio A. Piga
Keele University
Rimini Centre Economic Analysis

June 1, 2016

Abstract

An often disregarded, albeit central, aspect of the airline pricing's problem consists in assigning a fare to all the available seats on an airplane at the beginning of and during the whole booking period. We show how a flight's fare distribution is set in practice and how it changes over time using evidence from a leading European low-cost carrier. Such pricing behavior is consistent with the main predictions from the theoretical model we present. First, fare distributions are increasing across seats because a lower fare for the seat on sale enhances the likelihood of selling the subsequent seats. Second, over time fare distributions move, on average, downward to reflect the perishable nature of a flight's seats. Third, due to the increasing profile of the fare distributions across seats, we find that the price observed by prospective buyers tends to increase as the date of departure nears.

JEL Classification: D22, L11, L93.
Keywords: dynamic pricing, option value, seat inventory control, low-cost carriers.

1 Introduction

The definition of dynamic pricing (DP) in airline markets, both in the economic and operational research academic literature, as well as in the press, has been so far intrinsically related to the description of how fares on sale evolve over time (McAfee and te Velde, 2007). The world-wide success of Low Cost Carriers (LCCs), whose fares are easier to observe and compare than those offered by legacy carriers, has reinforced the view that

*We wish to thank Rajesh Aggarwal, Mian Dai, James Dana, Patrick Legros, the participants of the 2016 IIOC Conference in Philadelphia and the seminar participants at Northeastern University, Lancaster University and Swansea University. Usual disclaimer applies.

the temporal fluctuations of observed fares constitute the central part of a carrier's Revenue Management (RM) system (McGill and Van Ryzin, 1999; Weatherford and Bodily, 1992).

Three major factors are assumed to shape the fares' temporal profile. First, airlines sell a highly perishable service. While the fare set today has to account for the cost of the foregone option of selling the seat later on the same flight for a higher fare, for perishable seats such option value goes to zero as the take-off approaches, thus leading to the prediction of fares falling over time (McAfee and te Velde, 2007; Talluri and van Ryzin, 2004).

Second, carriers may want to discriminate the business travelers' segment from other lower demand travelers, e.g., those traveling for leisure or for visiting friends and family. Because the former are more likely to learn about their need to travel only a few days before the departure date and their demand is quite inflexible, fares are expected to rise over time (Gaggero and Piga, 2011; Alderighi et al., 2016).

Third, a similar increasing fares profile can emerge when customers are strategic and may postpone the purchase in anticipation of last-minute discounts (Deneckere and Peck, 2012; Sweeting, 2012). A commitment to raise fares over time is often necessary to discourage such behavior, unless the probability of a stock-out is high (Möller and Watanabe, 2010).

These three factors operate in different directions, so their relevance should be based on the extent by which their expected impact conforms to the actual temporal patterns of fare data. The literature provides overwhelming evidence in favor of a temporally increasing fare profile (Gaggero and Piga, 2010; Bergantino and Capozza, 2015; Bilotkach et al., 2010); . However, when commenting the findings from their own data analysis, McAfee and te Velde (2007) forcefully state that standard theoretical models predicting a declining time-path due to a falling option value are empirically false.

One of the major contributions of this paper is to show that a temporally declining option value is, contrary to the above literature, a major factor driving a carrier's RM decisions, and hence DP. To unveil such a result, it is necessary to abandon the analysis based on a single fare (notably, that of the seat on sale) so far used in the literature and adopt, as a building block, the notion of fare distribution as in Dana (1999). Loosely speaking, the focus on a fare distribution implies that at each point in time, the airline does not only define the fare of the seat on sale, but also of all the remaining seats on the flight. This approach allows us to extend the literature in a number of directions.

First, we develop a theoretical dynamic model where each seat in the distribution is characterized by a declining value as the departure date approaches, although at each point in time fares increase over seats. By doing so, we extend the results in Dana (1999), by allowing for the carrier's possibility to modify its fare distribution in different, but discrete, time intervals.

Second, we show for the first time in the empirical literature how such distributions are shaped. When seats are ordered with respect to their sequence of sale, we find that the theoretical prediction of fares increasing over seats is confirmed by the data, although the empirical distribution is non-strictly monotonous, that is, the airline arranges seats into groups, denoted as "buckets", where each bucket is defined by an increasing price tag and a variable size.

Third, through the characterization of such distributions at a flight's level, we can determine how DP is implemented in practice. Our assessment of what constitutes DP is different from the one used so far in the literature. Indeed, we do not classify fare increases over time as DP, when such increases are generated from a movement along the distribution.[1] Instead, we consider as an instance of DP only a situation involving an identifiable change in the fare distribution. In a sense, we rule out the fare variations that so far have taken a central role in the literature on DP. More interestingly, we show that DP takes many forms and shapes, involving not only fare variations but, most importantly, changes in the distribution that result in *variations of the buckets' size*, as well as the creation/deletion of new buckets. It turns out that DP associated with changes in the bucket sizes is quantitatively more relevant than changes involving modification of the buckets' fare levels, which tend to remain rather invariant over time.

Fourth, our econometric approach is divided into two parts. In the first, we model the fares of the first seat on sale. Our findings are consistent with previous results in the literature: the fare of the seat on sale follows an increasing temporal profile. To drive this result, however, is the increasing shape of the distribution and the fact that buckets tend to disappear when they are sold out. Although the theoretical model assumes away that consumers act strategically, the increasing fare distribution over seats help to discourage strategic behavior on the customers' side.

In the second part of the econometric analysis, we track the evolution over time of the fare of all the seats, by defining a seat identifier which is both independent from the number of available seats and time invariant. The analysis provides strong empirical support for the theoretical models predicting a declining option value. While similar support has been

[1] Most fare increases can occur without any change in the distribution: when a bucket is sold out, the seats allocated to the next higher bucket are put on sale.

shown in Sweeting (2012) for the price of single baseball tickets sold on the second-hand market, a crucial difference here is to show that, at the same time, *i*) the fare of the seat on sale tends to increase and, *ii*) the airline engages in DP to accommodate the declining value of all the seats in the distribution. This is done by shifting seats initially allocated to the higher-priced buckets to lower-priced buckets, a mechanism that may lead to the disappearance of the higher-priced bucket from the distribution. Because the upper buckets are not normally observed by customers, the carrier can thus engage in "hidden" DP in ways that effectively reduce, if necessary, the average selling fare of all remaining seats, without revealing to have done so.

The rest of the paper is structured as follows. The next section revises the main contributions of both theoretical and empirical literature. Then the theoretical model is presented. The collection of fare data is described in Section 4 followed by a descriptive analysis on dynamic pricing. Section 6 carries on the econometric investigation, testing the predictions of the theoretical model described in Section 3. Finally Section 7 summarizes and concludes.

2 Literature review

Dynamic pricing (DP) is a quite vague concept in the extant literature. Although it usually encompasses any change in prices occurring over time, its various definitions appear to be a direct consequence of the different theoretical, methodological and empirical approaches developed in order to account for the pricing behavior of firms.

In some cases, DP is associated to a price change that is directly linked to at least one intervening factor or event that induces a revision of the pricing approach followed by the firm. For instance, based on this approach, the decreasing prices of Major League Baseball tickets in secondary markets in Sweeting (2012) constitute a clear indication of an active DP intervention by sellers in the form of the decision to relist the ticket at a lower price.

On the contrary, in Abrate et al. (2012) the prices of hotel rooms are found to be either increasing or decreasing over the booking period for stays during, respectively, weekends and week-days. While these differences certainly denote distinct "inter-temporal pricing" profiles, they cannot be classified as instances of DP in terms of the previous definition since they arise from an empirical model that does not specify the source of price variation over time; i.e., because the hotels may have determined them at the start of the booking season, the decreasing or increasing profiles may be the result of a purely time-invariant pricing approach.

In the airline markets, pricing policies are central for any empirical analysis; Borenstein and Rose (1994) distinguish between systematic and stochastic peak-load pricing as sources of price dispersion in the US market. In the former, the price variation is based on systematic, i.e., foreseeable and anticipated, changes in shadow costs known before a flight is opened for booking, while the latter reflects a change in the probability during the selling season that demand for a flight exceeds capacity. In this sense, the former definition of DP and stochastic peak-load pricing may be considered as synonymous. More importantly, the distinction in Borenstein and Rose (1994) can be related to carriers' RM activity, intended broadly as a process of *i*) setting ticket classes, i.e., fare levels and associated restrictions (refundability, advance purchase, business vs. economy, etc.) and *ii*) defining the number of seats available at each fare.[2] RM thus encompasses both a systematic and a dynamic pricing dimension, where the former can be seen as the outcome of the process just before a flight enters its booking period, and the latter represents subsequent changes over time to the initial composition of ticket classes both in terms of fare levels and number of seats in each class.

As far as the systematic approach is concerned, Dana (1999) illustrates how, in a theoretical model with demand uncertainty and costly capacity, it is optimal for firms in to commit to an increasing fare distribution, where each fare reflects the fact that the shadow cost of capacity is inversely related with a seat's probability to be sold. The main ensuing testable prediction from Dana's model is that the fare charged should reflect the ranked position of the seat on sale in the fare distribution. To implement such a test, it is therefore necessary to know a flight's load factor at the time a fare is either posted online or a ticket is sold. This issue has been empirically tackled either by the use of web crawling methods (Alderighi et al., 2015), or of seat maps posted by online travel agents (Clark and Vincent, 2012; Escobari, 2012; Williams, 2013). All these works provide evidence in support to the hypothesis of fares increasing as a flight fills up. Interestingly, Alderighi et al. (2015) derive their results by using two fares, the seat on sale and the last seat in the distribution; their approach is further extended in the present work, where we model the fare for all the seats in the fare distribution.

Because in Dana (1999) firms cannot change the initial distribution they set, the model cannot provide any theoretical prediction on how firms would modify the price distribution

[2] RM involves a number of ancillary activities and techniques useful in the process (McGill and Van Ryzin, 1999; Weatherford and Bodily, 1992).

over time. That is, would all fares start low and then increase or start high and then decrease? The question of the optimal temporal profile of fares is generally addressed in the operational research literature surveyed in Talluri and van Ryzin (2004) and in McAfee and te Velde (2007). A drawback in this literature is that, unlike Dana (1999), either prices or seat inventory levels are treated as exogenous. In fare setting models the focus is on the opportunity cost of selling one unit of capacity, i.e., the value not-to-sell the unit today and reserve it for a future sale. As shown in Sweeting (2012), under standard conditions common to most models, the value of the option not-to-sell is expected to fall over time, leading to a similar prediction for fares. However, because such a prediction arises from models that treat seat inventory as exogenous, it is not possible to extend it directly to the case where the airlines adopt, as the empirical literature suggests, a pricing system based on the definition of a fare distribution over capacity units. In the theoretical model of the next Section, we show that if airlines can revise the fare distribution more than once, then under standard assumptions of demand, customers' evaluations and arrival rates being constant over time, the fares of all the seats are expected to decline over time.

There are a number of reasons proposed in the airline literature as to why fares could increase over time. First, offering advance-purchase discounts can be an optimal strategy when both individual and/or aggregate demand is uncertain (i.e., individuals learn their need to travel at different points in time and airlines cannot predict which flight will enjoy peak demand), and consumers have heterogenous valuations (e.g., they either incur different "waiting costs" if they take a flight that does not leave at their ideal time or they simply value the flight differently).[3] Second, the revenue management models that predict a declining option value assume a constant distribution of willingness to pay, and therefore do not account for the fact that business travelers tend to book at a later stage (Alderighi et al., 2016). Furthermore, those models assume an exogenous demand process and thus abstract from the presence of strategic buyers, i.e., those who maximize long-run utility by considering whether to postpone their purchases hoping to obtain a lower price. In a model characterized by uncertainty, advance production and inter-temporal substitutability in demand induced by strategic behavior, Deneckere and Peck (2012) predict that the prices set by competitive firms are martingales, i.e., they do not follow a predictable pattern. An often observed approach to discourage strategic waiting is to commit to a nondecreasing price temporal path; the counterfactual analysis in Li et al. (2014), based on the estimates on the fraction of strategic consumers in the buyers' population, suggests that commitment

[3]See Gale and Holmes (1993, 1992), and Dana (1998) and Möller and Watanabe (2010).

to a nondecreasing pricing strategy is likely to be more beneficial in business markets than in leisure ones. However, Board and Skrzypacz (2016) show that when a strategic buyer's value declines throughout the selling season, a seller can optimally set time-decreasing prices.

The present work does not aim to distinguish among competing theories of nondecreasing prices in airline markets, but makes the novel point that the adoption of a pricing strategy centered around a fare distribution across seats may have the (possibly indirect) effect for the firm to combine the need to take into account the presence of strategic consumers with the need to reduce fares due to a declining option value. Indeed, we show that, on the one hand, the fare for the seat on sale (i.e., the visible side of DP) tends to follow the non-strictly increasing shape of the distribution, although with occasional markdowns consistent with the prediction in Deneckere and Peck (2012); and, on the other, the firm tends to adjust downward the structure of the fare distribution, regardless of whether they are immediately for sale or not (i.e., the hidden side of DP).

3 Theoretical model

A carrier operates a single flight with $N > 1$ seats on a monopolistic route. The flight is sold over $T \geq 1$ booking periods: $t = T, T-1, \ldots, 2, 1$ describes the number of periods remaining before departure ($t = 1$ is the last booking period and $t = T$ is the first one), and $t = 0$ is the departure date. For each t, the carrier commits to a sequence of fares for all the $M \leq N$ remaining seats of the flight. Thus, until seat $m = M, \ldots, 2, 1$ has not been sold, each traveler presenting in booking period t faces $p(m, t)$. Within the booking period t, once seat m has been sold, then the next fare on offer becomes $p(t, m-1)$. At the end of the booking period t, the unsold seats are offered in the next period, $t-1$, until $t = 1$. Seats available at the end of the last booking period remain unsold.[4]

In each period t a set of consumers $h = 0, 1, 2, .., \infty$ arrives sequentially. The probability that the first consumer arrives in t is $\varphi_{1,t} \in (0,1)$, and that consumer $h+1$ arrives conditional on the fact that consumer h has already appeared is $\varphi_{h+1,t} \in (0,1)$. Consumer (h,t) is myopic and her willingness to pay is a random variable $\theta_{h,t}$, with cumulative

[4]The use of reverse indexes for both periods and seats simplifies the notation and the proofs. It also establishes a direct link to the empirical part of the paper, where the position of seats is counted by starting from the last one. Indeed, contrary to prevailing literature on airline pricing aiming at explaining the change of the first available seat as a function of time before departure and remaining seats, our goal is to trace the evolution of the fare of each seat over time. Thus, our indexing choice is such that each seat has the same index independent of how many seats remain.

distribution $F_{h,t}$ on the support $\left[0, \bar{\theta}_{h,t}\right]$, with $\bar{\theta}_{h,t} < \infty$.

We make the following simplifying assumptions: for any $h = 0, 1, 2, .., \infty$ and $t = 1, .., T$, $\varphi_{h,t} = \varphi_{h+1,t} = \varphi \in (0, 1)$; $F_{h,t} = F_{h+1,t} = F$, with $\bar{\theta}_{h,t} = \bar{\theta}$. Thus, we assume that the process is memoryless and consumers have the same ex-ante evaluation. The probability of selling the first available seat at the fare p is:

$$q(p) = \varphi(1 - F(p)) \sum_{h=0}^{\infty} (\varphi F(p))^h = \frac{\varphi(1 - F(p))}{1 - \varphi F(p)}, \tag{1}$$

where $\varphi(1 - F(p))$ is the probability that consumer h arrives and buys at price p provided that consumers $1, .., h-1$ have previously refused to buy at the same price; and $(\varphi F(p))^h$ is the probability that consumers from 1 to h arrived and did not buy.

The maximization problem of the carrier can be summarized by the following Bellman equation:

$$V(t, M) = \max_p \{q(p)[p + V(t, M-1)] + (1 - q(p)) V(t-1, M)\}, \tag{2}$$

with boundary conditions $V(t, 0) = 0$ and $V(0, M) = 0$, for any $t \in \{0, .., T\}$ and $M \in (0, .., N)$. Unlike the existing literature, the novel approach in equation (2) assumes the possibility that more than one seat can be sold within each t: this implies the need to set always a (possibly different) fare for all the seats on an airplane, which is precisely how carriers operate in practice.[5]

Note that equation (2) entails a trade-off between selling now at least one seat (gaining p and the revenue flow coming from the remaining seats, $V(t, M-1)$), and keeping the capacity intact and postpone the sale to the next period, gaining $V(t-1, M)$.

First order conditions imply that:

$$\Psi(p(t, M)) = V(t-1, M) - V(t, M-1), \tag{3}$$

where $p(t, M)$ is the optimal fare when there are t periods and M seats; and $\Psi(p) \equiv p + q(p)/q'(p)$.[6]

[5] The description of how the resulting fare distribution changes over time is one of the novel aspects in the empirical part of this paper, where we use fares posted by easyJet, a leading European low-cost carrier. Alderighi et al. (2015) find that a similar approach based on a fare distribution characterizes the revenue management of Ryanair. More generally, the need to set a distribution of fares over all available seats appears not to be restricted to low-cost carriers, as hinted by information posted, for instance, on a website like `http://www.expertflyer.com/`.

[6] Second order conditions are satisfied when: $\Psi'(p) = 2 - q(p) q''(p)/q'(p)^2 > 0$. The condition holds

Under $\Psi' > 0$, $p(t, M)$ is unique and can be easily found by inverting (3). Moreover, since $p(t, M)$ only depends on $V(t-1, M)$ and $V(t, M-1)$, the problem described by equation (2) can be easily solved recursively by using equation (3) with the boundary conditions $V(t, 0) = 0$ and $V(0, M) = 0$ (see the Appendix). This property of the model also implies that $p(t, m)$ is independent of the number of available seats at the start of each t.

Define: $\Delta_1 V(t, M) = V(t, M) - V(t-1, M)$ and $\Delta_2 V(t, M) = V(t, M) - V(t, M-1)$. The following proposition provides some standard results in the pricing literature (Gallego and van Ryzin, 1994; McAfee and te Velde, 2007).

Proposition 1 *The value function $V(t, M)$ is increasing in t and M, i.e. $\Delta_1 V(t, M) > 0$ and $\Delta_2 V(t, M) > 0$.*

Consider the two following special cases. First, assume that $N = 1$. Using Proposition 1, equation (3) and the fact that $\Psi' > 0$, it follows that $p(t, M)$ is increasing in t. This result captures the perishable nature of the airline service, and the fact that the option value decreases over time. Second, assume that $T = 1$; using the same line of argument, it follows that $p(t, M)$ is decreasing in M. This result captures the fact that a higher price reduces the likelihood to sell a given seat, and, consequently, the following seats.

In the next Proposition we show that these properties also hold for the general case of any T and N. We assume that $\Delta_1 V(t, M)$ is decreasing in t and increasing in M; and $\Delta_2 V(t, M)$ is increasing in t and decreasing in M. In the Appendix we show that these assumptions are satisfied when the willingness-to-pay of travelers is uniformly distributed.

Proposition 2 *The fare profile $\{p(t, m), t = 1, \ldots, T; m = 1, \quad \ldots, N\}$ has the following properties:*

1. *Invariance: $p(t, m)$ is independent of M.*

2. *Ascending fare profile: $p(t, m)$ is decreasing in m.*

3. *Decreasing fares over booking periods: $p(t, m)$ is increasing t.*

***** **Insert Table 1 around here** *****

when θ has an uniform or an exponential distribution.

The results of Proposition 2 are illustrated in Table 1, which presents the simulated fares in three different cases: one period ($T = 1$), three periods ($T = 3$), and five periods ($T = 5$).[7]

Result 1 of Proposition 2 implies that, conditional on seat m being available, its fare is not affected by the number of seats available on the airplane. Table 1 therefore always reports the fare distribution for all N seats: the Proposition indicates that the optimal fare of, say, seat $m = 9$ at $t = 3$ when $T = 5$ is always 0.516 regardless of whether at t the number of available seats is greater or equal to 9. This result depends on the fact that travelers' arrivals are independent and therefore, within each period t, only subsequent fares, $p\,(m-1, t, \ldots, p\,(1, t))$, but not previous fares, $p\,(M, t, \ldots, p\,(m+1, t))$, if any, affect the optimal level of $p\,(m, t)$.

Moving from the top (first available seat, i.e., seat $m = 12$) to the bottom (last available seat, i.e. seat $m = 1$) of each column, it appears that the fare distribution is increasing both in the one-period and in the multi-period cases. Thus, in any period, consumers who arrive first pay less than those showing up later (Result 2). This is a notable difference from Board and Skrzypacz (2016), where a seller can charge only a single posted price in each period.

An ascending fare profile is not novel in the theoretical economic literature, but the explanation proposed here provides interesting extensions. For example, in Dana (1999), where updates to the equilibrium price distribution are not considered, fares are proportional to the implicit cost of capital, and therefore, they are inversely proportional to the probability of selling. Since the last seats are less likely to be sold, their implicit cost is higher, and so is their fare. In this setup, however, an increasing fare distribution comes from the fact that the higher the fares, the more unlikely the sales of both current and subsequent seats; that is, a high fare for the seat on sale increases the opportunity cost of having to sell tomorrow all the subsequent seats.

The third result in Proposition 2 is illustrated in Table 1 by values of $p\,(t, m)$ declining over t for any m. This result extends the one-period case considered in Dana (1999) by showing that the carrier's option value decreases as the departure date approaches. This is standard for highly perishable services, as illustrated in Sweeting (2012), where however the analysis is limited to the case of a single ticket and not to a full distribution of prices as in the present case.

[7]In the simulations, we set: $N = 12$; θ uniformly distributed over $[0, 1]$; $\varphi = \{0.9796, 0.9412, 0.9057\}$ for, respectively, $T = \{1, 3, 5\}$. The values of φ are defined such that the expected number of consumers is the same in the three cases and equal to $4N = 48$.

The results in Proposition 2 offer several new empirical implications that we test in the remaining part of the paper. There are however two issues that the theoretical model assumes away: the possibility of strategic consumers and the fact that there is no learning on actual demand during the booking period. In the final part of this Section we discuss under what conditions each aspect is likely to play a minor role in driving the empirical results.

Strategic consumers. Proposition 2 shows two contrasting trends as far as the seat on sale is concerned. On the one hand, according to Result 2 within the same period the fare of the next seat is higher than the one on sale. On the other hand, Result 3 states that the fare of a given seat reduces over periods. Thus, the fare of the seat on sale moves up during the same period and down over period switches.[8] Therefore, the price reductions may be potentially conducive to strategic behavior because the consumer arriving when seat M is on sale at time t would always prefer to buy it at $t-1$. However, by postponing the purchase, the consumer faces the risk that, at time t, other consumers may arrive and buy M and some or all subsequent seats. That is, if a consumer expects that the fare of seat $M-1$ will be, on average, higher than that of seat M, then strategic behavior is discouraged.

Let $\pi(m,t)$ be the probability of selling seat m in period t, and

$$\bar{p}(m) = \sum_{t=1}^{T} \pi(m,t)\, p(m,t) \Bigg/ \sum_{t=1}^{T} \pi(m,t) \ , \tag{4}$$

the average fare paid for a given seat across all periods. Table 2 reports both $\pi(m,t)$ and, in the last column, $\bar{p}(m)$, based on the simulation values of Table 1. Here we notice that the average paid fare, $\bar{p}(m)$, is increasing over seats across periods.[9] Thus, the incentive to postpone a purchase is hindered by the increasing trend of the seat on sale. There are two extra practical reasons indicating that the empirical robustness of our theoretical predictions are only weakly affected by the presence of strategic consumers in real markets. First, all prices close to the departure date would be higher under the assumption that $\theta_{h,t}$ increases in t, e.g., because business travellers arrive in later periods (simulations of this version of the model are available on request). Second, from an empirical viewpoint, there is indeed evidence suggesting that the proportion of strategic consumers in airline

[8]Price drops are often observed and their impact on realized load factors is studied in Bilotkach et al. (2015).

[9]The result of an increasing average fare profile for $\bar{p}(m)$ is robust to changes of the number of periods as well as φ.

markets is rather limited, around 12 percent on average Li et al. (2014). Overall, our model is robust to the presence of a small fraction of strategic consumers.

***** **Insert Table 2 around here** *****

New information. Usually carriers set their fares on the basis of three different sources of information: historical data; internal data collected during the booking period; and external data. Historical data are information available to a carrier before its price setting decision; in the model, they are used to determine φ and θ. External data are information on demand shifters (e.g., such events as concerts, football matches, etc.) revealed during the booking period. If such information corresponds to an unexpected demand shock, it can be easily accommodated in the model by assuming that a carrier, after receiving it, redesigns a new fare profile based on new values for θ and φ. Basically, external data produces a positive or a negative shift of the fare profile from the moment the carrier processes the information onwards.

A different situation occurs when a carrier adjusts its fare profile based on internal data collected during the booking period, i.e., the number of the sold seats at each given point in time. Consider a simple case with two periods, $T = 2$, and two states of the world: high demand (φ_H) and low demand ($\varphi_L < \varphi_H$). At time t=2 (initial period), the carrier has a prior subjective probability of being in the high demand state: $\nu_{2,H} = \Pr(\varphi_2 = \varphi_H)$. Let $p(2, m; \nu_{2,H})$ be the fare profile chosen in $t = 2$. At $t = 1$, depending on the available seats, M, the carrier updates its expectations on the demand, $\nu_{1,H}(M)$, and sets a new pricing profile accordingly. That is, the fares set at $t = 1$ for any available seat $m = M, \ldots, 1$ are no longer independent on the number of available seats M observed in the previous period, i.e. $p(1, m; \nu_{1,H}(M))$. In particular, since a higher M signals a lower demand, then $p(1, m; \nu_{1,H}(M))$ is decreasing in M. Thus, for any available m, the larger number of seats before m, namely $M - m$, the lower the fare. Although in our theoretical model fares are set independently of the previous selling history (Result 1 of Proposition 2), in our econometric analysis we control for the possible endogeneity of fares with respect to a flight occupancy at the time fares are posted.

In the remainder of the paper, we develop an empirical analysis aimed at testing the predictions of the model summarized by Proposition 2: fares are decreasing over M (Result 2) and increasing in t (Result 3). Moreover, we also discuss the assumption of independency of fares (Result 1) by showing that fare profile is weakly affected by the number of seats before $M - m$.

4 Data collection

Our collected sample comprises a total of 37,501 flights scheduled to depart during the period May 2014 - June 2015, covering 74 European bi-directional routes. The fares for those flights whose outward journey originates in the UK are expressed in British Pounds and represent about 99% of the entire sample. The residual 1%, which refers to European routes outside the UK, is collected in euro.[10]

The data collection was carried out by means of a web crawler, as widely used in the literature.[11] Every day, the crawler automatically connected to the website of easyJet, the second largest European LCC, and issued queries specifying the route, the date of departure and the number of seats to be booked. Because European LCCs charge each leg independently (Bachis and Piga, 2011), to double the data size, the query was for a return flight, with a return date 4 days after the departure.[12]

The query dates were set such that a flight entered our database about four months before departure; it was then surveyed at 10-days distanced intervals until 30 days, and subsequently at more frequent intervals (21, 14, 10, 7, 4 and 1) to get a better understanding of the price evolution as the date of departure nears. The website response to the query included flight information, for each leg, for three different dates: the set date, the day before and after. Overall, each query allowed the saving of three consecutive days' information for each leg. For each flight, the crawler saved the date of departure and of the query (to calculate the number of days separating the query date from take-off), the time of the day the flight was due to depart and arrive, the departure and arrival airports (the route), the price for the number of seats specified in the query. The crawler also saved an important information published by the carrier: the number of seats available at a given posted fare. This is central for the validation of the data treatment implemented

[10] When necessary fares in euro are converted in pounds using the daily Eurostat exchange rate of the day when the fare is collected. See http://ec.europa.eu/eurostat/web/exchange-rates/data/database. Saturdays and Sundays adopt the exchange rate of the previous Friday.

[11] See Li et al. (2014), Gaggero and Piga (2011), Clark and Vincent (2012), Obermeyer et al. (2013), Escobari (2012), Escobari and Jindapon (2014), Bilotkach et al. (2015), Alderighi et al. (2015) and Alderighi et al. (2016), amongst others.

[12] As in the case of Ryanair in Alderighi et al. (2015), easyJet, the low cost carrier under the present analysis offers seats where buyer's name and dates can be changed only by paying a fixed fee which is often as high as the fare itself. The carrier also offers a "Flexi" fare, corresponding to the basic fare we retrieve plus a set of add-ons (extra luggage, cancelation refunds etc), which however can also be bought independently. Furthermore, there is no pricing-in-network considerations to account for, because the carrier only sell tickets for point-to-point services. Data from a low cost carrier thus originate from an environment that more closely resembles the assumptions made in many theoretical models of airline pricing.

to derive the price distributions from the posted fares, as illustrated in the Appendix.[13]

To the best of our knowledge, the empirical literature on airline pricing focuses on the price of one seat, that corresponding to the seat being on sale at the time of the query. A central contribution of this paper is to show that this is not sufficient to test the implications of theoretical models of DP in airline markets. Based on the model presented in Section 3, our data collection incorporates an experimental design explicitly aimed at recovering a flight's price distribution, as it is actually stored on the carriers' web reservation system. In practice, this entailed the implementation of the following procedure. For each flight and departure date, the crawler started by requesting the price of one seat, and then continued by sequentially increasing the number of seats by one unit. The sequence would stop either because the maximum number of seats in a query, equal to 40, was reached or at a smaller number of seats. As in Alderighi et al. (2015), the latter case directly indicates the exact number of seats available on the flight on a particular query date, which we store in a variable called *Available Seats* to track how a flight occupancy changes as the departure date nears. The former case corresponds to a situation where we know that at least 40 seats still remain to be sold on a given query date; i.e., the number of available seats is censored at 40.

After applying the treatment described in the Appendix to the retrieved fares, we obtained the flights' distribution of posted fares over the available seats on a query date. An example of such distributions is shown in Figure 1, which is based on the data of a randomly selected flight.[14]

**** **Insert Figure 1 around here** *****

Figure 1 is central for the whole analysis. Each graph represents the price distribution retrieved, respectively, 90, 50, 30 and 9 days to departure. In all cases, the number of available seats is censored to 40; i.e, the graphs fail to show the extreme right tail of the price distribution. As it appears clear, at each point in time the distributions across seats conform with the second theoretical result in Proposition 2, although in a non-strictly monotonous way, i.e., the carrier generally assigns the same price to more than one seat.[15]

[13] The possibility that posted fares could be affected by the number of queries executed was managed as follows. First, the cookie folder was cleaned every day; second, we checked a sample of fares retrieved by the computers in our university office with queries made on the same day from computers outside that university. No noticeable differences between the queries made from different computers could be found.

[14] This is the flight code U25293 leaving London Gatwick and going to Milan Malpensa on 19 May 2014. We will consistently refer to this flight as an example throughout the paper.

[15] The step-wise shape in Figure 1 can be easily reconciled with the strictly monotonous distribution

We denote as a "bucket" the set of seats carrying the same price tag.[16] Interestingly, such tags seem to remain very stable over the booking temporal horizon, an aspect that we investigate later on.

A visual inspection is sufficient to establish some interesting features of the distributions and their evolution over time. Ninety days to departure, the carrier had allocated three seats for sale at the price of £41 (the per-seat price that a customer buying up to 3 seats would pay), four seats at the price of £47, and so on and so forth. Due to the data censoring, we cannot ascertain the precise size of the last "bucket" valued at £156. Similarly, the size of the £41 bucket is likely not correct, since there may be missing seats from that bucket that were previously sold. Forty days later, the first two buckets have disappeared; only one seat is available at the price of £65 and the size of the £156's top bucket has clearly increased to at least 23 seats, although the censoring still prevents us to precisely measure its size. Interestingly, twenty days later, the size of the first bucket has increased to four seats, and that of the top bucket has fallen to 20 seats (still censored); the size of the intermediate buckets did not vary. Nine days prior to departure, the carrier is still offering eight seats at the price of £65, but noticeably, the size of also the intermediate buckets has increased while that of the top one has shrunk.

It is stressworthy that in our database *all* the series of fares associated to the 40 (or less in the case of non-censored flights) seats of a query assume a shape qualitatively similar to the ones shown in Figure 1: this is the case for *all* flight-codes and date of departure combinations. Such overwhelming evidence implies that the airline's pricing approach always defines a fare for all the seats of a flight, as predicted in our Proposition 2. This is an important contribution to the literature on airline pricing, that so far has empirically neglected this aspect of the airlines' fare setting. It could be argued that this is a peculiar approach followed by the carrier we used to create our sample. However, the empirical findings in Alderighi et al. (2015) also suggest that another European low cost carrier, Ryanair, defines a similar fare distribution across seats. As far as legacy carriers are concerned, the analysis is complicated by their adoption of a nested-classes

derived in Proposition 2 and reported in Table 1, if we assume that the former is a discrete version of the latter. The reason to make such a transformation can be a combination of technological and marketing factors, due to the way the computer reservation systems and RM systems have been historically designed. Thus, for each flight, a carrier usually uploads in the reservation system from 10 to 15 fare levels and use a revenue management system to determine how many seats to make available at each fare level over time.

[16] The term is drawn from the revenue management literature (McGill and Van Ryzin, 1999; Weatherford and Bodily, 1992). It is noteworthy that the buckets satisfy the Result 2 in Proposition 2. The possibility of bulk discounts is ruled out by the identification of the same bucket prices regardless of the number of seats in the query.

system, where the same seat can belong to different classes, each with different ticket restrictions; however, various papers present graphical evidence of fares whose temporal path also follows a step-wise pattern, with each step representing a class, i.e.,"bucket", level (Escobari, 2012; Lazarev, 2013; McAfee and te Velde, 2007; Puller et al., 2009).

Finally, the analysis of Figure 1 suggests the following considerations. First, the buckets' price levels tend to be quite fixed over time; second, the variation in the distribution of seats across buckets, especially in terms of movements from the top buckets towards the lower ones, appears to be a central aspect of the carrier's revenue management. So far the literature has failed to notice both considerations, which we will investigate further in the remainder of the paper.

5 Descriptive analysis

The descriptive analysis in this Section aims to provide some insights into the following two interrelated questions:

- how is DP implemented when the price equilibrium is in distribution?
- How is the fare distribution affected by the simultaneous presence of two conflicting incentives for the firm, one induced by the perishable nature of seats in a flight, the other arising from the presence of strategic consumers who would quickly learn of possible last-minute discounts?

The first question is addressed by first defining the various forms of DP that we can identify in our data and then providing a descriptive quantification of their importance; the second question is directly related to the third prediction in Proposition 2 that the opportunity cost for all the seats in the distribution is decreasing over time, an aspect that would be certainly exploited by strategic consumers.

5.1 Defining Dynamic Pricing

As Figure 1 suggests, DP clearly goes beyond the mere fluctuation of the price of the first seat in the distribution. One of the novel aspects of this paper is to show that DP entails a restructuring of the fare distribution, and that this practice may involve either a modification of the buckets sizes (i.e., a reallocation of remaining seats across buckets) or the creation/deletion of bucket levels, or both. To fully capture this behavior, we still refer to the flight data used in Figure 1, as reported in Table 3. Each cell contains the bucket size, with columns identifying the days prior to departure and rows the bucket price. The last row in each sub-panel indicates whether the number of available seats is censored (that is, there are 40+ seats left on the flight) or the precise number of available seats (this is visible only from six days onwards in Panel B, when the maximum number of prices observed is for 30 seats).

We define as dynamic pricing any change in the distribution of seats across two sequential query dates. That is, we do not restrict DP to be associated to price fluctuations over time of a single seat, but explicitly consider variations in all bucket sizes as relevant forms of DP, and therefore of revenue management activity.

Table 3 provides examples of some of the forms of dynamic pricing implemented by the carrier. To better identify them visually, we use circles to denote cases of DP associated with bucket size changes, and with rectangles the more standard DP cases of creation or deletion of a bucket price level. Note that we do not identify as DP the price increases that automatically occur when the first available bucket becomes sold out and the system moves to the next available bucket level.[17] Thus, our definition admits cases in which there is no DP even if the fare of the seat on sale varies, as well as situations in which there is dynamic pricing even if the fare of the seat on sale does not change.

******Insert Table 3 around here*******

Based on the content of Table 3, we can distinguish the following forms of DP.

- <u>Size increase of first bucket</u>: the bucket on sale has at least one seat more than in the previous observation. E.g., at 29 days before departure, the size of the £65's bucket increases from 4 to 5 seats. We likely underestimate this form of DP, because we cannot detect a size increase when the number of seats sold between two consecutive observations is larger than or equal to the size growth and the number of available seats is censured.

- <u>Price decrease of first bucket</u>: the fare of the seat on sale drops from the previous observation. For instance, at 90 days to departure the fare of the first bucket falls from £47 to £41. This could be also construed as a form of "size increase", because the size of the £41's bucket goes from zero to 3. However, to emphasize the fact that

[17] That is, if we observed that from, say 89 days to departure, the sequence of the first seat's price followed the bucket price levels $(41, 49, \ldots, 136, 156)$ without any change in the buckets sizes, then we would argue that no DP has taken place because the fare distribution has not been altered: simply the first fare evolved according to the sequence embedded in the fare distribution.

the fare changes of the first seat are more easily observed by customers, we prefer to denote this form of DP as a price change. Furthermore, it can also indicate the creation of new buckets, as in the case of the £58's bucket, which is not observed before 13 days from departure.

- Price decrease of last bucket: the fare of the last seat drops from the previous observation. This form of DP can be identified only when the data are not censored, as in the case of 2 days before departure, where the top bucket valued at £156 disappears.

- Size decrease of last bucket:the last bucket has at least one seat less than in the previous observation. As in the previous case, this form of DP can be traced only when the data are not censored. One example occurs 3 days to departure, when the £156's bucket size drops from 13 to 5 seats.

- Size increase of the second bucket: the second bucket has at least one seat more than in the previous observation. See for instance the second bucket size at 2, 4 and 21 days before departure.

- Size change of the penultimate (observed) bucket: it denotes either an increase or a decrease in the number of seats assigned to the penultimate or penultimate observed bucket, where the latter indicates that, when data are censored, the penultimate bucket we observe may not actually be the second-to-last. Indeed, in the case of censored data, there could be other higher buckets that the crawler could not access due to the censoring restriction. Examples occurs at 61, 21 and 13 days before departure.

- Any change to intermediate buckets: it accounts for any form of price and/or size change of any bucket between the second and the penultimate (observed) bucket. See for example 101, 91 or 61 days before departure.

Furthermore, in our sample we observe other additional cases of DP interventions that are not represented in Table 3.

- Price increase of first bucket only: the fare of the first bucket increases, whilst all the other buckets remain unchanged both on price and size. For the sake of exposition, suppose that in Table 3, at 59 day before departure the four seats in the £55's bucket are moved to the (previously unobserved) £58's bucket while the rest of the fare distribution remains unchanged.

- Price increase of last bucket: the fare of the last bucket is higher than that of the previous observation. It is identifiable only for uncensored data.

- Size increase of last bucket: This form of DP is symmetrical to the case of size decrease of last bucket. It is identifiable only for uncensored data.

- Size decrease of the second bucket: This form of DP is symmetrical to the size increase of the second bucket.

5.2 Descriptive statistics on Dynamic Pricing

Tables 4, 5 and 6 report the probability of observing the various forms of DP defined in the previous subsection. These were calculated by considering only variations between query dates separated by one day (e.g., in Table 3, between 3 and 2 days to departure). The qualitative results do not change if the probabilities were obtained considering variations as highlighted in Table 3, i.e., between any two consecutive, but not adjacent, query dates. An overall measure of DP, "Overall DP", which groups all the forms of DP into one case, is also included in the Tables. The probabilities in each table are broken down by different categories that identify different sub-samples: final number of left seats (Table 4); days to departure (Table 5); and fixed and varying flight occupancy (Table 6).

Table 4 defines the sub-samples based on final number of left seats, i.e. denotes the minimum observed number of available seats for a flight: this would be equal to 22 in Table 3.[18] The table suggests the following considerations. First, most DP treatments appear to be applied with a similar frequency across flights, regardless of their impact on the final occupancy rate; a notable exception are the "Price increase" and the "Size increase" of the first bucket, with the former (latter) more likely to be observed in flights with higher (lower) final occupancy rates.

Second, the "hidden side" of DP is revealed by the fact that "Size decreases" of the last bucket tend to be the most frequent type of DP; as previously discussed, seats initially allocated in top buckets are transferred to lower buckets: in the table, this does not seem to be different across flights with a different final number of left seats.

Third, if we use "Overall DP" as a generic proxy of the intensity of DP treatments, the table suggests that flights with different final occupancy rates have received similar treatments' intensity levels, with a frequency of 42% per cent. This result supports the idea that changes in the fare distribution usually occur, on average, every two-three days.

[18] Note that all the flights in the sample at some point reveal an uncensored number of available seats.

Finally, a change in bucket size is more likely observed than changes in fare. This result provides support to our approach that DP should not be based on fare fluctuations only.

***** **Insert Table 4 around here** *****

Table 5 is based on the distance between the query date and the departure date. Table 5 indicates that dynamic pricing interventions are more likely observed between 42 to 8 days before departure. As we have already mentioned, they are largely represented by a size decrease of the last bucket and an increases in the size of first and other buckets. DP is implemented more intensively in such period because there is still some time left for it to have an effect on sales. Moreover, the probability to observe a price decrease in the first bucket falls sharply in the last week, when fare increases are more likely applied.

***** **Insert Table 5 around here** *****

Table 6 defines the sub-samples based on the number of available seats at the query date; the "same occupancy" case refers to when the number of available seats does not vary between two consecutive, one-day distant, query dates. In both cases, the sample is determined only by uncensored observations, so as to control whether the number of available seats has changed. Table 6 points at the following considerations.

First, DP (i.e. a change in the fare distribution) may occur even when the carrier has not sold any seat, or, conversely, it may not occur when one or more seats are sold. In both cases, the pattern of the change in the fare distribution is quite similar for the intermediate buckets, while there is a different behavior concerning the first bucket and, partially, the last bucket. This suggests that general (automatic) changes are independent from the selling situation and that a direct intervention of RM analysts is mainly focused on the first and the last buckets.

Second, as the available seats reduce, most forms of DP are less likely observed, suggesting that modifications to the fare distribution are likely driven by the need to boost a flight's occupancy rate. The direction of these findings is in line with standard interpretation: e.g., when occupancy is unchanged, the probability of a "Price decrease" of the first bucket is two and a half times higher when at least 21 seats remain on the flight.

Third, "Price increases" of the first bucket are again notable exceptions: the selling price is more likely pushed up as the number of available seats lowers. This is consistent with the idea that the RM analysts adjust the fare upward for the seat on sale if the flight is selling well. Symmetrically, the probability of a “Price decrease” in the first bucket drops substantially when only few seats remain.

Finally, “Size increases” of the first bucket appear to be inversely related with the occupancy rate but highly correlated with the “Size decreases” of the last bucket: the “hidden side” of DP conforms to the general interpretation that modifications to the fare distribution implemented by transferring seats from top to lower buckets is motivated by a downward revision of the analyst’s belief on the flight’s demand. Indeed, this form of DP is more likely observed when occupancy is unchanged and there are at least 21 available seats.

*****Insert Table 6 around here *****

A combined analysis of Tables 4-6 suggests that:

- DP is driven by the need to manage inventory effectively; it does not however lead necessarily to identical level of final occupancy rates across flights.
- DP takes many forms and shapes, all aimed at redesigning the distribution of fares uploaded on the carrier’s reservation system.
- In particular, a decrease in the size of the last bucket suggests that the carrier’s option value to continue to retain seats in that bucket is a decreasing function of time, up to the point that the carrier may decide to eliminate a top bucket from the fare distribution, as it occurs 2 days prior to departure in Table 3.

5.3 Option Value in the presence of strategic consumers

An airline seat is a perishable product: because there is no value for the firm to take-off leaving some seats unoccupied, the incentive to carry out last-minute sales is strong. Similar to the case described in Sweeting (2012), we would therefore expect that the carrier’s fare distribution is decreasing over time. This is also one of the main predictions from our theoretical analysis in Section 3.

However, as shown in Li et al. (2014), the growing presence of strategic consumers complicates the carrier’s pricing problem: if it systematically dropped its prices a few days before departure, customers would react by postponing their purchase decision. The incentive to drop prices is likely to be even stronger in the airline market than in the case described in Sweeting (2012), where potential buyers of second-hand tickets for a baseball match face a high risk of a stock-out, a contingency less likely for travelers, especially those whose main purpose is leisure, who can substitute across dates or even destinations and still satisfy their wish to travel.

The previous analysis suggests the solution adopted by the carrier to manage this dilemma: the reshaping of the fare distribution, with seats reallocated from top buckets to lower ones, is consistent with the firm's belief that demand is not sufficiently high to fill the plane with enough customers willing to pay the higher prices set for those seats initially positioned in the higher buckets. At the same time, the way the distribution is designed, with its increasing profile over seats (see Figure 1 and Table 2), guarantees a commitment not to lower prices if demand is low: e.g., in Figure 1 and Table 3, the selling price gravitates around the value of £65 for about 40 days, simply because the firm during that period moved seats initially allocated to higher buckets down to the £65 bucket and, occasionally, also to lower ones. If 50 days from departure, when only one seats remained in the £65's bucket, the carrier had kept the distribution unchanged, and allowed the price to automatically increase to the higher prices of £75 and then £87, most likely fewer seats would have been sold. We can indeed infer that demand for this flight might not have been very high since two days prior to departure the flight still had 22 seats to sell. Furthermore, letting the fare go up to, say, £87 and then drop it back to the more suitable level of £65 would show customer's that the carrier is very likely to engage in price reductions, thus providing more justification for customers to behave strategically and postpone purchase.

To test formally that the above solution is widely applied, we need to show that:

1. The fare of the first seat on sale (i.e., the one customers normally observe) follows a generally increasing time trend, which is not incompatible with some occasional, and largely unpredictable, fare decreases (Tables 4, 6 and 5). This would takes care of strategic behavior: if customers systematically observe the increasing trend, as predicted by Table 2 of our theoretical model, the best response is to buy at the time they become aware of their need to travel.

2. The seats allocated to the top buckets are systematically transferred to lower buckets, i.e., the carrier over time reduces their selling fare to reflect the changing probability they will be sold. This is largely unobserved by customers visiting the carrier's website (*"the hidden side of DP"*).

To test this latter aspect, we exploit a feature of our data collection strategy. With uncensored observations, we can establish a seat's position in the distribution. Imagine that only 39 seats remain on a flight on a given query date. If we look at the distribution from the bottom up, the first seat is the one on sale, and the 39^{th} identifies the "last" seat that would be put up for sale.[19] Imagine that a few days later, the number of available seats drops to 30. In this case, the first seat is indeed the one that occupied the 10^{th} position in the previous query date, and the last seat would be now the 30^{th}. That is, it is not possible to use the bottom-up perspective to uniquely identify seats. However, if we assign the position using a top-down approach, it turns out that in both cases the last seat would be assigned a position equal to 1, the 10^{th} seat would take a position equal to 30, and the first seat in the first case a position of 39. We report these values in a variable denoted as *Position*.[20] That is, over different query dates, we can track the evolution of each seat's fare, as long as the observation is uncensored.

Table 7 reports the mean fare of the same seat (i.e. with the same value of the variable *Position*) at various clusters of days to departure. The numbers clearly indicate a decreasing pattern of the mean fare as we approach the departure day. Interestingly, and in line with the prediction of our model, the monotonicity of such decreasing pattern is confirmed for the last seats (i.e those with a low value of *Position*), while it is strong for the seats in lower positions from the top. Quite interestingly, the decline appears to be inversely proportional to the position. That is, the last seat ($Position = 1$) drops from an average fare of £182 to £139; the 20th seat from the top also has a starting mean of £181, which falls drastically down to £82. This is consistent with what we observe in Table 3: while the top seat is moved only one bucket down (2 days from departure), many seats that are originally allocated to the top bucket of £156 are moved down by several buckets.

***** **Insert Table 7 around here** *****

Although the values in Table 7 strongly suggest the impact of hidden DP to be mostly geared towards a downward shift of seats allocated to higher buckets, it is important to point out to at least one counter-example in our dataset, where we show fares moving in precisely the opposite direction. This is likely to happen in unexpected situations of high demand. The example in this case is based on the final of the football European Champions' Leagues, played by Real and Atletico Madrid in Lisbon (Portugal) on May 24^{th} 2014. We therefore look at the flight connecting Madrid with Lisbon the day before.

[19] Recall that the seats are all homogenous.

[20] We are using the same notation of the theoretical model where *Position* is identified by the reverse index m.

Real Madrid and Atletico Madrid qualified on, respectively, April 29^{th} and 30^{th}, i.e., 24 and 23 days before the flight departure. As Figure 2 shows, on April 30^{th} the crawler collected fares for the first forty seats on sale, that is, the flight was still censored with the first fare being set at €97. Noticeably, five days later, all the low priced seats disappear; the lowest fare moves up to €184, which was the highest bucket only five days before, and about 20 seats are allocated to the new highest bucket of €307. Although some seats from the top bucket are shifted down between eighteen and twelve days to departure, an example of increase in the last bucket size occurs between twelve and eight days, when the number of seats in the top bucket is increased from five to ten. Eventually, the flight departed with only four empty seats in the top bucket. Overall, both the levels of fares involved, which are much higher than those in Table 7 even after the conversion in British Pound is applied, and the increase in the size of the top bucket suggest that the rule of a declining option value may present many exceptions when new information induces the carrier to adjust its beliefs on demand upward. It is not possible, however, to identify all these exceptions in the data, given the temporal interval and the several routes covered. In the remainder of the analysis, therefore, we will focus the investigation on the prevailing effect of hidden DP on the time profile of all the fares in the distribution.

***** **Insert Figure 2 around here** *****

6 Econometric analysis

The econometric strategy is designed to test formally the two aspects discussed in the previous sub-section.

First, we want to show that the fare of the first seat on sale follows an increasing temporal profile determined by the structure of the bucket fare levels in the distribution. That is, generally the carrier tends to close a bucket fare once the seats in that bucket are sold out, so that automatically the fare of the next bucket becomes the advertised one on the site. Second, we formally test the hypothesis that the fare distribution decreases over time because the option value lowers as the departure date approaches.

Because, as explained before, the first seat is defined by taking the first value on the left tail of the distribution of remaining seats, while the option value necessitates the opposite approach where *Position* is counted starting from the top of the right tail, the two issues are tackled separately.

6.1 First seat on sale

Table 8 reports various estimates derived by restricting the sample to the first available seat; i.e., the one with the lowest fare.

The dummies *Days to departure* aggregate fare observations having similar temporal intervals between the query and the departure date. The reference group corresponds to the set of query dates that are furthest from departure. If fares tend to follow an increasing, albeit non-strictly monotonic, profile, then the coefficients of all the dummies should increase as the query date approaches the departure date.

The standard errors are clustered by route and week to take into account the possibility of flight-specific demand shocks on a given day affecting the demand for all the flights on the route in a given week.[21]

Model (1) estimates an order probit model in which the regressand is the *Bucket order*. This variable represents the rank of the seat on sale (lowest fare has rank 1) within a given flight In our illustrative example of Table 3, *Bucket order* is equal to 1 when fare is £36, 2 when fare is £41, and so on and forth; finally, it equals 8 when the fare is £87, the sale fare two days before take-off. Departure time and departure day of the week fixed effects are included in the model, but not reported in the table to save space.

Model (1) supports the view that the rank of the bucket to which the first seat belongs increases over time; i.e., the temporal sequence in which seats are sold is determined by the design of the bucket levels in the fare distribution. The predicted effects of the model are represented in Figure 3 which reports the predicted probability of observing the bucket orders 1, 2 and 8 as a function of the distance from departure date. As the diagram shows, seats allocated to the lower bucket orders (i.e., the low-fare buckets) are more likely put up for sale in the early booking period and are never sold a few days from departure, whilst seats in higher bucket orders (e.g., 8 in the Figure) are more likely put up for sale towards the end of the booking period, but never at an early stage.

***** **Insert Figure 3 around here** *****

Models (2)-(5) derive from a panel OLS fixed effect estimation of the first seat fare, in logs, with the panel identifier corresponding to the combination of flight-code plus day of departure, while the temporal dimension is denoted by the number of days prior to departure. That is, each panel tracks the fares posted over the booking period of a flight

[21] For instance, a large group booking for a Wednesday morning flight raise fares for this flight and may induce other customers to select alternative flights on nearby days.

on a specific route that departs on a specific date and time of the day.[22] Although in these models the actual fare replaces the bucket order, we still obtain that the inter-temporal profile of the first seat follows an increasing trend.

While models (2) and (5) use the full sample, models (3) and (4) restrict the analysis of model (2) to the case of flights in, respectively, "Leisure" and "Business" routes. Following Alderighi et al. (2016) and Gaggero and Piga (2011), the routes' classification is based on data derived from the "International Passenger Survey" (IPS), a quarterly survey collected by the UK Office of National Statistics.[23] Routes are classified based on the passengers' stated travel motivations. For each flight, we computed the share of business travelers carried by all companies on the city-pair comprising the route where the flight operates. Depending on whether such a share is below or above the value of 16 per cent, routes are respectively labeled as "Leisure" or "Business". Based on the findings in Alderighi et al. (2016), we should expect a steeper temporal profile for flights in business routes, due to the larger proportion of price-inelastic travelers who are more likely to book only a few days before departure. Indeed, based on the estimates of the *Days to departure* dummies, we find that in both types of routes evolve in a similar way until about 29 days to departure, but then increase more sharply in business routes. Such a robustness check does not impinge on the main message of this Section, i.e., that the fare of the first seat follows an increasing trend over time, which the finding in model (1) indicates to be strongly associated with the bucket design of the fare distribution.

***** **Insert Table 8 around here** *****

To sum up, all models in Table 8 are consistent with existing evidence in the literature supporting the view that fares increase as the departure date approaches (Bilotkach et al., 2010; Gaggero and Piga, 2010; Koenisgsberg et al., 2008; Mantin and Koo, 2009; Alderighi et al., 2015). On the one hand, our approach suggests that this is mainly due to the allocation of seats into buckets of increasing fare: once all the seats in a bucket are sold, the bid fare moves to the next higher bucket level. Our results thus provide a so far undetected perspective, that is, they directly relate the evolution of the selling fare to the design of the fare distribution for all the seats available on a flight at each point in time.

On the other hand, a time-increasing time profile may be beneficial in a situation where customers are likely to behave strategically and may be heterogeneous in their

[22] If there is more than one flight in a day, each is treated with a separate panel identifier.

[23] The IPS does not cover routes with both endpoints outside the UK; hence, the combined number of observation in models (3) and (4) is lower than in model (2) and (5).

willingness to pay: for instance, business travelers may learn about their need to travel only a few days before a flight's departure. That is, the pricing based on an increasing fare distribution over seats is capable to provide an effective mechanism to manage, at the same time, the presence of both strategic and late arriving high evaluation customers, even if the mechanism is consistent with a theoretical model such as ours that assumes away customers' heterogeneity.

6.2 Option value and the hidden side of dynamic pricing

To test whether the option value of each seat declines over time, as predicted in Result 3 of Proposition 2, we regress the log of fare of each seat in the distribution against the *Days to departure* dummies, the variable *Position* used in Table 7, and interactions of both. The panel identifier continues to be the flight-code; however, unlike the case of Table 8, the panel's temporal dimension is represented by a sequential counter that uniquely identifies all the possible combination of *Position* with the *Days to departure* dummies.[24] As before, we set the earliest day to departure dummy (Days to departure 36+) as reference group and we cluster the standard errors by route and week.

Because *Position* is identified precisely only when an observation is uncensored, we need to correct for two sources of sample selection, one of which arises because, as before, we restrict the sample to only those observations of flights that, on a given query date, have fewer than 40 seats left to sell. Relatedly, seats in lower buckets have a higher probability to be sold and disappear from the sample at an earlier stage, thus biasing the estimated relationship of a seat's fare over time. To tackle both, we adopt Procedure 17.1 in Wooldridge (2002). First, we run a probit on the full sample, where the dependent variable is a binary value equal to one if seat i, ($i = 1, \dots, 39$), is still available t days before departure: in the case of censored flights, the last 39 seats are *a fortiori* available.[25] We include, among the regressors, *Position* as well as dummies for the number of days to departure, the day of the week of the departure date, the route, the departure slot time (morning, afternoon, evening, etc.), the season (Winter and Summer). Second, we calculate the Inverse Mill's ratio and, third, we include it in the FE regressions, both OLS

[24] Alternatively, we could have incorporated either the variable *Position* or the *Days to departure* dummies into the fixed effect identifier: in these cases, only the interaction model could be identified in the FE estimation. The results would not change. Estimates available on request.

[25] Because, as discussed in the previous analysis, the seats in the bucket on sale are more easily observed by potential strategic consumers and tend therefore to follow an increasing time trend, we set the dependent variable to zero for those seats belonging to buckets containing the next available seat for sale (i.e., the one whose position is identified by the value of the variable *Available Seats*).

and IV, on the selected sample, with the log fare as the dependent variable.

***** **Insert Table 9 around here** *****

Table 9 reports the results. As far as the main variables of interest, *Position* and the *Days to departure* dummies are concerned, the OLS estimates in columns (1) and (2) suggest that their effects are qualitatively similar to the ones in columns (3)-(6) obtained using an instrumental variable approach where the variable *Seat Before=Available Seats - Position* is treated as endogenous. First of all, the coefficient of *Position* is, as expected, negative, because higher values of this regressor correspond to seats in lower buckets. That is, the econometric evidence indicates that the distributions of all flights are structured similarly to Figure 1, as also predicted in Proposition 2, Result 2. Second, and relatedly, the *Position* coefficient provides a rough estimate of the linear average gradient of the fare distribution: such a value varies from 1.8% to 2.5% in the Leisure routes sample.

Third, and more interestingly, the *Days to departure* dummies are also negative and their coefficients increase in absolute value as the departure date nears. Considering that the reference category corresponds to seats in early posted observations, the dummies' coefficients suggest a downward trend for the average fare of all the seats in the fare distribution. This finding is consistent with the view that the carrier generally faces strong incentives to move the seat down to lower buckets as the departure date nears and that such a move reflects a declining option value, as predicted in Proposition 2, Result 3.

Fourth, to get a better appreciation of whether the intensity of the decline over time varies with the seat's position, model (2) and (4) interact *Position* with the set of *Days to departure* dummies. Because the interaction coefficients are all negative, it can be inferred that the decline is stronger as the position value increases: the further a seat is positioned from the top one, the larger the fall in the bucket order (and in fare) it experiences.

Figure 4 shows the predicted effects from model (4) of Table 9. Each line, which represents the predicted relationship between fare and position, keeping the temporal dummies fixed, defines a stylized, smooth version of the fare distributions in Figure 1. The slope varies to reflect the interaction terms in model (4). When the position is fixed, each point depicts the extent by which the average fare falls over time during the booking period. Based on interaction coefficients in Table 9, the drop is larger as the position increases, as also shown descriptively in Table 7.[26]

[26] For instance, the fare of seat 39 drops, on average, from a value around $e^{4.9} = 134$ to about $e^{4.0} = 55$; of seat 25 from about $e^5 = 148$ to about $e^{4.3} = 74$, while for seat 1, the last one to be sold, the predicted fare moves from $e^{5.1} = 164$ to about $e^{4.8} = 121$.

In models (5) and (6) in Table 9 we restrict the analysis to the case of flights in, respectively, "Leisure" and "Business" routes. We do so to test whether the estimates from model (3) for the full sample hold in sub-samples of more homogeneous flights. Overall, the estimates in models (5) and (6) do not differ qualitatively from those in model (3), in particular as far as the variable *Position* and its interaction with the *Days to departure* dummies are concerned. That is, the DP approach resulting in the movement of seats from higher to lower buckets appears to be prevalent in both leisure and business routes. However, in the "Business" sample, the coefficients of the temporal dummies are considerably smaller in absolute magnitude, suggesting that in business routes the carrier tends not to drop its fares over time as much as it does in leisure routes; this results is in line with the standard characterization of business travelers as customers with a higher willingness to pay, whose need to travel is revealed only at a later stage of the booking period. Such a characterization could be easily accommodated in our theoretical model in Section 3 if we assumed that the average θ increases at some point during the booking period.

To study whether the fare drops are affected by the number of available seats at each point in time (that is, whether larger drops tend to occur when the number of sold seats is lower than expected at each point of the booking period), we include the variable *Seats Before* in every model, which captures the number of seats preceding the seat identified by *Position*. We first treat this variable as exogenous, and obtain highly significant and positive, albeit relatively small in magnitude, coefficients. Because in the presence of an endogenous variable the OLS estimates can only be considered in terms of correlation between the regressor and the dependent variable, our findings suggest that seats allocated to upper buckets tend to have, *a fortiori*, more seats preceding them: hence the positive correlation between the fare and the value of *Seat Before*. To test for causality, we use two instruments. The first one,*Lag Seat Before* , is simply the mean of the two weekly lagged values of *Seat Before* variable. The second one, *holiday period*, proposed by Alderighi et al. (2015) is a dummy variable indicating whether the query date falls within a holiday period (Chrismas, Easter, school breaks, etc) and captures possible differences on the demand side. The application of an instrumental variable approach indicates that the coefficient of *Seat Before* becomes negative, i.e., the causality runs in the expected direction, and takes a larger absolute magnitude in models (3) - 0.003 - and (5) - 0.006. The same coefficient are however insignificant in models (4) and (6). Most importantly, the impact of *Seat Before* does not modify the qualitative structure of the *Days to departure* dummies. The econometric analysis therefore provides compelling evidence of the persistent effects of the hidden aspects of DP that the carrier implements to manage the declining value of a set of perishable seats. To our best knowledge, no prior investigation of these aspects can be found in the literature.

6.3 Robustness

An alternative way to test for the decreasing option value would be to evaluate how each bucket size evolve over time. If the redesign of the distribution is such that seats in higher buckets tend to be moved down, then we should observe the size of top buckets to decrease over time, and viceversa for the lower-priced buckets. An advantage of this approach is that it can be estimated using both censored and uncensored observations, after having excluded the first bucket on sale since its size could be reduced by unobserved sales taking place between query dates, and the last bucket when the observation is censored, since in this case the size is likely to be measured incompletely and is thus biased downward.

To define the relative position of a bucket, we take the following approach. Using a route-month combination, we divide all the fares in each combination into quintiles, and thus create five sub-samples over which we run a FE panel regression of each bucket size over the Days-to-Departure dummies.

As Table 10 indicates, in Columns (1)-(3) the size of the lower priced buckets tends to increase as the departure date approaches. The effect decreases as we move up the distribution, with coefficients dropping as we move from model (1) to model (3). Conversely, the buckets in the top quintile clearly behave in the opposite way: they start large and then shrink in size (and may as well disappear, as previously discussed).

***** **Insert Table 10 around here** *****

Finally, in line with Table 7, a further robustness check to test for the decreasing option value is to consider the average fare of the n left seats, where n is the number of available seats at the latest collected day prior to departure. We regress the log of such average fare on our set of *Days to departure* dummies. The results, not reported to save space, clearly indicate a decreasing trend of the dependent variable as departure date approaches and therefore confirm the decreasing option value, which is the main finding of this paper.

7 Conclusion

This paper provides a set of important contributions to the existing literature on airline pricing.

First, we present a theoretical model where the airline sets in equilibrium a distribution of increasing fares similar to Dana (1999); an innovation in this paper is to allow the possibility for the carrier to modify the distribution as the date of departure approaches, thus allowing to derive a prediction of how the distribution is likely to evolve over time. Similar to the various models of revenue management surveyed in McAfee and te Velde (2007), our model also predicts a declining temporal profile of fares. Although those models' predictions have received empirical support in Sweeting (2012), their validity for the airline market is largely rejected in many papers and in McAfee and te Velde (2007) in particular, who deem them as "empirically false". A first important contribution is to show that the incentive to drop fares over time is indeed observed once the analysis is extended to consider all the fares characterizing the equilibrium distribution.

Furthermore, we show that focusing the empirical analysis on the fare for the first seat on sale is not a valid way to conduct a test of theoretical predictions on the fare time-path for at least two reasons. One, the theory we present predicts the equilibrium fare distribution to be monotonically increasing. If, in theory, each seat in a flight is assigned a different fare, then tracking the fare time path of the first seat on sale implies tracking the fare of different seats. The proper way to test for a declining fare is always to refer to the same seat's fare; this is something we do by establishing a seat's fixed position in the fare distribution. Two, airlines are aware of customers' strategic behavior and have an incentive to limit the posting of fare drops by showing that the temporal profile of the observed fare follows a upward trend.

More generally, the focus on a fare distribution allows the investigation of many so far neglected aspects of Dynamic Pricing (DP). Another important contribution of this paper is to show, for the first time in the literature, how these fare distributions are shaped. We define changes over time in the distribution as instances of DP, regardless of whether the actual observed fare has changed. Our analysis of DP in our sample of 37,501 flights operated by easyJet, a leading European low-cost carrier during the period May 2014 – June 2015, unveiled the following results.

Dynamic pricing takes many forms and shapes, involving not only fare variations but, most importantly, changes in the distribution that result in variations of the bucket size, as well as the creation of new buckets. The fare of the next seat on sale tends to increase, due to the increasing shape of the distribution and the fact that buckets disappear when they are sold out, although there can be some fare drops. More generally, over the observed booking period, the firm shifts seats initially allocated to the higher-priced buckets to lower-priced buckets, a mechanism that may lead to the disappearance of the higher-priced bucket from the distribution. This is what we call "*the hidden side of DP*".

The overall outcome of this pricing mechanism is an increasing profile of fares each period of booking but a decreasing profile of the fares assigned to the seats that remain to be sold. Hence, over time the slope of the distribution changes: initially, it is steeper because more seats are allocated to higher-priced buckets but it flattens as seats are reassigned to lower-priced buckets.

Our analysis provides important insights on how dynamic pricing may be relevant for such firms as hotels, cruise ships, car rentals, which have to set their prices facing conditions similar to those of airlines. While we believe that the approach based on fare distribution is widely applied in the airline industry, and provides a central aspect of the revenue management implemented by LCCs, future research needs to investigate how the findings in this work need to be adapted to the more complex revenue management systems adopted by traditional, legacy carriers.

References

Abrate, G., Fraquelli, G., and Viglia, G. (2012). Dynamic pricing strategies: Evidence from european hotels. *International Journal of Hospitality Management*, 31(1):160–168.

Alderighi, M., Nicolini, M., and Piga, C. A. (2015). Combined Effects of Load Factors and Booking Time on Fares: Insights from the Yield Management of a Low-Cost Airline. *Review of Economics and Statistics*, 97:900–915.

Alderighi, M., Nicolini, M., and Piga, C. A. (2016). Targeting leisure and business passengers with unsegmented pricing. *Tourism Management*, 50:502–515.

Bachis, E. and Piga, C. A. (2011). Low-Cost Airlines and online Price Dispersion. *International Journal of Industrial Organization*, 26(6):655–667.

Bergantino, A. S. and Capozza, C. (2015). Airline Pricing Behaviour under Limited. Intermodal Competition. *Economic Inquiry*, 53:700–713.

Bilotkach, V., Gaggero, A. A., and Piga, C. A. (2015). Airline pricing under different market conditions: Evidence from European Low-Cost Carriers. *Tourism Management*, 47:152–163.

Bilotkach, V., Gorodnichenko, Y., and Talavera, O. (2010). Are airlines price-setting strategies different? *Journal of Air Transport Management*, 16:1–6.

Board, S. and Skrzypacz, A. (2016). Revenue management with forward-looking buyers. *Journal of Political Economy*, forthcoming.

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

Clark, R. and Vincent, N. (2012). Capacity-contingent pricing and competition in the airline industry. *Journal of Air Transport Management*, 24:7–11.

Dana, J. D. (1998). Advance-Purchase discounts and price discrimination in competitive markets. *Journal of Political Economy*, 106(2):395–422.

Dana, J. D. (1999). Equilibrium price dispersion under demand uncertainty: the roles of costly capacity and market structure. *Rand Journal of Economics*, 30(4):632–660.

Deneckere, R. and Peck, J. (2012). Dynamic Competition with Random Demand and Costless Search: a Theory of Price Posting. *Econometrica*, 80(3):1185–1247.

Escobari, D. (2012). Dynamic pricing, advance sales, and aggregate demand learning in airlines. *Journal of Industrial Economics*, 60(4):697–724.

Escobari, D. and Jindapon, P. (2014). Price discrimination through refund contracts in airlines. *International Journal of Industrial Organization*, 34(3):1–8.

Gaggero, A. A. and Piga, C. A. (2010). Airline competition in the British Isles. *Transportation Research Part E: Logistics and Transportation Review*, 46:270–279.

Gaggero, A. A. and Piga, C. A. (2011). Airline Market Power and Intertemporal Price Dispersion. *Journal of Industrial Economics*, 59(4):552–577.

Gale, L. I. and Holmes, T. J. (1992). The efficiency of advanced-purchase discounts in the presence of aggregate demand uncertainty. *International Journal of Industrial Organization*, 10:413–437.

Gale, L. I. and Holmes, T. J. (1993). Advance-Purchase Discounts and Monopoly Allocation capacity. *American Economic Review*, 83(1):135–146.

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

Koenisgsberg, O., Muller, E., and Vilcassim, N. (2008). easyJet pricing strategy: should low-fare airlines offer last-minute deals? *Quantitative Marketing and Economics*, 6:279–297.

Lazarev, J. (2013). The Welfare Effects of Intertemporal Price Discrimination: An Empirical Analysis of Airline Pricing in U.S. Monopoly Markets. *mimeo*. 2013.

Li, J., Granados, N., and Netessine, S. (2014). Are Consumers Strategic? Structural Estimation from the Air-Travel Industry. *Management Science*, 60:2114–2137.

Mantin, B. and Koo, B. (2009). Dynamic price dispersion in airline markets. *Transportation Research Part E: Logistics and Transportation Review*, 45:1020–1029.

McAfee, R. P. and te Velde, V. (2007). Dynamic pricing in the airline industry. In Hendershott, T. J., editor, *Handbook of Economics and Information Systems*, volume Vol 1. Elsevier Science, New York, NY, USA.

McGill, J. I. and Van Ryzin, G. J. (1999). Revenue Management: Research Overview and Prospects. *Transportation Science*, 33:233–256.

Möller, M. and Watanabe, M. (2010). Advance Purchase Discounts and Clearance Sales. *Economic Journal*, 120:1125–1148.

Obermeyer, A., Evangelinos, C., and Püschel, R. (2013). Price dispersion and competition in European airline markets. *Journal of Air Transport Management*, 26:31–34.

Puller, S. L., Sengupta, A., and Wiggins, S. (2009). Testing theories of Scarcity pricing and Price Dispersion in the Airline Industry. N. 15555, December, NBER Working Paper Series.

Sweeting, A. (2012). Dynamic Pricing Behavior in Perishable Goods Markets: Evidence from Secondary Markets for Major League Baseball Tickets. *Journal of Political Economy*, 120(6):1133–1172.

Talluri, K. T. and van Ryzin, G. J. (2004). *The Theory and Proctice of Revenue Management*. Springer, New York.

Weatherford, L. R. and Bodily, S. E. (1992). A Taxonomy and Research Overview of Perishable-Asset Revenue Management: Yield Management, Overbooking, and Pricing. *Operations Research*, 5:831–844.

Williams, K. R. (2013). Dynamic Airline Pricing and Seat Availability. *mimeo*.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge MA, 1^{st} edition.

Table 1: Simulated optimal fares p_{jt} in the case of $T = 1, 3, 5$ periods

Seats	$T = 1$	$T = 3$			$T = 5$				
m	$t = 1$	$t = 3$	$t = 2$	$t = 1$	$t = 5$	$t = 4$	$t = 3$	$t = 2$	$t = 1$
12	0.656	0.691	0.624	0.552	0.703	0.661	0.613	0.563	0.521
11	0.665	0.704	0.636	0.559	0.719	0.676	0.626	0.573	0.526
10	0.676	0.719	0.650	0.567	0.735	0.693	0.642	0.585	0.531
9	0.687	0.735	0.665	0.577	0.752	0.710	0.659	0.598	0.538
8	0.700	0.752	0.681	0.589	0.771	0.730	0.677	0.614	0.547
7	0.714	0.771	0.700	0.602	0.790	0.750	0.699	0.633	0.558
6	0.730	0.791	0.721	0.619	0.811	0.773	0.722	0.655	0.572
5	0.748	0.813	0.745	0.639	0.833	0.798	0.749	0.680	0.590
4	0.770	0.837	0.773	0.664	0.857	0.825	0.779	0.711	0.614
3	0.796	0.864	0.805	0.696	0.883	0.855	0.814	0.749	0.646
2	0.829	0.895	0.844	0.739	0.911	0.889	0.854	0.796	0.692
1	0.875	0.932	0.894	0.805	0.944	0.928	0.904	0.860	0.765

Table 2: Simulated optimal probabilities of selling (π_{mt}) and average paid fare (P_m), $T = 5$ periods

Seats	Probability of selling, π_{mt}						Average
m	$t = 5$	$t = 4$	$t = 3$	$t = 2$	$t = 1$	Total	paid fare
12	0.747	0.185	0.052	0.014	0.004	1.000	0.688
11	0.567	0.269	0.109	0.046	0.009	0.999	0.689
10	0.421	0.307	0.168	0.075	0.025	0.996	0.690
9	0.301	0.303	0.224	0.120	0.041	0.989	0.691
8	0.202	0.296	0.253	0.162	0.061	0.972	0.694
7	0.129	0.272	0.263	0.193	0.093	0.948	0.699
6	0.085	0.210	0.267	0.227	0.136	0.924	0.703
5	0.052	0.155	0.251	0.249	0.177	0.883	0.711
4	0.028	0.109	0.220	0.253	0.215	0.824	0.724
3	0.015	0.075	0.173	0.249	0.238	0.749	0.744
2	0.008	0.040	0.126	0.236	0.256	0.664	0.774
1	0.002	0.019	0.074	0.180	0.265	0.539	0.822

Table 3: Number of seats in each bucket price across days to departure

Panel A: days to departure 109-49

Days to dep.	109	101	100	99	91	90	89	81	80	79	71	61	60	59	50	49
Bkt price																
36	5															
41	5	7	7	7		$\boxed{3}$	3									
47	5	5	5	5	5	4	4	1	1	1						
55	6	5	5	5	5	4	4	4	4	4	3	4	4	4		
58																
65	4	6	6	6	5	4	4	4	4	4	4	4	4	4	1	5
75	6	4	4	4	6	4	4	4	4	4	4	4	4	4	3	3
87	4	6	6	6	4	4	4	4	4	4	4	5	5	5	3	3
101	5+	4	4	4	6	4	4	4	4	4	4	3	3	3	3	3
117		3+	3+	3+	4	4	4	4	4	4	4	5	5	5	3	3
136					5+	4	4	4	4	4	4	3	3	3	3	3
156						5+	5+	11+	11+	11+	13+	12+	12+	12+	24+	20+
Av. seats	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+

Panel B: days to departure 48-2

Days to dep.	48	43	42	41	31	30	29	21	13	12	10	9	6	4	3	2
Bkt price																
36																
41																
47																
55																
58									$\boxed{4}$	3						
65	1				$\boxed{4}$	4	5	2	4	4	1	6				
75	3	1	1	5	3	3	3	4	4	4	4	4	2			
87	3	3	3	3	3	3	3	4	4	4	4	4	4	5	5	4
101	3	3	3	3	3	3	3	4	4	4	4	4	4	6	6	7
117	3	3	3	3	3	3	3	4	4	4	4	4	4	6	6	8
136	3	3	3	3	3	3	3	4	3	3	3	3	3	3	3	$\boxed{3}$
156	24+	27+	27+	23+	21+	21+	20+	18+	13+	14+	20+	15+	13	5	3	
Av. seats	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+	30	25	23	22

(a) n+ means that n or more seats are available for the observed bucket.

Table 4: Dynamic pricing by final number of left seats on a flight

Forms of dynamic pricing	Full sample	Final available seats 1-10	11-20	21-30	31-39
First bucket					
Price decrease first bkt	0.08	0.08	0.08	0.07	0.07
Price increase of first bkt only	0.05	0.07	0.05	0.03	0.02
Size increase first bkt	0.17	0.16	0.18	0.19	0.20
Last bucket					
Price increase last bkt	0.01	0.02	0.01	0.01	0.02
Price decrease last bkt	0.04	0.04	0.05	0.05	0.03
Size increase last bkt	0.04	0.04	0.05	0.04	0.04
Size decrease last bkt	0.23	0.23	0.24	0.24	0.21
Intermediate buckets					
Size increase 2nd bkt	0.06	0.05	0.06	0.07	0.07
Size decrease 2nd bkt	0.04	0.04	0.04	0.04	0.04
Size change 2nd-last observed bkt	0.08	0.08	0.09	0.09	0.08
Size/price changes intermediate bkts	0.22	0.20	0.23	0.25	0.26
Overall dynamic pricing	0.42	0.41	0.42	0.42	0.42
Observations	454,311	202,970	137,712	74,302	39,327
Observations (Last bucket)	133,011	74,435	39,817	14,577	4,182

Table 5: Dynamic pricing by days to departure

Forms of dynamic pricing	Days to departure				
	0-7	8-21	22-42	43-63	64-130
First bucket					
Price decrease first bkt	0.03	0.09	0.11	0.09	0.06
Price increase of first bkt only	0.11	0.09	0.05	0.03	0.01
Size increase first bkt	0.09	0.22	0.29	0.19	0.12
Last bucket					
Price increase last bkt	0.01	0.01	0.01	0.03	0.03
Price decrease last bkt	0.05	0.04	0.00	0.00	0.00
Size increase last bkt	0.04	0.05	0.06	0.01	0.00
Size decrease last bkt	0.10	0.32	0.57	0.27	0.11
Intermediate buckets					
Size increase 2nd bkt	0.03	0.10	0.08	0.06	0.03
Size decrease 2nd bkt	0.01	0.07	0.05	0.03	0.02
Size change 2nd-last observed bkt	0.05	0.14	0.09	0.07	0.06
Size/price changes intermediate bkts	0.10	0.36	0.27	0.21	0.15
Overall dynamic pricing	0.33	0.62	0.56	0.37	0.24
Observations	61,365	118,694	65,205	57,065	151,982
Observations (Last bucket)	59,628	67,012	5,848	415	108

Table 6: Dynamic pricing by available seats

Forms of dynamic pricing	Varying occupancy Available seats		Same occupancy Available seats	
	1-20	21-39	1-20	21-39
First bucket				
Price decrease first bkt	0.03	0.06	0.06	0.15
Price increase of first bkt only	0.17	0.15	0.03	0.01
Size increase first bkt	0.07	0.14	0.09	0.34
Last bucket				
Price increase last bkt	0.02	0.01	0.01	0.01
Price decrease last bkt	0.05	0.05	0.03	0.03
Size increase last bkt	0.04	0.04	0.03	0.06
Size decrease last bkt	0.12	0.33	0.07	0.37
Intermediate buckets				
Size increase 2nd bkt	0.02	0.07	0.02	0.09
Size decrease 2nd bkt	0.02	0.05	0.03	0.07
Size change 2nd-last observed bkt	0.05	0.12	0.05	0.14
Size/price changes intermediate bkts	0.09	0.32	0.09	0.33
Overall dynamic pricing	0.39	0.63	0.22	0.59
Observations	44,981	46,303	18,630	23,092

Table 7: Mean fares by Position in fare distribution and days to departure

Position in distribution	Days to departure 36+	35-29	28-22	21-15	14-11	10-8	7-4	3-0
1	182	171	172	167	158	152	146	139
2	182	171	172	167	158	152	145	137
3	182	171	171	167	156	150	143	134
4	182	171	171	165	152	143	134	124
5	182	171	170	164	150	141	131	120
6	182	170	170	163	146	137	127	116
7	182	170	169	161	143	135	125	113
8	182	170	168	160	141	133	122	111
9	182	170	167	158	138	131	120	109
10	182	169	166	156	136	128	117	107
11	182	169	164	154	134	126	115	105
12	186	169	162	151	131	121	110	101
13	185	168	161	149	128	119	108	99
14	185	167	159	147	126	117	106	97
15	185	166	156	144	124	115	104	96
16	184	164	154	141	122	114	102	93
17	183	163	151	139	120	111	99	90
18	183	161	148	136	117	106	94	86
19	182	160	146	133	116	105	93	84
20	181	157	143	131	114	103	90	82
21	179	155	140	128	111	100	88	80
22	177	152	138	126	110	99	86	79
23	175	149	135	123	108	97	84	77
24	173	146	132	121	105	94	81	75
25	170	142	130	118	104	92	79	73
26	167	139	127	115	102	88	76	71
27	164	136	125	113	100	86	74	69
28	161	133	122	111	98	84	72	68
29	158	130	120	109	96	82	71	67
30	154	128	117	106	95	80	69	66
31	150	125	115	104	92	78	68	65
32	147	122	113	102	90	76	65	64
33	144	120	111	100	88	74	63	63
34	141	117	108	98	86	71	62	62
35	139	115	106	96	84	69	60	61
36	137	115	104	94	82	67	58	60
37	134	112	102	93	80	65	56	60
38	131	111	99	90	78	64	55	58
39	127	111	94	90	77	62	54	60

Table 8: Regression analysis of the price of the first seat on sale

	(1)	(2)	(3)	(4)	(5)
Dependent variable	Bkt order	log(p)	log(p)	log(p)	log(p)
Estimation technique	Order probit	OLS-FE	OLS-FE	OLS-FE	OLS-FE
Sample	All routes	All routes	Leisure	Business	All routes
Days to departure 0-3	2.629***	0.805***	0.719***	0.857***	0.787***
	(0.028)	(0.007)	(0.011)	(0.010)	(0.007)
Days to departure 4-7	2.306***	0.686***	0.597***	0.735***	0.668***
	(0.028)	(0.007)	(0.011)	(0.010)	(0.007)
Days to departure 8-10	1.756***	0.510***	0.451***	0.544***	0.499***
	(0.027)	(0.007)	(0.010)	(0.009)	(0.006)
Days to departure 11-14	1.494***	0.424***	0.377***	0.453***	0.421***
	(0.025)	(0.006)	(0.009)	(0.009)	(0.006)
Days to departure 15-21	1.330***	0.371***	0.333***	0.400***	0.366***
	(0.025)	(0.006)	(0.009)	(0.009)	(0.006)
Days to departure 22-28	1.187***	0.325***	0.296***	0.351***	0.323***
	(0.024)	(0.006)	(0.009)	(0.009)	(0.006)
Days to departure 29-35	1.137***	0.298***	0.276***	0.315***	0.291***
	(0.022)	(0.005)	(0.008)	(0.008)	(0.005)
Days to departure 36-49	0.812***	0.198***	0.183***	0.209***	0.196***
	(0.020)	(0.005)	(0.007)	(0.008)	(0.005)
Days to departure 50-63	0.501***	0.110***	0.106***	0.112***	0.108***
	(0.020)	(0.004)	(0.007)	(0.007)	(0.004)
Days to departure 64-77	0.211***	0.040***	0.040***	0.039***	0.044***
	(0.020)	(0.004)	(0.007)	(0.007)	(0.004)
Days to departure 78-91	-0.033**	-0.007**	-0.007	-0.009*	-0.003
	(0.015)	(0.003)	(0.006)	(0.005)	(0.003)
Price decrease 1^{st} bkt					-0.099***
					(0.001)
Price increase 1^{st} bkt only					0.087***
					(0.002)
R^2 (or Pseudo R^2)	0.107	0.574	0.536	0.615	0.594
Observations	900,861	900,861	350678	375775	900,861

(a) Robust standard errors clustered by route and week of departure.
(b) Model (1) includes departure time dummies and departure day of the week dummies.
(c) Models (2)-(5) flight-code fixed effects.
(d) ***, ** and * denote statistical significance at 1%, at 5% and at 10% level.

Table 9: Regression analysis of the price of all seats in the distribution (Option value). Dependent variable: log(p)

	(1)	(2)	(3)	(4)	(5)	(6)
Estimation technique	OLS-FE	OLS-FE	IV-FE	IV-FE	IV-FE	IV-FE
Sample	All routes	All routes	All routes	All routes	Leisure	Business
Days to departure 0-3	-0.388***	-0.263***	-0.689***	-0.251*	-0.914***	-0.446*
	(0.061)	(0.060)	(0.119)	(0.118)	(0.158)	(0.221)
Days to departure 4-7	-0.358***	-0.191***	-0.599***	-0.182	-0.785***	-0.389*
	(0.048)	(0.048)	(0.095)	(0.095)	(0.127)	(0.178)
Days to departure 8-10	-0.319***	-0.117***	-0.485***	-0.111	-0.620***	-0.322**
	(0.033)	(0.035)	(0.066)	(0.066)	(0.089)	(0.121)
Days to departure 11-14	-0.255***	-0.063*	-0.369***	-0.059	-0.470***	-0.238**
	(0.024)	(0.027)	(0.047)	(0.047)	(0.064)	(0.084)
Days to departure 15-21	-0.139***	0.043*	-0.207***	0.046	-0.265***	-0.115*
	(0.017)	(0.021)	(0.030)	(0.031)	(0.041)	(0.053)
Days to departure 22-28	-0.084***	0.067***	-0.122***	0.068**	-0.150***	-0.072*
	(0.013)	(0.018)	(0.019)	(0.022)	(0.027)	(0.033)
Days to departure 29-35	-0.015	0.073***	-0.031*	0.073***	-0.044*	-0.004
	(0.011)	(0.016)	(0.013)	(0.017)	(0.019)	(0.020)
Position	-0.018***	-0.004***	-0.022***	-0.004**	-0.025***	-0.018***
	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
Seats before	0.001***	0.001***	-0.003*	0.002	-0.006***	0.001
	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	(0.003)
Pos*Days to dep. 0-3		-0.017***		-0.017***		
		(0.001)		(0.001)		
Pos*Days to dep. 4-7		-0.017***		-0.017***		
		(0.001)		(0.001)		
Pos*Days to dep. 8-10		-0.016***		-0.016***		
		(0.001)		(0.001)		
Pos*Days to dep. 11-14		-0.013***		-0.013***		
		(0.001)		(0.001)		
Pos*Days to dep. 15-21		-0.011***		-0.011***		
		(0.001)		(0.001)		
Pos*Days to dep. 22-28		-0.009***		-0.009***		
		(0.001)		(0.001)		
Pos*Days to dep. 29-35		-0.005***		-0.005***		
		(0.001)		(0.001)		
Heckman's λ	0.045	0.156***	0.172**	0.151*	0.285***	0.112
	(0.044)	(0.043)	(0.062)	(0.061)	(0.084)	(0.106)
Hansen J-stat			0.244	1.06	0.187	1.35
R2	0.633	0.659	0.628	0.659	0.612	0.654
Observations	2,962,064	2,962,064	2,962,049	2,962,049	1,355,145	1,087,653

(a) Robust standard errors clustered by route and week of departure.
(b) Flight-code fixed effects.
(c) ***, ** and * denote statistical significance at 1%, at 5% and at 10% level.

Table 10: Regression analysis of buckets' sizes by bucket position in the distribution

	(1) Low price buckets	(2) Low-Med price buckets	(3) Med price buckets	(4) Med-High price buckets	(5) High price buckets
Dependent variable	Bkt size	Bkt size	Bkt size	Bkt size	Bkt size
Estimation technique	OLS-FE	OLS-FE	OLS-FE	OLS-FE	OLS-FE
Days to departure 0-3	1.528***	1.526***	1.290***	0.343***	-2.252***
	(0.109)	(0.068)	(0.059)	(0.049)	(0.056)
Days to departure 4-7	1.671***	1.466***	1.064***	0.190***	-1.867***
	(0.077)	(0.047)	(0.043)	(0.038)	(0.054)
Days to departure 8-10	1.832***	1.431***	0.850***	0.223***	-0.957***
	(0.059)	(0.039)	(0.032)	(0.030)	(0.055)
Days to departure 11-14	1.749***	1.425***	0.979***	0.517***	-0.167***
	(0.051)	(0.035)	(0.031)	(0.028)	(0.052)
Days to departure 15-21	0.054	0.030	0.065***	0.171***	0.708***
	(0.034)	(0.025)	(0.024)	(0.021)	(0.038)
Days to departure 22-28	-0.029	-0.044*	0.102***	0.489***	1.273***
	(0.034)	(0.025)	(0.025)	(0.025)	(0.044)
Days to departure 29-35	0.022	-0.095***	0.013	0.401***	1.242***
	(0.031)	(0.021)	(0.021)	(0.022)	(0.046)
Days to departure 36-49	0.025	-0.184***	-0.129***	0.146***	0.406***
	(0.029)	(0.019)	(0.019)	(0.018)	(0.028)
Days to departure 50-63	-0.034	-0.287***	-0.256***	-0.082***	0.043*
	(0.025)	(0.018)	(0.019)	(0.017)	(0.024)
Days to departure 64-77	-0.099***	-0.335***	-0.340***	-0.230***	-0.164***
	(0.022)	(0.014)	(0.015)	(0.015)	(0.021)
Days to departure 78-91	-0.078***	-0.277***	-0.291***	-0.229***	-0.209***
	(0.017)	(0.011)	(0.012)	(0.011)	(0.016)
R^2	0.119	0.175	0.091	0.019	0.093
Observations	549,102	754,021	763,483	803,686	643,972

(a) Robust standard errors clustered by route and week of departure.
(b) Flight-code and bucket-price fixed effects.
(c) ***, ** and * denote statistical significance at 1%, at 5% and at 10% level.

Figure 1: Fare distribution at various days to departure

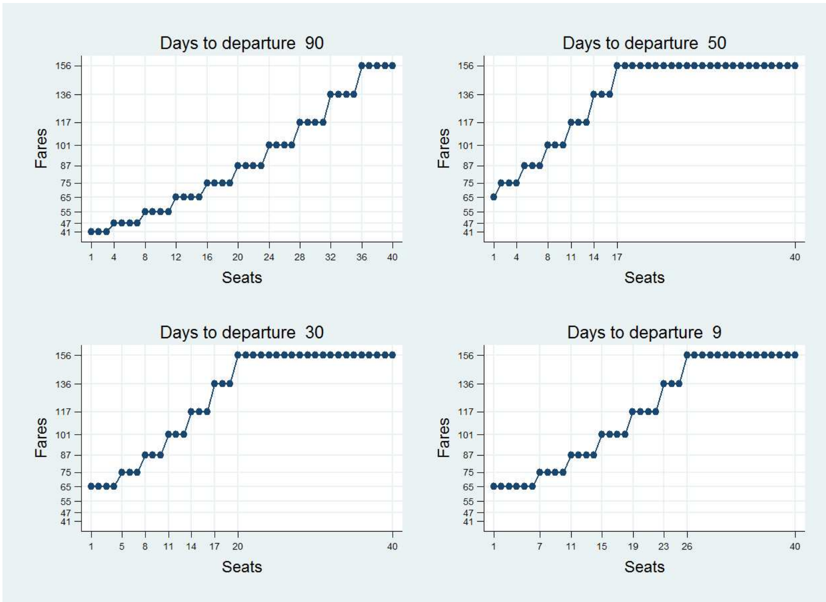


Figure 2: Football Champions' League final (Fare distribution)

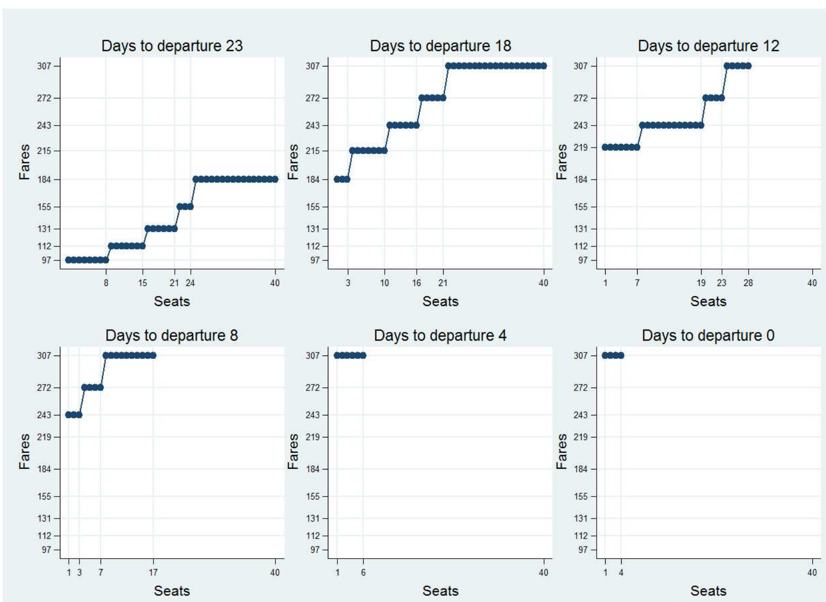


Figure 3: Probability of a bucket to include the first seat on sale, over days to departure

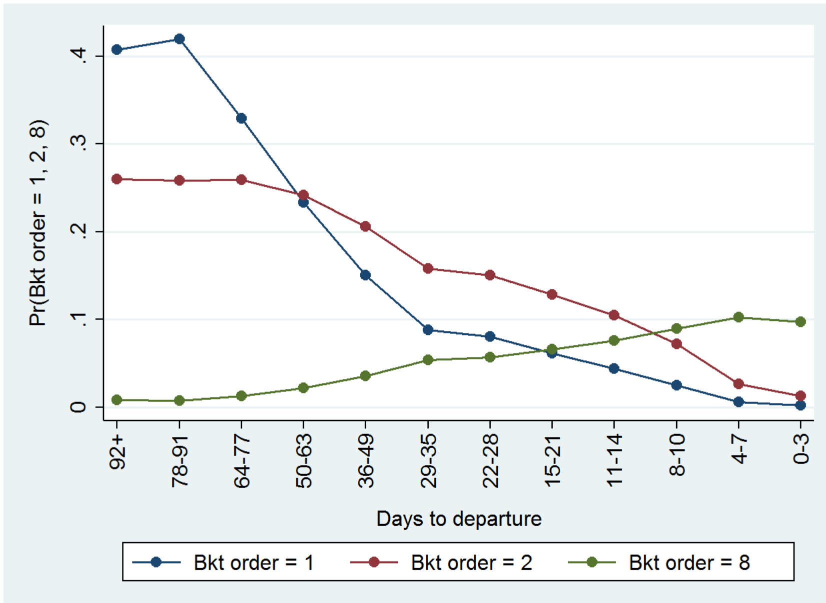
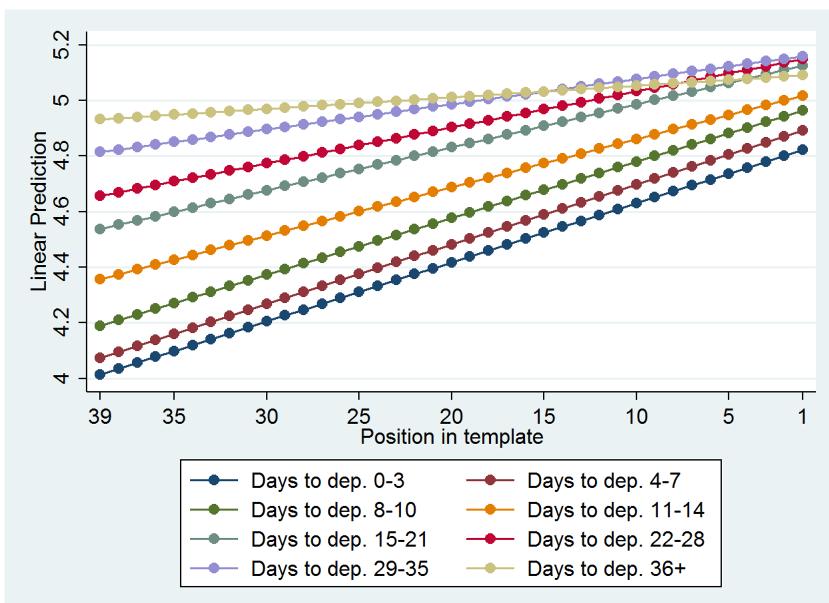


Figure 4: Predicted marginal effects of Position and days to departure on prices. Note: based on Model (4) in Table 9.



A Appendix

A.1 Algorithm

First note that (2) can be written as:

$$V(t,M) = \max_p \{q(p)[p + V(t,M-1) - V(t-1,M)]\} + V(t-1,M) \tag{A.1}$$

with boundary conditions $V(t,0) = 0$ and $V(0,M) = 0$, for any $t \in \{0,\ldots,T\}$ and $M \in \{0,\ldots,N\}$. To find a solution for the problem described in (A.1), we consider the following steps.

Step 1. Find the solution for $\max_p q(p)(p+x)$. Since F is bounded in $\left[0,\bar{\theta}\right]$, there exists a solution for the problem. When θ is uniformly distributed in $[0,1]$, there is a closed form solution given by:

$$p = (1 - \sqrt{(1-\varphi)(1+x\varphi)})/\varphi. \tag{A.2}$$

Step 2. Set $t = 1$ and $M = 1$.

Step 3. Compute $x = V(t,M-1) - V(t-1,M)$ and use Step 1 to get $p(t,M)$. Replace it in (A.1) to obtain $V(t,M)$.

Step 4. Set $m = m+1$. Repeat Step 3 until $m = N$.

Step 5. Set $t = t+1$ and m=1. If $t < T$, then go back to Step 3.

A.2 Proofs

Proof. of Proposition 1

We show that $V(t,M) > V(t-1,M)$. By contradiction assume that $V(t,M) \leq V(t-1,M)$. Let $p^*(\tau,m)$ with $\tau = 1,\ldots,t-1$ and $m = 1,\ldots,M$, be the set of fares that solves $(A.1)$ when there are $t-1$ periods and M seats. Define $\hat{p}(\tau,m)$ with $\tau = 1,\ldots,t$ and $m = 1,\ldots,M$, as a set of fares (not necessarily the optimal one) that is chosen when there are t periods and M seats: $\hat{p}(\tau+1,m) = p^*(\tau,m)$, for $\tau = 1,\ldots,t-1$ and $\hat{p}(1,m) = \bar{p} \in \left(0,\bar{\theta}\right)$. Then, under this fare profile the expected return gained in the first $t-1$ periods is $V(t-1,M)$. Because $\varphi < 1$, there is a positive probability that some seats are available in the last period ($t = 1$), and they generate positive expected revenue, which contradicts our assumption. The proof that $V(t,m) > V(t,M-1)$ is similar to the previous case and is omitted to save space. ■

Proof. of Proposition 2

Part (1): Clear, from the text. **Part (2)**: From (3) and the fact that $\Psi' > 0$, it follows that $p(t,N)$ is increasing in $\Delta_{21}(t,M) := \Delta_2 V(t,M) - \Delta_1 V(t,M)$. Moreover since that $\Delta_1 V(t,M)$ is increasing in M and $\Delta_2 V(t,M)$ is decreasing in M, it follows that $\Delta_{21}(t,M)$, and therefore $p(t,N)$, is decreasing in M. **Part (3)**: As showed above, $p(t,N)$ is increasing in $\Delta_{21}(t,M)$. Moreover since that $\Delta_2 V(t,M)$ is increasing in t and $\Delta_1 V(t,M)$ is decreasing in t, it follows that $\Delta_{21}(t,M)$, and therefore $p(t,N)$, is increasing in t. ■

A.3 Uniform distribution

In this part, we discuss the properties of the value function when the willingness to pay of consumers is uniformly distributed. Table A1 completes our previous example based on the simulation values of Table 1. It shows that the marginal return of having an additional seat or an additional period are decreasing over time, and at the same time, that there are increasing returns from doubling both variables. In order to verify the robustness of the result, we have simulated the case $T = N = 200$ and for any φ in the range $[0.01, 0.99]$, with a step of 0.01.

Table A1: Simulated value function, $V(t,m)$

Seats	Value function, $V(t,m)$				
m	$t=5$	$t=4$	$t=3$	$t=2$	$t=1$
12	7.421	6.577	5.474	4.035	2.201
11	7.014	6.255	5.248	3.909	2.159
10	6.577	5.903	4.996	3.763	2.108
9	6.107	5.518	4.712	3.593	2.045
8	5.602	5.097	4.395	3.396	1.968
7	5.060	4.638	4.040	3.168	1.873
6	4.480	4.137	3.643	2.902	1.757
5	3.857	3.591	3.199	2.593	1.613
4	3.191	2.995	2.701	2.232	1.432
3	2.476	2.345	2.143	1.810	1.205
2	1.710	1.635	1.516	1.312	0.913
1	0.887	0.857	0.808	0.719	0.530

A.4 Data treatment

This Section contains further details on the procedure we applied to derive the fare distributions from the posted fares.

Through data visual inspection, we learnt that the carriers' posted fare follow this rule:

$$PF(s) = \frac{C + \sum_{j=1}^{s} p_j}{s}, \tag{A.3}$$

where s denotes the number of seats in the query, $PF(s)$ the corresponding posted fare, p_j the fare of each seat, starting from the first one available for sale and C is a fixed charge which we interpret as a fixed commission per booking. The presence of C implies that the distribution of posted fares over seats is generally U-shaped, with the decreasing part due to the commission being spread over more seats and the increasing part due to the increasing values of buckets, as in Figure 1.

To find C, we rely on the fact that in most cases the first and the second seat are likely to belong to the same bucket. Therefore C (and the value of the first bucket) can be obtained by solving the following system of two linear equations in two unknowns, using the identity $p_1 = p_2 = p$:

$$\begin{aligned} PF(1) =& C + p \\ PF(2) =& (C + 2p)/2 \end{aligned}$$

The commission changed over the sampling period: it amounted to £5.5 until 25 June 2014, then to £6 until 6 May 2015 and subsequently to £6.5. For flights priced in euro the corresponding values are €7, €7.5 and €8.5 with changes taking place simultaneously to the fares in British Pounds. The values in the two currencies are highly related to the exchange rate in the various periods.

After finding C, using (A.3) it is straightforward to derive the bucket fare tags, P_j:

$$P_j = j * PF(j) - (j-1) * PF(j-1) \text{ with } j \in [2, 40], \tag{A.4}$$

with $P_1 = PF(1) - C$.[27]

Two aspects are noteworthy. First, the procedure to derive the bucket values does not impose any restriction on the monotonicity of the distribution. Second, and most importantly, the distributions we derive correspond exactly to the distributions advertised

[27] For simplicity, cents and pennies are rounded to unity.

on the carrier's website. As discussed in the Data Collection section, for each query the crawler retrieved the information that appears on the booking page regarding the "number of seats available at that fare".[28] We can then gauge the extent to which the size of each bucket, obtained from (A.4), conforms with the information provided by the carrier. It turns out that the above procedure generates buckets' sizes that perfectly correspond to the sizes implied by the information posted by the carrier on the number of seats available at a given fare. We take this as a strong indication that we succeeded in reverse-engineering the carrier's pricing approach.

[28]This and the other website's features illustrated in the paper were still operative at the date this paper was completed.