ISSN: 2279-7807

**Quaderni di Dipartimento**

# Combined Effects of Load Factors and Booking Time on Fares: Insights from the Yield Management of a Low-Cost Airline

Marco Alderighi
(Università della Valle d'Aosta e Università Bocconi)

Marcella Nicolini
(Università di Pavia e FEEM)

Claudio A. Piga
(Loughborough University e Rimini Centre Economic Analysis)

# 171 (05-12)

Dipartimento di economia politica
e metodi quantitativi
Università degli studi di Pavia
Via San Felice, 5
I-27100 Pavia

Maggio 2012

# Combined Effects of Load Factors and Booking Time on Fares: Insights from the Yield Management of a Low-Cost Airline*

Marco Alderighi†
Università della Valle d'Aosta
Università Bocconi

Marcella Nicolini
Università di Pavia
FEEM, Milan

Claudio A. Piga
Loughborough University
Rimini Centre Economic Analysis

May 1, 2012

**Abstract**

Based on two strands of theoretical research, this paper provides new evidence on how fares are jointly affected by in-flight seat availability and purchasing date. As capacity-driven theories predict, it emerges that fares monotonically and substantially increase with the flights occupancy rate. Moreover, as suggested in the literature on intertemporal price discrimination, the adoption of advance purchase discounts is widespread as the departure date nears, but it may be part of a U-shaped temporal profile, where discounts are preceded by periods of relatively higher fares. Finally, the intervention of yield management analysts appears to play a substantial role.

**JEL Classification**: D22, L11, L93.

**Keywords**: Pricing policy, Panel Data, Ryanair, Yield Management.

*We wish to thank seminar participants at Texas A&M University, Tilburg University, London City University, University of Nottingham, and at the 2010 EARIE conference in Instabul for their comments. Thanks to Maksim Belitski, Makoto Watanabe and Chris Wilson for their feedback. The responsibility for the arguments and results expressed in this paper is entirely ours. Piga gratefully acknowledges financial support from the Leverhulme Trust Research Fellowship Programme.

†Corresponding author. Università della Valle d'Aosta, Grand Chemin 75/77, Saint Christophe 11020, Italy; and CERTeT, Università Bocconi, Roentgen 1, 20136 Milan, Italy; telephone +39 0165 066 723; fax: +39 0165 066 748. E-mail: m.alderighi@univda.it.

# 1 Introduction

Yield management (henceforth, YM) refers to a broad set of techniques that are profitably used by such companies as airlines, hotels, car retails, cruise shipping, etc., to implement a price discrimination policy when customers are heterogeneous, demand is uncertain and capacity is hardly modifiable. In its simplest formulation, it entails a trade-off between accepting now a booking request at a low price or refusing it to leave room tomorrow for a potential customer willing to pay a higher price (Weatherford and Bodily, 1992; McGill and Van Ryzin, 1999).

In the airline sector, YM implementation usually requires that seats are grouped into different booking classes, each having a definite fare and, in most cases, specific restrictions (e.g., ticket refunding, advance purchase restrictions, valid travel days, or stay restrictions). YM activity, in practice, consists in setting fares and/or managing the number of seats allocated to each class. Although YM operations are heavily computerized, the human intervention (carried out by a "yield manager" or "analyst") still remains very important. It may occur when the observed sales are not aligned to the forecasted ones, or be due to a rapid change in market conditions such as an unexpected peak demand or a strategic action of rivals. In economic terms, YM can be interpreted as a very sophisticated way to implement pricing policies, which may produce a wide range of fares even for the same flight, so that two passengers sitting next to each other are likely to have paid different prices for their tickets.[1]

This paper aims at providing new evidence on the sources of such a difference by using an original database combining detailed information on fares and seat availability obtained from the website of Europe's largest Low-Cost airline (henceforth, LCA), Ryanair. The relatively simple pricing behaviour of a LCA helps us to identify the combined impact on fares of both in-flight seat availability and the time separating the purchase from the departure date. This, in turn, allows us to provide a test for the predictions of two theoretical strands of research on airline pricing: the capacity-driven and the time-driven approach, respectively.[2]

[1] In Borenstein and Rose (1994), the expected difference in fares between two random passengers on a given flight is on average 36% of the airline's average ticket fare; this percentage increases to 44% in Gerardi and Shapiro (2009) and to 66% in Gaggero and Piga (2011).

[2] Previous studies on pricing behaviour in the U.S. Airlines industry have used different cohorts of the same database, i.e., the Databank of the U.S.A. Department of Transportation's Origin and Destination Survey, which is a 10 percent yearly random sample of all tickets that originate in the United States on U.S. carriers (Borenstein, 1989; Kim and Singal, 1993; Evans and Kessides, 1993, 1994; Borenstein and Rose, 1994; Lederman, 2008; Gerardi and Shapiro, 2009). None of these studies addresses the issues of this paper.

Dana (1999a) provides a theoretical model that addresses the link between fares and seat availability. The basic idea is that the optimal fare is given by a constant mark-up over the capacity cost. Because the shadow cost of a capacity unit increases as the probability of selling a ticket decreases, the pricing profile increases with aircraft capacity utilization. In other words, intra-firm dispersion arises not because an airline is trying to segment its market, but because demand is uncertain, and the probability of selling an extra seat decreases with in-flight seat utilization. In equilibrium, the airline defines a fare distribution where the cheapest fares are assigned to seats with the highest probability of sale and the highest fares are associated to seats that are seldom occupied.[3]

In this paper, we provide a direct test of the relevance of the capacity-driven approach, i.e., of the extent by which fares increase with capacity utilization. A main practical difficulty in carrying out this test is the availability of data on capacity utilization at the time a fare is offered on an airline reservation system. Another complication, usually associated with fares by full service airlines, may arise because different booking classes, each with a different set of restrictions and fares, may be simultaneously available to travellers at a given point in time, thus making it necessary to account for all ticket characteristics (Stavins, 2001). A notable innovation in this study is the possibility to combine fares with the number of seats available at the time the fare was retrieved from the airline's website. Moreover, using data from Ryanair rules out any difference in seats' characteristics, because the airline imposes the same set of restrictions on all its fares. Furthermore, by using flight fixed-effect panel data techniques, where the temporal dimension is obtained by tracking a flight's fares and seat availability over a 70-day period, we also control for possible unobserved heterogeneity across flights. Our estimates indicate that, on average, an extra sold seat induces an increase of about 2.6% in offered fares. This effect increases in the sample of flights that: *i)* realized a high load factor early during the booking period; *ii)* operate in less competitive routes; *iii)* are scheduled in periods of higher demand, e.g., during Summer. These results show the relevant role played by a capacity-driven approach to airline pricing in explaining airline price dispersion.

The previous evidence on this issue is rather mixed. On the one hand, Puller et al. (2009) find only modest support for the capacity-driven approach, and illustrate that much of the fare variation may be associated with second-degree price discrimination (i.e. ticket characteristics). On the other, Escobari and Gan (2007) find that price quotes are on average higher in fully occupied flights, as predicted by the capacity-driven approach. However,

[3]Dana (1999a), which assumes homogeneous seats, shows that the equilibrium distribution of fares varies by market structure: fare dispersion is found to increase with competition.

this is obtained using the probability that a flight is sold-out, and not through the association of fares with the capacity utilization at the time the fare is posted. Both these studies, however, rely on data generated by the more complex process used by traditional (legacy) carriers, whose properties are only partly aligned with the assumptions adopted by any of the models in the theoretical literature.

In addition to the capacity-driven approach, the literature has indicated that airlines may use inter-temporal price discrimination to exploit customer's heterogeneity in terms of willingness-to-pay and uncertainty about departure time (Gale and Holmes, 1992, 1993; Dana, 1999b; Möller and Watanabe, 2010). Advance-purchase discounts (APD, hereafter) and clearance sales (CS, hereafter) provide a practical way to implement this strategy.

In Dana (1999b), firms cannot distinguish between peak and non-peak flights and travelers differ in their disutility to fly at their least preferred time; in equilibrium, firms commit to a distribution of monotonically increasing fares over time for each flight. Gale and Holmes (1992) show that a monopolist, and a social planner, can use APD to spread uncertain peak demand more evenly between two flights. Gale and Holmes (1993) show that in a monopoly with capacity constraints and perfectly predictable demand, APD arises from a mechanism design setting where consumers self-select so that demand is diverted from peak periods to off-peak periods. Möller and Watanabe (2010) study the conditions under which, over two consecutive periods, prices may either decline (i.e., firms offer CS) or increase (i.e., firms engage in APD). They demonstrate that the former (the latter) is more appropriate when a consumer's demand uncertainty is absent (present) and the risk of being rationed is high (low).

This study sheds light on Ryanair's time-driven pricing policy. The idea is that if a temporal profile is coded into the carrier's reservation system or is the result of the analyst's intervention, it can be identified by tracking the evolution of each flight's fares over time (Mantin and Koo, 2009). A novel feature of the present work is that we do so after controlling for capacity utilization; thus, we are able to separate fare variations due to purely capacity-driven motivation from those induced by the willingness to discriminate between customers booking at different times before departure. The evidence reveals that, in general, fares increase monotonically over the last three weeks before departure. However, a more complex fares' dynamics consistent with a combined use of both CS and APD is also found over the entire booking period we take into consideration: in the two months preceding departure the intertemporal profile of a standard flight's fares often appears to be U-shaped, especially in flights that fill up well in advance of departure. While the existing

literature has already identified that fares generally tend to increase as the departure date nears (**?**Puller et al., 2009), to our knowledge no previous contribution has illustrated a U-shaped temporal profile of fares.

In sum, this paper offers the first combined study of two testable implications derived from the theoretical economic literature on airline pricing. Both implications relate to the pricing profile of carriers suggesting that fares *i*) should increase as a flight fills up and *ii*) should grow over time (APD), but may have more complex U-shaped temporal pattern combining APD and CS.

A notable innovation of this study is that it addresses both of these features simultaneously. Given the parallel movement that both effects induce on fares, studying one without the other is likely to bias the analysis. Furthermore, the joint investigation of both properties sheds lights on the relative importance of two classes of theoretical airline pricing models, one emphasizing a capacity-driven motivation, the other focussing on intertemporal price discrimination and customers' segmentation (Alderighi, 2010; Puller et al., 2009). This does not imply that a discriminatory motive is revealed only by the analysis of the intertemporal profile of fares. It is precisely because we control for the intertemporal effects that we can tease out a discriminatory motive in the way the airline designs the relationship between fares and occupancy rate. Indeed, we find that the fare profile is highly affected by market structure; it is steeper in less competitive routes, implying that in such routes the last groups of seats are sold at a higher fare. The lack of competitive pressure thus facilitates the extraction of surplus from consumers who learn at a late stage that they have to fly on a specific date (which makes their demand inelastic) and who therefore end up buying when the flight is already quite full.

The remainder of the paper is structured as follows. Section 2 illustrates the importance of the airline whose fares we used in the study. In Section 3, we explain how we could retrieve the information on the flight's occupancy at the time the fares were posted. Section 4 provides some descriptive statistics on the fare profile. The econometric model used to tackle both the censoring and the endogeneity of the number of sold seats is presented in Section 5, which is followed by the comments to the main findings in Section 6. Section 7 concludes.

## 2 Ryanair's business model

Drawing on the business model established by Southwest Airlines in the US, Ryanair pioneered the low-cost strategy in Europe. The low-cost carrier business model that Ryanair adopts has several notable features: (i) a simple pricing structure with one cabin class (with optional paid-for in-flight food and drink); (ii) direct selling through internet bookings with electronic tickets and no seat reservations; (iii) simplified point-to-point routes often involving cheaper, less congested airports; (iv) intensive aircraft usage (typically with 25-minute turnaround times); (v) employees working in multiple roles (e.g. flight attendants, cleaning the aircrafts and acting as gate agents); and (vi) a standardized fleet made up of only Boeing 737-800 aircrafts, with a capacity of 189 seats.

Founded in 1985 and based in Dublin, Ryanair expanded its route network rapidly following liberalization of intra-EU air services, increasing its passenger numbers from 5.6 million in 2000 to 33.6 millions in 2005, reaching over 71.2 million by 2010. For comparison, in the same year the number of passengers flying with Lufthansa (44.4m), EasyJet (37.6m), Air France and Emirates (both 30.8m) and British Airways (26.3m) was considerably lower.[4] Not only has Ryanair been remarkably successful in growing passenger numbers and revenues, it has also been a consistently profitable business in a sector in which many airlines have struggled to make profits from one year to another: its operating revenues (profit) in 2000, which amounted to 370 (72.5) million euros, escalated to the value of 3,629 (374.6) million euros in the financial year ending on the 31 March 2011. The size and importance of this carrier, and its ability to attract customers, make it a key player in the European airline industry. Thus, the study of its pricing behaviour can shed empirical light on the pricing policies of airline carriers and on the validity of the related theoretical works.

### 2.1 Insights into Ryanair's YM practices

Unlike most full service carriers, Ryanair employs a relatively simple pricing structure with no price discrimination based on multiple service and cabin classes and on specific restrictions like minimum stay requirements and Saturday night stay-overs. Furthermore, all its tickets carry the same penalties for a name, date and/or route variation and permit the same free in-flight hand baggage allowance (10 kg) with a fixed fee for boarded baggage (max 15 kg per person in a single luggage). None of these impinge on the YM aspects on

[4]The information in this section is drawn from material, including yearly Financial Reports, available from Ryanair's website www.ryanair.com/en/about.

which we focus on in this paper, since they are unaffected by capacity utilization and/or by temporal aspects.[5] Thus, the use of Ryanair data offers some advantages for the empirical analysis, since its ticket characteristics (identical restrictions) are close to the modelling assumption of many theoretical works.

Interesting insights into how Ryanair operates its YM system are given in European Commission (2007), which provides details of the investigation that lead to the decision to block the takeover of Ryanair over Aer Lingus. Both companies adopt a certain number of standard "templates", which describe "the number of places that should be available in a given price category ("booking class") (p. 109, item 439)". Templates are chosen by a yield manager with the aid of a software that tracks the booking status of each flight. The analyst can compare the actual booking status (or "load factor") of a flight with the booking forecast which is provided by the system. When the load factor falls behind forecast, the analyst may decide "to stimulate the demand, normally by making more seats available in the cheaper price categories (p. 109, item 440)". The analysis in European Commission (2007) thus suggests that (*i*) the initial number of seats allocated to each booking class is set through the choice of a template defining the distribution of cheaper and more expensive seats; (*ii*) fare adjustments are generally designed as a switch or shift of the template. Such adjustments are generated by an "analyst's intervention" in response to either a misalignment between actual and forecasted load factors, or external factors, e.g., the need to match a rival's fare.

More importantly, the foregoing discussion highlights two important YM activities, both contributing to the determination of the airline's fares. On the one hand, the airline follows a standard, "routinary" approach when it sets the fares for a flight with specific characteristics (e.g., route, time and day of departure etc.). On the other, idiosyncratic, discretional interventions by an yield analyst may be due to either external (e.g., matching rivals' prices, new qualitative information on future demand etc.) or internal (a promotional policy required by the marketing manager, etc.) factors. One of the main contributions of this paper is to show how the routinary activity is captured by an augmented version of the template, which simultaneously details how fares are related to seat occupancy and how they are designed to change as the time to departure nears.[6] By doing so, we shed

[5]The charges for a ticket variation are so high relative to the average price of a ticket that it is often cheaper to buy a new ticket. This, combined with the fact that Ryanair does not practice overbooking, may explain why we practically observe no cases where capacity utilization decreases over two consecutive periods. Further, the luggage charge was started after our sample period ended.

[6]In Section 4.2 we show how changes over time appear to be coded into the airline's computerized reservation system.

light on the role played by the capacity and time-driven theories in explaining the airlines' fare setting process.

## 3 Data Collection

Our analysis is based on primary data on fares collected using an "electronic spider" linking to the website of Ryanair. The database includes daily flights information from January 2004 up to, and including, June 2005. In order to account for the heterogeneity of fares offered by airlines at different times prior to departure, every day we instructed the spider to collect the fares for departures due, respectively, 1, 4, 7, 10, 14, 21, 28, 35, 42, 49, 56, 63 and 70 days from the date of the query. Henceforth, these will be referred to as *BookingDays*.[7] Thus, for every daily flight we managed to obtain up to 13 prices that differ by the time interval from the day of departure. Thus, we can identify the evolution of fares over time - from more than two months prior to departure to the day before departure.

Data collection was carried out everyday at the same time and included: the price of one seat, which in the remainder of the paper is denoted as *Fare1*, the number of seats available at each booking day, denoted as *Seats*, and the corresponding unit price for a query involving that number of seats, referred to as *TopFare* since it is never smaller than *Fare1* (see subsection 3.2 for a discussion of both fares and their role within a template). We also collected the time and date of the query, the departure date, the scheduled departure and arrival time, the origin and destination airports and the flight identification code, which will be used as controls in the econometric analysis.

In addition to UK domestic fares, routes to the following countries were surveyed: Austria, Belgium, France, Germany, Ireland, Italy, Netherlands, Norway, Spain, Sweden. For consistency, the procedure considered only flights departing from an airport within the UK, and arriving at either a domestic or an international airport. We have data for 82 of the 154 routes that Ryanair operated to these countries over the sample period; in some cases, we consider more than one flight code per route when the airline operated more than one daily flight. All fares, which do not include tax and handling fees, are for a one-way flight and are quoted in Sterling.[8] Some descriptive statistics are reported in Table 1 and

[7] For instance, assume the queries were carried out on March $1^{st}$ 2004. The spider would retrieve the fares for flights whose departures was due on 2/3/04, 5/3/04, 8/3/04, 11/3/04 etc. The routine was repeated every day over the data collection period.

[8] Focusing only on the outward leg from the UK emerges as a valid data collection strategy since it is widely acknowledged that European LCAs price each leg independently (Bachis and Piga, 2011). Moreover, excluding taxes and fees does not affect the results for the following reasons. First, Ryanair started charging

2, whose comment is deferred to section 3.2.

### 3.1 Retrieving data on *Seats* and *TopFare*

The collection strategy exploited a feature of the Ryanair's website: during the sample period, Ryanair allowed purchases of up to 50 seats using a single query. This made it possible to learn if, at the time of the query, fewer than 50 seats were available on a flight with a specific identification code. The spider worked using the following algorithm:

- issue a query for $S = 50$ seats for a specific flight identified by a unique flight code on a route. The flight was due to depart $D$ days from the date of the query, where $D$ assumes the values of the *BookingDay* previously introduced.
- If the airline's site returned a valid fare for that flight code, then we interpreted this finding as follows: $D$ days prior to departure, there were at least 50 seats available on the flight. We could not however retrieve any more precise information regarding the actual number of available seats, which is thus censored at the level of 50. The spider would then save the value of *Seats*$= 50$, and the corresponding value of *TopFare*, as well as the value of $D$ and all the other flight's details (see above).
- If the site failed to return a valid fare for that flight, the programme inferred that there were fewer than 50 seats available and then started a search to obtain the highest number of seats in a query that returned a valid fare. This corresponds to the number of seats available $D$ days before a flight's departure, a value which was saved in *Seats*. In this case, *TopFare* corresponds to the unit price at which the airline was willing to sell all the $S$ remaining seats in a single transaction.

By repeating this procedure every day, we could track the seats and the associated fare for each value of *BookingDay*.

### 3.2 Interpretation of retrieved fares

When $Seats < 50$, *TopFare* corresponds to the fare of a transaction whose completion would fill the flight to capacity.[9] For this reason, *TopFare* presents two important characteristics.

a fixed fee for check-in and luggage only in 2006, that is, after our sample period. Second, the fixed per-passenger tax that contributes to the full cost of the ticket would not impinge on the evaluation of how a flight's fare change relative to the flight's occupancy rate or over time.

[9] If, for example, the spider returned 26 left seats for a given booking day, then the retrieved fare would correspond to the posted fare for a booking of 26 seats, i.e., for the number of seats that would close the flight.

First, as Table 1 shows, it hardly varies with the number of seats. Indeed, for all the routes in the Table, it appears clear that, despite the wide sample period covered by the data, the distribution of *TopFare* is highly concentrated. In many routes, its maximum value coincides with the median and the mode values, which are in turn only marginally above the mean value, thus suggesting a very limited number of cases where *TopFare* assumes values below the mode. In other routes, the maximum value is higher, but not more than £10 or £20 than the median/mode. Overall, it appears that *TopFare* is largely insensitive to the number of seats that remain to be sold as well as to the number of days that separate the fare retrieval from the flight's departure. This is supported also by the low standard deviations reported.[10]

Second, and relatedly, if $Seats < 50$, in line with the capacity-driven approach, *TopFare* represents the maximum fare of a flight. When a query that closes the flight is issued, the Ryanair's reservation system always retrieves the fare that is associated with the value of the last seat.[11] This implies that the value of *TopFare* and *Fare1* can coincide when only a few seats remain on a flight. Table 2 reports the maximum and mean values for the *Fare1* variable. The maximum value for one seat is generally either identical to or slightly below the equivalent value for *TopFare* when there are less than 50 seats available. Therefore, we have cases where the values of *TopFare* coincide with the highest values of *Fare1*. Relatedly, the mean values of *Fare1* across all routes are well below those in Table 1, even when we condition on observations with less than 50 seats available. Indeed, conditioning for $Seats < 50$, *Fare1* is more dispersed than *TopFare*, given the wider gap between the maximum and the mean values of *Fare1* relative to those of *TopFare*. An implication is that *TopFare* does not represent an average of the remaining "forward" values of *Fare1*. If this were the case, *TopFare* would change with the number of remaining seats.

Always with reference to Table 2, with 50 seats or more available the fares for one seat cannot, *a fortiori*, refer to the last seats available on a flight, and indeed we do not observe any coincidence between equivalent values of *Fare1* and *TopFare* between the two Tables. Furthermore, the value of one seat when there are at least 50 available is expected to be

[10] That is, the spider retrieved the same value of *TopFare* (or very similar ones) both when the "booking day" was, say, 28 days from departure and there were 45 seats available to sell, and when the "booking day" was 10 and the spare capacity was 22.

[11] The capacity of a Ryanair's flight is 189 seats. When $Seats = N < 50$, issuing a query for $N$ seats always retrieves the value of the $189^{th}$ seat in the template. Consistently, when $S = 50$, i.e., when we do not know the exact number of available seats, *TopFare* indicates the fare of the $50^{th}$ seat ahead of the one being available.

not higher than the fare for one seat when 49 ones remain to be sold. This is clearly borne out by the difference in the mean values of *Fare1* when the remaining number of seats is either below the value of 50 or not.

# 4 Preliminary evidence

The results drawn from the descriptive analysis in this section help gain better insights into the airline's pricing policy and its relation with both the in-flight seat availability and the purchasing date. They also provide a useful guide for the specification of the econometric model and the interpretation of its findings.

## 4.1 The fare-occupancy rate schedule

Figure 1 shows the median spline plot of *Fare1* on *Sold Seats*, which represents the complement to 50 for the number of available seats retrieved by the spider (i.e. $50 - Seats$) and is thus available only for those observations where the number of available seats is strictly less than 50. The values in the Figure refer to the London Gatwick - Dublin route: each line represents a different flight code. The lowest fare is about £25, while the highest is just below £150. In all periods the plot shows, on average, a monotonically increasing relationship; however, there are a number of instances where *Fare1* marginally falls as occupancy increases. A smoother increasing relationship is obtained in Figure 2, where we use the *Log* of *Fare1* in a non-parametric fit with the last 50 seats' occupancy.

To generalize the evidence from one route to the entire sample, in Figure 3 we follow the approach used in Puller et al. (2009). We first calculate, for each flight-code/booking day combination, the mean value of both *Fare1* and *Sold Seats* in a given month; then, we derive the percentage deviation of each daily observation from each respective mean value. We then aggregate the pairs of percentage deviations across three categories of booking days: early, middle and late. In the first category, a percentage increase (decrease) of 20% of *Sold Seats* from its mean, as reported on the horizontal axis, is associated with a percentage increase (decrease) of about 100% (60-70%) in *Fare1* from its mean (as can be read on the vertical axis). As the date of departure approaches, fares become more responsive to increases in a flight's occupancy rate. Indeed, the same increase of 20% over the mean of *Sold Seats* is associated with almost a 200% increase in *Fare1* over its mean, when we consider the middle and late booking days' categories. Interestingly, ten to one days to departure, the same 20% decrease from the mean of *Sold Seats* is met only by

a similar decrease in the mean of *Fare1*. It is worth stressing here that the evidence in Figure 3 does not necessarily imply any temporal effect and may be easily reconciled with the capacity-driven approach. Indeed, it is obvious that as the date of departure nears, the flight becomes fuller (and therefore the fares higher). If the fare/occupancy function is non-linear as Figure 1 may suggest, and its gradient becomes steeper as the flight fills up, then the middle and late booking days' categories will include a greater proportions of high fares, and thus drive the findings reported in Figure 3. Incidentally, this view lends support to the notion that seat availability is a crucial source of fare dispersion. Nonetheless, a complementary possibility could be that the large percentage increases may arise due to the combined effect of temporal as well as occupancy rate changes.

A number of considerations can be drawn from the foregoing graphical analysis. First, the evidence reported in Figure 3 suggests that YM techniques designed by airlines to manage capacity constitute an important factor driving fares' dispersion. Interestingly, despite the methodological similarity, Puller et al. (2009) reach an opposite conclusion in their study of U.S. airline markets.[12] Second, it introduces the need to combine capacity concerns with at least two other aspects of YM: *1)* the fares' temporal profile, i.e., the possibility that fares may change regardless of the flight's occupancy rate and *2)* the discretional intervention of an yield manager to tackle unexpected contingencies. The latter point will be considered in the econometric analysis, where we employ instrumental variable techniques to isolate the carrier's routinary pricing behaviour net of such discretional interventions. Given the crucial role of advance purchase discount (APD) in the literature, in the next section we delve deeper into the existence and the characteristics of the temporal profile (Gale and Holmes, 1992, 1993; Dana, 1999b; Möller and Watanabe, 2010).

### 4.2 Do fares increase over time all the times?

The descriptive analysis in the previous sub-section highlighted a positive relationship between fares and sold seats, that appears to hold on average over a range of dates and routes. In this sub-section, we extend the analysis by focusing on possible intertemporal effects in the airline's pricing structure. Our objective is to separate fare changes induced by variations in the flight's load factor from temporal effects that are unrelated with the actual evolution of sold seats. This is also important in terms of modelling testing, because the literature on APD leads to predictions that may be confused with those derived in the case

[12] Differences may be due to the different type of airlines considered (legacy vs. low cost) and to the different methods used to obtain the flights' occupancy rate.

where fares are only driven by the shadow cost of capacity. That is, in both cases, fares may increase over time. Indeed, under APD, there is a moment when the airlines increase their fares to exploit costumer heterogeneity.

Table 3 reinforces the previous analysis, and shows that, when we hold the booking day fixed and look at the fares in each line of the table, fares in our sample on average decrease as more seats on the plane are available. More interestingly, when we condition on capacity utilization to see how fares on average change with the booking day, we observe that the intertemporal profile of fares assumes a U-shaped form, with the minimum fares occurring 21 to 14 days prior to departure. Indeed, the evidence in every column suggests that during the last fortnight fares return to the level they assumed about 35-28 days before departure. Thus, Table 3 provides *prima facie* evidence that a more appropriate representation of the intertemporal profile of the fares in our sample should include both APD (because fares increase in the last fortnight) and discount sales, since fares appear to drop between 21 and 14 days from departure (Möller and Watanabe, 2010).

However, it might be possible that the temporal profile in Table 3 is due to the aggregation of fares from heterogenous routes and the extensive sample period used. Therefore, Table 4 focuses on economically significant (i.e., worth at least £5) fare changes that occur within a single flight. It illustrates the likelihood of a fare drop over two consecutive booking days conditional on available seats remaining stable or decreasing. Under such circumstances, we should not observe any drop in fares if the template is decided once and for all, as discussed in Dana (1999a). Conversely, the airline adjusts its fares downward quite frequently, and in ways that appear to be consistent with an active intervention by the yield manager, as suggested in European Commission (2007). First, in each row the likelihood of observing a price drop generally increases as more seats are available, especially when the departure time is not within a week.[13] This is consistent with the expectation that drops are likely meant to stimulate demand. The Total row indicates that 13% of observations with at least 40 seats available report fare drops, while this occurs in only 6% of observations where less than 10 seats are recorded. Second, the highest probabilities of observing a drop are found in the 28-14 days period, after which they diminish sharply and are hardly observed a few days prior to a flight's departure.

Table 4 can only identify cases of decreases, but not increases, over time. However, for

[13]The fact that $4-5\%$ of late booking cases report a drop when less than ten seats are available indicates an active intervention, which may be explained by the carrier's desire to fill a flight to capacity to generate ancillary revenues and boost market shares. This incentive is however offset by the need not to offer "last-minute discounts", which customers may learn to anticipate: hence the lower probability of observing a drop within a week from departure.

the large majority of observations, fares increase between two consecutive booking days, and at the same time, available capacity reduces. In such a case, using descriptive statistics, it is not easy to separate the variation due to capacity utilization from the temporal variation. In Table 5 we show a *pure* temporal variation, since we hold in-flight occupancy fixed between two consecutive booking days by considering only those observations where the number of available seats has not changed over two consecutive booking periods. Any change in price is thus not due to a change in the occupancy rate. We distinguish between *Large* and *Moderate* changes, the former being greater than £20.0 in absolute terms. As the first row in the Table indicates, the average value of a change tends to be the same for each category of decreases and increases. We also consider the case of no change, which, in line with the capacity-driven theory, accounts for the largest majority of observations (about 73%). Interestingly, this also implies that 27% of fare changes are generated by a pure temporal effect, with increases ($N = 1905$) being more than twice as many as decreases ($N = 919$). The way changes are distributed across flight characteristics does not appear to differ significantly, with some minor exceptions. First, the proportion of increases (decreases) is above (below) the sample mean when the booking day is (is not) within two weeks from departure. That is, it is more likely to observe a fare increase as the date of departure nears. By the same token, large increases are hardly observed during the early booking period. Second, more temporal variability (i.e., both more increases and decreases) is found in flights that have more than 20 seats available and are operated in routes with low competition.[14]

Overall, the evidence in Tables 3-5 suggest that fares are affected by a combination of capacity and temporal considerations. These will be further investigated in the econometric analysis.

# 5 The econometric model

We aim at estimating a pricing equation linking a flight's seat occupancy and time before departure with offered fares to identify the standard pricing behaviour of the airline. However, a standard OLS regression of the price (obtained from a query for a single seat) on the number of sold seats is not appropriate because *Sold Seats* has two features which need special attention. A first obvious issue is the endogeneity of this variable, since some unobserved determinants of the airline pricing behaviour may be correlated with a specific

[14] See section 6.3 for a formal definition of routes with high and low competition.

flight's time-invariant factors (an issue which could be dealt with the standard fixed-effects panel technique) and, more importantly, with the idiosyncratic, *discretional* intervention of the airline's yield manager. This problem calls for an instrumental variables estimator. A second, more subtle, issue, is that the number of sold seats is censored due to the retrieving procedure. Indeed, the number of sold seats may range from 0 to 189, i.e., the airline's aircrafts capacity. However, we can only detect the number of available seats when they are less than 50. This censoring, therefore, induces a bias in the estimates, and needs to be corrected.

Consider a simple model where $y$ is a function of a vector of explanatory variables, $\mathbf{x}$, and $\mathbf{z}$ is a vector of instruments such that:

$$y = \mathbf{x}\beta + u \tag{1}$$

$$E(u \mid \mathbf{z}) = 0$$

The key assumption underlying the validity of two stage least squares (2SLS) on the selected sample is $E(u \mid \mathbf{z}, s) = 0$, where $s$ is a selection indicator. This assumption holds if we observe a random sample selection: $s$ is independent of $(\mathbf{z}, u)$, and a sufficient condition for this is that $s$ is independent of $(\mathbf{x}, y, \mathbf{z})$. Therefore, it can be proven that the 2SLS estimator on the selected subsample is consistent for $\beta$.

However, if the selection indicator is not independent from $\mathbf{x}$, as in our case, things are different. Suppose that $\mathbf{x}$ is exogenous, and $s$ is a nonrandom function of $(\mathbf{x}, v)$, where $v$ is a variable not appearing in equation (1). If $(u, v)$ is independent of $\mathbf{x}$, then $E(u \mid \mathbf{x}, v) = E(u \mid v)$ and we may write:

$$E(y \mid \mathbf{x}) = \mathbf{x}\beta + E(u \mid \mathbf{x}, v) = \mathbf{x}\beta + E(u \mid v)$$

Specifying a functional form for $E(u \mid v) = \gamma v$, we can rewrite:

$$E(y \mid \mathbf{x}) = \mathbf{x}\beta + \gamma v + e$$

where $e = u - E(u \mid v)$. As $s$ is a function of $(\mathbf{x}, v)$, $E(e \mid \mathbf{x}, v, s) = 0$ and $\beta$ and $\gamma$ can be consistently estimated by ordinary least squares (OLS) on the selected sample. Thus, including $v$ in the regression eliminates the sample selection problem and allows to consistently estimate $\beta$. Of course, if some variable in $\mathbf{x}$ is endogenous, the procedure to correct for sample selection is the same, while to consistently estimate $\beta$ we need 2SLS.

In our specific case, one of the explanatory variables, *Sold Seats*, is expected to be correlated with the error term $u$, and therefore instrumental variables are required. Moreover, we need to specify the selection mechanism, which in this case is determined by a censoring

of the data. The model in the population is:

$$Fare1 = \mathbf{z}_1\delta_1 + \alpha\, Sold\ Seats + u \tag{2}$$

where *Sold Seats* is the endogenous regressor, and $\mathbf{z}_1$ are the other exogenous regressors.[15] Equation (3) is a linear projection for the endogenous and censored variable, while equation (4) describes the censoring induced by the data retrieving procedure:

$$Sold\ Seats = \mathbf{z}\delta_2 + v_2 \tag{3}$$

$$Sold\ Seats^* = \max(0, \mathbf{z}\delta_3 + v_3) \tag{4}$$

We allow correlation among the three error terms. We assume: a) $(\mathbf{z}, Sold\ Seats^*)$ is always observed, but $(Fare1, Sold\ Seats)$ is observed when *Sold Seats* is not censored, i.e., when $Sold\ Seats^* > 0$; b) $(u, v_3)$ is independent of $\mathbf{z}$; c) $v_3$ is normally distributed; d) $E(u \mid v_3) = \gamma_1 v_3$; e) $E(\mathbf{z}\prime v_2) = 0$ and $\mathbf{z}\delta_2 = \mathbf{z}_1\delta_{21} + \mathbf{z}_2\delta_{22}$ where $\delta_{22} \neq 0$. Defining $e \equiv u - E(u \mid v_3) = u - \gamma\, v_3$ equation (2) can be written as

$$Fare1 = \mathbf{z}_1\delta_1 + \alpha\, Sold\ Seats + \gamma v_3 + e \tag{5}$$

Since $(e, v_3)$ is independent of $\mathbf{z}$ by assumption b), we have that $E(e \mid \mathbf{z}, v_3) = 0$. As discussed above, if $v_3$ were observed we could estimate equation (5) by 2SLS on the selected sample using as instruments $\mathbf{z}$ and $v_3$. However, we can estimate $v_3$ when $Sold\ Seats^* > 0$, since $\delta_3$ can be consistently estimated by Tobit of $Sold\ Seats^*$ on $\mathbf{z}$, on the entire sample. To sum up, we proceed as follows:

1. We estimate a Tobit specification for equation (4) using all observations;
2. We retrieve the residuals: $\hat{v}_3 = Sold\ Seats^* - \mathbf{z}\hat{\delta}_3$ for the selected subsample;
3. On the selected subsample, we estimate a modified version of (5), where instead of $v_3$, which is not observed, we include $\hat{v}_3$ among the regressors. As *Sold Seats* is endogenous, we adopt an Instrumental Variable 2 Stage Fixed Effect (IVFE) estimator, using as instruments $\mathbf{z}_1$ and $\hat{v}_3$.[16]

[15] The *true* fare setting model should also include the analyst's discretional intervention, $AYM$, so that $Fare1 = \mathbf{z}_1\delta_1 + \alpha\, Sold\ Seats + \lambda AYM + \varepsilon$. Because $AYM$ is unobserved and its effect is included in $u$, endogeneity is thus due to an omitted variable problem resulting from the positive correlation between *Sold Seats* and $AYM$. Indeed, the analyst is more likely to *discretionally* reduce (increase) fares when *Sold Seats* is low (high).

[16] Our approach therefore strictly follows the Procedure 17.4 in Wooldridge (2002, p.574).

It is possible to test if the selection bias is statistically significant by observing the $t$ statistic on $\hat{v}_3$ in the IVFE model: when $\gamma_1 \neq 0$ standard errors should be corrected. We do so by means of a bootstrapping procedure.

### 5.1 Model specification

To estimate (5), given the structure of our data, we focus on a panel where the identifier is the single flight (defined by a combination of departure date and flight code) and the time dimension is given by the time interval from the day of departure (i.e., the booking day). This panel structure allows to control for all unobserved characteristics which are specific to the single flight, such as, for instance, market structure and distance. Furthermore, focussing on a single flight using a fixed-effects approach allows to control for possible strategic effects at the route level, where, e.g., the airline can opt to implement temporary capacity limits, i.e., reduce the number of daily flights.

With regards to the regressors in (5), $\mathbf{z}_1$ includes a set of booking days dummies and month of booking dummies. These exogenous regressors are part of the set of explanatory variables, $\mathbf{z}$, in the first stage estimation. To these we add the residuals from the Tobit procedure, $\hat{v}_3$, to account for the sample selection. As exclusion restrictions, we consider two instruments. Their validity depends on the extent they are correlated with *Sold Seats* and uncorrelated with the residuals $e$ of the pricing equation. The first is a dummy indicating whether the day the fare was posted is during a holiday period (i.e., main UK Bank Holidays and the week before and after Christmas and Easter). Its effect on *Sold Seats* may be driven by the fact that the ticket purchasing activity in such periods is likely to be different from non-holiday periods (e.g., when on holiday a person is less likely to spend time planning future trips), and that this difference does not bring about a discretional intervention by the yield manager (e.g., because there are fewer staff working during holidays). The second instrument is derived by building upon the interpretation of *Top Fare* and *Fare1* presented in Section 3.2. Their difference captures a relevant feature of the flight's pricing template, that is, how the gap between the two fares narrows as the flight fills up; such a gap is clearly highly correlated with the occupancy rate, but it is also correlated with $e$. However, under the assumption that template changes are specific to each daily flight, using the lagged and the forward values of this difference would still retain the important information about the template, without any correlation with other flights' idiosyncratic shocks. To capture the fact that templates may change with the day of the week, the instrument denoted "*L-FW Mean* $\Delta$ *Fare*" is constructed by taking, for each day of week-flight code combination, the

difference between *Top Fare* and *Fare1* in the two weekdays preceding (Lag) and following (Forward) the combination, and then by taking the mean of these four values.

Notice that in principle the same set of exogenous variables, $\mathbf{z}$, could appear in the selection equation and in the first stage of the IV procedure.[17] However, in practice, the two sets of regressors should differ, otherwise a severe problem of multicollinearity between $\hat{v}_3$ and $\mathbf{z}_1$ may affect the results. Therefore, in the Tobit specification for model (4) we exclude the dummy for the booking during a holiday period and instead we include the number of UK airports serving the destination airport: this is not correlated with $v_3$, since the decision to open a route is generally taken in the preceding quarter, but it captures that a higher demand destination is more likely to be served by more than one UK airport. Furthermore, dummies for the weekday of booking are included in the Tobit, but not in the IVFE model. Finally, a set of week, route, and daytime of departure dummies are included. These would be dropped in the IVFE procedure used in (5).

The validity of the chosen instruments is confirmed by a number of tests presented in Tables 7-11. The first one is the Hansen's J statistic for overidentifying restrictions: the joint null hypothesis is that the instruments are valid. If the test fails to reject the null hypothesis, then all instruments used are considered exogenous. The second one is the Kleibergen-Paap LM statistic, which tests whether the equation is identified. A rejection of the null indicates that the matrix of reduced form coefficients is full column rank and the model is identified.[18] To anticipate our results, both tests, as well as the weak instruments tests not reported, strongly support our choice of instruments.

Finally, Table 3 shows that there is a more than proportional increase in fares for any additional seat sold, suggesting a log-linear specification of the model.

# 6 Results

Table 6 reports the Tobit and the first stage estimates, respectively. As discussed above, although in principle the two sets of regressors should be identical, problems of multicollinearity require the two groups to differ. Additionally, the two estimation samples differ, as the Tobit model is estimated on all available observations, while the IVFE model is run on the non-censored subsample. Notwithstanding these two differences, we observe

[17] The *econometric* identification of the main equation can rely on the non-linearity of the auxiliary regression. However, for *economic* identification, different instruments are required (Wooldridge, 2002).

[18] The tests for weak instruments are reported only in Table 6, for the full sample IVFE estimates in Table 7. As for the specifications presented in Tables 8-11, the tests are not reported but available upon request.

similar results in the two specifications. This suggests that we are correctly accounting for censoring in the dependent variable of the Tobit and for its possible bias in the IVFE estimates.[19]

Table 7 shows the second step of the two stage least squares estimation. We compare the results with an OLS specification which corrects for selection, but not for the endogeneity of *Sold Seats*. Notice that the IV approach yields a lower coefficient for *Sold Seats*: an extra sold seat induces a 3.10% increase in fares if we do not correct for endogeneity, while the IVFE specification indicates a 2.56% increase.[20] The magnitude of the *Sold Seats* coefficient suggests that a considerable proportion of a flight's fare dispersion can be attributed to a load factor effect. Indeed, if we apply a 2.56% change rate per seat to the mean value of *Fare1* (£65.35) when *Sold Seats* changes from its mean sample value (23) to either its maximum (49) or minimum value (1), we obtain a prediction for the fare of about £122.94 and £37.47 , respectively.[21] These results provide strong support to the capacity-driven approach and therefore shed empirical light on the relevance of the theoretical set-up developed in Dana (1999a). In addition, this contrasts with the conclusions in Puller et al. (2009) where fares appear to be insensitive to a flight's occupancy rate.

The temporal profile in the two estimations are also quite different. The coefficients of the "Booking Days" dummies in the IV specification suggest a steeper (relative to the OLS) increase in prices in the last days before the flight. Relative to the base case of prices posted 70 days from departure, for fares posted 7, 4 and 1 days we record percentage increases of about 24%, 58% and 114%, respectively.[22] The evidence supports an important role for APD, at least as far as the booking period comprising the 2 weeks before departure is concerned. Notice however that the IVFE estimates do not provide any support for the U-shaped temporal profile. This may be due to a compositional effect in the full sample. Indeed, the U-shaped profile is found in more homogenous subsamples (see below), which suggests that it may be part of a standard pricing policy only in specific circumstances.

All in all, after controlling for endogeneity (i.e., after purging the estimates from the effect due to the *discretional* analyst intervention), we obtain strong evidence supporting a

[19] As a further check, we have run the IVFE regression by omitting the Tobit residuals: the first stage estimates - not reported to save space but available on request - are considerably different from the ones reported in Table 6.

[20] The upward bias of the OLS coefficient for *Sold Seats* comes from the fact that it includes both the direct impact of *Sold Seats* due to the airline pricing policy and the indirect impact due to an analyst intervention, which is positively related with *Sold Seats*. See fn.15.

[21] Note that the $5^{th}$ and the $95^{th}$ percentile values of *Fare1* are, respectively, 13.99 and 149.99.

[22] E.g., for the "Booking Day7" dummy, the percent change relative to the base period is found by using $exp(0.2171) - 1 \cong 0.24$.

crucial role for both capacity and time-driven approaches in driving the airline's standard (i.e., routinary) pricing policy. It is important to note, however, that the results in Table 7 are obtained using data from an heterogenous sample featuring 82 routes that differ in terms of market structure and length. In turn, each route includes flights which vary by departure time, day of the week, seasonal period etc. Indeed, the summary statistics reported in Tables 1 and 2 indicate that the pricing policy of the airline could vary across routes (e.g. substantial differences in terms of mean and maximum value of *Fare1* and *Top Fare*); furthermore, Tables 4 and 5 suggest additional complexity in the pricing behaviour that is compatible with variability at the flight level.

In the remainder of the paper, we study how the *average* pricing policy depicted in Table 7 changes as we take these sources of heterogeneity into account. The evidence we present also reveals the robustness of the importance of capacity and temporal effects in the determination of the airline's standard pricing policy. In subsection 6.1 we consider flights that filled up either early or late, but both realized a final high load factor. In subsection 6.2, we further investigate how the pricing policy adjusts to different flights characteristics, i.e., route distance, time of day and season. Finally, in subsection 6.3 we analyze the pricing policy differences driven by market structure.

### 6.1 Early and late full flights

In this subsection we study whether differences in the pricing profile arise when we compare flights that reached a sufficiently high load factor early rather than late. This comparison may capture the airline's different expectations about the underlying demand conditions, which, in turn, may induce the application of different standard pricing policies.

We classify flights as either being full early or not. In Table 8, we consider the former category, i.e., flights that had less than 30 seats available at least 14 days prior to departure, while flights that did not fill up early are considered in Table 9.[23] Most importantly, both Tables report the estimates from samples that realized an equivalent level of the final load factor. Thus, in each column, we consider flights whose lowest recorded value of available seats was strictly below, respectively, the value of 30 or 15. Therefore, an alternative interpretation of the samples in Tables 8 and 9 is that they differ only because they are made up of flights that sold a larger proportion of seats prior to, or within, the last two weeks before departure, respectively.

[23]Results, not reported but available on request, are robust when we consider flights with less that 30 seats available 28 or 21 days before departure.

The coefficients of *Sold Seats* is about $0.14\% - 0.3\%$ larger for flights that filled up at an early stage. Such differences do not however impinge on the conclusion that routinely fares increase as the plane becomes fuller. As far as the temporal effects are concerned, large increases in the last week prior to departure are also observed in all specifications. More interestingly, and in line with the descriptive evidence in Table 3 and the OLS estimates in Table 7, over the two months preceding departure the temporal effects appear to be U-shaped in the case of flights that filled up early, which reports significant drops in fares between 42 and 14 days from departure. This contrasts with the coefficients in Table 9, where the only significant temporal effects are those for late booking days. Overall, while the analysis in this section confirms the important role played by both the capacity and the temporal dimensions, it also highlights how the airline may vary their combined structure depending on some of the underlying characteristics at the flight or route level.

### 6.2 Time of the day, seasonality and route length

Table 10 offers further insights into the nature of the effects of in-flight occupancy and booking days on fares. First, we use the samples of morning and evening flights, since the departure time is likely to vary with the passengers' travel motivation and their flight's convenience.[24] The coefficient of *Sold Seats* is found to be larger in the evening sample; as Figures 4 and 5 clearly suggest, this is likely due to the fact that evening flights include a larger proportions of observations with both a higher number of sold seats and, hence, such a larger demand is met with a pricing policy designed to manage a higher shadow cost of capacity.

Second, the last two columns of Table 10 consider the two samples of flights operated in the Winter (Nov-Mar) and the Summer (Apr-Oct) periods. Because Ryanair serves many Mediterranean destination whose demand is obviously larger in the Summer, the higher coefficient for *Sold Seats* is again due to the adoption of a pricing policy which weighs capacity issues more heavily. Interestingly, the U-shaped intertemporal profile is only found in the Summer flights, possibly because larger demand is also accompanied by larger customers' heterogeneity. In such a situation, the airline faces a stronger incentive to adopt a U-shaped intertemporal profile to attract price-sensitive consumers with high demand uncertainty who would not book their flights too in advance.

Finally, in Table 11 we investigate whether the different cost structure that characterizes

[24] Morning flights are from 6am to 11am; evening ones from 4pm to 10.15pm. We thus exclude late morning and afternoon flights.

routes of varying length may affect the carrier's pricing approach. Indeed, short haul flights are subject to a greater incidence of fixed costs on total costs, due to the greater fuel consumption during take-off and landing. In-flight occupancy appears to play a similar role in both types of routes; however, the intertemporal effect is U-shaped only in short-haul flights.

### 6.3 Market structure

As previously discussed, Dana (1999a) characterizes an equilibrium in price distributions where higher prices are associated with higher occupancy rates. An important prediction of Dana's model is that the price distribution's domain expands as competition increases: unlike a monopolist, competitive firms pass through all of their cost increases and therefore they should exhibit more intrafirm price dispersion. However, Gerardi and Shapiro (2009) argue that in less competitive markets it may be easier to implement price discrimination tactics: their estimates support the hypothesis that overall price dispersion should decrease with competition. By focussing on particular forms of online price discrimination strategies by European LCAs, Bachis and Piga (2011) also show that such strategies are more likely found in less competitive markets.

To study how the coefficient of *Sold Seats* changes with market structure, we have distinguished between markets with low and high competition, where a market is identified at both the route and the city-pair level.[25] In lowly competitive markets, Ryanair is at most a duopolist at either the route or the city-pair level, while in highly competitive ones travellers may substitute Ryanair's services with those of at least two or more of its direct competitors in that route/city-pair.[26]

Table 11 reports the estimates from the low and the high competition subsamples, and shows that the coefficient of *Sold Seats* is larger in markets with low competition. Thus, when travellers find it more difficult to substitute Ryanair's services with those of competitors, Ryanair appears to adopt a pricing policy where a larger proportion of seats are assigned higher fares, and therefore the gradient of *Sold Seats* is on average, steeper than those in more competitive markets. Our findings therefore suggest that competitive pressure may prevent Ryanair from extracting more surplus from those travellers' segments whose demand is more inelastic because their need to travel on a specific flight is revealed

[25] A city-pair defines the airline market for two cities (e.g., London and Milan). It generally includes more than one route, each identified by a unique airport-pair combination (e.g., London Heathrow/Milan Malpensa and London Stansted/Milan Linate).

[26] Data on market structure are obtained from the UK Civil Aviation Authority.

only a few days before its departure, when therefore fewer seats remain available.

As far as the intertemporal profile is concerned, the estimates confirm the previous finding of significant fare increases during the last ten days of booking. However, market structure does not seem to be an explanatory factor for the presence of a U-shaped relationship. More precisely, flights exhibiting significant price drops six to three weeks before departure do not feature prominently or exclusively in either subsamples in Table 11.

# 7 Conclusions

This study builds on the extensive and well developed theoretical literature on airline pricing and sheds new empirical light on two of its predictions. It thus fills a gap in the literature, since there are very few studies that managed to overcome the scarcity of appropriate data. To do so, we rely on data obtained from the website of Ryanair, whose business model very closely aligns with the assumptions used in the theoretical literature.

Both the descriptive and the econometric evidence lend strong support to the hypothesis of fares becoming higher as fewer seats remain available on a flight. On average, we find that each extra sold seat induces a 2.56% increase in a flight's fare. This drives to the conclusion that the capacity-driven motivation is an important determinant of airline pricing. This inference appears to be robust to the criticism that capacity-driven theories are derived assuming a perfect commitment by the airlines not to revise their pricing curve as they gather new information about a flight's actual demand.

The study also reveals novel evidence regarding the temporal profile of fares. All econometric specifications indicate a sharp increase in fares in the last ten days prior to departure, which is consistent with the idea that late bookers are less willing to substitute a flight with another departing on a different time or date. This leads to the conclusion that Ryanair's pricing policy appears to be designed to include late increases in fares regardless of the actual load factor realization. That is, higher late fares are part of an ex-ante YM decision by the airline.

More importantly, the descriptive evidence points to a more complex, U-shaped temporal profile, where early bookers (those booking at least 49 days prior to departure) appear to pay a higher fare than those booking between 35 and 14 days from departure. The econometric evidence captures a similar effect only for those flights that filled up relatively quickly. Although the empirical evidence shows that the advance purchase discount strategy (on which the theoretical literature has largely focussed) is often complemented (and

preceded) with a clearance sale one (Möller and Watanabe, 2010). This appears to be the case especially in periods of high demand, i.e., during the Summer season. Overall, the evidence indicates that a monotonic temporal profile, which is typical of the APD approach, is not necessarily observed after capacity utilization is controlled for.

To conclude, in addition to providing a test for two strands of literature on airline pricing, this paper provides the foundation for an investigation of the theoretical prediction, reported in Dana (1999a), that fare dispersion is expected to be larger in competitive markets. Although this issue has been widely studied, the prediction has received mixed support when dispersion is measured at the route-level (Borenstein and Rose, 1994; Gerardi and Shapiro, 2009). The flight-level analysis in this study supports the findings in Gerardi and Shapiro (2009) that the lack of competitive pressure allows Ryanair to extract more surplus from consumers with more inelastic demand. This is revealed in our estimates by a steeper pricing curve in less competitive markets, implying that the last seats are sold at higher fares. Because the last seats on a flight are generally purchased by travellers with a highly inelastic demand, this study suggests that the lack of competitive pressure appears to facilitate the implementation of price discrimination tactics in European routes.

## References


Alderighi, M. (2010). Fare dispersions in airline markets: A quantitative assessment of theoretical explanations. *Journal of Air Transport Management*, 16:144–50.

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

Borenstein, S. (1989). Hubs and High Fares: Airport Dominance and Market Power in the U.S. Airline Industry. *Rand Journal of Economics*, 20:344–365.

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

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

Dana, J. D. (1999b). Using yield management to shift demand when peak time is unknown. *Rand Journal of Economics*, 30(4):456–474.

Escobari, D. and Gan, L. (2007). Price Dispersion under Costly Capacity and Demand Uncertainty. N. 13075, NBER Working Paper Series.

European Commission (2007). Case No COMP/M. 4439 – Ryanair / Aer Lingus. Regulation (EC) No 139/2004 - Merger Procedure, Directorate General for Competition, Brussels, June.

Evans, W. N. and Kessides, I. N. (1993). Localized Market Power in the U.S. Airline Industry. *Review of Economics and Statistics*, 75:66–75.

Evans, W. N. and Kessides, I. N. (1994). Living by the "Golden Rule": Multimarket Contact in the U.S. Airline Industry. *The Quarterly Journal of Economics*, 109:341–366.

Gaggero, A. and Piga, C. A. (2011). Airline Market Power and Intertemporal Price Dispersion. *Journal of Industrial Economics*, 26(6):655–677.

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

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

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

Kim, E. H. and Singal, V. (1993). Mergers and Market Power: Evidence from the Airline Industry. *American Economic Review*, 83:549–569.

Lederman, M. (2008). Are Frequent-Flyer Programs a Cause of the "Hub-Premium"? *Journal of Economics and Management Strategy*, 17:35–66.

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

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.

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.

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

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

Table 1: Distribution of *TopFare*, by route.

| Route | Max | Median | Mode | Mean | S.D. | Route | Max | Median | Mode | Mean | S.D. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| BLK-DUB | 149.99 | 149.99 | 149.99 | 143.5 | 25.5 | STN-CCF | 199.99 | 169.99 | 169.99 | 162.4 | 26.6 |
| BHX-DUB | 159.99 | 149.99 | 149.99 | 148.1 | 21.8 | STN-NOC | 189.99 | 169.99 | 169.99 | 157.5 | 27.6 |
| BRS-DUB | 159.99 | 149.99 | 149.99 | 146.9 | 19.7 | STN-DUB | 149.99 | 139.99 | 139.99 | 134.8 | 22.8 |
| CWL-DUB | 159.99 | 149.99 | 149.99 | 144.1 | 23.2 | STN-EIN | 139.99 | 139.99 | 139.99 | 134.9 | 21.1 |
| EDI-DUB | 169.99 | 149.99 | 149.99 | 144.9 | 22.5 | STN-FRL | 199.99 | 199.99 | 199.99 | 185.8 | 37.4 |
| LGW-DUB | 149.99 | 139.99 | 139.99 | 136.9 | 16.9 | STN-GOA | 189.99 | 169.99 | 169.99 | 163.6 | 30.7 |
| LBA-DUB | 169.99 | 149.99 | 149.99 | 146.9 | 21.7 | STN-GRO | 199.99 | 189.99 | 189.99 | 168.8 | 46.8 |
| LPL-DUB | 169.99 | 159.99 | 159.99 | 156.8 | 20.9 | STN-GSE | 189.99 | 179.99 | 179.99 | 170.7 | 35.8 |
| LTN-BGY | 249.99 | 159.99 | 159.99 | 159.9 | 44.9 | STN-HHN | 159.99 | 145.99 | 145.99 | 139.9 | 23.9 |
| LTN-DUB | 149.99 | 139.99 | 139.99 | 137.0 | 17.3 | STN-HAU | 169.99 | 169.99 | 169.99 | 157.4 | 39.0 |
| MAN-DUB | 189.99 | 179.99 | 179.99 | 176.3 | 24.4 | STN-LBC | 139.99 | 139.99 | 139.99 | 130.3 | 31.8 |
| MME-DUB | 159.99 | 149.99 | 149.99 | 145.1 | 23.0 | STN-MMX | 169.99 | 159.99 | 159.99 | 147.1 | 39.7 |
| NCL-DUB | 179.99 | 169.99 | 169.99 | 166.1 | 24.5 | STN-MPL | 199.99 | 189.99 | 189.99 | 172.7 | 30.6 |
| PIK-BVA | 159.99 | 139.99 | 139.99 | 140.6 | 19.4 | STN-MJV | 199.99 | 179.99 | 179.99 | 150.6 | 43.5 |
| PIK-CRL | 139.99 | 129.99 | 129.99 | 128.4 | 22.1 | STN-AOI | 179.99 | 149.99 | 149.99 | 151.1 | 24.7 |
| PIK-DUB | 159.99 | 149.99 | 149.99 | 145.4 | 23.7 | STN-VBS | 179.99 | 129.99 | 129.99 | 133.9 | |
| PIK-GRO | 199.99 | 189.99 | 189.99 | 177.2 | 37.6 | STN-VLL | 199.99 | 189.99 | 189.99 | 171.5 | 45.9 |
| PIK-NYO | 159.99 | 139.99 | 139.99 | 125.9 | 37.3 | STN-PSA | 209.99 | 189.99 | 189.99 | 181.3 | 25.9 |
| STN-EGC | 189.99 | 179.99 | 179.99 | 175.9 | 25.2 | STN-PIK | 149.99 | 129.99 | 129.99 | 119.6 | 28.9 |
| STN-SXF | 149.99 | 149.99 | 149.99 | 140.9 | 28.4 | STN-CIA | 209.99 | 199.99 | 199.99 | 187.0 | 44.4 |
| STN-LRH | 189.99 | 169.99 | 169.99 | 160.9 | 30.1 | STN-REU | 199.99 | 189.99 | 189.99 | 163.0 | 52.7 |
| STN-LIG | 189.99 | 179.99 | 179.99 | 174.2 | 27.3 | STN-PUF | 189.99 | 179.99 | 179.99 | 156.2 | 37.9 |
| STN-PIS | 189.99 | 179.99 | 179.99 | 172.5 | 32.0 | STN-PGF | 199.99 | 169.99 | 169.99 | 164.8 | 26.9 |

Note: The Table includes a selection of routes with more than 1000 observations in our estimation sample of flights with less than 50 seats available.

Table 2: Distribution of $Fare1$, by route and flight occupancy.

| | Available Seats | | | | | Available Seats | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Less than 50 | | 50 or more | | | Less than 50 | | 50 or more | |
| Route | Max | Mean | Max | Mean | Route | Max | Mean | Max | Mean |
| BLK-DUB | 149.99 | 48.00 | 119.99 | 10.29 | STN-CCF | 199.99 | 69.25 | 149.90 | 26.38 |
| BHX-DUB | 159.99 | 62.95 | 109.90 | 18.23 | STN-NOC | 189.99 | 75.46 | 139.99 | 37.14 |
| BRS-DUB | 159.99 | 62.87 | 119.99 | 18.63 | STN-DUB | 144.99 | 49.52 | 119.99 | 15.31 |
| CWL-DUB | 159.99 | 56.01 | 109.90 | 22.26 | STN-EIN | 139.99 | 60.54 | 99.99 | 9.46 |
| EDI-DUB | 169.99 | 67.06 | 129.90 | 16.72 | STN-FRL | 199.99 | 71.64 | 139.99 | 16.60 |
| LGW-DUB | 144.99 | 55.89 | 104.99 | 19.09 | STN-GOA | 189.99 | 72.79 | 149.90 | 14.78 |
| LBA-DUB | 169.99 | 56.93 | 149.99 | 16.48 | STN-GRO | 199.99 | 74.26 | 159.90 | 22.72 |
| LPL-DUB | 169.99 | 60.81 | 139.99 | 13.55 | STN-GSE | 189.99 | 91.54 | 179.90 | 24.51 |
| LTN-BGY | 249.99 | 78.63 | 179.99 | 22.53 | STN-HHN | 159.99 | 59.37 | 85.99 | 12.80 |
| LTN-DUB | 139.99 | 55.55 | 99.99 | 14.23 | STN-HAU | 169.99 | 53.42 | 169.99 | 16.03 |
| MAN-DUB | 189.99 | 61.08 | 149.90 | 10.82 | STN-LBC | 139.99 | 62.88 | 139.99 | 13.55 |
| MME-DUB | 159.99 | 48.89 | 119.99 | 13.95 | STN-MMX | 169.90 | 66.82 | 129.99 | 16.16 |
| NCL-DUB | 179.99 | 60.83 | 139.99 | 20.08 | STN-MPL | 199.99 | 77.35 | 159.90 | 25.55 |
| PIK-BVA | 159.99 | 55.07 | 109.99 | 22.59 | STN-MJV | 199.99 | 87.82 | 159.90 | 54.73 |
| PIK-CRL | 139.99 | 49.66 | 89.99 | 17.15 | STN-AOI | 179.99 | 72.73 | 139.90 | 25.53 |
| PIK-DUB | 149.99 | 61.12 | 149.99 | 11.31 | STN-VBS | 179.90 | 58.25 | 129.99 | 18.55 |
| PIK-GRO | 189.99 | 79.75 | 189.90 | 50.55 | STN-VLL | 189.99 | 73.80 | 159.99 | 19.09 |
| PIK-NYO | 159.99 | 69.73 | 79.99 | 17.29 | STN-PSA | 189.99 | 90.56 | 169.99 | 32.28 |
| STN-EGC | 189.99 | 81.47 | 179.90 | 31.77 | STN-PIK | 149.99 | 48.77 | 89.99 | 9.29 |
| STN-SXF | 149.99 | 62.63 | 149.99 | 18.99 | STN-CIA | 199.99 | 66.83 | 139.90 | 19.98 |
| STN-LRH | 189.99 | 69.58 | 109.99 | 29.47 | STN-REU | 199.99 | 67.04 | 139.99 | 20.99 |
| STN-LIG | 189.99 | 75.64 | 149.99 | 25.81 | STN-PUF | 179.99 | 65.43 | 129.90 | 23.11 |
| STN-PIS | 189.99 | 65.72 | 149.99 | 24.23 | STN-PGF | 199.99 | 75.46 | 159.99 | 25.75 |

Note: The Table includes a selection of routes with more than 1000 observations in our estimation sample.

Table 3: Mean *Fare1* by available seats and booking day

| Booking | Available Seats | | | | | | |
|---|---|---|---|---|---|---|---|
| Day | 1-9 | 10-19 | 20-29 | 30-39 | 40-49 | ≥50 | Total |
| 1 | 125.5 | 95.4 | 83.7 | 78 | 74.2 | 64.3 | 84.5 |
| 4 | 114.3 | 75.3 | 57.8 | 49.4 | 43.6 | 36.1 | 57.2 |
| 7 | 110.9 | 69.5 | 49.1 | 37.9 | 31.1 | 19.4 | 40.6 |
| 10 | 109.3 | 68.8 | 48.2 | 37.7 | 31.3 | 19.7 | 36.3 |
| 14 | 106.4 | 72.5 | 48.1 | 35.9 | 28.0 | 13.5 | 27.3 |
| 21 | 116.4 | 82.1 | 56.2 | 41.8 | 32.7 | 15.4 | 24.1 |
| 28 | 130.9 | 92.9 | 64.3 | 47.0 | 36.9 | 16.5 | 21.6 |
| 35 | 135.6 | 97.6 | 71.3 | 53.0 | 41.9 | 17.3 | 20.4 |
| 42 | 128.0 | 97.9 | 74.9 | 57.1 | 49.4 | 18.0 | 20.0 |
| 49-70 | 124.5 | 107.4 | 88.6 | 66.1 | 54.9 | 18.4 | 19.3 |
| Total | 116.9 | 78.6 | 58.8 | 47.1 | 39.5 | 20.0 | 31.1 |

Note: *Fare1* is the fare obtained from a query for one seat.

Table 4: Percentage mean of observations with a price drop in *Fare1* of at least £5.00 between two consecutive booking periods.

| Booking | Available seats | | | | | | |
|---|---|---|---|---|---|---|---|
| Period | 1-9 | 10-19 | 20-29 | 30-39 | 40-49 | Total | N |
| 4-1 | 0.04 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 26,632 |
| 7-4 | 0.05 | 0.05 | 0.05 | 0.04 | 0.04 | 0.05 | 26,281 |
| 10-7 | 0.07 | 0.09 | 0.09 | 0.10 | 0.09 | 0.09 | 24,904 |
| 14-10 | 0.09 | 0.11 | 0.10 | 0.11 | 0.10 | 0.10 | 22,340 |
| 21-14 | 0.14 | 0.15 | 0.16 | 0.19 | 0.21 | 0.18 | 18,382 |
| 28-21 | 0.09 | 0.13 | 0.13 | 0.16 | 0.19 | 0.16 | 11,899 |
| 35-28 | 0.10 | 0.11 | 0.13 | 0.15 | 0.18 | 0.15 | 6,717 |
| 42-35 | 0.06 | 0.06 | 0.12 | 0.15 | 0.14 | 0.13 | 3,691 |
| 49-42 | 0.09 | 0.14 | 0.10 | 0.14 | 0.14 | 0.13 | 2,107 |
| 63-49 | 0.02 | 0.10 | 0.10 | 0.10 | 0.13 | 0.11 | 2,420 |
| Total | 0.06 | 0.08 | 0.09 | 0.11 | 0.13 | 0.09 | |
| N | 22,434 | 30,147 | 30,973 | 31,363 | 30,456 | | 145,373 |

Note: *Fare1* is the fare obtained from a query for one seat. The price drop is calculated conditional on the number of available seats being less than 50 and non-increasing between two consecutive periods.

Table 5: Fare changes between two consecutive booking periods when flight occupancy remains unchanged. (Percentage values), by flight characteristics.

| | | Fare Change | | | | | |
|---|---|---|---|---|---|---|---|
| | | Large Drop | Moderate Drop | No Change | Moderate Increase | Large Increase | N |
| Average Change in £ | | -46.21 | -12.45 | 0 | 14.27 | 49.78 | |
| Available Seats > 20 | (% row) | 3.94 | 6.45 | 64.98 | 13.09 | 11.54 | 4,141 |
| Available Seats <= 20 | (% row) | 3.63 | 4.13 | 78.19 | 5.68 | 8.36 | 6,301 |
| Booking Day > 14 | (% row) | 5.49 | 8.89 | 74.56 | 6.61 | 4.45 | 1,529 |
| Booking Day <= 14 | (% row) | 3.46 | 4.39 | 72.68 | 8.96 | 10.51 | 8,913 |
| Winter | (% row) | 5.37 | 5.50 | 70.25 | 8.88 | 10.00 | 3,129 |
| Summer | (% row) | 3.06 | 4.85 | 74.11 | 8.51 | 9.46 | 7,313 |
| High Competition | (% row) | 2.88 | 4.83 | 74.89 | 7.93 | 9.47 | 6,496 |
| Low Competition | (% row) | 5.20 | 5.40 | 69.77 | 9.76 | 9.88 | 3,946 |
| N | (% row) | 3.75 | 5.05 | 72.96 | 8.62 | 9.62 | |
| N | | 392 | 527 | 7,618 | 900 | 1,005 | 10,442 |

Note: Large (Moderate) increases/drops refer to changes strictly greater than (smaller than) £20.0 in absolute terms.

Table 6: Tobit and First Stage estimates. Dependent Variable: *Sold Seats*

| | Tobit | | First stage | |
|---|---|---|---|---|
| L-FW mean $\Delta$ *Fare* | 0.180 | (0.004)*** | 0.163 | (0.000)*** |
| Booking Day1 | 69.178 | (0.678)*** | 66.025 | (0.122)*** |
| Booking Day4 | 62.310 | (0.691)*** | 59.202 | (0.119)*** |
| Booking Day7 | 55.925 | (0.700)*** | 53.201 | (0.12)*** |
| Booking Day10 | 50.632 | (0.699)*** | 47.730 | (0.119)*** |
| Booking Day14 | 43.823 | (0.704)*** | 41.380 | (0.117)*** |
| Booking Day21 | 33.374 | (0.692)*** | 31.282 | (0.115)*** |
| Booking Day28 | 24.537 | (0.684)*** | 22.842 | (0.113)*** |
| Booking Day35 | 16.910 | (0.671)*** | 15.603 | (0.111)*** |
| Booking Day42 | 10.702 | (0.661)*** | 9.822 | (0.103)*** |
| Booking Day49 | 6.218 | (0.658)*** | 5.751 | (0.098)*** |
| Booking Day56 | 3.552 | (0.653)*** | 3.254 | (0.091)*** |
| Booking Day63 | 3.088 | (0.578)*** | 2.830 | (0.087)*** |
| N. UK airports serving arrival | -1.255 | (0.195)*** | | |
| Tobit residual | | | 0.908 | (0.002)*** |
| Booking Day is in Holiday period | | | -0.200 | (0.028)*** |
| Constant | 112.274 | (5.419)*** | | |
| DUMMIES: | | | | |
| Month booking | Yes | | Yes | |
| Week | Yes | | No | |
| Route | Yes | | No | |
| DOW Booking | Yes | | No | |
| Time Departure | Yes | | No | |
| Number of obs. | 547,543 | | 113,535 | |
| Pseudo $R^2$ | 0.1663 | | 0.9674 | |
| Test excluded instruments: | | | $F(2, 4652) = 67344.48^{***}$ | |
| Underidentification K-P LM Test | | | $\chi^2(2)$=1676.50*** | |
| Anderson-Rubin Wald test | | | $F(2, 4652)$= 488.43*** | |
| Anderson-Rubin Wald test | | | $\chi^2(2)$=977.29*** | |
| Stock-Wright LM S statistic | | | $\chi^2(2)$=580.76*** | |

Note: "*Top Fare*" denotes the fare obtained by using the highest possible number of seats in a query. $\Delta Fare = Top\ Fare - Fare1$, where *Fare1* is the fare for one seat. The means are obtained by taking their 7 and 14 days lagged (L) and forward (FW) values. Coefficients *** statistically significant at 1%, ** at 5% and * at 10%. K-P=Kleibergen-Paap.

Table 7: Pricing equation results using the full sample and different estimation methods. Dependent Variable: *LnFare1*

| | IVFE | | FE-OLS | |
|---|---|---|---|---|
| Sold seats | 0.0256 | (0.001)*** | 0.0309 | (0.001)*** |
| Booking Day1 | 0.7630 | (0.051)*** | 0.4578 | (0.049)*** |
| Booking Day4 | 0.4577 | (0.05)*** | 0.1716 | (0.047)*** |
| Booking Day7 | 0.2171 | (0.047)*** | -0.0451 | (0.045) |
| Booking Day10 | 0.1642 | (0.046)*** | -0.0712 | (0.044) |
| Booking Day14 | -0.0096 | (0.043) | -0.2143 | (0.042)*** |
| Booking Day21 | -0.0191 | (0.04) | -0.1716 | (0.039)*** |
| Booking Day28 | 0.0020 | (0.037) | -0.1079 | (0.037)*** |
| Booking Day35 | 0.0085 | (0.036) | -0.0652 | (0.037)* |
| Booking Day42 | -0.0162 | (0.037) | -0.0614 | (0.036)* |
| Booking Day48 | -0.0108 | (0.036) | -0.0374 | (0.036) |
| Booking Day56 | -0.0068 | (0.037) | -0.0211 | (0.038) |
| Booking Day63 | 0.0081 | (0.038) | -0.0039 | (0.035) |
| Tobit residual | 0.0046 | (0.000)*** | 0.0012 | (0.000)*** |
| DUMMIES: | | | | |
| Month booking | YES | | YES | |
| Number of obs. | 113,535 | | 113,535 | |
| Centered $R^2$ | 0.5688 | | 0.5697 | |
| Excluded instruments: | 2 | | | |
| Underidentification K-P LM Test | $\chi^2(2) = 1676.50^{***}$ | | | |
| Hansen J statistic | $\chi^2(1) = 1.962$ | | | |

Note: *Fare1* is the fare obtained from a query for one seat. Bootstrap Standard Errors (SE) are reported in parenthesis, clustered by route and week. 250 repetitions. Coefficients *** statistically significant at 1%, ** at 5% and * at 10%. K-P=Kleibergen-Paap.

Table 8: Pricing equation results in flights that filled up early, i.e., that had less than 30 seats available 14 days before their departure. Dependent Variable: *LnFare1*

| | Less than 30 final left seats | | Less than 15 final left seats | |
|---|---|---|---|---|
| Sold seats | 0.0264 | (0.001)*** | 0.0273 | (0.001)*** |
| Booking Day1 | 0.5405 | (0.067)*** | 0.4361 | (0.074)*** |
| Booking Day4 | 0.3601 | (0.063)*** | 0.2664 | (0.07)*** |
| Booking Day7 | 0.2202 | (0.06)*** | 0.1239 | (0.067)* |
| Booking Day10 | 0.1021 | (0.058)* | 0.0015 | (0.064) |
| Booking Day14 | -0.0714 | (0.055) | -0.1753 | (0.06)*** |
| Booking Day21 | -0.1408 | (0.048)*** | -0.2262 | (0.053)*** |
| Booking Day28 | -0.0998 | (0.042)** | -0.1707 | (0.049)*** |
| Booking Day35 | -0.0810 | (0.038)** | -0.1412 | (0.044)*** |
| Booking Day42 | -0.0801 | (0.038)** | -0.1262 | (0.041)*** |
| Booking Day48 | -0.0310 | (0.037) | -0.0774 | (0.039)** |
| Booking Day56 | -0.0255 | (0.039) | -0.0573 | (0.041) |
| Booking Day63 | -0.0101 | (0.038) | -0.0361 | (0.049) |
| Tobit residual | 0.0017 | (0.001)*** | 0.0017 | (0.001)** |
| DUMMIES: | | | | |
| Month booking | Yes | | Yes | |
| Number of obs. | 42,814 | | 36,222 | |
| Centered $R^2$ | 0.5468 | | 0.5601 | |
| Excluded instruments: | 2 | | 2 | |
| Underidentification K-P LM Test | $\chi^2(2)$ =986.96*** | | $\chi^2(2)$=907.23** | |
| Hansen J statistic | $\chi^2(2)$=0.753 | | $\chi^2(2)$=0.781 | |

Note: The two samples are built by selecting all the observations for those flights whose lowest number of available seats was, respectively, less than 30 and less than 15. In both samples, flights had less than 30 seats available 14 days prior to departure. *Fare1* is the fare obtained from a query for one seat. Bootstrap Standard Errors (SE) are reported in parenthesis, clustered by route and week. 250 repetitions. Coefficients *** statistically significant at 1%, ** at 5% and * at 10%. K-P=Kleibergen-Paap.

Table 9: Pricing equation results in flights that had at least 30 seats or more available 14 days before their departure. Dependent Variable: *LnFare1*

| | Less than 30 final left seats | | Less than 15 final left seats | |
|---|---|---|---|---|
| Sold seats | 0.0250 | (0.001)*** | 0.0243 | (0.001)*** |
| Booking Day1 | 0.6936 | (0.126)*** | 0.6224 | (0.183)*** |
| Booking Day4 | 0.3736 | (0.124)*** | 0.3138 | (0.181)*** |
| Booking Day7 | 0.0926 | (0.122) | 0.0016 | (0.179) |
| Booking Day10 | 0.0577 | (0.122) | -0.0567 | (0.175) |
| Booking Day14 | -0.1270 | (0.12) | -0.1380 | (0.173) |
| Booking Day21 | 0.0059 | (0.116) | -0.0071 | (0.167) |
| Booking Day28 | 0.0730 | (0.117) | 0.0669 | (0.167) |
| Booking Day35 | 0.1397 | (0.112) | 0.1517 | (0.165) |
| Booking Day42 | 0.1026 | (0.114) | 0.1254 | (0.169) |
| Booking Day48 | 0.0105 | (0.124) | 0.1308 | (0.173) |
| Booking Day56 | 0.0211 | (0.133) | 0.1408 | (0.172) |
| Booking Day63 | 0.0570 | (0.122) | 0.2086 | (0.166) |
| Tobit residual | 0.0063 | (0.001)*** | 0.0047 | (0.001)*** |
| DUMMIES: | | | | |
| Month booking | Yes | | Yes | |
| Number of obs | 52,778 | | 25,865 | |
| Centered $R^2$ | 0.5822 | | 0.6494 | |
| Underidentification K-P LM Test | $\chi^2(2)$ =1435.2*** | | $\chi^2(2)$=1012.2** | |
| Hansen J statistic | $\chi^2(2)$=0.001 | | $\chi^2(2)$=0.110 | |

Note: The two samples are built by selecting all the observations for those flights whose lowest number of available seats was, respectively, *i*) less than 30; *ii*) less than 15. In both samples, flights had more than 30 seats available 14 days prior to departure. *Fare1* is the fare obtained from a query for one seat. Bootstrap Standard Errors (SE) are reported in parenthesis, clustered by route and week. 250 repetitions. Coefficients *** statistically significant at 1%, ** at 5% and * at 10%. K-P=Kleibergen-Paap.

Table 10: Pricing equation results by time of the day and season. Dependent Variable: *LnFare1*

| | Morning | | Evening | | Winter | | Summer | |
|---|---|---|---|---|---|---|---|---|
| Sold seats | 0.0250 | (0.001)*** | 0.0274 | (0.001)*** | 0.0236 | (0.001)*** | 0.0253 | (0.001)*** |
| Booking Day1 | 0.9558 | (0.095)*** | 0.7716 | (0.09)*** | 1.0537 | (0.132)*** | 0.5724 | (0.061)*** |
| Booking Day4 | 0.6410 | (0.092)*** | 0.4931 | (0.087)*** | 0.6924 | (0.128)*** | 0.2942 | (0.058)*** |
| Booking Day7 | 0.3549 | (0.089)*** | 0.2959 | (0.083)*** | 0.374 | (0.125)*** | 0.0863 | (0.056) |
| Booking Day10 | 0.2924 | (0.085)*** | 0.2454 | (0.079)*** | 0.2696 | (0.123)** | 0.0538 | (0.053) |
| Booking Day14 | 0.0888 | (0.081) | 0.0995 | (0.073) | 0.0756 | (0.118) | -0.1198 | (0.050)** |
| Booking Day21 | 0.0513 | (0.075) | 0.0706 | (0.065) | -0.0232 | (0.112) | -0.1113 | (0.046)** |
| Booking Day28 | 0.0605 | (0.07) | 0.0636 | (0.057) | 0.0006 | (0.107) | -0.1008 | (0.041)** |
| Booking Day35 | 0.1003 | (0.069) | 0.0532 | (0.055) | 0.0059 | (0.107) | -0.0949 | (0.038)** |
| Booking Day42 | 0.0671 | (0.067) | -0.0297 | (0.054) | -0.1101 | (0.103) | -0.0868 | (0.038)** |
| Booking Day48 | 0.0390 | (0.069) | 0.0413 | (0.048) | -0.173 | (0.108) | -0.0653 | (0.038)* |
| Booking Day56 | 0.0657 | (0.067) | -0.0017 | (0.055) | -0.2762 | (0.109)** | -0.029 | (0.033) |
| Booking Day63 | 0.0584 | (0.072) | 0.0178 | (0.048) | -0.104 | (0.114) | -0.0055 | (0.038) |
| Tobit residual | 0.0063 | (0.001)*** | 0.0034 | (0.001)*** | 0.0236 | (0.001)*** | 0.0057 | (0.001)*** |
| DUMMIES: | | | | | | | | |
| Month booking | YES | | YES | | YES | | YES | |
| Number of obs. | 44,403 | | 30,807 | | 37,504 | | 76,031 | |
| Centered $R^2$ | 0.590 | | 0.574 | | 0.560 | | 0.580 | |
| Underidentification K-P LM Test | $\chi^2(2)$=936.9*** | | $\chi^2(2)$=753.1*** | | $\chi^2(2)$ =661.1*** | | $\chi^2(2)$=1047.4*** | |
| Hansen J statistic | $\chi^2(2)$=0.002 | | $\chi^2(2)$=1.090 | | $\chi^2(2)$=0.454 | | $\chi^2(2)$=0.001 | |

Note: Morning=6am-11am; Evening=4pm-10.15pm. Winter= Nov-Mar; Summer= Apr-Oct. Bootstrap Standard Errors (SE) are reported in parenthesis, clustered by route and week. 250 repetitions. Coefficients *** statistically significant at 1%, ** at 5% and * at 10%. K-P=Kleibergen-Paap.

Table 11: Pricing equation results in Short and Medium Haul routes with Low and High Competition. Dependent Variable: *LnFare1*

| | Short Haul | | Medium Haul | | Low Competition | | High Competition | |
|---|---|---|---|---|---|---|---|---|
| Sold seats | 0.0282 | (0.001)*** | 0.0253 | (0.001)*** | 0.0281 | (0.001)*** | 0.024 | (0.001)*** |
| Booking Day1 | 0.7532 | (0.091)*** | 0.6481 | (0.064)*** | 0.6621 | (0.086)*** | 0.8217 | (0.066)*** |
| Booking Day4 | 0.4179 | (0.089)*** | 0.3515 | (0.061)*** | 0.3837 | (0.082)*** | 0.4981 | (0.064)*** |
| Booking Day7 | 0.1335 | (0.086) | 0.1558 | (0.058)*** | 0.1523 | (0.08)* | 0.251 | (0.061)*** |
| Booking Day10 | 0.0667 | (0.083) | 0.1219 | (0.055)** | 0.1153 | (0.077) | 0.1867 | (0.057)*** |
| Booking Day14 | -0.1571 | (0.079)** | -0.0004 | (0.051) | -0.0494 | (0.075) | 0.0074 | (0.055) |
| Booking Day21 | -0.1679 | (0.074)** | 0.0020 | (0.046) | -0.0249 | (0.069) | -0.0258 | (0.048) |
| Booking Day28 | -0.1181 | (0.069)* | 0.0220 | (0.042) | 0.0109 | (0.065) | -0.013 | (0.043) |
| Booking Day35 | -0.008 | (0.066) | -0.0146 | (0.04) | 0.0195 | (0.063) | -0.0023 | (0.042) |
| Booking Day42 | -0.0095 | (0.066) | -0.0482 | (0.039) | 0.0204 | (0.061) | -0.0464 | (0.042) |
| Booking Day48 | -0.0301 | (0.067) | -0.0093 | (0.039) | 0.0138 | (0.061) | -0.0309 | (0.041) |
| Booking Day56 | -0.0449 | (0.075) | -0.0043 | (0.04) | -0.0509 | (0.068) | 0.021 | (0.04) |
| Booking Day63 | -0.0047 | (0.072) | 0.0015 | (0.042) | 0.0316 | (0.064) | -0.0088 | (0.043) |
| Tobit residual | 0.0047 | (0.001)*** | 0.0016 | (0.001)** | 0.0035 | (0.001)*** | 0.0052 | (0.001)*** |
| DUMMIES: | | | | | | | | |
| Month booking | YES | | YES | | YES | | YES | |
| Number of obs. | 43,496 | | 59,661 | | 46,934 | | 66,601 | |
| Centered $R^2$ | 0.646 | | 0.512 | | 0.550 | | 0.584 | |
| Underidentification K-P LM Test | $\chi^2(2) = 615.1^{***}$ | | $\chi^2(2) = 870.8^{***}$ | | $\chi^2(2) = 947.6^{***}$ | | $\chi^2(2) = 857.8^{***}$ | |
| Hansen J statistic | $\chi^2(1) = 1.953$ | | $\chi^2(1) = 0.312$ | | $\chi^2(1) = 0.961$ | | $\chi^2(1) = 0.808$ | |

Note: *Fare1* is the fare obtained from a query for one seat. Short (Medium) Haul routes are less than 370 (more than 500) miles long. Low Competition includes flights in routes/city-pairs where Ryanair is at most a duopolist. In High Competition Ryanair operates with two or more other carriers at either the route or the city-pair level. Bootstrap Standard Errors (SE) are reported in parenthesis, clustered by route and week. 250 repetitions. Coefficients $^{***}$ statistically significant at 1%, $^{**}$ at 5% and $^{*}$ at 10%. K-P=Kleibergen-Paap.


Figure 1: Median Spline of $Fare1$ and sold seats, by timetable season.

Route: London Gatwick - Dublin. Each line refers to a different flight code, defined in the legenda.


Figure 2: Non parametric fit of $LnFare1$ and sold seats, by timetable season.

Route: London Gatwick - Dublin. Each line refers to a different flight code, defined in the legenda.

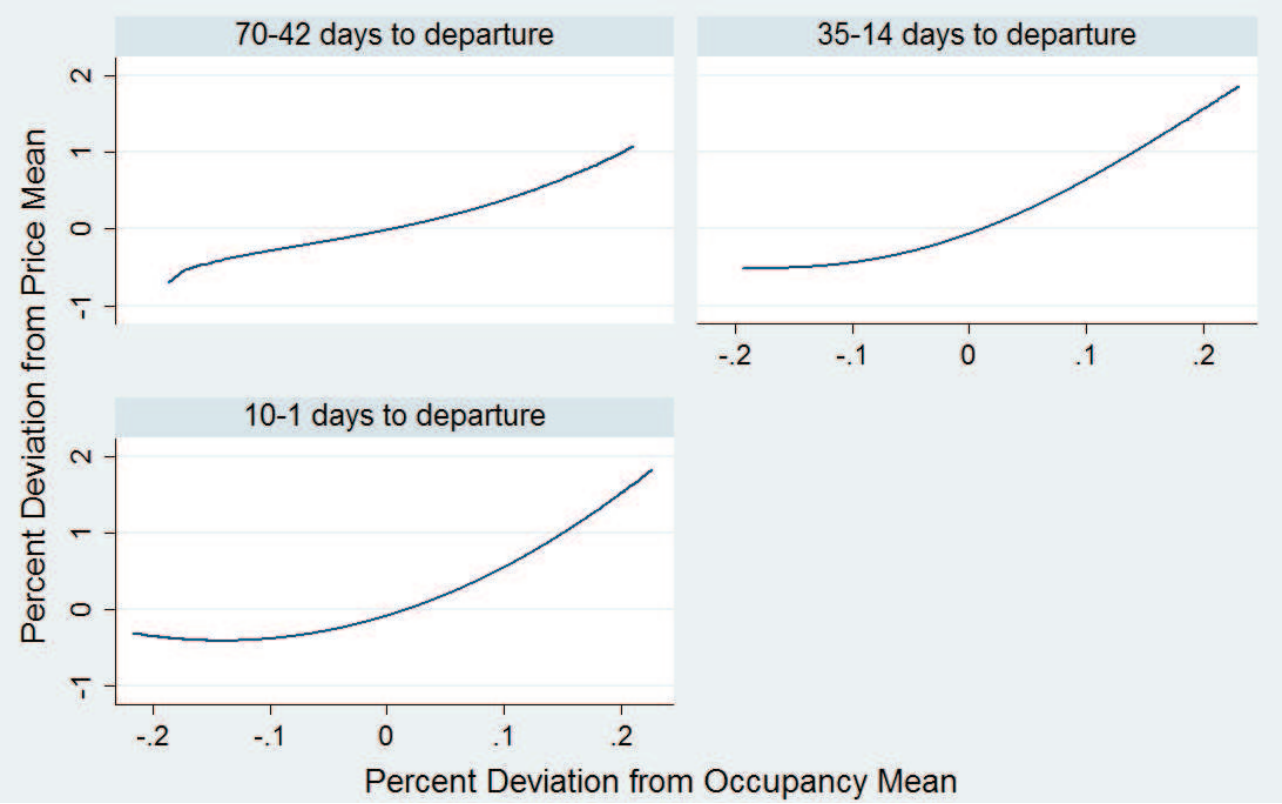


Figure 3: Non parametric fit between percentage deviation from mean $Fare1$ and percentage deviation from mean occupancy.

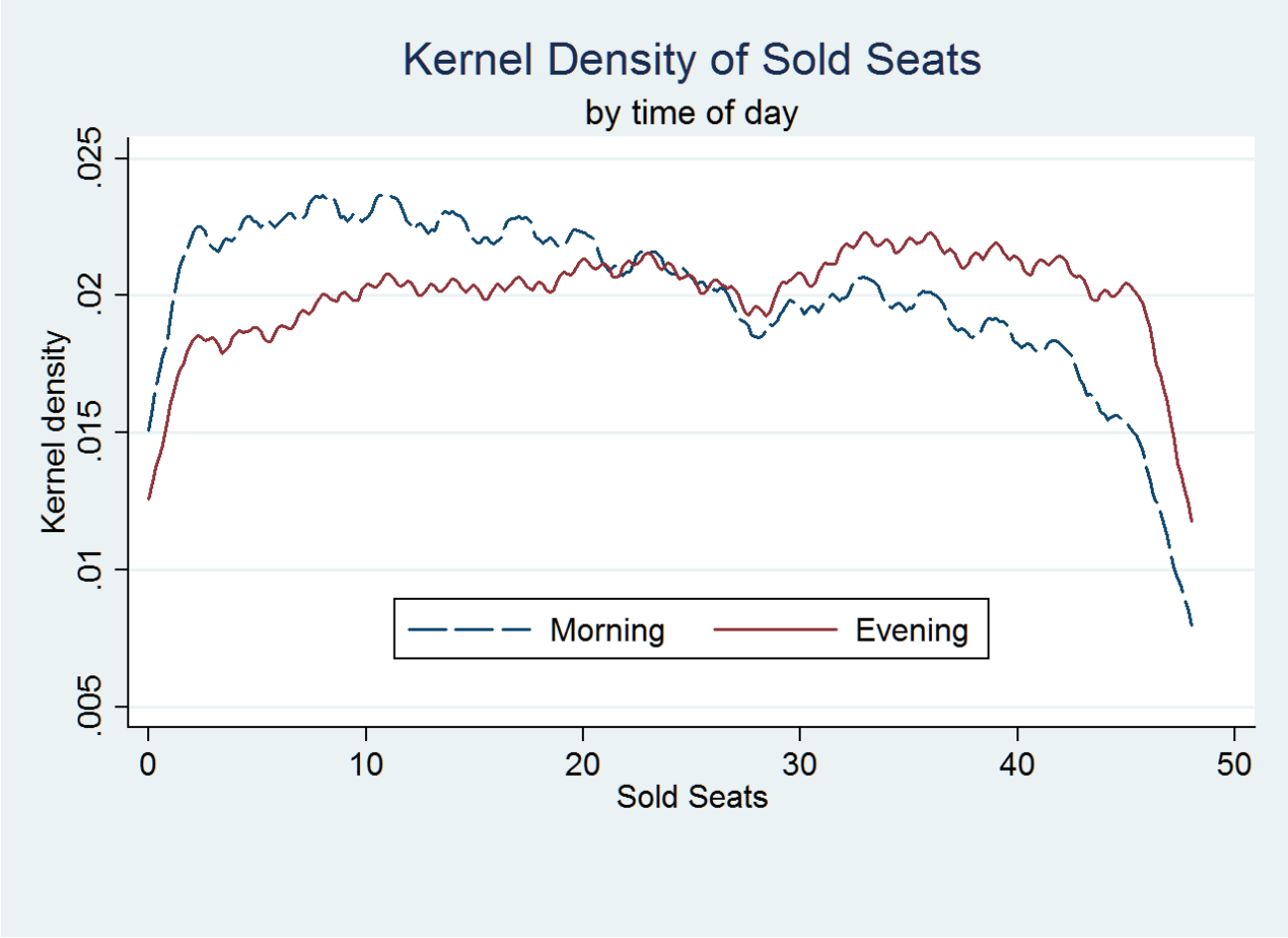


Figure 4: Distribution of Sold Seats in morning and evening flights.

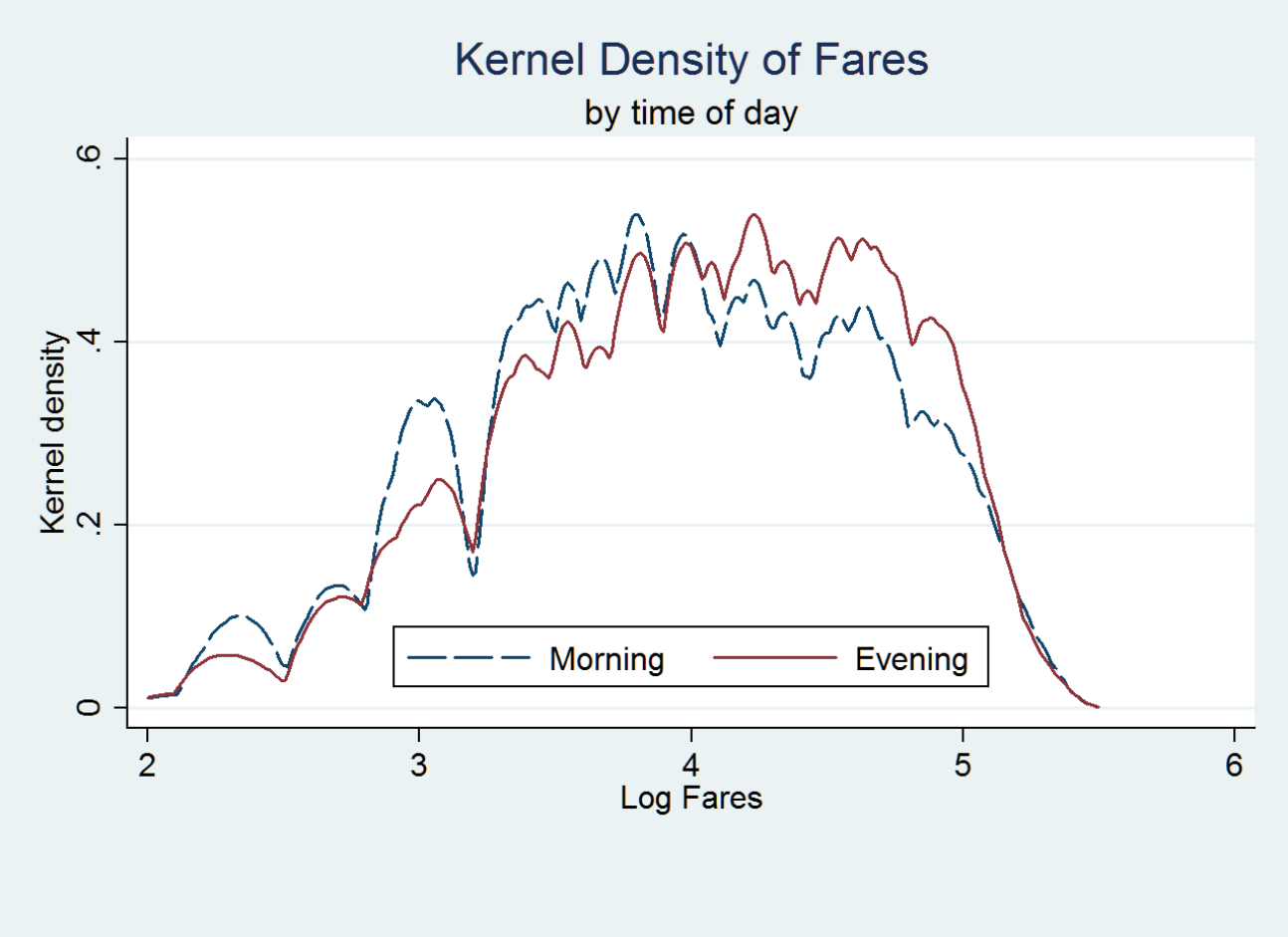


Figure 5: Distribution of Fares in morning and evening flights.