The Hidden Side of Dynamic Pricing: Evidence from the Airline Market*

Marco Alderighi
Università della Valle d'Aosta
Università Bocconi

Alberto A. Gaggero
University of Pavia

Claudio A. Piga†
Keele University

November 24, 2017

Abstract

Both our theoretical analysis and empirical evidence describe how airlines price in distribution: they set a possibly different fare for each seat on an aircraft. Focussing on fare distributions unveils some hidden dynamic pricing strategies. We show how a flight's fare distribution is set in practice and is consistent with two main theoretical predictions. First, fare distributions are increasing over the seats' order of sale. Second, over time fare distributions move downward to reflect the perishable nature of seats, although inter-temporal price discrimination may attenuate such a reduction. Overall, fares for the seat on sale increase over time.

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, Jordi Perdiguero, Patrick Legros, Alicia Rambaldi, Chris Tsoukis, the participants of the 2016 IIOC Conference in Philadelphia, the 2016 EARIE Conference in Lisbon, the 2017 Royal Economic Society Annual Conference in Bristol, the 2017 Workshop on Air Transport in Barcelona, and the seminar participants at Lancaster University, Northeastern University, University of Queensland, University of Technology Sydney and Swansea University. This article 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, Keele University, ST5 5BG, UK. email: c.piga@keele.ac.uk

1 Introduction

The definition of dynamic pricing (DP) in airline markets, both in the academic literature and 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). Very little attention has been paid to how DP is connected to the peculiar features of a carrier's Revenue Management (RM) system (McGill and Van Ryzin, 1999; Talluri and van Ryzin, 2004). In this article we show that once this connection is made, the extant definition of DP in the literature is either lacking, to the extent that it fails to consider cases of DP not involving an observable price variation, or misleading, because it allows for instances of price variation inconsistent with DP.

More precisely, we argue that, to capture how DP operates, it is first essential to understand a fundamental principle driving airlines' price setting behaviour. We propose a novel empirical approach based on the theoretical work by Dana (1999) and Gallego and van Ryzin (1994), where we abandon the focus on a single fare so far used in the literature and replace it with the analysis of a fare distribution. Loosely speaking, focussing 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.

This study is the first in the literature to show how fare distributions are shaped in practice. In all the 43,275 flights in our sample, they are found to be stepwise increasing: the airline arranges seats into groups, denoted as 'buckets', each defined by an increasing price tag and a variable number of seats. 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. For instance, we do not classify fare increases over time as DP when such increases arise from a movement along the distribution; i.e., when a bucket is sold out, automatically the seats allocated to the next higher bucket are put on sale, without causing any modification in the distribution. Instead, we consider as an instance of DP only a situation involving an identifiable change in the fare distribution; e.g., at least one seat's fare is increased or decreased (i.e., the seat is moved to a higher or lower bucket). None of these forms of DP in airline markets has been previously considered in the literature. Thus, we can also gauge how frequently distributions vary, finding that, on average, they remain unchanged for about 2-3 consecutive days.

In addition to showing that the use of a fare distribution is consistent with an airline's optimizing behavior, through our theoretical model we can tease out the role of two antagonistic effects, which are at work simultaneously to determine the fares of every seat. First, airlines sell a limited number of seats. Thus, fares should increase as the number of seats reduces, that is, as scarcity increases ("capacity effect") (Puller et al., 2009; Talluri and van Ryzin, 2004). Second, 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 a seat, i.e., the expected value of having an additional seat in a period ("temporal effect"). Overall, 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.

In an extension to the theoretical model, we investigate the role of a price discriminatory motive as a counter-force to the temporal effect. Assuming that the proportion of consumers with a higher willingness to pay (e.g., business-people) increases during the last few periods prior to departure, our simulations indicate that the presence of a more heterogeneous customer basis has a dampening impact on the intensity of the temporal effect. Thus, a stronger incentive to pursue an inter-temporal price discrimination strategy operates across periods by either slowing down the shift of seats from upper (higher-priced) to lower buckets, or by moving seats from a lower to an upper bucket.

The econometric analysis gauges the impact on the shape of fare distributions induced by both the capacity and the temporal effect, and the time-varying distribution of willingness to pay (WTP). We find that the capacity effect 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 the case of Ryanair. Furthermore, unlike McAfee and te Velde (2007), the analysis provides strong empirical support in favor of the theoretical models, including ours, predicting a declining temporal effect. Although a similar finding has been shown in Sweeting (2012) for the second-hand market price of a single baseball ticket, the focus on a fare distribution reveals that the temporal effect is a force that effectively pushes downward *all* the seats in the distribution.

In sum, the analysis posits the central role of the fare distribution as the starting point to investigate the presence of DP in airline markets and how it manifests itself in "hidden" forms involving seats not directly on sale. It also emphasizes the intertwined relationship involving a "standard" pricing approach in the industry, DP and price discrimination, to suggest that the latter cannot be simply inferred by an inter-temporal price profile that is increasing over time, due to the strong confounding impact of the capacity effect (Stavins, 2001). More generally, the study highlights the impact of both the capacity and the temporal effect as fundamental drivers of airlines' pricing behaviour.

The rest of the paper is structured as follows. The next section revises the main contributions of both theoretical and empirical literature. The collection of fare data is described in Section 3, followed by real-world examples of fare distributions. The theoretical model in Section 5 generalises the capacity and temporal effects, and highlights the moderating role of inter-temporal price discrimination on the latter effect. Section 6 develops a descriptive analysis on how distributions and dynamic pricing are related to the capacity and temporal effects. The econometric investigation testing the properties of the theoretical model's equilibrium solution is carried out in Section 7. Finally, Section 8 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, whereas the latter reflects a change during the selling season in the probability that demand for a flight exceeds capacity. More importantly, the systematic and stochastic peak-load pricing 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 (McGill and Van Ryzin, 1999; Talluri and van Ryzin, 2004).[1] RM thus encompasses both a systematic and a dynamic pricing component, 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. In this sense, DP and stochastic peak-load pricing may be considered as synonymous.

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

[1] RM involves a number of ancillary activities and techniques useful in the process.

is that the fare charged should reflect the ranked position of the seat on sale in the fare distribution. To implement such a test, it is therefore necessary to know a flight's load factor at the time a fare is either posted online or a ticket is sold. This issue has been empirically tackled either by the use of web crawling methods (Alderighi et al., 2015), or of seat maps posted by online travel agents (Clark and Vincent, 2012; Escobari, 2012; Williams, 2017). 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 expected value of holding the unit in the next period. As shown in Sweeting (2012), under standard conditions common to most models, the opportunity cost not to sell a ticket 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 a pricing system based on the definition of a fare distribution over capacity units. In this article's theoretical model, 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 effect).

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 RM models that predict a declining option value assume a constant distribution of WTP, and therefore do not account for the fact that business travelers tend to book at a later stage

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

(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 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 finds a positive, but somewhat limited, impact of fare hikes introduced a few days before departure; instead, it makes the novel point that the capacity effect 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).

More generally, this is the first article to analyse and model simultaneously the joint impact of the capacity and the temporal effect together with the mediating role on the latter by the inter-temporal price discriminatory motive. Pang et al. (2015) also consider both effects in a model that predicts bid prices to be first increasing and then decreasing, once a certain level of capacity is reached. They do not allow for customer heterogeneity and, hence, for a possible discriminatory motive. Furthermore, they do not put their theoretical prediction to an empirical test. Williams (2017, p.36) develops an empirical structural model based on the assumption that DP is generated by two broad rationales: segmentation of consumers (i.e., intertemporal price discrimination) and changes in scarcity. That is, the temporal effect, and its interplay with price discrimination, is not considered, despite the relevance it has received in the literature (McAfee and te Velde, 2007). There are many methodological differences between William's approach and the one we develop in this article. His theoretical model disregards the crucial practical and institutional role of fare distributions, by assuming an airline that offers a single price to all customers (Williams, 2017, p.13). In addition, his analysis focuses entirely on the monopoly case, and does not extend to consider other market structures. Furthermore, unlike the present study where the information on capacity and fares is retrieved simultaneously from the airline's website, Williams (2017) merges data from an aggregator travel agent, Kayak, for the fares and an industry consultancy site for sales (expertflyer.com), thus raising two types of concerns: one, there may be a temporal mismatch between the fares posted on the travel agent and on the airlines' website; two, for each single flight and point in time, the number of available seats reported on a seat map may again be outdated or imprecisely measured.[3]

[3]The controls discussed in William's article refer to values aggregated on a monthly basis, a procedure that does not correct for possible measurement imprecisions arising at a single daily flight's level.

3 Data

The data collection employed a web crawler, as widely used in the literature.[4] 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; fares are all denoted in British Pounds as the first leg originated in a British airport (Gaggero and Piga, 2011).[5]

For each departure date, the data collection started about four months in advance; it was then repeated 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. In total, we surveyed 43,275 daily flights scheduled to depart during the period May 2014 - June 2015, covering 67 European bi-directional routes. 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.[6]

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 article is to show that this is not sufficient to test the implications of theoretical models

[4]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) compares online and offline prices in multi-channel retailers, whereas Cavallo et al. (2014) use online prices posted by international retailers to investigate deviations from the Law of One Price across countries with different currency regimes.

[5]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.

[6]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, consistent with the case in Cavallo (2017) where web scraping does not generate dynamic pricing.

of DP in airline markets. Based on the model presented in Section 5, 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 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.

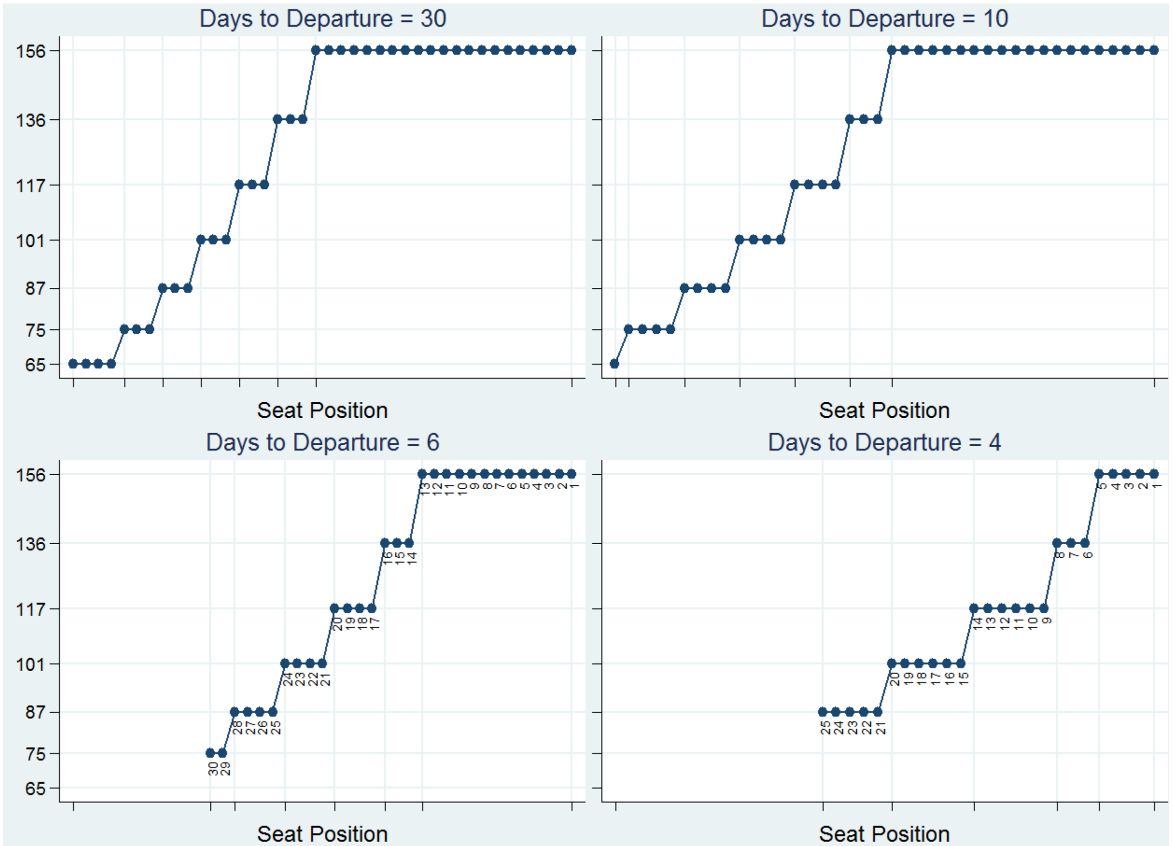
After applying the treatment described in the Appendix A.1 to the retrieved fares, we obtained the flights' distribution of posted fares over the available seats on a query date. An example of such distributions is shown in Figure 1, which is based on the data of a randomly selected flight.

4 Properties of fare distributions

Figure 1, which includes a flight's fare distributions retrieved 30, 10, 6 and 4 days to departure, is central for the whole analysis. Each graph, where a dot denotes a seat, shows that the fare sequence includes a series of 'buckets of seats' – that is, groups of seats that have the same price tag. The size of each bucket varies, and in all the panels, the seat on sale corresponds to the one positioned at the extreme left. Considering that such distributions are found for all the flights in our sample in all booking periods, the Figure indicates that a fundamental aspect of Revenue Management consists in the definition of a (possibly non-strictly) monotonically increasing sequence that assigns a fare, starting from the cheapest and ending with the dearest, to each seat on a plane; in Section 5, we show that such distributions are consistent with the results of a revenue optimization problem.

In the top panels of Figure 1, 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, 30 seats remain to be sold, reducing to 25 four days before departure (right lower panel). Interestingly, the same bucket fares, ranging from £156 down to £65, are repeatedly found over the booking temporal horizon, thus suggesting that they tend to be used throughout most of the booking period and that new buckets may be only occasionally created during the selling period.

Figure 1: Fare distribution at various days to departure



Legenda - Flight EZY5293 from London Gatwick (13.05) to Milan Malpensa (16.00) on 19 May 2014

A visual inspection is sufficient to establish some interesting features of the distributions and their evolution over time. Thirty days to departure, the carrier had allocated four seats for sale at the price of £65, three seats at the price of £75, and so on and so forth. Due to the data censoring, we cannot ascertain the precise size of the last observed "bucket" valued at £156. Twenty days later, the first bucket includes only one seat; the size of the buckets £75-£117 has increased to four seats, and the size of the £156's top observed bucket is still unmeasurable due to censoring. Six days prior to departure, there is no censoring and we can precisely gauge the number of available seats on the flight, as well as uniquely identify the position of each seat, and its corresponding fare.[7] Noticeably, the size of the top bucket (£156) includes thirteen seats, whereas the other buckets have maintained the size they had four days before. Two days later, five seats have been sold, and the size of the top bucket has further shrunk because some of its seats were moved down to increase the size of most

[7]To guarantee that the positions in the Figure are time-invariant, because seats are sold over time, they need to be counted from right to left; we follow this procedure throughout the article, in both the theoretical and empirical part.

intermediate buckets.

4.1 Connecting fare distributions to dynamic pricing

The focus on a fare distribution implies that dynamic pricing should not be defined exclusively in terms of the evolution of the fare of the seat on sale, but by considering how all the fares in the distribution evolve. We argue that dynamic pricing corresponds to a modification of the distribution structure as revealed by a reallocation of seats to a higher or lower bucket, and that some instances of price increases may not be dynamic in nature, in the sense that they may not reflect the airline's revision of the underlying distribution's structure.

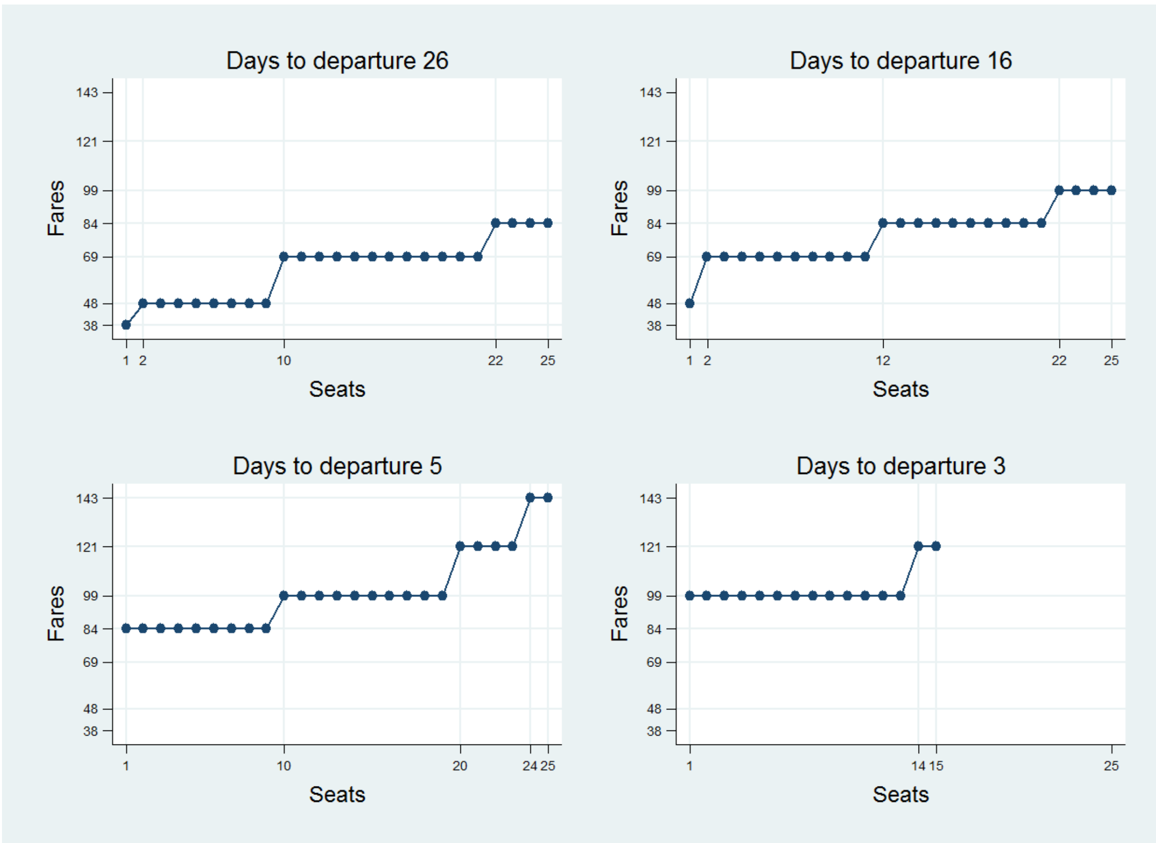
More importantly, we can explain how the "capacity" and "temporal" effects discussed in the Introduction (and further theoretically developed in Section 5) are responsible for the evolution of not only the price of the first seat on sale, but also of all the seats available on a plane. To illustrate both, we refer again to the bottom panels of Figure 1.

The "capacity" effect implies that the fare goes up as the plane fills up. Four days to departure, the number of available seats dropped to 25, that is, the seats denoted with numbers 30 to 26 in the right bottom panel were sold, each at the fare indicated by their bucket. Thus, the offered fare increased from £75 to £87 simply by following the predetermined sequence, and in this sense, we do not count this as an example of dynamic pricing because the bucket allocation of the sold seats did not change over time.

The "temporal" effect is related to the perishable nature of the airline service, which implies that the airline faces a strong incentive to reduce fares in the attempt to minimise the number of empty seats at take-off. It may engender hidden forms of dynamic pricing, leading to dynamic pricing taking place even if online customers cannot observe any change in the offered fare. This happens when the carrier shifts some seats from higher to lower-priced buckets, or viceversa, thus generating a modification of the fare sequence that is not associated with a change in the offered fare. Between six and four days to departure, the seats labelled 24 to 6 were moved to a lower bucket, relative to their previous bucket position. As a result, the £87 bucket ended up containing 5 seats, whereas it only had 4 two days before. The £101 and £117 buckets also increased in size, but the top bucket shrank to contain only 5 seats. Such variations are indeed forms of dynamic pricing. Interestingly, the replenishment of the £87 bucket implies that the offered fare will remain at this level longer, thereby slowing down the rate of increase of the offered fare due to the force illustrated in the first force. Thus, contrary to the common belief that airlines rely on dynamic pricing to charge higher fares, we highlight how dynamic pricing can achieve the opposite effect.

To strengthen the argument that easyJet's pricing approach, exemplified by the distribution in Figure 1, is largely adopted of the industry, in the next subsection we generalize the analysis by presenting similar fare distributions derived from data collected from the websites of another European and a U.S. Low-Cost carriers, and in the Appendix A.2 we discuss how the present analysis also offers interesting insights into Full Service Carriers (FSCs) pricing.

Figure 2: Fare distribution at various days to departure (Ryanair)



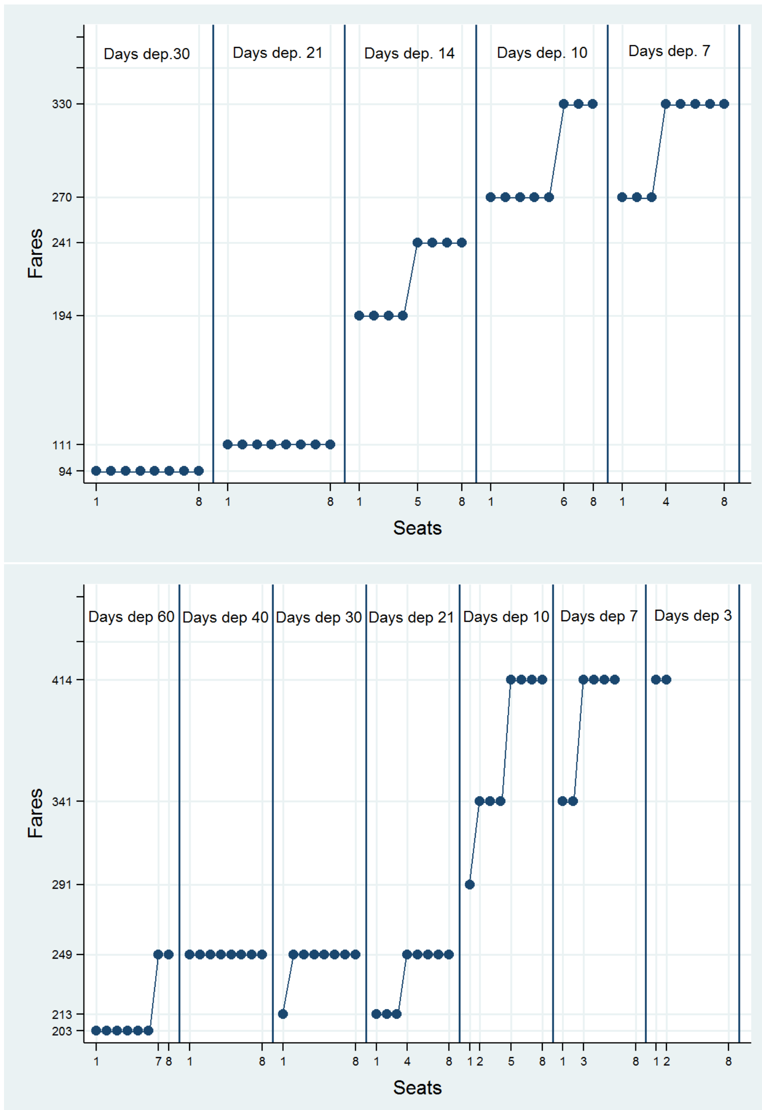
Legenda - Flight FR 8547 from Berlin Schonefeld (21:55) to London Stansted (22:40) on 21 Oct 2011

4.2 Examples of fare distributions from other airlines

Figure 2 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 Ryanair's 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. Whereas the price of the seats allocated in higher buckets has clearly fallen (temporal effect), the price of the seat on sale has increased from £84 to

£99, consistent with the "capacity" effect according to which fares increase as the buckets are sold out. That is, the main implications of this study could easily be extended to at least another large player in the industry.[8]

Figure 3: 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

Southwest allows queries with 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

[8]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 2 is derived, led us to the decision to focus on easyJet only.

as in the case of easyJet and Ryanair. Nonetheless, the analysis of Figure 3 reveals striking similarities between Southwest's pricing approach and that of its European counterparts.[9] On the one hand, when holding the query date fixed, for the majority of flights the eight seats carry the same fare, as it is shown for instance in the first two top left vertical panels. On the other, the data include several examples where the fare of the first eight seats exhibits 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, which features various instances of the temporal effect, thus suggests that Southwest also 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.

5 Theoretical set-up

In this Section we offer a stylized model of RM which translates some of the previously discussed key elements of RM practices into economic terms. First, carriers price in distribution, that is, in each period they assign a fare to all the seats in a flight (Flig et al., 2010). This is because in each period a carrier can sell more than one seat and possibly all the seats of the flight. Second, carriers charge a very limited number of fares during the entire selling period (Talluri and van Ryzin, 2004). Third, fare distributions remain fixed over discrete time intervals of one or more days, that is, they are not instantaneously updated. Escobari et al. (2016) report evidence suggesting full service airlines revise their prices overnight; in our data, distributions stay unchanged for two-three days on average.

5.1 A stylized model of revenue management

A carrier operates a single flight with $N > 1$ seats on a monopolistic route.[10] 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 $p(t, m)$. 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

[9] The data were collected using the same web crawling technique and is part of work in progress.

[10] In our empirical analysis we show that the main implications of the model hold regardless of the actual market structure.

available at the end of the last selling period remain unsold.[11]

In each period t, consumers $h = 1, 2, \dots, \infty$ arrive sequentially. The probability that the first consumer arrives in t is $\varphi_{1,t} \in (0, 1)$, and that consumer $h + 1$ arrives conditional on the fact that consumer h has already appeared is $\varphi_{h+1,t} \in (0, 1)$. Consumer (h, t) is myopic and her willingness to pay is a random variable $\theta_{h,t}$, with (right-continuous) cumulative distribution $F_{h,t}$ on the support Θ, with $\underline{\theta} = \inf \Theta > 0$ and $\bar{\theta} = \sup \Theta < \infty$.[12]

We assume that the arrival process is memoryless and consumers have the same ex-ante evaluation; i.e., for any h and t, $\varphi_{h,t} = \varphi_{h+1,t} = \varphi \in (0, 1)$; $F_{h,t} = F_{h+1,t} = F$. Thus, 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 $1, \dots, h - 1$ 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 $t \in \{0, \dots, T\}$ and $M \in \{0, \dots, N\}$. 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 $V(t - 1, M)$.

Because problem (2) is solved recursively by starting from the last period, the optimal fare of seat $m \leq M$ in period t when there are M seats available, $\tilde{p}(t, m, M)$, is independent of the total number of available seats M in period t, i.e. $\tilde{p}(t, m, M) = \tilde{p}(t, m, M + 1)$, for any $M = 1, \dots, N - 1$ and $t = 1, \dots, T$. 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 would imply that the optimal fare is also affected by the

[11] 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 article, where the position of seats is counted by starting from the last one.

[12] 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.

total number of available seats at the beginning of the period and, in general, $\tilde{p}(t, m, M)$ is not necessarily equal to $\tilde{p}(t, m, M+1)$. 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 because it plays no role with current assumptions.

Definition 1 *$V(t, M)$ has decreasing returns in t and M, respectively, if and only if, for any t and M:*

$$\begin{aligned} V(t, M) - V(t-1, M) &\leq V(t-1, M) - V(t-2, M) \\ V(t, M) - V(t, M-1) &\leq V(t, M-1) - V(t, M-2). \end{aligned}$$

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

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

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 returns 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 and M, which is a standard results in the pricing literature (Gallego and van Ryzin, 1994; McAfee and te Velde, 2007). Second, periods and seats are two resources, 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. Note that complementarity is equivalent to stating that the option value of any seat M in any period t, declines as the departure approaches:

$$V(t, M) - V(t, M-1) \geq V(t-1, M) - V(t-1, M-1). \tag{3}$$

From these properties we derive Corollary 1, which is essential for the characterization of the optimal fare $p^*(t, m)$.

Corollary 1 *Let $X(t, M) = V(t, M-1) - V(t-1, M)$ be the marginal value of substituting a seat with a period, then, for any t and M:*

$$X(t, M) \leq X(t-1, M) \tag{4}$$

$$X(t, M) \geq X(t, M-1) \tag{5}$$

Corollary 1 states that the value of giving up a resource in exchange for the other tends to increase as the latter becomes scarcer. For instance, when the total number of periods available to sell the seats reduces, the value of giving up a seat increases because the value of that additional seat reduces. This property comes directly from complementarity between t and M, or, equivalently, from the fact that the option value of a seat is declining across periods. Assuming, now, that the optimal fare $p^*(t,m)$ which solves (2) is unique, then:

Proposition 2 *For any t and M, the optimal fare $p^*(t,m)$ has the following properties:*

A. (capacity effect): $p^*(t,m) \leq p^*(t,m-1)$,

B. (temporal effect): $p^*(t,m) \geq p^*(t-1,m)$.

To provide an intuition of this result, with F continuous and q differentiable, the internal solution $p^*(t,m)$ from the f.o.c. in (2) satisfies:

$$q(p^*(t,m)) = -q'(p^*(t,m))(p^*(t,m) + X(t,m)). \tag{6}$$

The left-hand side and the right-hand of the equation capture, respectively, the expected gains and the expected losses from a marginal increase of p. Because $q(p)$ is decreasing in p, and, from Corollary 1, $X(t,m)$ is increasing in m, we have that the larger a seat's position, the lower the optimal fare in any period (Property A). Thus, within a given period, seats are sold by setting a sequence of fares (i.e. a fare distribution) which is increasing. This result extends the cost-based justification of an increasing equilibrium fare distribution considered in Dana (1999).

Similarly, because $X(t,m)$ is decreasing in t, it follows that the larger the number of periods, the higher the optimal fare for any seat m (Property B). Thus, 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.

5.2 Simulating and extending the model

To gauge how robust the theoretical results in Proposition 2 are, we simulate the model by first retaining and then relaxing the assumption of a time-invariant $F(\theta)$.[13] By doing so, we can gain insights into how the capacity and the temporal effects are affected by the standard assumption that business-people tend to purchase only a few days before departure, or, equivalently, how dynamic pricing is applied in the presence of an inter-temporal price discriminatory motive.

Table 1: Simulated observed number of seats for each bucket fare across booking periods, under a time-invariant or time-varying distribution of WTP over time.

$F(\theta)$		Periods to departure - time-invariant $F(\theta)$											Periods to departure - varying $F(\theta)$										
Prob.	fares	11	10	9	8	7	6	5	4	3	2	1	11	10	9	8	7	6	5	4	3*	2*	1*
14/64	50																						
12/64	65	6	6	5									7	6	6								
8/64	80	6	5	5	4	4	4	3	3	2	2		5	5	5	4	3	3	2	2			
6/64	95	6	6	5	5	5	4	3	2	2	1	1	6	6	5	5	4	3	3	2	$\boxed{5}$		
6/64	110	5	5	4	4	3	3	3	2	2	1	1	5	4	4	3	$\boxed{4}$	3	2	2	1	1	
5/64	130	6	5	5	4	4	3	3	3	2	2	1	5	5	4	4	3	3	3	2	2	2	2
5/64	150	4	4	3	3	3	2	2	1	1	1	0	4	4	4	3	3	3	2	2	2	2	2
4/64	175	4	3	4	3	2	2	2	2	1	0		4	3	3	3	2	2	2	1	2	1	0
4/64	200	2	2	1	1	1	1	0					3	3	2	2	2	1	1	1	0		
Avail.	seats	39	36	32	24	22	19	16	13	10	7	3	39	36	33	24	21	18	15	12	12	6	4

* The time-varying $F(\theta)$ is obtained by shifting, in each of the last three periods, $1/64^{th}$ of probability from each of the three smallest buckets (i.e., 50, 65 and 85) to the three largest ones (i.e., 150, 175 and 200). The circles (squares) denote buckets with seats moved to a lower (upper) bucket.

We set the number of periods $T = 11$; available seats $N = 39$; and the average total number of prospective travellers $L = 1.2N$.[14] We suppose $F(\theta)$, the distribution of consumers' WTP, is discrete with its probability density function reported in the left columns of Table 1. A time-varying $F(\theta)$ is obtained by shifting, in each of the last three periods, $1/64^{th}$ of probability from each of the three smallest buckets (i.e., 50, 65 and 85) to the three largest ones (i.e., 150, 175 and 200). That is, we simulate the impact of a larger proportion of buyers with a higher willingness to pay (business-people) on the fares distributions and on dynamic pricing. Each cell reports the resulting bucket size (i.e., number of seats) in each period, with bucket fares corresponding to a subset of the support of the distribution; note that no seat

[13] In the empirical analysis, we test whether the theoretical framework can be generalized to market structures other than monopoly.

[14] From T and L we obtain $\varphi = L/(L+T) \simeq 0.81$. The numerical simulation is based on the algorithm described in Section B.1 of the online Appendix to solve problem (2) recursively.

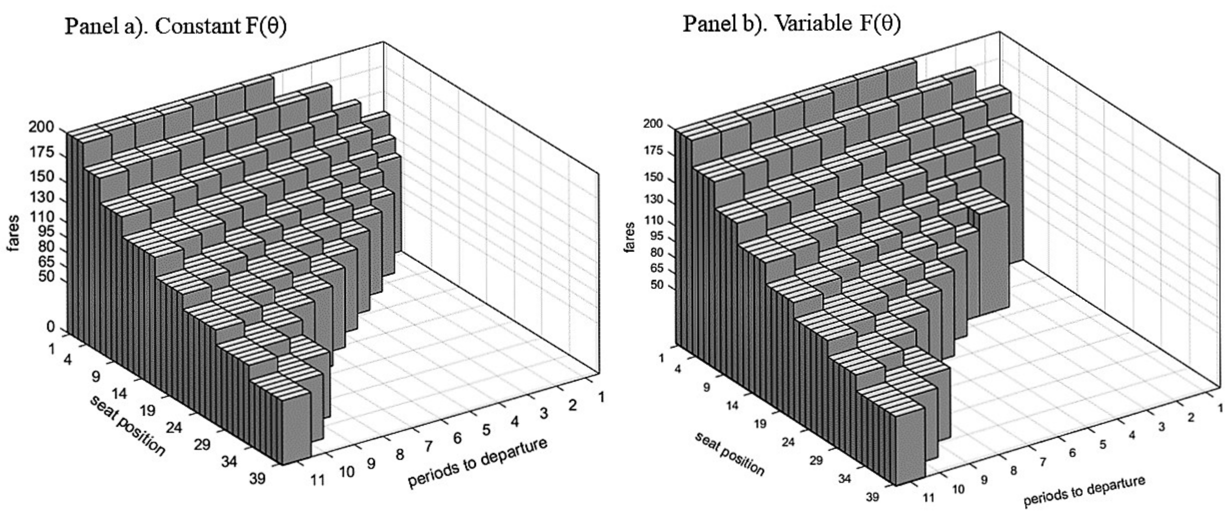
is offered at the value of 50, despite the large mass of consumers in the distribution. Bucket sizes denote the modal value of the simulated number of seats at the beginning of each period based on 50.000 runs. We do so because each simulated flight exhibits a possibly different selling path across periods, so the Table refers to a *representative* simulated distribution by focusing on the more likely evolution of the selling path, as reported in the last row that indicates the seats that remain to sell.

A visual presentation of Property A is given in both panels of Figure 4, which graphically depicts the content of Table 1: in each period and under time-invariant $F(\theta)$, the fare distribution has a stepwise increasing profile over seats, equivalent to the observed ones in Figure 1. For instance, in the initial selling period (period 11), the first six seats are put on sale at 65, the next six at 80 and so on and so forth up until the last two seats, which are in the 200 bucket. Similarly shaped distributions arise under a time-varying $F(\theta)$, leading to the conclusion that Property A is largely independent of the assumption on $F(\theta)$.

As far as Property B is concerned, the temporal effect is present both under time-invariant and time-varying $F(\theta)$, although at different degrees. As previously illustrated, such an effect is revealed by seats being moved from upper to lower buckets. Take the two most expensive buckets, priced at 200 and 175. Under time-invariant(varying) $F(\theta)$, the former disappears in period 5 (period 3), and the latter in period 3 (period 1). That is, having a larger mass of consumers with a high WTP late in the booking period mitigates and slows down the temporal effect, but it does not invalidate the main implication of Property B that the temporal effect operates to reshape the fare distribution by moving seats from upper to lower buckets. This is shown by a comparison of the two parts of Table 1. On the one hand, relative to the time-varying case, there appear to be more circles in the time-invariant $F(\theta)$ simulation that identify an instance of buckets with seats moved down. On the other, under time-varying $F(\theta)$, the impact of the temporal effect continues to be noticeable, even if the simulation also reveals cases, denoted by a square, of seats being moved to an upper bucket: for instance, the two seats allocated in period 4 to bucket 80 are moved up by one bucket in the next period (see also Panel b. of Figure 4). Overall, upward movements of seats appear to be in general less likely than downward movements.

Concerning the behaviour of the first seat on sale across periods, the temporal evolution is the same in both simulations until period 4. Subsequently, in the time-varying $F(\theta)$ case, the selling fare increases in every period, reaching the value of 135 in the last period, which is much higher than the value of 95 resulting under time-invariant $F(\theta)$. The simulation thus suggests that under a time-varying $F(\theta)$, the ensuing incentive to pursue an intertemporal price discriminatory motive is consistent with a faster growth of the selling fare in the periods immediately preceding departure. As far as dynamic pricing is concerned, such a faster growth can be implemented via movements of seats to higher buckets and by reducing the downward movements.

Figure 4: Fare distribution by periods to departure

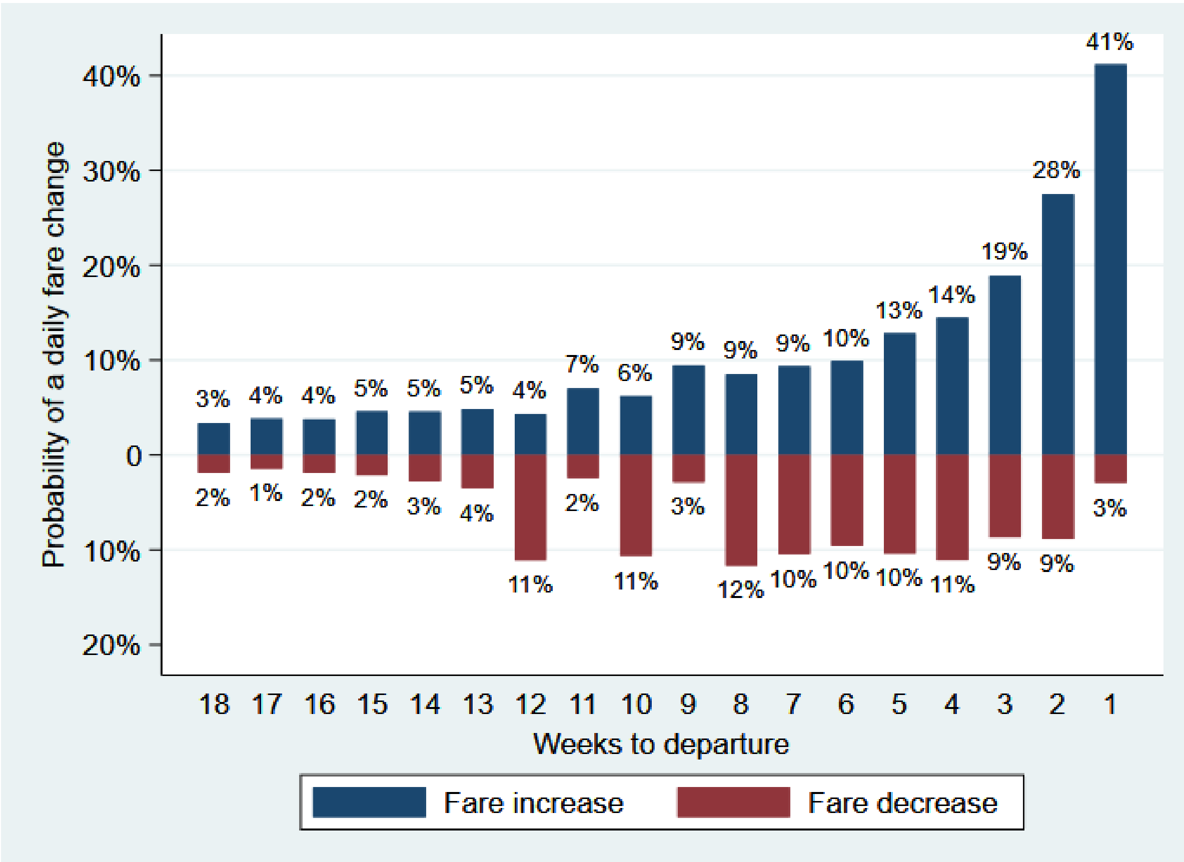


6 Descriptive analysis of DP

Figure 5 reports the probability of observing the fare of the seat on sale increase or decrease between two consecutive booking days during the booking period. It thus represents a measure of the fare variability over time that would result from tracking the posted fare from a query for one seat. The Figure highlights how the posted fare remains invariant on average in at least 80% of flights until six weeks to departure, and that such a proportion continues to be larger than 60% until the last week.

The theoretical analysis has highlighted the sources of such variability. On the one hand, the price of the seat on sale may vary due to upward movements along the fare distribution when seats are sold (capacity effect) On the other, all seats' fares could change because of modification in the fare distribution induced by the temporal effect and the countervailing impact of a time-varying composition of the willingness to pay distribution F_t. This is important because, as previously discussed in the analysis of Figure 1, the fare of the seat on sale may be made more stable over time if its bucket is replenished by moving seats from upper buckets. Disentangling such effects' individual impact on the overall evolution of all the seats' fares, but in particular on the fare of the seat on sale is the central focus of this and the next section. To this purpose, we classify as DP any case where a seat in any position is either moved up or down to an existing (i.e., previously observed as part of a flight's fare distribution) or a new bucket.

Figure 5: Probability of fare changes induced by any effect. Full sample of censored and uncensored observations.



To tease out the net capacity effect, Table 2 reports the probability of observing the fare of the seat on sale increase or stay the same, as a function of the number of seats sold in two consecutive booking days, after excluding fare changes due to any form of DP. Because the capacity effect can operate only by pushing the fare upward, the probability of downward fare movements is always zero. On average, every time one seat is sold, the fare of the seat on sale increases in about 20% of observations; such a probability rises to about 37% in the case of two seats, and reaches almost 66% for three or more seats. A similar progression, which is observed throughout the various parts of booking period, suggests that, on the one hand, the capacity effect is an important determinant of the fare increases in 5; on the other, because of the stepwise shape of the fare distribution, the sale of one or two seats is unlikely to lead to a fare change.

As Figures 1-2 and the theoretical analysis suggest, DP clearly goes beyond the mere

Table 2: Probability of the price of the seat on sale to increase, stay the same or decrease, by booking period and number of seats sold between two consecutive days. Pure capacity effect.

Days to departure	Fare variation	Number of seats sold 0	1	2	3+	Overall
0-7	up	0.0%	19.5%	36.8%	65.6%	31.7%
	same	100.0%	80.5%	63.2%	34.4%	68.3%
8-14	up	0.0%	21.7%	37.6%	66.6%	26.1%
	same	100.0%	78.3%	62.4%	33.4%	73.9%
15-35	up	0.0%	28.0%	40.9%	63.9%	12.8%
	same	100.0%	72.0%	59.1%	36.1%	87.2%
36+	up	0.0%	25.0%	27.8%	60.0%	4.3%
	same	100.0%	75.0%	72.2%	40.0%	95.7%
Overall	up	0.0%	20.5%	37.2%	65.8%	28.8%
	same	100.0%	79.5%	62.8%	34.2%	71.2%
	down	0.0%	0.0%	0.0%	0.0%	0.0%

fluctuation of the price of the first seat in the distribution and has to include any modification of the fare distribution. Because our definition of DP excludes any price increase due to the capacity effect, 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, as discussed in the comment to the bottom panels in Figure 1.

Table 3 reports the probability that each seat in the distribution is treated with one of the forms of DP, calculated by considering only variations between query dates separated by one day.[15] The first four columns investigate whether a seat has moved up or down and 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 a seat's own bucket has increased or decreased.

Table 3 provides several insights into how DP reshapes the fare distributions. First, 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. This provides further support to the previous claim that the structure of all the distributions is largely fixed and is not subject to drastic redesigns. Third, and relatedly, the

[15]The qualitative results do not change if the probabilities were obtained considering variations between any two consecutive, but not adjacent, query dates.

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

Position in distribution	Seat moves to Higher bkt	Higher new bkt	Lower bkt	Lower new bkt	Bkt size increase	Bkt size decrease	Obs.
1	1.7%	1.2%	4.4%	0.4%	7.4%	26.7%	124,441
2	1.4%	1.1%	4.3%	0.4%	7.6%	27.3%	121,578
3	1.5%	1.0%	5.1%	0.5%	6.7%	27.4%	119,069
4	3.3%	1.0%	7.4%	0.5%	7.8%	28.1%	116,774
5	3.3%	0.9%	8.4%	0.5%	8.5%	27.5%	114,439
6	2.8%	1.0%	9.2%	0.6%	9.1%	26.8%	111,907
7	2.6%	0.9%	9.7%	0.6%	9.9%	26.6%	109,387
8	2.5%	0.8%	9.7%	0.6%	10.3%	26.1%	106,752
9	2.3%	0.8%	10.0%	0.7%	10.9%	25.9%	103,852
10	2.3%	0.9%	11.0%	0.8%	11.9%	25.2%	100,811
11	2.5%	0.8%	11.2%	0.8%	12.6%	24.4%	97,683
12	3.4%	1.0%	13.3%	1.0%	13.1%	23.6%	94,168
13	4.1%	0.3%	14.1%	0.4%	13.5%	22.7%	90,936
14	4.0%	0.4%	14.1%	0.5%	13.8%	21.8%	87,745
15	3.9%	0.4%	14.2%	0.5%	14.3%	20.4%	84,330
16	3.8%	0.4%	14.4%	0.5%	14.6%	19.6%	80,584
17	3.8%	0.5%	14.9%	0.6%	14.8%	18.6%	76,840
18	4.7%	0.7%	16.4%	0.8%	15.1%	16.8%	72,814
19	4.4%	0.7%	16.1%	0.8%	15.3%	15.7%	69,061
20	4.2%	0.8%	16.0%	0.9%	15.7%	14.6%	65,436
21	4.3%	0.9%	16.7%	1.1%	16.3%	13.4%	61,537
22	4.1%	0.8%	16.3%	1.1%	16.7%	12.6%	57,721
23	4.0%	0.9%	16.3%	1.2%	16.9%	11.8%	54,058
24	4.2%	1.1%	17.4%	1.6%	17.8%	10.9%	50,125
25	4.8%	0.6%	17.4%	0.9%	17.9%	10.0%	46,471
26	5.0%	0.8%	17.6%	1.1%	18.5%	9.1%	42,864
27	4.6%	0.9%	17.5%	1.2%	18.9%	8.2%	39,306
28	4.5%	1.0%	17.6%	1.6%	19.7%	7.1%	35,715
29	4.4%	1.2%	16.5%	1.9%	20.8%	6.3%	32,281
30	4.1%	1.3%	16.4%	2.3%	21.6%	5.2%	28,905
31	3.8%	1.6%	16.2%	2.8%	22.0%	4.4%	25,396
32	3.8%	1.9%	14.9%	3.1%	22.2%	3.8%	22,014
33	3.4%	2.1%	13.6%	3.4%	23.1%	3.3%	18,612
34	3.2%	2.4%	12.6%	4.2%	24.3%	3.0%	15,275
35	2.8%	2.5%	11.4%	4.8%	25.8%	2.5%	11,954
36	2.2%	2.2%	11.3%	5.2%	27.5%	2.4%	8,782
37	2.0%	1.8%	10.0%	7.4%	30.6%	2.0%	5,836
38	1.9%	1.3%	8.7%	12.1%	34.2%	1.1%	3,305
39	1.4%	0.5%	3.8%	14.0%	45.1%	0.0%	1,316

size of buckets in the right tail of the distribution (i.e., those with low positions) tends to shrink, whereas 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. That is, seats in top buckets are highly likely moved to lower buckets, thus increasing their size. Overall, Table 3 provides strong descriptive support to the role of the temporal effect in driving down the option value of all the seats in the fare distribution, and that such an effect is only partially offset by upward movement of seats.

6.1 DP over time

The theoretical analysis focussed on the role of the capacity and temporal effect, showing how the latter may be offset or slowed down by the incentive to engage in discriminatory pricing. As the simulations suggested, in our empirical setting we should then expect that the probability to observe a seat moving to a lower (higher) bucket reduces (increases) as the date of departure nears. Table 4 reports the probability of whether, during the various booking periods, at least one of the seats in the distribution has moved to a higher or lower, possibly new, bucket. The "Any fare move" column reports the probability that the distribution has changed due to a fare movement in either directions.[16] 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 change in the 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. Finally, in the three weeks before departure, the probability of observing seats being transferred to higher buckets (both existing and new) increases up to about 21%, with a combined fall in the probability of observing downward movements during the last ten days; both aspects are consistent with the view that a discriminatory motive can be revealed by both a weakening of the temporal effect as well as by upward movements, consistent with the simulations in Table 1.

It could be argued that DP, as well as price discrimination, mostly likely concerns the first

[16] This does not coincide with the row total because, in each period, different seats may be moved in different directions

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

	Seat moves to				
Days to departure	Higher bkt	Higher new bkt	Lower bkt	Lower new bkt	Any fare move
0-3	17.6%	3.9%	14.3%	1.3%	33.1%
4-7	10.9%	4.0%	20.6%	2.2%	32.8%
8-10	13.6%	4.2%	23.5%	3.0%	37.4%
11-14	13.0%	4.7%	74.4%	10.6%	82.3%
15-21	16.1%	5.1%	73.2%	8.9%	84.3%
22-28	12.1%	5.8%	79.4%	10.7%	88.1%
29-35	10.5%	3.7%	62.6%	10.9%	73.3%
36-50	5.8%	4.0%	40.6%	4.8%	46.5%
51+	2.7%	5.4%	14.9%	2.1%	20.7%
Overall	14.1%	4.2%	35.3%	4.5%	48.7%

seat on sale, the one with the lowest position in the distribution. Table 5, which reports the same information as Table 4 only for the first seat on sale, highlights two main aspects. One, it shows that the shifts downward in the distribution in Table 4 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% in Table 4, but it is less than 20% for the fare of the seat on sale. Once a seat becomes visible (i.e., its price is immediately revealed by an online query), it is less subject to a downward movement.[17] Two, a large proportion of movements to higher buckets shown in Table 4 involve the first seat, especially during the last ten days prior to departure. Although this is consistent with the hypothesis that a discriminatory motive moderates the overall impact of the temporal effect, it is noteworthy that in Table 5 the first seat is moved upward in at most 17% of observations taken three or less days before departure, and that the same percentage is lower in the two preceding weeks. If, on the one hand, the seat on sale is moved quite rarely to a higher seat, the third and the fourth columns of both Tables 4 and 5 clearly indicate that the probability to observe all seats being moved down falls sharply during the ten days before departure. It would seem therefore that price discrimination does not appear to lead to price increases of the first seat on sale but, mostly, manifests itself through a reduction in the strength with which the temporal effects pushes all fares down. Most importantly, the analysis of the first seat on sale highlights the importance to clean DP of the upward fare movements' induced by the capacity effect, because failing to do so would

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

confound late price increases as manifestation of price discrimination.

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

	Seat moves to				
Days to departure	Higher bkt	Higher new bkt	Lower bkt	Lower new bkt	Any fare move
0-3	13.8%	2.9%	5.4%	0.6%	22.7%
4-7	8.7%	3.0%	7.4%	1.1%	20.2%
8-10	10.1%	2.9%	7.5%	1.4%	21.8%
11-14	7.3%	3.3%	19.8%	6.3%	36.7%
15-21	7.2%	4.0%	11.8%	4.9%	27.9%
22-28	5.6%	4.6%	11.0%	6.4%	27.5%
29-35	4.2%	3.1%	7.3%	8.3%	23.0%
36-50	3.0%	2.7%	7.0%	2.6%	15.3%
51+	2.2%	1.6%	4.8%	1.6%	10.1%
Overall	9.8%	3.1%	9.3%	2.5%	24.7%

7 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 effect 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, whereas the temporal effect operates in a "hidden" way. Both regressions lend strong support to property B, after the role of the capacity effect is taken into account.

7.1 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:

$$\ln Fare_{jdt}^{m} = \sum_{t} \beta_t D_t + \gamma m + \sum_{t} \omega_t D_t * m + \zeta_{jd} + \varepsilon_{jdt}, \tag{7}$$

where D_t defines a set of dummy variables *Days to departure*, with t defining the intervals between the query and the departure date; m denotes the *Position* variable. The ωs denote the coefficients of their interaction. In a model without interaction terms, as far as property B is concerned, the dummies D_t 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, whereas Property A would be supported by a negative and significant coefficient of γ (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. First, *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. Second, 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:

$$FNC_{jdt} = 1[z_1\theta_1 + \nu_1 > 0] \tag{8}$$

$$s^m_{jdt} = 1[z_1\theta_2 + \theta_3 m + \nu_2 > 0] \text{ if } FNC_{jdt}=1. \tag{9}$$

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] In (8)-(9), ν_1 and ν_2 denote the respective error terms; under the assumptions $(\nu_1, \nu_2) \sim N(0, 1)$ and $\text{corr}(\nu_1, \nu_2) = \rho$, (8)-(9) 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 Inverse Mills ratios (IMR) for the selected observations are: $\hat{\lambda}^m_{jdt}(\hat{\theta}_2, \hat{\theta}_3) = \phi(z_1\hat{\theta}_2 + \hat{\theta}_3 m) \Big/ \Phi(z_1\hat{\theta}_2 + \hat{\theta}_3 m)$. We can then estimate an augmented version of (7):

$$\ln Fare^m_{jdt} = \sum_t \beta_t D_t + \gamma m + \sum_t \omega_t D_t * m + \hat{\lambda}^m_{jdt}(\hat{\theta}_2, \hat{\theta}_3) + \zeta_{jd} + \xi_{jdt}, \tag{10}$$

by panel OLS fixed-effects.[19]

[18] Imagine that at t we only retrieve fares for, say, the last 20 seats; these would have $s_{jdt} = 1$. To estimate (8)-(9), we would append observations for seats 21-39 and set $s_{jdt} = 0$.

[19] The approach draws from procedure 17.1 in Wooldridge (2002).

The panel identifier corresponds to the combination of flight-code plus day of departure; the panel's temporal effect is represented by a sequential counter that uniquely identifies all the possible combinations of *Position* for all query dates t.[20] We set the earliest day to departure dummy (Days to departure 51+) as reference group and we 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.

Table 6 reports the results. Models (1) and (2) use the full sample, whereas the others focus on flights in different market structures. The distinction by market structure provides an assessment of whether the results of the theoretical model, derived assuming a monopoly firm, can be generalised to other types of markets. In our sample, we define as competitive those routes with an HHI no higher than 0.5, as monopoly if the index is at least equal to 0.90 and as oligopolistic all the routes with intermediate values.[21]

All models in Table 6, except the first, include the interaction between the temporal dummies and the Position variable. In the first column, the coefficient of *Position* is, as expected, negative. That is, the econometric evidence indicates that the fare distributions are structured as predicted in property A of Proposition 2. Second, and relatedly, the *Position* coefficient provides an estimate of the fare distribution's gradient: such a value is about 1.6%, that is, a unit decrease in *Position* is associated with an equivalent expected percentage increase in fare. Third, and more importantly, the *Days to Departure* (DtD) 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 fare of all the seats in the fare distribution, holding the position fixed. This finding is consistent with the view that the carrier moves the seats down to lower buckets as the departure date nears and that such a move reflects a reduction in the expected value of the seats (Property B).

In the second column, the temporal dummies' coefficients also decrease in magnitude as the departure date nears; importantly, because the interaction coefficients are all negative, the temporal decline is stronger as the position value increases: the farther a seat is positioned from the top one, the larger the fall in the bucket order (and in fare) it experiences.

Models (3) to (5) indicate that Proposition 2 is robust to variations based on market structure. Indeed, each sub-sample leads to estimates that are qualitatively similar to those

[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] The HHI was derived using the punctuality statistics published by the UK Civil Aviation Authority on their website. Using the HHI based on the citypair classification does not change the qualitative nature of the results.

Table 6: OLS Regression analysis of the price of all seats in the distribution - Flight-code fixed effects. NB: Dtd=Days to Departure. Standard Errors in parentheses

	(1) Full sample	(2) Full sample	(3) Compet. routes	(4) Oligop. routes	(5) Monop. routes
DtD 0-3	-0.273***	-0.015	0.021	-0.130***	0.033
	(0.043)	(0.061)	(0.122)	(0.038)	(0.098)
DtD 4-7	-0.267***	0.021	0.059	-0.094*	0.069
	(0.043)	(0.061)	(0.122)	(0.038)	(0.098)
DtD 8-10	-0.256***	0.053	0.085	-0.058	0.101
	(0.043)	(0.061)	(0.122)	(0.038)	(0.098)
DtD 11-14	-0.202***	0.085	0.110	-0.029	0.137
	(0.043)	(0.061)	(0.122)	(0.038)	(0.098)
DtD 15-21	-0.099*	0.185**	0.222	0.068	0.234*
	(0.043)	(0.061)	(0.122)	(0.037)	(0.098)
DtD 22-28	-0.047	0.201***	0.266*	0.072*	0.241*
	(0.042)	(0.061)	(0.122)	(0.037)	(0.097)
DtD 29-35	0.017	0.207***	0.276*	0.062	0.255**
	(0.042)	(0.061)	(0.123)	(0.036)	(0.097)
DtD 36-50	0.031	0.142*	0.213	0.026	0.177
	(0.039)	(0.057)	(0.121)	(0.032)	(0.091)
Position	-0.016***	-0.002	0.000	-0.005***	-0.001
	(0.000)	(0.002)	(0.003)	(0.001)	(0.003)
Pos*DtD 0-3		-0.020***	-0.021***	-0.016***	-0.021***
		(0.002)	(0.003)	(0.001)	(0.003)
Pos*DtD 4-7		-0.021***	-0.022***	-0.017***	-0.022***
		(0.002)	(0.003)	(0.001)	(0.003)
Pos*DtD 8-10		-0.020***	-0.021***	-0.017***	-0.021***
		(0.002)	(0.003)	(0.001)	(0.003)
Pos*DtD 11-14		-0.017***	-0.018***	-0.014***	-0.019***
		(0.002)	(0.003)	(0.001)	(0.003)
Pos*DtD 15-21		-0.017***	-0.018***	-0.013***	-0.017***
		(0.002)	(0.003)	(0.001)	(0.003)
Pos*DtD 22-28		-0.014***	-0.017***	-0.010***	-0.014***
		(0.002)	(0.003)	(0.001)	(0.003)
Pos*DtD 29-35		-0.011***	-0.013***	-0.006***	-0.011***
		(0.002)	(0.003)	(0.001)	(0.003)
Pos*DtD 36-50		-0.006***	-0.009**	-0.003***	-0.007**
		(0.002)	(0.003)	(0.001)	(0.002)
IMR λ	-0.100***	-0.015***	-0.028*	-0.018*	-0.007
	(0.003)	(0.004)	(0.012)	(0.007)	(0.005)
Constant	5.197***	4.930***	4.917***	5.029***	4.881***
	(0.043)	(0.061)	(0.122)	(0.037)	(0.098)
R^2	0.698	0.710	0.709	0.710	0.712
Observations	5,443,535	5,443,535	1,118,566	1,458,918	2,866,051

produced using the full sample, indicating that the insights generated by our theoretical model in relations to the main features of fare distributions can be generalized and hold in many market settings.

Figure 6: Predicted effects of Position and days to departure on fares.
Note: based on Model (2)-(5) in Table 6.

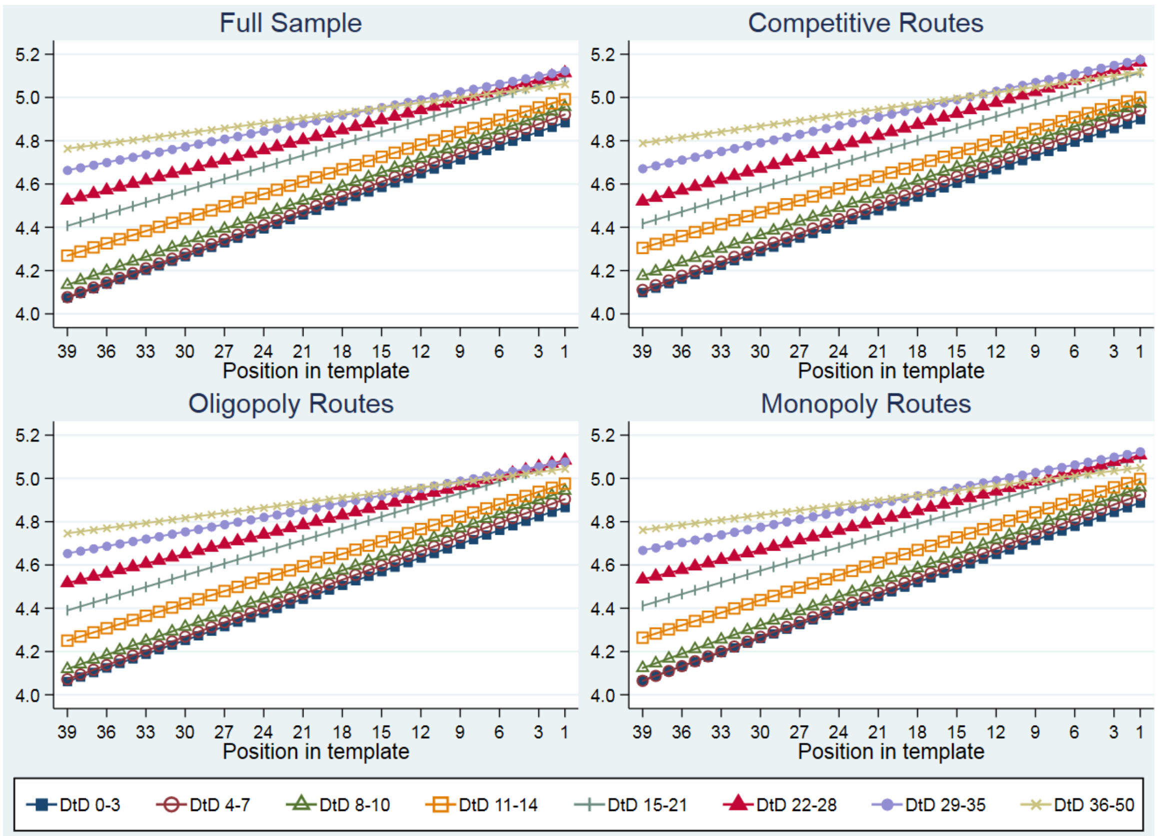


Figure 6 shows the predicted values from models (2)-(5) of Table 6. Each line, which represents the predicted relationship between fare and position keeping the temporal dummies fixed, defines a stylized, smooth version of the fare distributions in Figure 1. The slope varies to reflect the interaction terms. Remarkably, Figure 6 shows that each model generates predicted values that are very similar in each sub-sample, suggesting that the RM approach that easyJet adopts is very consistent in all its routes.

Furthermore, Figure 6 can shed some light on how the discriminatory motive may intervene to counteract the temporal effect. Indeed, as the analysis of Tables 4 and 5 illustrated, price discrimination is implemented by either moving seats to upper buckets, or by slowing down their rate of descent to lower buckets (which could imply keeping them on the same bucket). Intuitively, within a fortnight from departure, the temporal effect would be particularly strong for seats on the left side of the distribution, if still available to be sold. By the same token, as the previous descriptive evidence showed, seats on the right side of the distribution should be less likely to fall or more likely to be moved up. This mechanism results in distributions becoming steeper as the departure date nears, something that is clearly depicted in Figure 6, where distributions collected within a fortnight, but especially during the last seven days, are indeed steeper than earlier ones. Because this is the case in all market structures, a further implications is that, at least in our sample, price discrimination does not appear to be strongly related to market concentration, as hypothesized in Borenstein and Rose (1994); Gerardi and Shapiro (2009); Gaggero and Piga (2011).

7.2 The temporal profile of the first seat on sale

Studying the fare of the first 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 articles reporting graphical and/or econometric evidence of fares increasing as the departure date nears.[22] 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).

Using the insights offered by the foregoing theoretical and empirical analysis, in this section we investigate the extent by which the behavior of the fare of the seat on sale in our dataset can be reconciled with and related to the evidence reported in the existing literature. To this purpose, the econometric strategy hinges on testing properties A and B of Proposition 2 on the first seat on sale, using the specification in equation (10) modified to take into account that for such a seat the censoring process can be modeled using equation (8) only.

Models (1) and (2) in Table 7 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. 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, the increasing temporal profile suggests that the capacity effect appears to be a stronger driving force than the temporal effect, especially when the latter is further weakened by the price discriminatory motive. Interestingly, as far as the presence of strategic consumers is concerned, the first two models in Table 7 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

[22]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.

Table 7: Regression analysis of the price of the first seat on sale - Flight-code fixed effects. NB: Dtd=Days to Departure. Standard Errors in parentheses

	(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.
DtD 0-3	0.783***	0.327***	-0.339***	-0.376***
	(0.007)	(0.041)	(0.078)	(0.105)
DtD 4-7	0.665***	0.210***	-0.349***	-0.382***
	(0.007)	(0.041)	(0.077)	(0.105)
DtD 8-10	0.489***	0.056	-0.347***	-0.372***
	(0.006)	(0.041)	(0.076)	(0.104)
DtD 11-14	0.402***	-0.000	-0.285***	-0.303**
	(0.005)	(0.041)	(0.075)	(0.103)
DtD 15-21	0.350***	-0.022	-0.190*	-0.202*
	(0.005)	(0.041)	(0.074)	(0.103)
DtD 22-28	0.304***	-0.030	-0.101	-0.109
	(0.005)	(0.041)	(0.073)	(0.103)
DtD 29-35	0.276***	0.006	-0.001	-0.010
	(0.004)	(0.041)	(0.072)	(0.102)
DtD 36-50	0.169***	0.002	0.045	0.037
	(0.003)	(0.039)	(0.070)	(0.098)
Position=Available Seats			-0.019***	-0.020***
			(0.000)	(0.001)
IMR λ			-0.020	-0.024
			(0.011)	(0.014)
Kleibergen-Paap rk LM stat				825.897
Hansen J-stat				.842
R^2	.569	.414	.623	.629
Observations	887,671	249,125	249,125	172,411

to postpone purchase.

However, the first two models are misspecified. To tease out the possible separate impact of inter-temporal 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 first seat on sale, the number of available seats corresponds to the position of the first seat in the distribution (see, e.g., the bottom panels in Figure 1). Such a property has important implications because it sheds more light on the impact of the capacity effect. Indeed, unlike the estimates in Table 6 where 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.

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 Property B in this article. Relative to those posted fifty-one or more days from departure, fares posted twenty-eight days or later are ceteris paribus 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" DtD dummy is slightly larger than the previous one (-0.339 vs. -0.349), pointing to a U-shaped temporal profile, whose increasing part can be ascribed to the implementation of an inter-temporal price discrimination strategy (Alderighi et al., 2015; Bilotkach et al., 2010; Escobari, 2012).

The capacity effect, 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 6 and indicates that 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. Our results thus provide a so far undetected perspective, that is, they directly relate the evolution of the selling fare to the overall design of the fare distribution.

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. (10), 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. In other words, the instrument is correlated with *Position* because the airline has likely adopted for the past flights a similar distribution, as well as 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.

8 Conclusions

This article presents several strong reasons, both based on theoretical and empirical grounds, for modeling airline pricing using the concept of fare distribution. This research strategy allows to unveil and explain some relevant, and so far, neglected aspects of DP. First, the fare variations of the first seat on sale are the result of the pricing behavior of carriers that is largely unknown to consumers. Fare increases mainly emerge as seats are sold, whereas drops occur when carriers revise the fare distribution to account for the declining option value of the seats. Second, obfuscation of this pricing behavior also depends on the fact that the set of fares is discrete so that in some occurrences, when a seat is sold or when there is a downshift of the fare distribution, the fare of the first seat on sale remains unchanged. Third, although not the central focus of the study, our analysis suggests that to correctly identify inter-temporal price discrimination practices, it is necessary to distinguish between upward changes in the distribution from fare hikes driven by capacity considerations. Finally, our analysis helps to solve the contrast between theory and empirical evidence illustrated in McAfee and te Velde (2007).

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.

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*, 107(1):283–303.

Cavallo, A., Neiman, B., and Rigobon, R. (2014). Currency unions, product introductions, and the real exchange rate. *Quarterly Journal of Economics*, 129(2):529–595.

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.

Flig, T., Isler, K., Opperstad, C., and Belobaba, P. (2010). Optimization of mixed fare structures: Theory and application. *Journal of Revenue and Pricing Management*, 9(1):152–170.

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. (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.

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.

Pang, Z., Berman, O., and Hu, M. (2015). Up then down: Bid-price trends in revenue management. *Production and Operations Management*, 24(7):1135–1147.

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.

Williams, K. R. (2017). Dynamic Airline Pricing and Seat Availability. Discussion paper no. 3003, Cowles Foundation.

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

A Appendix - For Online Publication

A.1 Building the distributions from easyJet's posted fares.

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:

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

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

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

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

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

After finding C, using (A.1) it is straightforward to derive the bucket fare tags, P_j:

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

with $P_1 = PF(1) - C$.[23]

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,

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

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".[24] We can then gauge the extent to which the size of each bucket, obtained from (A.2), 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.

A.2 Fare distributions and Full Service Carriers' pricing

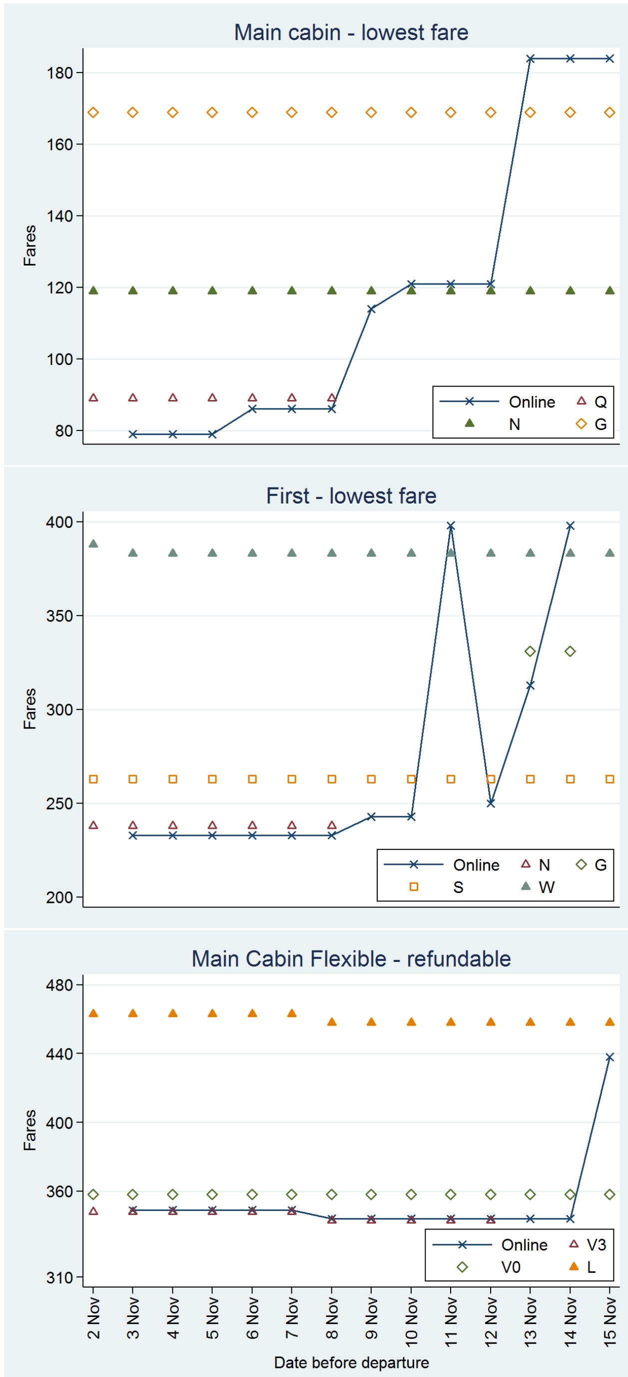
As far as FSCs are concerned, the analysis is complicated by their adoption of a nested-classes system, where the same seat can belong to different classes, each with different ticket restrictions; therefore, one would need to retrieve a distribution for each class category, with precise information 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 articles present graphical evidence of the temporal profile of fares by FSC, i.e., they report the fare of the first 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 flight 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 7, 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.[25] There are at least

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

[25] For instance, the full code for the class Q in Figure 7 is Q7ALKNN3. It is noteworthy that expertflyer.com reports a very large number of classes, and that we only report those whose value is close to that of the posted

Figure 7: Fare classes and online posted fares- American Airlines



flight AA2296 New York JKF (7:05) - to Chicago ORD (9:01) on 15 November 2016 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 classes N and G for a seat in the main cabin (top panel of Figure 7) were available on the computer reservation system during the whole period. Interestingly, the class Q in the top part of Figure 7 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 7 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).[26] 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 7 suggests that the notion of a fare distribution provides a useful starting point for any investigation of FSCs' pricing methods.

B Proofs

Proof of Proposition 1.

Non-negativeness. Non-negativity of V can be easily shown from (2) by induction because $V(t, M)$ 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 ones) 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 $\hat{p}(1, m) = \hat{p}(2, m)$. Then, under this fare profile the expected return gained in the first $t-1$ periods is $V(t-1, M)$. Moreover, because in the last period returns are non negative, we have $V(t, M) \geq V(t-1, M)$, which contradicts our assumption. The

online fares.

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

proof that $V(t,M) \geq V(t,M-1)$ is similar to the previous one.

Decreasing return 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 returns in t and M can be, respectively, defined as:

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

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

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

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 $\Delta_1(t_H,M_H) - \Delta_1(t_H,M_L) + \Delta_1(t_H-1,M_H) - \Delta_1(t_H-1,M_L) + \cdots + \Delta_1(t_L+1,M_H) - \Delta_1(t_L+1,M_L) \geq 0$. Because the previous inequality must hold for any $t_H > t_L$ and $M_H > M_L$, it is equivalent to (B.2)

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

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

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 because we have subtracted a constant term $V(t-1,M)$. Moreover, from the Envelope theorem, $\Delta_1(t,M)$ is increasing in X. Therefore, it is possible to state the following result:

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

Moreover:

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

Similarly, (2) can be rephrased as:

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

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:

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

Moreover:

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

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

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

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

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

Step 3. We prove that inequalities in (B.1) are satisfied by induction. We start to show that inequalities in (B.1) hold for any $(t,1)$ or $(1,M)$, with $t=1,2,..,T$ and $M=1,2,..,N$. When $M=1$, $X(t,1)=-V(t-1,1)$. Because $V(t-1,1)\ge V(t-2,1)$, using (B.4), we have that $X(t,1)\le X(t-1,1)$ and $\Delta_1(t,M)\le\Delta_1(t-1,M)$. Similarly, when $t=1$,

$Y(1, M) = -V(1, M-1)$. Because $V(t-1, M) \geq V(t-2, M)$, using (B.7), 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 (B.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 (B.1) hold for preceding values of (t, M), and we want to show that they hold for (t, M). Using as assumption that the first inequality of (B.1) holds for $(t, M-1)$ and that the second inequality of (B.1) holds for $(t-1, M)$, and thanks to the first part of (B.10), we obtain:

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

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

Similarly, using as assumption that the second inequality of (B.1) holds for $(t-1, M)$ and that the first inequality of (B.2) holds for $(t, M-1)$, and thanks to the first part of (B.9), we obtain

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

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

■

Proof of Corollary 1.

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

■

Proof of Proposition 2.

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

$$p^*(X) = \arg\max_{p \in \Theta} \{q(p)[p + X]\} \tag{B.13}$$

Let $\rho = \bar{\theta} - p$ and $H(\rho, X) = q\left(\bar{\theta} - \rho\right)\left[\bar{\theta} - \rho + X\right]$. 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:

$$\left[q(\bar{\theta} - \rho') - q(\bar{\theta} - \rho)\right](X' - X) \geq 0, \tag{B.14}$$

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

$$X' \geq X \implies \rho^*(X') \geq \rho^*(X) \iff p^*(X') \leq p^*(X). \tag{B.15}$$

From (B.15) and Corollary 1, we obtain the proof. ■

B.1 Algorithm

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

$$V(t, M) = \max_p \{q(p)[p + V(t, M-1) - V(t-1, M)]\} + V(t-1, M) \tag{B.16}$$

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 (B.4), we consider the following steps.

Step 1. Find the solution for $\max_{p\in\Theta} q(p)(p + X)$. Because 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 (B.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.