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

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

February 19, 2017

Abstract

To study dynamic pricing in airline markets it is essential to model the time variation of the distribution of fares that airlines assign to all the airplane's seats. We show how a flight's fare distribution is set in practice and its consistency with the properties of a theoretical model. First, fare distributions are increasing across seats. Second, over time fare distributions move downward to reflect the perishable nature of seats. Third, we find that the fare 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.

*We wish to thank Rajesh Aggarwal, Paolo Bertoletti, Mian Dai, James Dana, Philippe Gagnepain, Patrick Legros, Chris Tsoukis, the participants of the 2016 IIOC Conference in Philadelphia, the 2016 EARIE Conference in Lisbon and the seminar participants at Lancaster University, Northeastern University and Swansea University. This paper was partially written while Gaggero was visiting the Department of Economics at Northeastern University in Spring 2016, Gaggero gratefully acknowledges the warm hospitality provided by the Department. The usual disclaimer applies.
†Corresponding Author: Keele Management School, email: c.piga@keele.ac.uk

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) has reinforced the view that the temporal fluctuations of observed fares constitute the central part of a carrier's Revenue Management (RM) system (McGill and Van Ryzin, 1999; Talluri and van Ryzin, 2004).

Current literature provides overwhelming evidence in favor of a fare fluctuation, as well as of a temporally increasing fare path (Bergantino and Capozza, 2015; Bilotkach et al., 2010; Gaggero and Piga, 2010; Stavins, 2001). Such a finding is, however, extremely at odds with standard theoretical models predicting a declining time-path of fares, to the point that McAfee and te Velde (2007), when commenting the findings from their own data analysis, state that those models are empirically non-validated.

In this paper we address such a divergence head-on by developing a theoretical model whose equilibrium properties are consistent with the empirical findings based on an original dataset of airline fares with unique characteristics. Our new theoretical model emphasizes the role of two forces, which are at work simultaneously to determine the fares of every seat available on a flight. First, airlines sell a highly perishable service. As pointed out by McAfee and te Velde (2007) and Sweeting (2012), fares should decrease as the departure date approaches, because so does the option value of waiting to sell to only higher demand customers ("temporal dimension"). Second, airlines sell a limited number of seats. Thus, fares should increase as the number of seats tend to reduce, as scarcity increases ("capacity dimension") (Puller et al., 2009; Talluri and van Ryzin, 2004). These two forces 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.

A major limitation of the studies in the airline pricing, which is also at the core of the divergence identified but not solved in McAfee and te Velde (2007), rests on the fact that they only rely on transaction fares, which are the joint outcome of both the temporal and capacity dimensions. We argue that to study the problem in a more suitable setting, 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 a fare distribution as in Dana (1999) and Gallego and van Ryzin (1994). Loosely speaking, the focus on a fare distribution implies that, during the booking period, the airline does not limit itself to define only the fare of the seat on sale, but also of all the remaining seats on the flight. We document that this corresponds indeed to the practice of many airlines, which, on their computer reservation systems, post fares for all the seats available on a flight. The experimental design of our data collection exploits such a feature and generates a dataset that we use to model the pricing of as many seats in the distribution as we could obtain from the website of a large European Low Cost Carrier.

The equilibrium solution of the theoretical model is characterized by two main properties. First, to reflect the capacity dimension, the optimal fare distribution is increasing across seats. Second, the temporal dimension operates so that each seat in the distribution presents a declining value over the booking period. Thus, the model extends the theoretical results in Dana (1999), by allowing for the carrier's possibility to modify its fare distribution in different, but discrete, time intervals. To emphasize the empirical implications of the theoretical setting, this study is the first in the literature to show how fare distributions are shaped in practice. Our simulated results conform to our data showing that fare distributions are stepwise increasing: the airline arranges seats into groups, denoted as "buckets", where each bucket is defined by an increasing price tag and a variable size. Such distributions are found to be used, with no exception, in all the 37,489 flights in our sample.

Through the characterization of such distributions at a flight's level, we can extend and better define DP in airline markets. Our assessment of what constitutes DP is different from the one used so far in the literature. Indeed, due to the way the capacity dimension works in practice, we do not classify fare increases over time as DP when such increases arise from a movement along the distribution. This is because 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. Instead, we consider as an instance of DP only a situation involving an identifiable change in the fare distribution. That is, we rule out the fare variations that so far have taken a central role in the literature on DP. Based on this definition, distributions remain, on average, unchanged for about 2-3 consecutive days.

Furthermore, the data reveal 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, occasionally, the creation/deletion of new buckets. In particular, we show how the temporal dimension operates by giving rise to a form of DP that may not involve any change in the fare for the next seat on sale. This happens when the carrier shifts some seats from higher to lower-priced buckets, thus generating a decreasing profile for all the fares in the distribution, in line with theoretical predictions.

Our descriptive identification of new forms of DP provides a useful backdrop for the econometric analysis, which uses insights from the theoretical model to tease out the separate impact of the capacity and the temporal dimensions on online posted fares. With regards to the latter dimension, to our knowledge, this is the first study to provide a combined theoretical and empirical evaluation of the temporal dimension as revealed by a declining option value (McAfee and te Velde, 2007). Thanks to the focus on a fare distribution, we can track the evolution of the fare of all its seats over time, by defining a unique and time-invariant position of each seat in the distribution. The analysis provides strong empirical support in favor of the theoretical models predicting a declining option value. While a similar finding has been shown in Sweeting (2012) for the price of single baseball ticket sold on the second-hand market, a crucial difference here is to show that, at the same time, *i*) the carrier adopts a stepwise increasing fare distribution designed to induce the upward movement of fares consistent with the capacity dimension and, *ii*) it engages in DP to accommodate the declining value of *all* the seats in the distribution. This practice is carried out, as previously mentioned, by shifting seats initially allocated to the higher-priced buckets to lower-priced ones, and, hence, represents a mechanism that may lead to the disappearance of the higher-priced buckets from the distribution. Because the upper buckets are normally not observed by customers, the airline 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. Interestingly, contrary to the common belief that airlines rely on DP to charge higher fares, we highlight how DP can achieve the opposite effect.

To assess the impact of the capacity dimension, we test whether the fare distribution is increasing across seats' positions, as predicted by the theoretical model. We find that the capacity dimension plays a significant role in driving fares upwards: on average, the sale of an extra seat (i.e., a move to the right in the fare distribution) is accompanied by a fare increase of about 1.6-2.0 percent, depending on specifications (see also Alderighi et al. (2015) for a similar result).

Finally, similar in spirit to Gerardi and Shapiro (2009), we reconcile our approach to that traditionally followed in the literature of airline pricing, and in McAfee and te Velde (2007) in particular, by modelling only the fare of the first seat on sale, i.e., the one customers can easily observe when they issue the query for a ticket. In specifications where we omit the seat position, that is, the number of seats still available on the flight, we also find that fares of the seat on sale have an increasing temporal path. This is due to the effect of the capacity dimension: as time passes, the plane fills up and the fare moves up accordingly. Importantly, when we include the seat position, and thus control for a flight's load factor at the time the fare was posted, the estimates continue to provide support to the theoretical prediction of a declining option value.

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 real-world examples of fare distributions and then by a descriptive analysis on dynamic pricing. Section 6 carries the econometric investigation out, testing the properties of the theoretical model's equilibrium solution described in Section 3. Finally, Section 7 summarizes and concludes.

2 Literature review

In the economics literature, 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, 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.

In airline markets, the way fares are set plays a central role in any empirical analysis aimed at defining and identifying DP; Borenstein and Rose (1994) distinguish between systematic and stochastic peak-load pricing as sources of fare dispersion in the U.S. market. In the former, the fare variation is based on foreseeable and anticipated changes in shadow costs known before a flight is opened for booking, while the latter reflects a change during the selling season in the probability that demand for a flight exceeds capacity. In this sense, 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.[1] 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 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. Puller et al. (2009) refer to this as "scarcity-based" pricing. 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

[1] RM involves a number of ancillary activities and techniques useful in the process (McGill and Van Ryzin, 1999; Talluri and van Ryzin, 2004).

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). 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 fare distribution 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 fares 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 (temporal dimension).

Various reasons explain 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).[2] 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). Third, 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

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

hoping to obtain a lower fare. 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 (Li et al., 2014).

The present work makes the novel point that the capacity dimension is the driving force pushing the fare of the seat on sale upward, although with occasional markdowns consistent with the prediction in Deneckere and Peck (2012). It does also investigate the extent by which DP is applied by the carriers to take advantage of the larger proportion of business buyers during the last week before a flight's departure.

3 Theoretical background

In this Section we offer a stylized model of RM which translates some key elements of RM practices into economic terms. First, carriers sell multiple indivisible units (seats). Second, carriers charge a very limited number of fares (holding class fixed). Our data (see below) suggest that there are about 12 to 18 different economy fares in each flight over the entire selling period. Third, carriers price in distribution, that is, in each period they assign a fare to all the seats in a flight; this is because, in each period, a carrier can sell more than one seat and possibly all the seats of the flight. Fourth, fare distributions remain fixed over discrete time intervals. Escobari et al. (2016) report evidence suggesting airlines revise their prices overnight; in our data, distributions last unchanged for two-three days on average.

A carrier operates a single flight with $N > 1$ seats on a monopolistic route. The flight is sold over $T \geq 1$ selling periods: $t = T, T-1, \ldots, 2, 1$ describes the number of periods remaining before departure ($t = 1$ is the last selling 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 selling period t faces fare $p(m, t)$. Within the selling period t, once seat m has been sold, then the next fare on offer becomes $p(t, m-1)$. At the end of the selling period t, the unsold seats are offered in the next period, $t-1$, until $t = 1$. Seats available at the end of the last selling period remain unsold.[3]

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

[3]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.

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 (right-continuous) cumulative distribution $F_{h,t}$ on the compact support Θ, with $\underline{\theta} = \inf \Theta > 0$ and $\bar{\theta} = \sup \Theta < \infty$.[4]

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

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

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

The carrier's maximization problem is denoted by the following Bellman equation:

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

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 aircraft. Moreover, 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 $V(t-1, M)$.

Note that because the solution of the maximization problem can be reached backwards and recursively, the optimal fare of seat $m \leq M$ in period t when there are M seats available, $p^*(t, m, M)$, is independent of the total number of available seats M in period t, i.e. $p^*(t, m, M) = p^*(t, m, M+1)$, for any $M = 1, \ldots, N-1$ and $t = 1, \ldots, T$. This property is a consequence of the assumption that the arrival process is memoryless. Indeed, by having φ depending on the number of travellers already arrived during the period implies that the optimal fare is also affected by the total number of available seats at the beginning of the period and, in general, $p(t, m, M)$ is not necessary equal to $p(t, m, M+1)$. In what follows, we refer to the optimal fare of seat m at time t as $p^*(t, m)$ without indexing for the number of available seats since it plays no role with current assumptions.

[4]This guarantees the existence of a solution of the problem. Moreover, note that the random variable $\theta_{h,t}$ can be one of continuous, discrete or mixed type.

Definition 1 *$V(t,M)$ has decreasing differences in t and M, respectively, if and only if, for any $t = 1,..,T$ and $M = 1,..,N$:*

$$\begin{aligned} V(t,M) - V(t-1,M) &\leq V(t,M-1) - V(t-1,M-1) \\ V(t,M) - V(t,M-1) &\leq V(t-1,M) - V(t-1,M-1). \end{aligned}$$

Definition 2 *$V(t,M)$ has increasing differences in (t,M) if and only if for any $t_H > t_L$ and $M_H > M_L$, we have:*

$$V(t_H, M_H) - V(t_L, M_H) \geq V(t_H, M_L) - V(t_L, M_L).$$

The following proposition characterizes the value function described in (2).

Proposition 1 *The value function $V(t,M) : \{0,1,..,T\} \times \{0,1,..,N\} \to \mathbb{R}$ is non negative and exhibits positive but decreasing differences in t and M, and increasing differences in (t,M).*

Proposition 1 has important implications for our analysis. First, $V(t,M)$ is increasing in (t,M), which is a standard results in the pricing literature (Gallego and van Ryzin, 1994; McAfee and te Velde, 2007). Second, periods and seats can be seen as two factors affecting firm's profits, which generate positive but decreasing value: the additional impact of one period (or one seat) is lower when the number of periods (seats) increases. Third, increasing differences in (t,M) is a form of complementarity. The larger the selling periods and the higher the return from an additional seat, and vice versa. From these properties we derive Corollary 1, which is essential for the characterization of the optimal fare $p^*(t,m)$.

Corollary 1 *Let $X(t,M) = V(t-1,M) - V(t,M-1)$, then:*

$$X(t,M) \leq X(t-1,M), \text{ for any } t = 2,..,T \text{ and } M = 1,..,N \tag{3}$$

$$X(t,M) \geq X(t,M-1), \text{ for any } t = 1,..,T \text{ and } M = 2,..,N \tag{4}$$

Assuming that the optimal fare $p^*(t,m)$ which solves (2) is unique, then:

Proposition 2 *The optimal fare $p(t,m)$ has the following properties:*

A. (capacity dimension) $p(t,m) \leq p(t,m-1)$, for any $t = 0,..,T$ and $M = 1,..,N$,

B. (temporal dimension) $p(t,m) \leq p(t-1,m)$, for any $t = 1,..,T$ and $M = 0,..,N$.

Proof. From the maximization problem in (2), the optimal fare $p^*(t,m)$ can be written as a function of X:

$$p^*(X) = \arg\max_{p \in \Theta} \{q(p)[p + X]\} \tag{5}$$

Let $\rho = \bar{\theta} - p$ and $H(\rho, X) = q(\bar{p} - \rho)[\bar{p} - \rho + X]$. From Definition 2, after some computations, we obtain that H has increasing differences in (ρ, X), if and only if, for $\rho' \geq \rho$ (i.e. $p' \leq p$) and $X' \geq X$, we have:

$$[q(\bar{p} - \rho') - q(\bar{p} - \rho)](X' - X) \geq 0, \tag{6}$$

which is always satisfied seeing that q is decreasing in p. From the Topkis (1998)'s Theorem 2.8.2, when H has increasing differences in (ρ, X) then

$$X' \leq X \implies \rho^*(X') \geq \rho^*(X) \Longleftrightarrow p^*(X') \leq p^*(X). \tag{7}$$

From (7) and Corollary 1, we obtain the proof. ■

Proposition 2.A states that, within a given period, seats are sold by setting a sequence of fares (i.e. a fare distribution) which is (non-strictly) increasing, implying that the fare of the seat on sale may increase every time a seat is sold. This property of the fare distribution reflects the fact that the higher the fare, the lower the likelihood to sell a given seat, and, consequently, the following seats. Thus, a high fare for the seat on sale produces an expected loss of revenue that is increasing in the remaining seats. Since each fare is set on the basis of a balancing between filling up the flight and increasing margins on each seat, a carrier charges lower fares for the first seats on sale and higher fares for the next ones. This result extends the cost-based justification of an increasing equilibrium fare distribution considered in Dana (1999).

Proposition 2.B predicts that the fares of all the seats in the distribution tend to decrease over time. This result reflects the perishable nature of the airline service, and the fact that the option value decreases over time. When the number of periods is high, a carrier has multiple chances to sell seats, but approaching the departure date, the likelihood of selling each seat of the (remaining) fare distribution decreases and therefore, the carrier reduces the fares of all seats. 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 fare distribution as in the present case.

To further investigate the nature of the properties in Proposition 2, we solve equation (2) numerically using calibrated parameters derived on the basis of data employed in the econometric analysis below. In particular, we restrict the example to 39 seats which can be sold at the fare levels (or bucket prices) reported in Column 1 of Table 1. That is, we assume that travellers' willingness to pay (WTP) is drawn from a discrete distribution whose probability density function (pdf) is reported in Column 2 of the Table. Therefore, the fare distribution in every period is stepwise increasing and can be summarised by simply reporting the number of seats in each bucket. Consumers' arrival frequency is set so that the average number of available seats at the end of the flight is around eight.[5]

Each column in Table 1 conforms to property A. of Proposition 2. We denote as a "bucket" the set of seats carrying the same fare tag.[6] For instance, in the first period (period 11) the seats are allocated across buckets of different sizes: five at the fare 65, six at the fare 80 and so on and so forth up until fare 200 with three seats. Note that although there is a relatively high proportion of customers with a WTP of 50, no seats are allocated to this bucket. In the last period 1, each of the three remaining seats is allocated to a different bucket. Property B. is clearly revealed by the fact that the size of the upper buckets tends to shrink over time, leading to their disappearance. However, it is stress-worthy that Property B. also affects seats in lower buckets alike. If it did not, at period 7, when seventeen seats have previously been sold, the selling fare would be that of the seventeenth seat in period 11, that is, 95, while it is still 80 and remains so in subsequent period when extra seats are sold. As a combined result of both properties, the fare of the first seat on sale tends to increase, moving from 65 to 80 and eventually 95.

Figure 1 provides a graphical representation of the content of Table 1.

4 Data

Our collected sample comprises a total of 37,489 daily 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.[7]

[5]We set the average total number of prospective travellers $L = 5/4\,N = 48.75$, which implies $\varphi = L/(L+T) = 0.81$.

[6]The term is drawn from the revenue management literature (McGill and Van Ryzin, 1999; Talluri and van Ryzin, 2004)

[7]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.

4.1 Sample Collection

The data collection employed a web crawler, as widely used in the literature.[8] 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 and there is no pricing-in-network considerations to account for, to double the data size, the query was for a return flight, with a return date 4 days after the first leg (Bachis and Piga, 2011).[9]

The query dates were set such that a flight entered our database about four months before departure; it was then surveyed at 10-days intervals until 30 days before departure, 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's response to the query included, for each leg, flight information 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 dates 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 to derive the price distributions from the posted fares, as illustrated in the Appendix.[10]

To the best of our knowledge, the empirical literature on airline pricing focuses on the fare of one seat, namely, 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 fare 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

[8]For the airline market, see Li et al. (2014), Gaggero and Piga (2011), Clark and Vincent (2012), Obermeyer et al. (2013), Escobari (2012), Bilotkach et al. (2015), Alderighi et al. (2015) and Alderighi et al. (2016), amongst others. Cavallo (2017) and Gorodnichenko and Talavera (2017) make international comparisons of online prices in retail markets.

[9]As in the case of Ryanair in Alderighi et al. (2015), easyJet offers seats where the 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.

[10]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.

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., *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 2, which is based on the data of a randomly selected flight, which will be consistently referred to as an example throughout the paper.

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

4.2 An example of a fare distribution: easyJet

Figure 2 is central for the whole analysis. Each graph, where a dot denotes a seat, represents the fare distribution retrieved, respectively, 100, 35, 15 and 5 days to departure; the fare of the first seat in the lowest bucket corresponds to the fare of the seat on sale, i.e., the fare shown on the carrier's website after a query for one seat.[11] It is evident that these four bi-dimensional distributions are qualitatively identical to the three-dimensional ones shown in Figure 1. Considering that similar stepwise distributions characterize all the flights in our sample, we can conclude that our data collection design yields compelling descriptive evidence in support of property A. of Proposition 2.

In the top panels of Figure 2, the number of available seats is censored to 40; i.e, the graphs do not show the extreme right tail of the price distribution, which is instead represented in the two bottom panels, where, on the left, 38 seats remain to be sold, reducing to only 11 on the right. Interestingly, bucket fares are repeatedly found over the booking temporal horizon, thus suggesting that they tend to be used throughout most of the booking period, until they are sold out or, more occasionally, emptied.

A visual inspection is sufficient to establish some interesting features of the distributions and their evolution over time. One-hundred days to departure, the carrier had allocated five seats for sale at the price of £38 (the per-seat price, net of booking fee, that a customer buying up to 5 seats would pay), five seats at the price of £48, and so on and so forth. Due

[11]As discussed in the Appendix, the fares in the Figure are net of the booking fee.

to the data censoring, we cannot ascertain the precise size of the last "bucket" valued at £158. Similarly, the size of the £38 bucket is likely not correct, since there may be missing, previously sold, seats. Sixty-five days later, the first two buckets are not found; only two seats are available at the price of £58 and the size of the £158's top bucket has clearly increased to at least 18 seats, although the censoring still prevents us to precisely measure its size. Interestingly, twenty days later, even without censoring the distribution is made up of buckets whose fares are the same as the ones reported in the previous periods. Moreover, the size of the second bucket (£80) has increased to five seats, and that of the top bucket can now be precisely measured as equal to 19. Five days prior to departure, the carrier is offering six seats at the price of £112, but noticeably, the size of the top bucket (£158) has shrunk to only two seats.

It could be argued that Figure 2 exemplifies just the peculiar approach followed by easyJet but that it is not representative of the industry. Therefore, in the next subsection we generalize the analysis by providing several examples of similar fare distributions derived from data collected from the websites of many other European and U.S. carriers, both Low Cost and Full Service Carriers (FSCs).

4.3 Examples of fare distributions from other airlines

Figure 3, which is constructed using web crawling, shows the striking resemblance between the fare distributions of easyJet and Ryanair, the largest European LCC. The censoring point, which is caused by the limit on the maximum number of seats in a query imposed by the website's programming code, is in this case set at 25 seats. Interestingly, five days to departure there are at least two, four and nine seats in, respectively, the £143, £121 and £99 buckets. Two days later, the £143 bucket has disappeared, only two seats are allocated to the £121 one, and the size of the £99 bucket has increased to thirteen seats. While the price of the seats allocated in higher buckets has clearly fallen, in line with property B. of Proposition 2 , the price of the seat on sale has increased from £84 to £99, as predicted by Property A. That is, the main implications of this study could easily be extended to at least another large players in the industry.[12]

Southwest allows queries with only a maximum number of seats restricted to eight and it is therefore not possible to depict a fare distribution encompassing a number of buckets as high as in the case of easyJet and Ryanair. Indeed, when holding the query date fixed, for the majority of flights the eight seats carry the same fare, as it is shown for instance

[12]The original plan for this study was indeed to use data from both Ryanair and easyJet. However, the adoption by the former of Captcha techniques made web crawling impossible. The limited amount of data collected prior to this event, from which Figure 3 is derived, led us to the decision to focus on easyJet.

in the first two left vertical panels in the top part of Figure 4.[13] However, the data also includes several examples exhibiting a jump upward from one bucket to the next, as in the vertical panels for the departure dates 14, 10 and 7 in the top part of the Figure, and in most of the panels in the bottom part. The Figure suggests that also Southwest organizes the fares on its reservation system making use of a flight's fare distribution where fares tend to follow the sequence defined by the buckets' rank. Indeed, in the top part of Figure 4, the number of seats in the $270 bucket reduces from four to three seats between ten and seven days to departure. In the bottom part of the same Figure, something similar happens to the seats in the $341 bucket, which disappear three days before departure, when only two seats at $414 remain. Overall, Figure 4 suggests that also Southwest, the largest U.S. LCC, makes extensive use of fare distributions that are organized in a way similar to its European counterparts.

As far as FSCs are concerned, the analysis is complicated by their adoption of a nested-classes system, where the same seat can belong to different classes, each with different ticket restrictions; therefore, one would need to retrieve a distribution for each class category, with precise information on the number of seats (and classes) each category is designed to contain. It is however possible to connect some features of FSCs' pricing approach with the present analysis based on fare distributions. For instance, various papers present graphical evidence of the temporal profile of fares by FSC, i.e., they report the fare of the seat on sale and its evolution over time (Escobari, 2012; McAfee and te Velde, 2007; Puller et al., 2009). It turns out that such temporal paths also follow a step-wise pattern, which can be rationalised along the terms we use to define a fare distribution. Indeed, one could view each bucket as a different "fare class", which, like buckets, is stored in the reservation system, regardless of whether it is immediately available for sale or not. To shed light on this assumption, starting from 2^{nd} November 2016, we saved data from the website expertflyer.com, whose 'Pro' subscription allows access to the list of fare classes (and associated fare and ticket restrictions) an airline uses on a specific route (i.e., the list is not flight-specific). To minimize network pricing effects, we chose one direct flights departing on 15 November 2016 operated by American Airlines (AA), connecting New York JFK to Chicago ORD. In addition to the list of classes from www.expertflyer.com, starting from the 3^{rd} November 2016, we visited AA's website and recorded manually all the different fares therein reported.

In Figures 5, the posted fares are joined by a line; the other symbols refer to specific classes listed by expertflyer.com, of which we report only the first letter.[14] There are at least

[13] The data comes from a work in progress involving the authors and another scholar based in a U.S. institution.

[14] For instance, the full code for the class Q in Figure 5 is Q7ALKNN3. It is noteworthy that expertflyer.com

two main aspects worth highlighting. One, our analogy between buckets and classes appears to be supported by the fact that expertflyer.com reports most classes for the full period, regardless of the posted online fares. For instance, the non-refundable class N or G for a seat in the main cabin (top panel of Figure 5) was available on the computer reservation system during the whole period. Interestingly, the class Q in the top part of Figure 5 and the class N in the central part cease to appear on the 8^{th} November, i.e., seven days prior to the flight departure.

It could be argued that the fare classes in Figure 5 are not relevant because they are not specific to the flight under study; however, such a criticism is thwarted by the second aspect the Figure shows. Indeed, we find that the website's fares often perfectly match the class fares reported by expertflyer.com. This happens for the days 6-8 and 10-12 November (classes Q and N in the top part), 3-8 November (class N in central part), and 3-12 November (class $V3$ in bottom part).[15] Interestingly, for the case of the Main Cabin lowest fare, the posted fares depict a step-wise path with fare levels defined by predetermined fare classes. Although with the limitations due to matching data from different sources, the short period of analysis, and the fact that FSCs rely extensively on the traditional travel agents' channel, the overall analysis based on Figures 5 suggests that the notion of a fare distribution provides a useful starting point for any investigation of FSCs' pricing methods.

5 Descriptive analysis

To lay the foundations for the empirical strategy we will adopt to study the properties in Proposition 2, we need to shed more light on the link between the fare distribution and dynamic pricing (DP). Indeed, for property B., it is essential to describe "*the hidden side*" of DP, i.e., how DP can take place even if the selling price does not change; as for property A., we need to describe how the fare distribution operates like a template that, at each point in time, defines the sequence of fares as the flights fills up.

5.1 Defining Dynamic Pricing

As Figure 2 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 mod-

reports a very large number of classes, and that we only report those whose value is close to that of the posted online fares.

[15] Due to time zone difference, we could retrieve the fares on the date of departure when in the USA it was still nighttime.

ification 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 now refer to the entire set of data for the flight used in Figure 2, as reported in Table 2. 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 at least 40 or more seats left on the flight) or the precise number of available seats (this is visible from fifteen days onwards in Panel B, when the maximum number of prices observed is for 38 seats). The fare of the seat on sale corresponds to the lowest fare of the bucket with a strictly positive number of seats.

Table 2 provides examples of the various forms of DP implemented by the carrier. We define as DP any change in the distribution of seats across two sequential query dates. That is, we do not consider as DP the fare increase from £31 to £38 that takes place between 129 and 121 days from departure, because it corresponds to a movement along the distribution and is consistent with the selling out of the seats in the £31 bucket. Similarly, during the last fortnight, the fare of the seat on sale assumes the values £68, £80, £96, £112; such a movement does not count as DP because it automatically occurs when the first available bucket becomes sold out and the system moves to the next available bucket level.

For a better visual identification, 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, however, that the appearance or reappearance of a bucket fare is equivalent to an increase of its bucket size from zero to a positive number of seats; so effectively all forms of DP are equivalent to a reallocation of seats to an upper or lower bucket. For instance, we do consider as DP the fare drop observed between 119 and 111 days to departure, because it corresponds to a reopening of the £31 bucket. Similarly, between 2 and 1 days to departure when the number of available seats drops from 6 to 4, one seat is moved up from the £112 to the £133 bucket.

More interestingly, based on our definition, DP takes place even if the fare of the seat on sale remains unchanged (the “hidden DP”). This happens several times in the Table. For instance, between 70 and 69 days to departure, the size of the £48 bucket increases from four to six seats; another instance in which the bucket on sale is replenished is observed between fifteen and fourteen days to departure, when six out of nine seats from the top £158 bucket are moved down to the £68 bucket (each of the other three seats are reallocated to the £80, £96 and £112 buckets, respectively).

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

To quantify DP consistently in our dataset and identify whether a seat has received a

DP treatment, it is necessary to study what happens to that seat between two consecutive booking days. In its simplest form, DP occurs if we observe a seat has changed its fare either upwards or downwards; this implies a movement to a bucket that either has already been observed as part of the distribution, as in the case of the reopening of the £48 bucket twenty-one days to departure, or to an entirely new bucket. A complementary way is to look at whether a seat's bucket size has increased (e.g., 1 day to departure in Table 2), or decreased (5 days to departure). For the case of the first seat on sale, if its bucket size decreases we cannot distinguish whether that has happened due to a reduction of the available seats or to DP, and so we limit the analysis to bucket size's increases only. Overall, the distribution is deemed to have changed whenever we register any of the above movements for at least one seat.

The descriptive analysis of DP is carried out using only the non-censored observations because doing so allows the position of each seat to be precisely identified. Consider again the bottom left panel of Figure 2, when only 38 seats remain on the flight. If we look at the distribution from the bottom up (left to right), the first seat is the one on sale, and the 38^{th} identifies the "last" seat that would be put up for sale. In the bottom right panel, the number of available seats dropped to 11. In this case, the seat that occupied the 28^{th} position in the other panel is now the first seat (the seat has clearly dropped down two buckets); and the position of the last seat would be now the 11^{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 (that is, we count seats starting from the extreme right of the distribution), it turns out that in both panels the top right seat would be assigned a position equal to 1, the one immediately on its left position equal to 2, etc; the first seat in the left bottom panel would then take position 2. We report these values in a variable denoted as *Position*.[16] That is, over different query dates, we can track the evolution of each seat's fare, as long as the observation is non-censored.

5.2 Descriptive statistics on Dynamic Pricing

Table 3 reports the probability that each seat in the distribution is treated with one of the forms of DP defined in the previous subsection, obtained by considering only variations between query dates separated by one day (e.g., in Table 2, between 3 and 2 days to departure) and only non-censored observations. The qualitative results do not change if the probabilities were obtained considering variations between any two consecutive, but not adjacent, query

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

dates. The first four columns investigate whether a seat has moved up or down, that is, whether it has moved to a bucket previously observed as part of the distribution or to an entirely new one. The subsequent two columns report whether the size of the bucket where the seat is positioned, has increased or decreased. The Table provides several insights into how DP affects the fare distributions. First, consistent with the property B. of Proposition 2, the probability that a seat is moved to a lower bucket is much higher relative to that of being moved in the opposite direction; the maximum probability of moving to a higher bucket is about 5% for the seat in position 26, which also records a 17.6% likelihood to be shifted down to a previously observed bucket. Second, the design of a fare distribution is rarely altered by adding new buckets, given the generally low probability of observing the creation of a new bucket. Third, and relatedly, the size of buckets in the right tail of the distribution (i.e., those with low positions) tends to shrink, while seats in the left tail belong to buckets whose size is more likely to increase. Indeed, the buckets for the seats in positions 1 to 9 exhibit a probability of more than 25% to be shrunk; conversely, the probability of a size increase is larger for seats in lower positions 20 to 39. Overall, Table 3 provides strong descriptive support to the role of the temporal dimension in driving down the option value of all the seats in the fare distribution.

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

Table 4 is based on the distance between the query date and the departure date. It reports the probability of whether the overall distribution has received a DP treatment, that is, whether at least one of the seats in Table 3 has changed bucket fare or size. The "Any fare move" column reports the probability that the distribution has changed due to a fare movement in either directions; similarly, the "Any fare change" denotes a change in bucket size. The highest of these two values constitutes the probability that a distribution changes during the specified booking period. The "Overall" row provides a sample estimate: on average, a flight distribution has a probability of 48.6% of changing between two consecutive days; that is, distributions change less than once every two days. There are however important variations across the booking period. Distributions rarely change when more than fifty days separate the query date from the date of departure: the probability of 21.1% implies that distributions remain unchanged for about four out of five days. Between thirty-six and eleven days to departure, the likelihood of observing a fare distribution increases drastically, but, in line with property B that predicts a decreasing option value, this is largely due to seats being moved to lower, pre-existing buckets. Moreover, even when frequently applied, DP does not generally involve a drastic redesign of the fare distribution via the creation of new buckets; this finding lends support to the new approach in this paper linking airline pricing and fare distributions. Finally, within ten days to departure the probability of a fare distribution change drops again to values below 40%; that is, fares remain unchanged for about 2-3 days. This period appears to be characterized by a larger (lower) probability to observe a movement of seats towards higher (lower) buckets.

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

A comparison between Table 4 and Table 5 indicates that changes in the distribution do not necessarily involve the first seat on sale. For instance, between eleven and twenty-eight days to departure, the probability of a downward fare movement of any seat in the entire distribution is always higher than 70%, but it is less than 20% for the fare of the seat on sale.[17] Similarly, the size of the bucket where the first seat is placed increases less frequently than in the full distribution. A possible exception can be found in the fare movement to a higher bucket, where the two Tables present values which are similar but well below 20%; that is, in the last three days, less than one flight out of five receives the treatment. Relatedly, the total probability of a fare drop during the last three (seven to four) days is lower than 6.0% (9.0%) but it is reaches the value of about 20% between eleven and fourteen days to departure. The fact that, during the last three days, the first seat is moved up to a higher bucket consistently more than it is moved down suggests that DP can be used to pursue an inter-temporal price discrimination strategy aimed at capturing the higher proportion of customers with a higher willingness to pay, generally those flying for business purposes (Dana, 1998; Gale and Holmes, 1993). Overall, the type of DP most frequently applied to the seat on sale appears to be that involving the increase in its bucket size, consistent with the idea that the property B. of Proposition 2 is implemented by shifting seats initially allocated to the higher-priced buckets to lower-priced ones, what we term as "the hidden DP". Often, this results in the replenishment of the bucket where the first seat is placed, leading to a slowing down in the rate in which the selling fare would increase due to the pure capacity dimension effect. For instance, in Table 2, the first bucket's size increases significantly fourteen days to departure, thus postponing the increase in the posted fare from £68 to £80.

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

The foregoing descriptive analysis provides the necessary backdrop for Table 6, which 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, in line with the property B in Proposition 2. The decline appears to be inversely proportional to the position. That is, the last seat

[17]The fact that easyJet does not resort to last-minute deals to clear capacity was also noted in Koenisgsberg et al. (2008).

($Position = 1$) 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 Figure 2, where, between fifteen and five days to departure, the left seats in the £158 bucket end up being moved down by several buckets, unlike the last two on the right.

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

To sum up, a combined analysis of Tables 3-5, in addition to providing empirical support to the properties of the equilibrium solution in Proposition 2, highlights several practical aspects of DP in airline markets. First, DP takes many forms and shapes, all aimed at redesigning the distribution of fares uploaded on the carrier's reservation system; second, fares are seldom added to the original structure of a distribution; third, changes in the distribution are carried out at lumpy time intervals, so that a flight's distribution remain unaltered for an average of 2-3 days; fourth, DP treatments may not necessarily involve the first seat in the distribution, whose bucket is however highly likely to increase in size.

6 Econometric design and analysis

We now proceed to test formally the two properties characterizing the equilibrium solution in Proposition 2, by providing two sets of regressions. In the first, we consider the full sample, and focus on how the fare of each seat in the distribution is affected by its position and how it changes over time. The second regression sheds light on how the fare of the first seat on sale changes as its position changes over time. As far as property A. is concerned, we have already shown how the adoption of a fare distribution is pervasive and offers the carrier a practical way to implement DP. The second regression shows that the capacity dimension is responsible for the movement of the seat on sale along the distribution, leading to a temporally increasing profile of the "easily observable" fare on sale, while the temporal dimension operates in a "hidden" way. Both regressions lend strong support to property B., after the role of the capacity dimension is taken into account.

6.1 Full distribution analysis

To test both properties in Proposition 2, we estimate the following equation for the fare of seat with $Position = m$ on flight j departing on date d:

$$\ln Fare_{jd}^{m} = \sum_{t} \beta_t D_t + \gamma Position^{m} + \zeta_{jd} + \varepsilon_{jdt}, \tag{8}$$

where D_t defines a set of dummy variables *Days to departure*, with t defining the intervals between the query and the departure date. As far as property B. is concerned, they represent our variables of interest as they track the time evolution of the fare of a specified seat's position, which we expect to be declining, while Property A. would be supported by a negative and significant coefficient of *Position* (recall that we count the position by starting from the right of the distribution).

The econometric strategy takes into account two related sources of sample selection. One, *Position* is identified precisely only when an observation is non-censored, and so we have to restrict the sample to only those observations of flights that, on a given query date t, have fewer than 40 seats left to sell (see Alderighi et al. (2015) for a similar problem). Two, conditional on a flight being non-censored, 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. Formally:

$$FNC_{jdt} = 1[z_1\theta_1 + \nu_1 > 0] \tag{9}$$

$$s^m_{jdt} = 1[z_1\theta_2 + \theta_3 Position^m + \nu_2 > 0] \text{ if } FNC_{jdt}=1. \tag{10}$$

When $FNC_{jdt} = 1$, i.e., a flight jd is non-censored at booking day t, we can identify, out of the possible 39 seats that the distribution may potentially include, the seats s in positions m which are still available for sale.[18] Under the assumptions $(\nu_1, \nu_2) \sim N(0, 1)$ and $\text{corr}(\nu_1, \nu_2) = \rho$, (9)-(10) can be estimated using a bivariate probit with sample selection model (Greene, 2003, ch.21), where z_1 includes the following regressors: dummies for the number of days to departure, the day of the week of the departure date, the departure slot time (morning, afternoon, evening, etc.), the season (Winter and Summer), the route (estimates available on request). After obtaining the estimated coefficients $(\hat{\theta}_2, \hat{\theta}_3)$ using all observations, the estimated Mill ratios for the selected observations are: $\hat{\lambda}^m_{jdt}(\hat{\theta}_2, \hat{\theta}_3) = \dfrac{\phi(z_1\hat{\theta}_2 + \hat{\theta}_3 Position^m)}{\Phi(z_1\hat{\theta}_2 + \hat{\theta}_3 Position^m)}$. We can then estimate an augmented version of (8):

$$\ln Fare^m_{jd} = \sum_t \beta_t D_t + \gamma Position^m + \hat{\lambda}^m_{jdt}(\hat{\theta}_2, \hat{\theta}_3) + \zeta_{jd} + \xi_{jdt}, \tag{11}$$

by panel OLS fixed-effects.[19]

The panel identifier corresponds to the combination of flight-code plus day of departure;

[18] Imagine that at t we only retrieve fares for, say, the last 20 seats; these would have $s_{jdt} = 1$. To estimate (9)-(10), we would append observations for seats 21-39 and set $s_{jdt} = 0$.

[19] The approach we follow to correct for the simultaneous presence of sample selection draws from procedure 17.1 in Wooldridge (2002).

the panel's temporal dimension is represented by a sequential counter that uniquely identifies all the possible combination of *Position* for all query dates t.[20] We set the earliest day to departure dummy (Days to departure 51+) as reference group and cluster the standard errors 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]

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

Table 7 reports the results. Models (1) and (2) use the full sample, while the others focus the analysis to the case of flights in, respectively, "Leisure" and "Business" routes. We do so to test whether the estimates for the full sample hold in sub-samples of more homogeneous flights. 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.[22] 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 percent, routes are respectively labeled as "Leisure" or "Business".

In odd-numbered models, which do not include the interaction between our variables of interests, the estimates indicate qualitatively similar effects. First, the coefficient of *Position* is, as expected, negative. That is, the econometric evidence indicates that the distributions of all flights are structured as predicted in property A. of Proposition 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 1.6% to 1.7%. There appear to be no difference between Leisure and Business routes. Third, and more importantly, 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, holding the position fixed. 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 by property B. of Proposition 2. Interestingly, in the "Business" sample, the coefficients of the

[20]Alternatively, we could have incorporated either the variable *Position* into the fixed effect identifier so that only the interaction model could be identified in the Fixed Effects estimation. The results would not change. Estimates available on request.

[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.

[22]The IPS does not cover routes with both endpoints outside the UK; hence, the combined number of observation in models (3) to (6) is lower than in model (1)-(2).

temporal dummies are somewhat larger in absolute magnitude, suggesting that in business routes the carrier tends to drop its fares over time more than it does in leisure routes; this results is not 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 (Alderighi et al., 2016). However, evidence not reported to save space indicates that fares in business routes tend to be higher on average.

To get a better appreciation of whether the intensity of the decline over time varies with the seat's position, even-numbered models present an interaction of *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 6 shows the predicted effects from model (2) of Table 7. 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 2. The slope varies to reflect the interaction terms in model (2). When the position is fixed, each point depicts the extent by which the average fare of each seat falls over time. Based on interaction coefficients in Table 7, the drop over time is larger as the position increases, as also shown descriptively in Table 6. For instance, the fare of seat 39 drops, on average, from a value around $e^{4.8} = 121$ to about $e^{4.1} = 60$; of seat 25 from about $e^{4.9} = 134$ to about $e^{4.4} = 81$, while for seat 1, the last one to be sold, the predicted fare moves from $e^{5.1} = 164$ to only about $e^{4.9} = 134$.

***** **Insert Figure 6 around here** *****

The econometric analysis therefore provides compelling evidence of the persistent effects of the hidden aspects of DP that the carrier implements to manage its yield; furthermore, it strongly supports the joint operation of the capacity and temporal dimensions as drivers of DP interventions.

6.2 The temporal profile of the seat on sale

As the foregoing discussion has highlighted, the use of a fare distribution bears important implications on the first seat on sale, that is, the seat with the lowest fare and the largest value for *Position* still available on a flight. Studying the fare of the seat on sale is important, because all the existing empirical literature on airline pricing, whether it uses transacted or posted fares, focusses exclusively on it. There is general consensus that the overall temporal profile of such a fare is upward sloping, with many papers reporting graphical and/or econometric evidence of fares increasing as the departure date nears.[23] The pervasiveness of such a correlation is strongly at odds with the theoretical prediction of fares falling as the takeoff date approaches (Gallego and van Ryzin, 1994), as first highlighted in McAfee and te Velde (2007). Subsequent empirical research has shown that after controlling for the remaining capacity on a flight, the theoretical prediction of temporally declining fares largely holds (Alderighi et al., 2015; Escobari, 2012).

Using the insights offered by the foregoing theoretical and empirical analysis, in this section we investigate the extent by which the behaviour of the fare of the seat on sale conforms to the evidence reported in the existing literature. While this provides further validation to the approach adopted in the paper, the combined analysis of the capacity and the temporal dimensions also helps clarify and consolidate an empirical approach to airline pricing where both dimensions are always properly accounted for. To this purpose, the econometric strategy hinges on testing properties A. and B. of Proposition 2 on the seat on sale, using the specification in equation (11) modified to take into account that for such a seat the censoring process can be modelled using equation (9) only.

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

Considering their panel fixed-effect design, models (1) and (2) in Table 8 replicate the regressions in McAfee and te Velde (2007), by first using the full sample with all observations, and then only the non-censored sample, i.e., the one we use to estimate equation (11). Like McAfee and te Velde (2007), the temporal trajectory is clearly either increasing or non-declining, with sharp rises during the last week. In terms of our analysis, we could interpret such result by saying that the capacity dimension is a stronger driving force than the temporal dimension, that is, movements along the distribution more than offset the negative impact of the declining option value. Recall, however, that fares may increase due to intertemporal price discrimination, aimed at exploiting customers' heterogeneity in terms of demand uncertainty and willingness to pay, so that late fares may be pushed up to take advantage of the larger proportion of business-people among potential buyers.

To tease out the possible separate impact of intertemporal price discrimination, we need to control for the evolution of available capacity on the flight, as in model (3), which uses only the non-censored observations to identify the number of seats left on the flight at a given point in time. Importantly, for the seat on sale, the number of available seats corresponds to the position of the first seat in the distribution. Such a property has important implications since it allows a dual interpretation of the estimates. Indeed, unlike the estimates in Table 7 where

[23]see Alderighi et al. (2015); Bergantino and Capozza (2015); Clark and Vincent (2012); Escobari (2012); Gaggero and Piga (2010); Koenisgsberg et al. (2008); McAfee and te Velde (2007); Stavins (2001) inter alia.

each seat occupies a fixed position in the distribution, the position of the first seat varies over time, and thus captures how the fare changes as the seat moves along the distribution. Moreover, the dual interpretation in terms of available seats allows a comparison with the results in the previous literature that looked at how the fare changes as the plane fills up, as for instance Alderighi et al. (2015) and Escobari (2012).

The inclusion of *Position* in model (3) drastically alters the structure of the temporal dummies to reveal a declining time path for fares, consistent with the prediction B. in this paper. Relative to those posted fifty-one or more days from departure, fares posted twenty-eight days or later are significantly different, and show a constant decreasing trend which is minimally reversed in the last three days before departure. Indeed, the coefficient of the "$0-3$ days" dummy is slightly larger than the previous one (-0.363 vs. -0.373), hinting to a U-shaped temporal profile (Alderighi et al., 2015; Bilotkach et al., 2010; Escobari, 2012). Combined with the descriptive evidence reported in Table 5, showing a larger probability of observing the DP treatment of a shift to a higher bucket for the first seat on sale, we can conclude that the increasing part of the U-shaped temporal path can be ascribed to the implementation of an inter-temporal price discrimination strategy by means of DP techniques that increase fares above the natural progression due to the capacity dimension.

The latter, however, is by far responsible for the overall upward trend highlighted in models (1) and (2). Indeed, the *Position*'s coefficient of -0.019 is similar to the ones estimated in Table 7, which, as discussed above, in this case can be interpreted in two ways. One, the first seat on sale follows an increasing temporal profile determined by the structure of the distribution. That is, the carrier tends to close a bucket once all the seats in that bucket are sold out, so that automatically the fare of the next bucket becomes the one advertised on the site. On average, a one-position movement to the right of the distribution increases fares by about 1.9%. 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. Two, and equivalently, fares increase by the same amount as an extra seat is sold.

The fact that the position of the first seat on sale varies over time suggests that the variable *Position* is likely correlated with ξ_{jdt} in eq. (11), i.e., it is endogenous. So we use two instruments in our identification strategy, similar to those in Alderighi et al. (2015). The first one, *Lag Position*, is simply the mean of the two weekly lagged values of *Position*, where the lags are intended over d and not t, that is, we take values for the same flights departing on the same week day one and two weeks before. The use of lagged values guarantees the instrument is not correlated with the shock ξ_{jdt}; furthermore, fare distributions are flight-specific, and so is the ideal (from the airline perspective) rate of growth of a flight's load factor. That is, the instrument is correlated with *Position* because the airline has likely adopted for the past flights a similar distribution, as well pursued a similar booking curve for the temporal progression of the load factor. The second instrument, *holiday period*, is a dummy variable indicating whether the query date falls within a holiday period in UK (Christmas, Easter, school breaks, etc) and captures possible differences on the demand side. That is, the ticket purchasing activity in such periods is likely to be different from non-holiday periods (e.g., when on holiday, a person has less time to spend planning future trips), and thus seat fares are likely less affected by shocks. Despite the loss of observations due to the use of a lagged instrument, the estimates in model (4) are equivalent to those in model (3), and confirm the presence of a weak U-shaped temporal profile and a slightly stronger capacity effect, with fares expected to increase by 2.0% every time an extra seat is sold.

The overall evidence we provide offers some insights into two aspects mentioned in the literature review. As far as inter-temporal price discrimination is concerned, the estimates indicate, during the last three days from departure, a weakening of the downward pressure that fares receive due to the temporal dimension. We can link this result to the DP activity on the first seat on sale in the form of its movement to a higher bucket. Notably, the evidence suggests a connection between the central role played by the fare distribution as a building block of airline pricing and its modification to implement price discrimination. As far as the presence of strategic consumers is concerned, the first two models in Table 8 indicate that a consumer would generally observe fares following an increasing trend, which Li et al. (2014) describe as the standard way to curb the incentive to postpone purchase. Although the temporal dimension is largely responsible for the upward trend of fares throughout the entire booking period, we record an additional effect due to an inter-temporal price discrimination motive during the last week.

7 Conclusions

This paper presents several strong reasons, both based on theoretical and empirical grounds, for modelling airline pricing using the concept of a fare distribution. For instance, it allows the investigation of many so far neglected aspects of Dynamic Pricing. Furthermore, it helps solve the contrast between theory and empirical evidence illustrated in McAfee and te Velde (2007).

Although not the central focus of the study, its main findings have profound implications on the identification of inter-temporal price discrimination strategies in airline markets. For instance, carriers may want to discriminate the business travelers' segment from other lower demand travelers, e.g., those traveling for leisure. 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, carriers would implement price discrimination by raising fares at a specified interval before take-off. But because carriers use fare distributions, identifying the discriminatory motive is only possible if the researcher can distinguish higher fares driven by capacity considerations (i.e., the fare is high because only a few seats are left on the flight) from discriminatory upward changes in the distribution that increase all relevant fares but are not motivated by a change in a flight's load factor. Although we find the carrier in our study does not pursue an inter-temporal price discrimination strategy too intensively, the approach set out in the paper could be fruitfully applied in future research.

References

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.

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

Cavallo, A. (2017). Are online and offline prices similar? evidence from large multi-channel retailers. *American Economic Review*, forthcoming.

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., Rupp, N. G., and Meskey, J. (2016). Dynamic Price Discrimination in Airlines. Working paper, n.2248124, SSRN.

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.

Gerardi, K. and Shapiro, A. (2009). Does Competition Reduce Price Dispersion? New Evidence from the Airline Industry. *Journal of Political Economy*, 117:1–37.

Gorodnichenko, Y. and Talavera, O. (2017). Price Setting in Online Markets: Basic Facts, International Comparison, and Cross-Border Integration. *American Economic Review*, forthcoming.

Greene, W. H. (2003). *Econometric Analysis.* Prentice Hall, New York, 5th edition.

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.

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

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.

Stavins, J. (2001). Price Discrimination in the Airline Market: The Effect of Market Concentration. *The Review of Economics and Statistics*, 83(1):200–202.

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 Practice of Revenue Management*. Springer, New York.

Topkis, D. M. (1998). *Supermodularity and complementarity*. Princeton University Press, Princeton, NJ, 4^{th} edition.

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

Table 1: Simulated observed number of seats for each bucket fare across booking periods

bkt fares	pdf WTP	periods to departure, t 11	10	9	8	7	6	5	4	3	2	1
50	14/64											
65	12/64	5	5	4								
80	8/64	6	6	5	4	4	3	3	2	2	1	
95	6/64	7	6	6	4	4	4	3	3	2	2	1
110	6/64	5	5	4	4	4	3	3	2	2	1	1
130	5/64	6	5	5	5	4	4	3	3	2	2	1
150	5/64	4	3	3	3	2	2	2	1	1	1	
175	4/64	3	4	3	2	3	2	1	2	1		
200	4/64	3	2	2	2	1	1	1				
seats		39	36	32	22	22	19	16	13	10	7	3

Figure 1: Fare distribution and probability by periods to departure

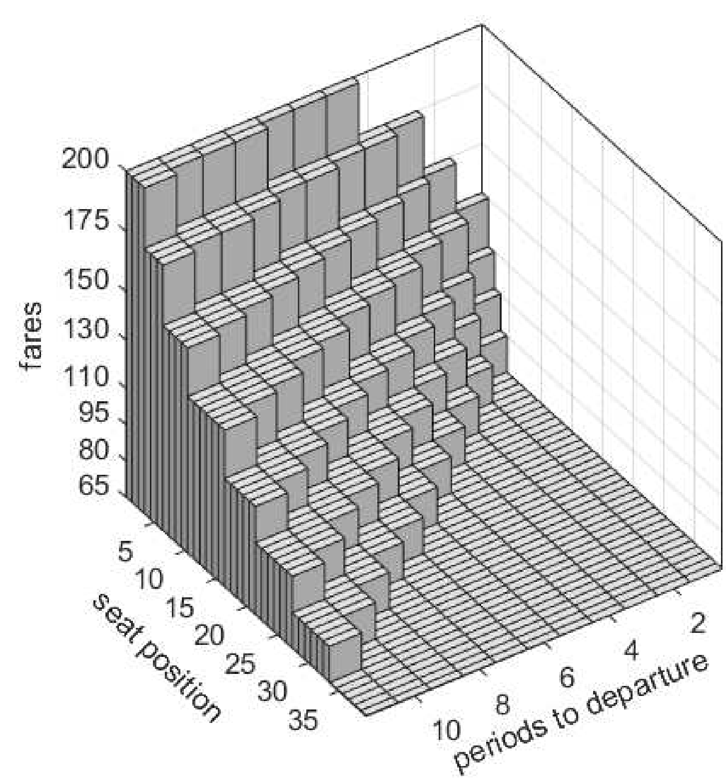


Table 2: Number of seats in each bucket price across days to departure

Panel A: days to departure 130-35

Days to dep.	130	129	121	120	119	111	110	109	100	81	79	70	69	50	49	48	36	35
Bkt. price																		
31	3	1				3	3	1										
38	5	5	5	5	5	5	5	5	5									
48	5	5	5	5	5	5	5	5	5			4	6	5	5	5		
58	6	5	5	5	5	6	6	5	5	4	2	4	4	5	5	5	2	2
68	4	5	6	6	6	4	4	5	6	4	4	4	4	5	5	5	4	4
80	6	6	4	4	4	6	6	6	4	4	4	5	5	5	5	5	4	4
96	4	4	6	6	6	4	4	4	6	5	4	3	3	5	5	5	4	4
112	4	4	4	4	4	4	4	4	3	3	5	5	5	5	5	5	4	4
133	3+	5+	5+	5+	5+	3+	3+	5+	4	4	3	3	3	5	5	5	4	4
158									2+	16+	18+	12+	10+	5+	5+	5+	18+	18+
Av. seats	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+	40+

Panel B: days to departure 34-1

Days to dep.	34	22	21	20	19	18	17	16	15	14	13	10	9	5	4	3	2	1
Bkt. price																		
31																		
38																		
48			1															
58	2	5	5	4	2	2	2	1										
68	4	5	5	5	5	4	4	5	1	7	3							
80	4	5	5	5	5	4	4	5	5	6	6	6	6					
96	4	5	5	5	5	4	4	5	5	6	6	6	6	1				
112	4	5	5	5	5	4	4	5	5	6	6	6	6	6	6	5	3	
133	4	4	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4
158	18+	11+	11+	13+	15+	19+	19+	16+	19	10	10	4	4	2	2	2	0	0
Av. seats	40+	40+	40+	40+	40+	40+	40+	40+	38	38	34	25	25	12	11	10	6	4

(a) The + sign indicates a censored value for either the bucket size or the number of available seats.

Table 3: Probability to observe specific forms of Dynamic Pricing applied to each seat in a fare distribution

	Fare Change to						
Position in distribution	Higher bkt	Higher new bkt	Lower bkt	Lower new bkt	Bkt size increase	Bkt size decrease	Obs.
1	0.017	0.012	0.043	0.004	0.074	0.267	126,108
2	0.014	0.011	0.043	0.004	0.076	0.273	123,188
3	0.015	0.010	0.051	0.004	0.067	0.273	120,613
4	0.033	0.010	0.074	0.005	0.079	0.280	118,288
5	0.033	0.010	0.084	0.005	0.085	0.275	115,919
6	0.028	0.010	0.092	0.006	0.091	0.268	113,364
7	0.026	0.009	0.097	0.006	0.099	0.266	110,815
8	0.025	0.008	0.097	0.006	0.103	0.260	108,156
9	0.023	0.008	0.100	0.007	0.109	0.258	105,217
10	0.023	0.009	0.110	0.008	0.119	0.252	102,138
11	0.025	0.008	0.112	0.008	0.126	0.244	98,979
12	0.034	0.010	0.133	0.010	0.130	0.237	95,416
13	0.041	0.003	0.141	0.004	0.135	0.228	92,138
14	0.040	0.004	0.141	0.005	0.138	0.218	88,913
15	0.039	0.004	0.142	0.005	0.142	0.205	85,466
16	0.038	0.004	0.144	0.005	0.146	0.197	81,680
17	0.038	0.005	0.149	0.006	0.148	0.187	77,882
18	0.047	0.007	0.164	0.008	0.151	0.169	73,808
19	0.044	0.007	0.161	0.008	0.153	0.157	70,005
20	0.042	0.008	0.159	0.009	0.157	0.147	66,351
21	0.043	0.009	0.167	0.011	0.163	0.134	62,392
22	0.041	0.008	0.163	0.011	0.167	0.127	58,544
23	0.041	0.009	0.163	0.012	0.169	0.119	54,831
24	0.042	0.011	0.174	0.016	0.178	0.110	50,852
25	0.048	0.006	0.174	0.009	0.179	0.101	47,133
26	0.050	0.008	0.176	0.011	0.185	0.091	43,497
27	0.047	0.009	0.175	0.012	0.189	0.083	39,881
28	0.045	0.010	0.176	0.016	0.197	0.071	36,253
29	0.044	0.012	0.165	0.019	0.208	0.063	32,768
30	0.041	0.013	0.164	0.023	0.217	0.052	29,364
31	0.038	0.016	0.162	0.028	0.221	0.044	25,811
32	0.038	0.019	0.149	0.031	0.223	0.038	22,395
33	0.033	0.022	0.137	0.034	0.231	0.034	18,919
34	0.032	0.024	0.127	0.042	0.244	0.031	15,559
35	0.028	0.025	0.114	0.048	0.258	0.025	12,162
36	0.022	0.022	0.113	0.052	0.276	0.024	8,948
37	0.020	0.018	0.100	0.074	0.305	0.020	5,945
38	0.019	0.013	0.086	0.121	0.341	0.011	3,386
39	0.014	0.005	0.038	0.139	0.446	0.000	1,334

Table 4: Probability to observe specific forms of Dynamic Pricing applied to a fare distribution, over booking periods

	Fare Change to							
Days to departure	Higher bkt	Higher new bkt	Lower bkt	Lower new bkt	Any fare move	Bkt size increase	Bkt size decrease	Any size change
0-3	0.176	0.038	0.143	0.013	0.330	0.198	0.160	0.229
4-7	0.109	0.040	0.206	0.021	0.327	0.182	0.196	0.231
8-10	0.136	0.042	0.235	0.030	0.374	0.218	0.243	0.280
11-14	0.131	0.047	0.743	0.106	0.822	0.675	0.776	0.799
15-21	0.161	0.051	0.729	0.090	0.840	0.707	0.804	0.820
22-28	0.120	0.059	0.786	0.106	0.874	0.631	0.839	0.861
29-36	0.106	0.036	0.621	0.106	0.726	0.544	0.677	0.700
36-50	0.058	0.040	0.406	0.046	0.464	0.341	0.444	0.457
51+	0.024	0.048	0.158	0.019	0.211	0.129	0.187	0.191
Overall	0.141	0.042	0.350	0.045	0.486	0.341	0.372	0.412

Table 5: Probability to observe specific forms of Dynamic Pricing applied to the seat on sale, over booking periods

	Fare Change to					
Days to departure	Higher bkt	Higher new bkt	Lower bkt	Lower new bkt	Any fare move	Bkt size increase
0-3	0.138	0.029	0.053	0.006	0.226	0.083
4-7	0.087	0.030	0.074	0.011	0.202	0.068
8-10	0.101	0.029	0.075	0.014	0.218	0.084
11-14	0.073	0.033	0.198	0.063	0.367	0.308
15-21	0.071	0.040	0.118	0.049	0.279	0.309
22-28	0.056	0.046	0.109	0.064	0.275	0.394
29-36	0.043	0.030	0.073	0.082	0.228	0.355
36-50	0.032	0.026	0.076	0.025	0.158	0.274
51+	0.019	0.014	0.057	0.015	0.105	0.105
Overall	0.098	0.031	0.093	0.025	0.247	0.150

Table 6: 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 7: OLS Regression analysis of the price of all seats in the distribution (Option value)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	log(p)	log(p)	log(p)	log(p)	log(p)	log(p)
Sample	All routes	All routes	Leisure	Leisure	Business	Business
Days to departure 0-3	-0.280***	-0.030	-0.272***	-0.005	-0.320***	-0.114*
	(0.001)	(0.057)	(0.063)	(0.090)	(0.041)	(0.053)
Days to departure 4-7	-0.275***	0.006	-0.272***	0.024	-0.304***	-0.064
	(0.001)	(0.057)	(0.063)	(0.090)	(0.041)	(0.053)
Days to departure 8-10	-0.265***	0.038	-0.266***	0.056	-0.282***	-0.022
	(0.001)	(0.057)	(0.062)	(0.089)	(0.040)	(0.053)
Days to departure 11-14	-0.211***	0.069	-0.215***	0.094	-0.220***	0.008
	(0.001)	(0.057)	(0.062)	(0.089)	(0.040)	(0.053)
Days to departure 15-21	-0.107***	0.170**	-0.106	0.205*	-0.115**	0.097
	(0.001)	(0.057)	(0.062)	(0.089)	(0.040)	(0.053)
Days to departure 22-28	-0.056***	0.185**	-0.047	0.228*	-0.079*	0.086
	(0.001)	(0.057)	(0.061)	(0.089)	(0.040)	(0.053)
Days to departure 29-35	0.008***	0.191***	0.019	0.233**	-0.019	0.078
	(0.001)	(0.056)	(0.061)	(0.089)	(0.039)	(0.052)
Days to departure 36-50	0.023***	0.127*	0.037	0.168*	-0.020	0.010
	(0.001)	(0.053)	(0.057)	(0.084)	(0.037)	(0.049)
Position	-0.016***	-0.002	-0.017***	-0.001	-0.017***	-0.006***
	(0.000)	(0.002)	(0.000)	(0.002)	(0.000)	(0.001)
Position*Days to dep. 0-3		-0.019***		-0.021***		-0.014***
		(0.002)		(0.002)		(0.001)
Position*Days to dep. 4-7		-0.020***		-0.022***		-0.016***
		(0.002)		(0.002)		(0.001)
Position*Days to dep. 8-10		-0.020***		-0.021***		-0.015***
		(0.002)		(0.002)		(0.001)
Position*Days to dep. 11-14		-0.017***		-0.019***		-0.012***
		(0.002)		(0.002)		(0.001)
Position*Days to dep. 15-21		-0.016***		-0.018***		-0.011***
		(0.002)		(0.002)		(0.001)
Position*Days to dep. 22-28		-0.013***		-0.016***		-0.008***
		(0.002)		(0.002)		(0.001)
Position*Days to dep. 29-35		-0.010***		-0.012***		-0.004***
		(0.002)		(0.002)		(0.001)
Position*Days to dep. 36-49		-0.006***		-0.007**		-0.000
		(0.001)		(0.002)		(0.001)
Heckman's λ	-0.102***	-0.015***	-0.092***	-0.000	-0.103***	-0.023***
	(0.000)	(0.004)	(0.005)	(0.006)	(0.006)	(0.006)
R^2	0.70	0.71	0.69	0.71	0.72	0.73
Observations	5,510,306	5,510,306	2,430,891	2,430,891	2,087,774	2,087,774

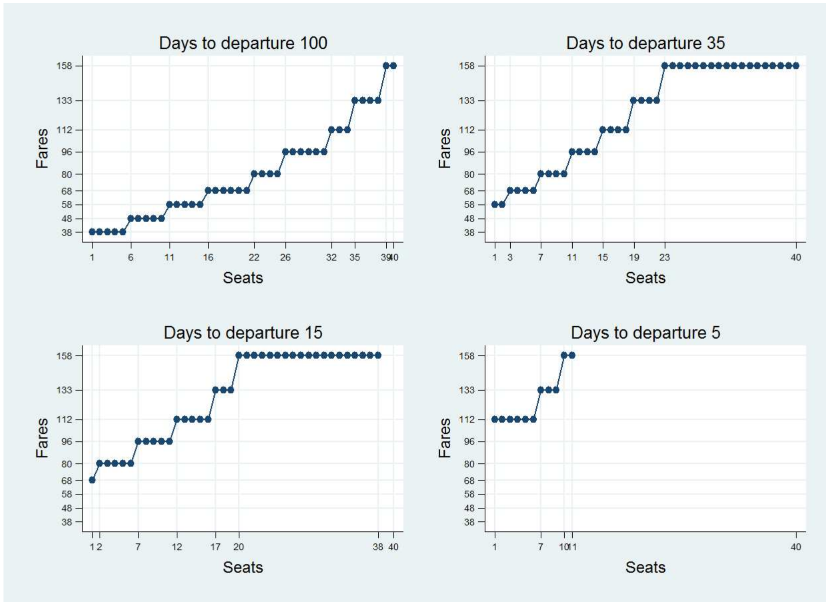
(a) Flight-code fixed effects.
(b) ***, ** and * denote statistical significance at 1%, at 5% and at 10% level.

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

	(1)	(2)	(3)	(4)
Dependent variable	log(p)	log(p)	log(p)	log(p)
Estimation technique	OLS-FE	OLS-FE	OLS-FE	IV-FE
Sample	All obs.	Not cens. obs.	Not cens. obs.	Not cens. obs.
Days to departure 0-3	0.785***	0.323***	-0.363***	-0.393***
	(0.007)	(0.038)	(0.074)	(0.101)
Days to departure 4-7	0.666***	0.205***	-0.373***	-0.399***
	(0.007)	(0.038)	(0.073)	(0.101)
Days to departure 8-10	0.490***	0.050	-0.370***	-0.388***
	(0.006)	(0.038)	(0.072)	(0.100)
Days to departure 11-14	0.403***	-0.005	-0.307***	-0.318***
	(0.005)	(0.038)	(0.071)	(0.099)
Days to departure 15-21	0.351***	-0.026	-0.211***	-0.216**
	(0.005)	(0.038)	(0.070)	(0.099)
Days to departure 22-28	0.305***	-0.035	-0.121*	-0.122
	(0.005)	(0.038)	(0.069)	(0.099)
Days to departure 29-35	0.277***	0.000	-0.020	-0.022
	(0.004)	(0.038)	(0.068)	(0.098)
Days to departure 36-50	0.170***	-0.005	0.027	0.026
	(0.003)	(0.037)	(0.065)	(0.094)
Position=Available Seats			-0.019***	-0.020***
			(0.000)	(0.001)
Heckman's λ			-0.022*	-0.024*
			(0.011)	(0.013)
Kleibergen-Paap rk LM stat				709.215***
Hansen J-stat				.00318
R^2	0.568	0.411	0.622	0.629
Observations	901,751	252,063	252,063	174,066

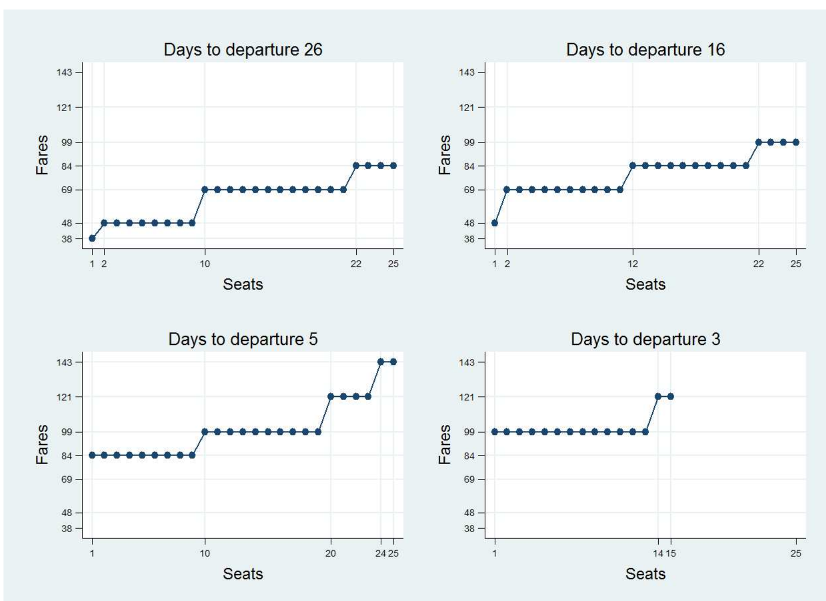
(a) Flight-code fixed effects.
(b) ***, ** and * denote statistical significance at 1%, at 5% and at 10% level.

Figure 2: Fare distribution at various days to departure



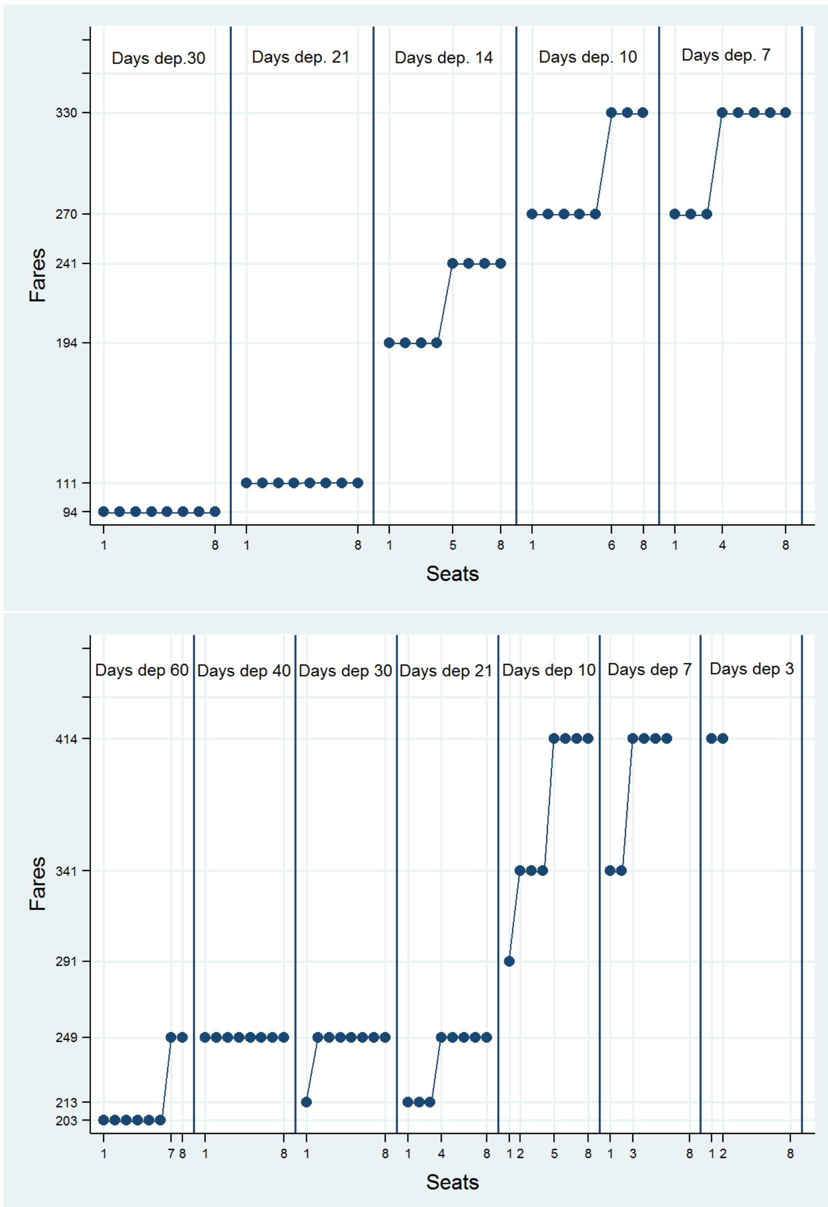
Legenda - Flight EZY8716 from Lisbon (6:45) to London Gatwick (9.25) on 22 Jun 2014

Figure 3: Fare distribution at various days to departure (Ryanair)



Legenda - Flight FR 8547 from Berlin Schonefeld (21:55) to London Stansted (22:40) on 21 Oct 2011

Figure 4: Fare distribution at various days to departure (Southwest). 'Wanna Get Away' fares.



Upper Flight - Chicago MDW (6:00) to New York LGA (9:05) on 9 Nov 2012

Lower Flight - Chicago MDW (18:20) to Los Angeles on 21 Sept 2012

Figure 5: American Airlines, JKF-ORD on 15 November 2016

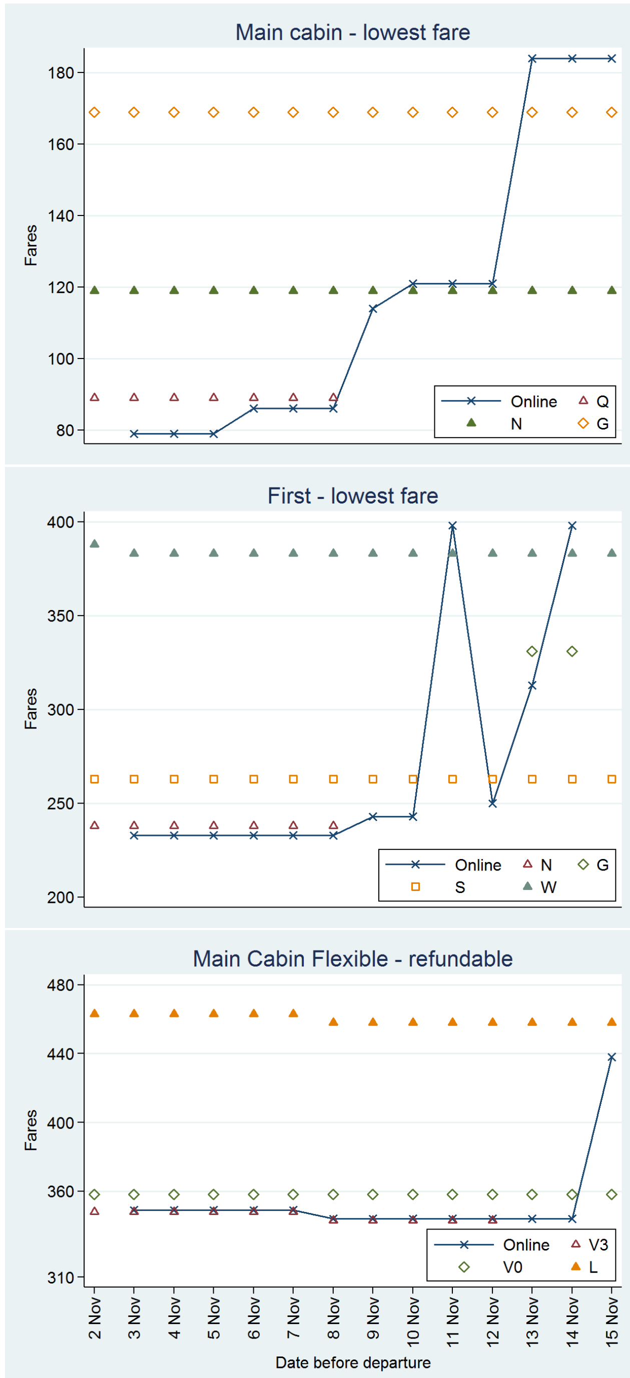
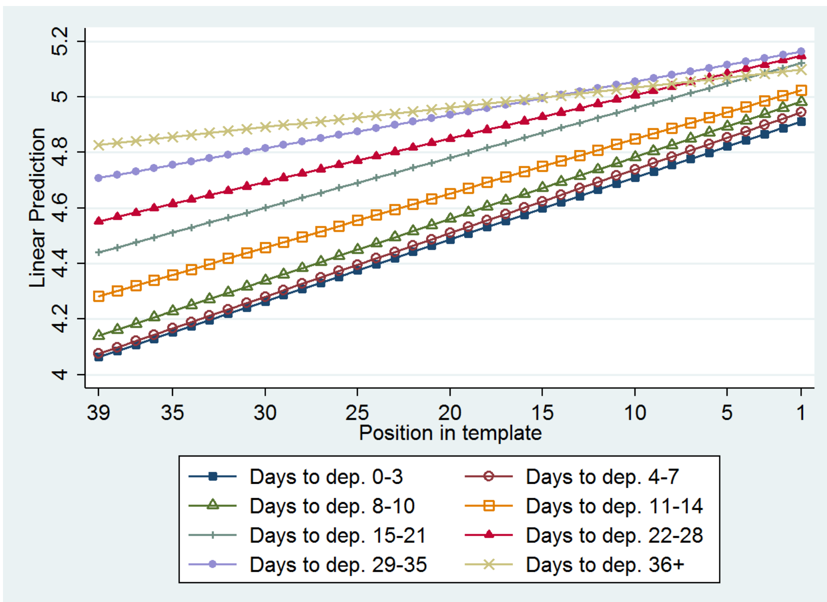


Figure 6: Predicted marginal effects of Position and days to departure on prices. Note: based on Model (4) in Table 7.



A Appendix - For Online Publication

A.1 Proofs

Proof. of Proposition 1

Non-negativeness. Non-negativity of V can be easily shown from (2) by induction since $V(t, M)$ comes from the maximization over p of sums and products of nonnegative terms.

Increasing in both arguments. We show that $V(t, M) \geq V(t-1, M)$. By contradiction assume that $V(t, M) < V(t-1, M)$. Let $p^*(\tau, m)$ with $\tau = 1, \ldots, t-1$ and $m = 1, \ldots, M$, be the set of fares that solves (2) 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 $\hat{p}(1, m) = \bar{p} \in (0, \bar{\theta})$. 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 $V(t, m) \geq V(t, M-1)$ is similar to the previous case.

Decreasing difference in t and M and increasing differences in (t, M). We organize this part of the proof in different steps.

Step 1. We introduce the following notation: $\Delta_1(t, M) = V(t, M) - V(t-1, M)$ and $\Delta_2(t, M) = V(t, M) - V(t, M-1)$. Note that decreasing differences in t and M can be, respectively, defined as:

$$\Delta_1(t, M) \leq \Delta_1(t-1, M), \qquad \Delta_2(t, M) \leq \Delta_2(t, M-1). \tag{A.1}$$

Moreover, increasing differences in (t, M) is guaranteed by one of these two equivalent expressions:

$$\Delta_2(t-1, M) \leq \Delta_2(t, M), \qquad \Delta_1(t, M-1) \leq \Delta_1(t, M). \tag{A.2}$$

Indeed, we can write: $V(t_H, M) - V(t_L, M) = \Delta_1(t_H, M) + \Delta_1(t_H - 1, M) + \cdots + \Delta_1(t_L + 1, M)$. Thus, increasing difference property as in Definition 2 requires that the following inequality holds $\Delta_1(t_H, M_H) + \Delta_1(t_H - 1, M_H) + \cdots + \Delta_1(t_L + 1, M_H) \geq \Delta_1(t_H, M_L) + \Delta_1(t_H - 1, M_L) + \cdots + \Delta_1(t_L + 1, M_L)$, or $\Delta_1(t_H, M_H) - \Delta_1(t_H, M_L) + \Delta_1(t_H - 1, M_H) - \Delta_1(t_H - 1, M_L) + \cdots + \Delta_1(t_L + 1, M_H) - \Delta_1(t_L + 1, M_L) \geq 0$. This is guaranteed by (A.2).

Step 2. We rewrite the Bellman equation in an useful way. First note that (2) can be rephrased as:

$$\begin{aligned} \Delta_1(t, M) &= \max_p \{q(p)[p + V(t, M-1) - V(t-1, M)]\} \\ &= \max_p \{q(p)[p + X(t, M)]\}, \end{aligned} \tag{A.3}$$

where $X(t, M) = V(t, M-1) - V(t-1, M)$. Note that the solution of the maximization problem

$p = \arg\max_p \{q(p)[p+X]\}$ does not change since we have added a constant term $X(t,M)$. Moreover, from the Envelope theorem, $\Delta_1(t,M)$ is increasing in X. Therefore, it is possible to state the following result:

$$\begin{aligned} \Delta_1(t,M) \le \Delta_1(t-1,M) &\iff X(t,M) \le X(t-1,M) \\ &\iff V(t,M-1) - V(t-1,M) \le V(t-1,M-1) - V(t-2,M) \\ &\iff \Delta_1(t,M-1) \le \Delta_1(t-1,M). \end{aligned} \tag{A.4}$$

Moreover:

$$\begin{aligned} \Delta_1(t,M-1) \le \Delta_1(t,M) &\iff X(t,M-1) \le X(t,M) \\ &\iff V(t,M-2) - V(t-1,M-1) \le V(t,M-1) - V(t-1,M) \\ &\iff \Delta_2(t-1,M) \le \Delta_2(t,M-1). \end{aligned} \tag{A.5}$$

Similarly, (2) can be rephrased as:

$$\begin{aligned} \Delta_2(t,M) &= \max_p \{q(p)\,p + [1-q(p)]\,[V(t-1,M) - V(t,M-1)]\} \\ &= \max_p \{q(p)\,p + [1-q(p)]\,Y(t,M)\}, \end{aligned} \tag{A.6}$$

where $Y(t,M) = V(t-1,M) - V(t,M-1)$. Also in this case, from the Envelope theorem, $\Delta_2(t,M)$ is increasing in Y. Therefore:

$$\begin{aligned} \Delta_2(t,M) \le \Delta_2(t,M-1) &\iff Y(t,M) \le Y(t,M-1) \\ &\iff V(t-1,M) - V(t,M-1) \le V(t-1,M-1) - V(t,M-2) \\ &\iff \Delta_2(t-1,M) \le \Delta_2(t,M-1). \end{aligned} \tag{A.7}$$

Moreover:

$$\begin{aligned} \Delta_2(t-1,M) \le \Delta_2(t,M) &\iff Y(t-1,M) \le Y(t,M) \\ &\iff V(t-2,M) - V(t-1,M-1) \le V(t-1,M) - V(t,M-1) \\ &\iff \Delta_1(t,M-1) \le \Delta_1(t-1,M). \end{aligned} \tag{A.8}$$

Previous results presented in (A.5), (A.6), (A.8) and (A.9) can be summarized as follows:

$$\Delta_1(t,M) \le \Delta_1(t-1,M) \iff \Delta_2(t-1,M) \le \Delta_2(t,M) \iff \Delta_1(t,M-1) \le \Delta_1(t-1,M) \tag{A.9}$$
$$\Delta_1(t,M-1) \le \Delta_1(t,M) \iff \Delta_2(t,M) \le \Delta_2(t,M-1) \iff \Delta_2(t-1,M) \le \Delta_2(t,M-1) \tag{A.10}$$

Note that inequalities presented in (A.1) are equivalent to those presented in (A.2). Thus, in order to show that $V(t,M)$ has decreasing differences in t and M and increasing differences in (t,M) we can only need to prove that inequalities presented in (A.1) are satisfied.

Step 3. We proof that inequalities in (A.1) are satisfied by induction. We start to show that inequalities in (A.1) hold for any $(t,1)$ or $(1,M)$, with $t = 1,2,..,T$ and $M = 1,2,..,N$. When $M = 1$, from (A.4), $X(t,1) = -V(t-1,1)$. Since $V(t-1,1) \geq V(t-2,1)$, using (A.5), we have that $X(t,1) \leq X(t-1,1)$ and $\Delta_1(t,M) \leq \Delta_1(t-1,M)$. Similarly, when $t = 1$, from (A.7), $Y(t,M) = -V(t,M-1)$. Since $V(t-1,M) \geq V(t-2,M)$, using (A.8), we have that $X(t,M) \leq X(t-1,M)$ and $\Delta_2(t,M) \leq \Delta_2(t,M-1)$.

Because we have two different indices (t,M), in order to provide a proof by induction we need to introduce an ordering, $((t,M),\prec)$, on the indexes $t = 1,2,..T$ and $M = 1,..,N$. We assume that there is a lexicographic order in (t,M), i.e. $(t',M') \prec (t,M)$ when $t' < t$ or when $t' = t$ and $M' < M$. Thus, we have to prove two different cases.

Case a. We assume that inequalities in (A.1) hold for $(t-1,N)$ and we want to show that they hold for $(t,1)$. This has been already done above.

Case b. We assume that inequalities in (A.1) hold for preceding values of (t,M), in particular for $(t-1,M)$ and $(t,M-1)$, and we want to show that they hold for (t,M). Using as assumption that the first inequality of (A.1) holds for $(t,M-1)$ and that the second inequality of (A.1) holds for $(t-1,M)$, we obtain:

$$\Delta_1(t,M-1) \leq \Delta_1(t-1,M-1) \leq \Delta_1(t-1,M). \tag{A.11}$$

Using (A.9), we obtain the proof that the first inequality in (A.1) is satisfied for (t,M).

Similarly, using as assumption that the second inequality of (A.1) for $(t-1,M)$ holds and that the first inequality of (A.2) holds for $(t,M-1)$, we obtain:

$$\Delta_2(t-1,M) \leq \Delta_2(t-1,M-1) \leq \Delta_2(t,M-1). \tag{A.12}$$

Using (A.10), we obtain the proof that the second inequality in (A.1) is satisfied for (t,M).

■

Proof. of Corollary 1

It directly follows from (A.1) and (A.2) and by the fact that $X(t,M) = \Delta_2(t,M) - \Delta_1(t,M)$. ■

A.2 Algorithm

As noted in proof of Proposition 1, (2) can be written as:

$$V(t, M) = \max_{p} \{q(p)[p + V(t, M-1) - V(t-1, M)]\} + V(t-1, M) \tag{A.13}$$

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.4), we consider the following steps.

Step 1. Find the solution for $\max_{p \in \Theta} q(p)(p + X)$. Since Θ is compact, there exists a solution for the problem.

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.4) 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.3 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:

$$PF(s) = \frac{C + \sum_{j=1}^{s} p_j}{s}, \tag{A.14}$$

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 2.

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 $p_1 = p_2 = p$:

$$PF(1) = C + p$$
$$PF(2) = (C + 2p)/2$$

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.14) it is straightforward to derive the bucket fare tags, P_j:

$$P_j = j * PF(j) - (j-1) * PF(j-1) \text{ with } j \in [2, 40], \tag{A.15}$$

with $P_1 = PF(1) - C$.[24]

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 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".[25] We can then gauge the extent to which the size of each bucket, obtained from (A.15), 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.

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

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