Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

Combined Effects of Capacity and 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
Keele University
Rimini Centre Economic Analysis

February 27, 2014

*We wish to thank seminar participants at the Texas A&M University, Tilburg University, London City University, Bocconi University, the University of Nottingham, and at the 2010 EARIE conference in Istanbul for their comments. Thanks also go to Makoto Watanabe and Chris Wilson for their feedback. The responsibility for the arguments and results expressed in this paper is entirely ours. Claudio Piga gratefully acknowledges financial support from the Leverhulme Trust Research Grant RPG-051.

[†]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.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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 the capacity-based theories predict, it emerges that fares monotonically and substantially increase with flight occupancy. After controlling for capacity utilization, our analysis also supports time-based theories, indicating a U-shaped temporal profile over a two-month booking period, as well as a sharp increase in fares in the two weeks prior to departure.

EconLit Codes: D22, L11, L93.

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

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

1 Introduction

Yield management (hereafter, 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 pricing policy when customers are heterogeneous, demand is uncertain, and capacity is hardly modifiable. In its simplest formulation, it entails a trade-off between accepting a booking request now at a low price or refusing it in the expectation that tomorrow a potential customer will be 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 of 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") 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 to provide 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 (hereafter, LCA), Ryanair. The relatively simple pricing behaviour of an LCA helps us to identify the combined impact on fares of both in-flight seat availability and the time separating the purchase

[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).

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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-based and the time-based theories, respectively. The *capacity-based theories* focus on the relationship between the evolution of fares and a flight's occupancy rate; Dana (1999a) postulates that such a relationship is defined by the airlines once and for all at the beginning of the planning horizon, while Deneckere and Peck (2012) extend the static analysis to allow for possible dynamic updating of the relationship. In both the static and the dynamic case, fares are predicted to be a non-strictly increasing function of the remaining capacity within each planning period.

In this paper, we provide a direct test of the relevance of capacity-based theories. 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 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 when the fare was retrieved from the airline's website. Moreover, using data from Ryanair rules out any difference in seat characteristics, because the airline sets the same set of restrictions on all its fares. Furthermore, by using flight fixed-effect panel data techniques, where the time 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 seat sold induces an increase of about 3.1% in offered fares. This effect increases in the sample of flights that: *i)* operate in less competitive routes; *ii)* are scheduled in Summer or depart in the evening, i.e., in periods of higher demand; *iii)* are short-haul and less volatile.

These results show the relevant role played by capacity-based theories 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-based theories,

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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 capacity-based theories. Both these studies, however, rely on data generated by the more complex process used by legacy carriers, whose properties are only partly aligned with the assumptions adopted by any of the models in the theoretical literature.

The *time-based theories* state that airlines may use inter-temporal price discrimination to exploit customer heterogeneity in terms of willingness-to-pay and uncertainty about departure time (Gale and Holmes, 1992, 1993; Dana, 1999b; Möller and Watanabe, 2010). On the one hand, the application of advance-purchase discounts (hereafter APD), i.e., fare reductions in the periods far from the departure date, plays in favor of an increasing temporal fare profile. On the other, clearance sale practices (Möller and Watanabe, 2010), i.e., fare reductions in the period immediately preceding departure, and the declining option value of waiting for customers with a higher willingness-to-pay (Gallego and van Ryzin, 1994) suggest that the opposite effect cannot be excluded.

This study sheds light on Ryanair's time-based pricing policy. 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-based motivation from those induced by the willingness to discriminate between customers booking at different times before departure. The evidence reveals that, on average, fares increase monotonically over the last three weeks before departure. However, a more complex price dynamics is also found over the entire booking period we take into consideration: in the two months preceding departure the temporal profile of fares often appears to be U-shaped. A similar finding is reported in Bilotkach et al. (2010), where, however, no control for capacity utilization is made.

In sum, this paper offers the first combined study of two testable implications derived from the theoretical economics literature on airline pricing. Both implications relate to

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

the pricing profile of carriers, suggesting that fares: *i*) should increase as a flight fills up; and *ii*) should grow over time, but may have a more complex U-shaped temporal pattern. 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, each focusing on, respectively, the capacity and the time dimension (Alderighi, 2010; Puller et al., 2009).

The remainder of the paper is structured as follows. Section 2 discusses the relevant theoretical and empirical literature, while Section 3 illustrates Ryanair's business model and its importance in the European airline market. In Section 4, we explain how we could retrieve the information on the flight's occupancy at the time when the fares were posted. Section 5 provides some descriptive statistics on the fare profile. The econometric model is presented in Section 6, which is followed by the comments to the main findings in Section 7. Section 8 concludes.

2 Literature Review

This section reviews the main theoretical and empirical works on YM, which are related to both the capacity-based and the time-based theories.[2] Dana (1999a) provides a theoretical model that addresses the link between fares and seat availability when capacity is fixed, by assuming that fares are set before demand is known. 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 unit of capacity increases as the probability of selling a ticket decreases, the

[2]Previous studies on pricing behavior 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 US 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.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

distribution of fares increases with capacity utilization. In other words, price dispersion arises not because an airline is trying to segment the 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 with seats that are seldom occupied. The analysis, however, assumes a commitment to the equilibrium price schedule, i.e., the airline cannot revise its ex-ante pricing decision as it learns new information about actual observed demand. As Deneckere and Peck (2012) indicate, such a commitment where prices cannot be adjusted over time is inefficient. More importantly, Deneckere and Peck (2012) develop a model where, on the supply side, current prices depend on the evolution of the aggregate quantity sold in previous periods; on the demand side, inter-temporal substitution is allowed so that, on the basis of their expectation of future prices and their private information about the state of demand, consumers may decide whether to purchase today or to delay buying in the hope of getting a better deal in later periods. In equilibrium, within each period transaction prices increase with the amount of sales made over the period, with the possibility at the beginning of each period of a markdown whose size is inversely related with the observed demand of the previous period. This is consistent with the empirical observation that prices within a route do not generally monotonically increase over time, but can indeed be sometimes observed to fall (Piga and Bachis, 2007). Furthermore, because the equilibrium prices in Deneckere and Peck (2012) are martingales, on average a flat temporal path over the booking period should be observed. However, such a prediction may not hold in a model incorporating the possibility of using inter-temporal pricing strategies to screen heterogeneous travelers, as the time-based theories suggest.

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

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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 or increase. 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).

Due to the difficulty of obtaining data on the occupancy rate at the time when a fare is offered either online or on a computer reservation system, only a limited number of empirical studies have tried to shed light on the accuracy and relevance of the theoretical predictions resulting from both the capacity- and time-based theories. Puller et al. (2009) test a number of implications derived from the theoretical works of Dana (1999a) and Gale and Holmes (1992, 1993). Using a proxy for a flight's load factor, they divide their sample into quartiles of expected demand (i.e., from expected full to expected empty): according to Dana (1999a), within each category the carriers are supposed to apply the same rule linking fares and occupancy rate. That is, after controlling for the expected demand, one should observe a greater proportion of higher fares (and therefore a higher mean fare and a greater measure of fare dispersion) as the realized load factor grows. The descriptive evidence in Puller et al. (2009) shows no support for these predictions.

To test the predictions from Gale and Holmes's works, Puller et al. (2009) evaluate whether discount tickets account for a smaller share of tickets in high-peak flights (defined as those expected to fly full and indeed realized to be full) and a larger share on off-peak ones (i.e., those expected to be empty and turned out to fly empty). After controlling for the number of days in advance of the flight a ticket was purchased, the evidence shows no significant difference among these shares across types of flights and carriers. Finally, Puller et al. (2009) illustrate how ticket restrictions explain a substantial amount of variation in fares, and conclude that standard price discrimination strategies appear to play a more crucial role in driving fare dispersion than those based on capacity management. Such a result may be due to the highly heterogeneous way with which airlines manage

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

their inventory using ex-ante mechanisms, as discussed in Bilotkach et al. (2010).

Interestingly, using panel data techniques, Escobari and Gan (2007) find evidence in support of the capacity-based theories. In their work, which is dual to ours, they derive an effective cost of capacity (ECC) by dividing an estimate of the fixed unit capacity cost by a calibrated measure of the probability a seat is sold; thus, the ECC increases as a seat's probability to sell decreases. Escobari and Gan (2007) test and find support for the hypothesis that a higher ECC should lead to higher prices, and that this effect should be larger in competitive markets.

Using the same data, Escobari (2012) confirms that, holding inventories constant, fares decrease until about 14 days from departure and subsequently increase. The declining profile is consistent with the theoretical model in Gallego and van Ryzin (1994), where the option value of waiting for the arrival of a customer with a high willingness-to-pay falls as the departure date approaches. However, the increasing time profile in the late booking period, also found in McAfee and te Velde (2007), suggests that other factors are at play, such as, the carriers' need to establish a reputation for consistently not offering last-minute deals which could lead to customers delaying their purchases. Furthermore, Escobari (2012) focuses on whether carriers adjust dynamically to unexpected shocks in demand, and finds strong evidence in favor of prices responding to new information about the pattern of sales: as the uncertainty on aggregate demand dispels, the airline intervenes to adjust its fares. This finding is thus complementary to our analysis, which focuses on the role played by the predetermined adjustments implied by the capacity-based theories.

3 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 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), and no discrimination between one-way and round-trip ticket fares; (ii) direct selling through Internet bookings with electronic tickets and no seat

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

reservations;[3] (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 aircraft and acting as gate agents); (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 the 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. Ryanair 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 31 March 2011.[4] The size and importance of this carrier, and its ability to attract customers, make it a key player in the European airline industry.

3.1 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 (max 10 kg) with a fixed fee for each checked baggage (max 15 kg per item). The applicability of our findings to legacy carriers might be limited by these differences, albeit none of these impinge on the YM aspects on which we focus on in this paper, since they are unaffected by capacity utilization

[3]Since 2011 Ryanair has been offering the opportunity to reserve a seat, for a fee. This happened after the time period considered in the present analysis.

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

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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 modeling assumption of many theoretical works.

Interesting insights into how Ryanair designs its YM system are given in European Commission (2007), which provides details of the investigation that led to the decision to block the takeover by Ryanair of Aer Lingus. Both companies may choose from a set of standard "templates", each describing "the number of places that should be available in a given price category ("booking class") (p. 109, item 439)"; each adopted template is the one that is expected to best match a specific flight's characteristics. The template thus appears to correspond to the practical implementation of the notion of an equilibrium distribution of fares across the full set of an aircraft's seats given in Dana (1999a).

The template is also instrumental in the carrier's adoption of a multifaceted pricing policy that, in addition to a pure capacity-based strategy, allows for time effects, as well as responses to unexpected shocks. Indeed, as suggested in European Commission (2007), the time dimension of the carrier's YM practice is obtained by changing the number of seats assigned to each booking class: a decrease or an increase of fares is brought about "by making more seats available in the cheaper [or more expensive] price categories (p. 109, item 440)". Note, however, that similar alterations of the standard template may be applied when the carrier needs to respond to an unexpected shock; the important difference lies in the fact that, on the one hand, time effects are designed to include pre-defined changes in the template that take place in a "routine" fashion, i.e., they occur systematically at specific points in time. On the other, responses to an unexpected shocks arise only after the template is set; the ensuing modification to the template is thus "discretionary" in nature, and is spurred by either external (e.g., new qualitative information on future demand, changes in the rivals' behavior, etc.) or

[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.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

internal (a promotional policy of the marketing department, etc.) factors that are known by the carrier but are unobserved by the econometrician.[6]

To sum up, fares are the results of three drivers: capacity, time, and shocks, where the former two give rise to an "augmented template" that captures the airline's routine YM operations, while the latter come about through a discretionary intervention of a YM analyst. One of the main contributions of this paper is to identify the routine activity, which simultaneously considers how fares are related to seat occupancy and how they are designed to change as the time to departure nears.[7] By doing so, we shed light on the role played by the capacity- and time-based theories in explaining the airlines' fare-setting process, after purging for the discretionary intervention.

4 Data Collection

Our analysis is based on primary data on fares collected using an "electronic spider" which is linked to the Ryanair website.[8] 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 one-way fares for departures due 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

[6]From an YM analyst's operational perspective, the distinction between routine and discretionary interventions implies that the former are carried out as part of a set of standard codified tasks that are regularly scheduled, e.g., the modification of the original template to implement a specified temporal price profile. As emerged from the authors' discussion with industry practitioners, routine tasks may be automated subject to the analyst's approval. Conversely, discretionary changes trigger an update of a flight's prior forecast and therefore may lead to a revision of the template.

[7]In Section 5.2 we show how changes over time appear to be coded into the airline's computerized reservation system.

[8]All fares are net of add-ons and other fees, i.e., charges for the use of some of the methods of payment, such as credit cards.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

as *BookingDays*.[9] Thus, for every daily flight we obtained up to 13 prices that differ by the time interval from the day of departure, and allow the identification of the evolution of fares over time.

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* (see Sections 4.1 and 4.2 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, all of 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, the 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.[10] Some descriptive statistics are reported in Tables 1 and 2, to be discussed in more detail in Section 4.2.

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

[10] 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 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 changes relative to the flight's occupancy rate or over time.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

4.1 Retrieving data on *Seats* and *TopFare*

The collection strategy exploited a feature of 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 observed 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 program 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*.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

4.2 Interpretation of retrieved fares

When $Seats < 50$, *TopFare* corresponds to the fare of a transaction whose completion would fill the flight to capacity.[11] For this reason, *TopFare* presents two important characteristics. First, as Table 1 shows, it exhibits some limited variation around its median value. Indeed, for all the routes in the table, 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. On other routes, the maximum value is higher, but by not more than £10 or £20 above 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 also supported by the low standard deviations reported.

Second, and relatedly, if $Seats < 50$, in line with the capacity-based theories, *TopFare* represents the maximum fare of a flight. When a query that closes the flight is issued, the Ryanair reservation system always retrieves the fare that is associated with the value of the last seat. 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 that being made available.

It follows that *TopFare* varies in a similar fashion as *Fare1*, and their two values 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

[11]If, for example, the spider returned 26 unsold 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.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

well below those of *TopFare* reported 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.

Table 2 shows also that, 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 no higher than the fare for one seat when 49 or less 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.

The previous analysis therefore interprets *TopFare* as the price of the last seat that can be purchased in a single query. Although *TopFare* exhibits limited variation when $Seats < 50$, when $Seats = 50$ (i.e., 50 or more seats are available) *TopFare* corresponds to the price of the 50th seat ahead, and therefore it varies in a similar fashion as *Fare1*.

5 Preliminary evidence

The results drawn from the descriptive analysis in this section help to gain better insights into the airline's pricing policy and its relation to 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.

5.1 Do fares increase as the flight fills up?

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

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

- 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 and for all flights, the plot shows, on average, a monotonically increasing relationship; fluctuations are probably due to idiosyncratic changes in the template and/or to lack of fare observations for specific occupancy values.[12]

To generalize the evidence from one route to the entire sample, in Figure 2 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. Next, we aggregate the pairs of percentage deviations across two categories of booking days: early-middle and middle-late, that is, 70 to 35 and 28 to 1 days from departure. 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 110% (60-70%) in *Fare1* from its mean (as can be read on the vertical axis). In the second, fares appear to be 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 an almost 200% increase in *Fare1* over its mean. Interestingly, the same 20% decrease from the mean of *Sold Seats* is met by a slightly lower deviation from the mean of *Fare1* than in the case of the early-middle booking period.

A number of considerations can be drawn from the foregoing graphical analysis. First, the evidence reported in Figure 2 suggests that YM techniques designed by airlines to manage capacity constitute an important factor driving price dispersion. Interestingly, despite the methodological similarity, Puller et al. (2009) reach an opposite conclusion in their study of the US airline markets.[13] Second, it introduces the need to combine

[12]A smoother increasing relationship can be obtained from a nonparametric plot of the *Log* of *Fare1* with the last 50 seats' occupancy. This is not reported to save on space but is available upon request.

[13]Differences may be due to the different type of airlines considered (legacy vs. low cost) and to the different methods used to obtain the in-flight's remaining capacity.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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 in-flight remaining capacity; and *2)* the discretionary intervention of a 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 routine pricing behavior net of such discretionary interventions. Given the crucial role of time 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).

5.2 Do fares increase over time all of the time?

The descriptive analysis in the previous sub-section highlighted a positive relationship between fares and available seats, which appears to hold on average over a range of dates and routes. In this sub-section, we extend the analysis by focusing on possible time effects in the airline's pricing structure. Our objective is to separate fare changes induced by variations in the flight's remaining capacity from time effects that are unrelated to the actual observed evolution of sales. This is also important in terms of econometric testing, because the time-based theories lead to predictions that may be confused with those of the capacity-based ones. That is, in both cases, fares may increase over time. In the latter case, fares are based on the shadow cost of capacity, while, in the former, airlines increase their fares to exploit consumer 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 the availability of seats increases. More interestingly, when we condition on capacity utilization to see how fares on average change with the booking day, we observe that the temporal 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.

However, it might be possible that the temporal profile in Table 3 is due to the aggrega-

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

tion of fares from heterogeneous 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), and predicted in Deneckere and Peck (2012). First, in each row the likelihood of observing a price drop generally increases as more seats become available, especially when the departure time is not within a week.[14] 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 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 that due to the time variation. In Table 5 we show a *pure* time 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

[14]The fact that $4-5\%$ of late booking cases report a price 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.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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 (latter) being greater (smaller) 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-based theories, accounts for the largest majority of observations (about 73%). Interestingly, this also implies that 27% of fare changes are generated by a pure time 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 approaches. By the same token, large increases are hardly observed during the early booking period. Second, more pure time variation (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.[15]

Overall, the evidence in Tables 3-5 suggests that fares are affected by a combination of capacity and time factors. These will be further investigated in the next Section.

6 The econometric model

The foregoing descriptive analysis provides evidence that both time and capacity are strong drivers of fares. As discussed in Section 3, fares are also affected by pricing decisions arising from an unexpected shock. We now aim to trace out an augmented template linking in-flight remaining capacity and time before departure with offered fares, net of human discretionary intervention. This is equivalent to estimating a pricing equation where the independent regressors are *Sold Seats* and *BookingDay* dummies, holding other factors fixed. To achieve this, the use of an OLS regression is inappropriate because *Sold Seats* has two features which need special attention. A first obvious issue is its endogeneity, since some unobserved determinants of the airline pricing behavior may

[15]See Section 7.2 for a formal definition of routes with high and low competition.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

be correlated with a specific flight's time-invariant factors (an issue which could be dealt with using the standard fixed-effects panel technique) and, more importantly, with the idiosyncratic, *discretionary* intervention of the airline's yield manager. This aspect calls for an instrumental variables estimator. A second, more subtle, issue, is that *Sold Seats* is censored due to the retrieving procedure. Indeed, the number of sold seats may range from 0 to 189, i.e., the aircraft's 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 sub-sample is consistent for β.

However, if the selection indicator is not independent of $\mathbf{x}$, 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 β and γ can be consistently estimated by ordinary least squares (OLS) on the selected sample. Thus,

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

including v in the regression eliminates the sample selection problem and allows us to consistently estimate β. Of course, if some variable in $\mathbf{x}$ is endogenous, the procedure to correct for sample selection is the same, while to consistently estimate β 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:

$$LnFare1 = \mathbf{z}_1\delta_1 + \alpha\, Sold\ Seats + u, \tag{2}$$

where the logarithmic transformation of $Fare1$ is meant to linearize its convex relationship with the endogenous regressor *Sold Seats* shown in Figure 2 and in Table 3, where there is a more than proportional increase in fares as the number of available seats falls; $\mathbf{z}_1$ are the other exogenous regressors, including dummy variables for booking days.[16] Our specification links the current fare with the current in-flight available capacity: this is a significant departure from Escobari (2012), where lagged values for both are used.

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,

[16]The booking days dummies capture the routine intervention of the yield manager. The *true* fare setting model should also include the analyst's discretionary intervention, H, so that $LnFare1 = \mathbf{z}_1\delta_1 + \alpha\, Sold\ Seats + \lambda H + \varepsilon$. Because H 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 H. Indeed, the analyst is more likely to *discretionally* reduce (increase) fares when *Sold Seats* is low (high). Therefore, estimation using OLS should produce an upward bias in the coefficient for *Sold Seats*.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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:

$$LnFare1 = \mathbf{z}_1\delta_1 + \alpha\, Sold\ Seats + \gamma v_3 + e. \tag{5}$$

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 obtain v_3 when $Sold\ Seats^* > 0$, since δ_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 sub-sample.

3. On the selected sub-sample, 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 Two-Stage Fixed Effect (IVFE) estimator, using as instruments $\mathbf{z}_1$ and $\hat{v}_3$.[17]

It is possible to test whether 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.

6.1 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 before departure (i.e., the booking day). This panel structure allows us to control for all unobserved characteristics which are specific to the single flight, such as, for instance, market structure and distance. Furthermore, focussing

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

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

on a single flight using a fixed-effects approach allows us to control for possible strategic effects at the route level, where, for example, the airline can opt to implement temporary capacity limits, i.e., reduce the number of daily flights.

With regard 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.

To deal with endogeneity, we propose the following identification strategy, which is based on 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 willing to spend time planning future trips); and it is less likely to observe a discretionary 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 4.2. As the convex relationship between fares and seat occupancy shown in Figure 2 suggests, the slope of the template is expected to increase with occupancy, and can therefore be considered as a valid candidate for an instrument. To derive it, we take the difference between *Top Fare* and *Fare1* and divide it by the number of available seats ($50 - \textit{Sold Seats}$); as previously discussed, the difference in the numerator tends to shrink (expand) as occupancy increases (decreases), and so does the denominator. However, the convex relationship indicates that, as the plane fills up, the denominator decreases at a faster rate than the numerator, which is sufficient to guarantee a positive correlation between the template's slope and *Sold Seats*, whose value, in the selected sub-sample, is about 0.383. Although the slope captures a relevant feature of the flight's pricing template, it is also correlated with the error e. However, under the

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

assumption that template changes are specific to each daily flight, using the lagged values of the slope 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 (e.g., Monday), the instrument denoted "*Lag Mean Slope*" is constructed by averaging, for each booking day, the value of the slope in flights with the same code departing on the same day (e.g., Monday) of the three preceding weeks. Some of these lagged values may belong to observations where $Seats = 50$, and for which therefore we cannot ascertain the exact number of available seats. Nonetheless, this would not impact on the interpretation of the instrument as the slope of the template. Indeed, as discussed in Section 4.2, in this case the instrument would include values of $TopFare$ indicating not the value of the last seat on the flight but that of the 50^{th} seat ahead of the one being available (whose value is captured by $Fare1$).

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. 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 (Wooldridge, 2002). 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 the fact that a higher demand destination is more likely to be served by more than one UK airport. Furthermore, dummies for the day of the week 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

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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.

7 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 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. As expected, *Sold Seats* is positively affected by "*Lag Mean Slope*". This is evidence that yield managers raise fares when past sales increase more than expected. This is probably because they anticipate that, at least some of the time, changes in the current sales are correlated with future sales, or more directly because they observe shocks in demand that are correlated with current and future sales. Moreover, the instrument that identifies booking during a holiday period has a negative coefficient, i.e., in those days sales tend to be lower than usual.

Table 7 shows the second step of the IVFE 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*: a unit increment induces a 3.43% increase in fares if we do not correct for endogeneity, while only a 3.11% increase in the IVFE case. 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 a human *discretionary* intervention, which

[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 are available upon request.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

is positively related to *Sold Seats* (see fn.16). The magnitude of the *Sold Seats* coefficient suggests that a considerable proportion of a flight's fare dispersion can be attributed to a capacity effect. Indeed, if we apply a 3.11% change rate per seat to the mean value of *Fare1* (£65.17) when *Sold Seats* changes from its intermediate value (25) to either its maximum (49) or minimum value (1), we obtain a prediction for the fare of about £137.53 and £30.88, respectively.[19] These results provide strong support to the capacity-based theories, 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 *BookingDays* 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 4 and 1 days from departure we record percentage increases of about 13% and 51%, respectively. Interestingly, in both specifications these last-minute increases, which highlight an important role for time effects, are part of an U-shaped temporal fare profile whose declining part reflects the option value of waiting for a high-demand traveler to book (Gallego and van Ryzin, 1994).

To get further insights on the price dispersion generated by the two dimensions, we use the estimates in Table 7 to compute both the price variation caused by a change over the time before departure keeping the available capacity fixed and the price variation caused by a change in available capacity keeping the time before departure fixed. The analysis is performed starting from the intermediate values of the two variables, i.e., 25 available seats and 28 days before departure. The maximum price difference is £54.87 moving along the time dimension (from 1 to 70 days before departure) and £106.65 considering the capacity variation (from 1 to 49 seats), which roughly corresponds to one-third and two-thirds of total variation.

Overall, after controlling for endogeneity (i.e., after purging the estimates from the

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

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

effect due to the *discretionary* analyst intervention), we obtain strong evidence for both capacity- and time-based theories in driving the airline's pricing policy. In Table 8 we check whether these two effects operate cumulatively by interacting *Sold Seats* with a dummy variable capturing all the booking days within 7, 10 and 14 days, respectively. The insignificant coefficient for the interaction term suggests that the two effects operate independently of one another.

The results in Tables 7 and 8 are obtained using data from a heterogeneous 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 behavior that is compatible with variability at the flight level. In the remainder of the paper, we study whether the *average* pricing policy depicted in Table 7 changes as we take these sources of heterogeneity into account.

7.1 Demand volatility and flight's characteristics

First, we study whether differences in the pricing profile arise in relation to the extent of aggregate demand uncertainty: a pricing policy based on available capacity may be less precisely applied when flights exhibit large rather than small demand volatility, thus leading to a larger role for the time dimension. In Table 9 high (low) volatility flights are those whose standard deviation of *Sold Seats* in a given month is larger (smaller) than the sample one. As expected, the impact of *Sold Seats* is weaker than average in the high volatility sample, where, in addition, the temporal profile is steeper.

Second, Table 10 offers further insights into the nature of the effects of in-flight remaining capacity 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.[20] The coefficient of *Sold Seats* is found to be larger

[20]Morning flights are from 6am to 11am; evening ones from 4pm to 10.15pm. We thus

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

in the evening sample; as evidence not reported here indicates, evening flights include a larger proportion of observations with 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.

Third, the last two columns of Table 10 consider the two samples of flights operated in the Winter (November-March) and the Summer (April-October) periods. Because Ryanair serves many Mediterranean destinations 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 temporal 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 temporal profile to attract price-sensitive consumers with high demand uncertainty who would not book their flights too far in advance.

In situations of higher demand, i.e., evening and summer routes, the capacity effect dominates. If we consider the price variation due to time and capacity independently, as discussed in the comment of Table 7, we find that the latter accounts for approximately three-quarters of the total. Vice versa, in lower demand conditions, the price variations induced by capacity and time are more similar.[21] However, the sum of the two effects is of a comparable magnitude across the different sub-samples.

Finally, in Table 11 we investigate whether the different cost structure that characterizes routes of varying length may affect the carrier's pricing approach. Indeed, short haul flights are subject to higher operative costs per kilometer, due to the greater fuel consumption during take-off and landing. This is likely to induce a pricing strategy that is

exclude late morning and afternoon flights.

[21]More precisely, in the evening (summer) sample we have a price variation induced by time of £41.66 (£43.77) and £126.00 (£115.81) for capacity variation. For morning (winter) flights the time variation is £68.49 (£94.89) and the capacity variation is £94.98 (£85.82).

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

more based on the capacity dimension. Consistent with this conjecture, we find that the coefficient of *Sold Seats* is larger in the short-haul sample. Furthermore, the prominent U-shaped temporal profile indicates a stronger reliance on a time-based strategy aiming at attracting passengers with a lower willingness-to-pay, who in other circumstances would not be targeted.

Overall, while the analysis in this section confirms the important role played by both the capacity and the time dimensions, it also highlights how the airline may vary its pricing policy mix depending on some of the underlying characteristics at the flight or route level.

7.2 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 intra-firm 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 to be 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.[22] In less 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

[22]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).

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

its direct competitors on that route/city-pair.[23]

Table 11 reports the estimates from the low and the high competition sub-samples, 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 thus suggest that competitive pressure flattens the relationship between fares and the remaining capacity, in contrast with the prediction in Dana (1999a) where fare dispersion increases with competition. However, as far as the temporal profile is concerned, the estimates confirm the previous finding of a significant time effect. Furthermore, flights exhibiting significant price drops six to two weeks before departure feature exclusively in the sub-sample of highly competitive markets.

8 Conclusions

This study has built 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 have 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, each extra sold seat induces a 3.11% increase in a flight's fare. Such a result indicates that the capacity dimension is an important determinant of airline pricing. The study also reveals novel evidence regarding the temporal profile of fares. All econometric specifications show a sharp increase in fares in the last few 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

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

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

appears to be designed to include late increases in fares regardless of the actual observed capacity utilization. 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. Indeed, the evidence captures a similar effect which is quite robust to variations in the sample composition. Overall, the evidence indicates that a monotonic temporal profile is not necessarily observed after capacity utilization is controlled for.

Furthermore, 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 template in less competitive markets, implying that the last seats are sold at higher fares.

Finally, it is worth recalling that our results relate to the pricing behaviour of the largest European low-cost carrier, Ryanair. It is left to future research to investigate the extent to which the YM approach we have illustrated is relevant for other airlines (with a similar or different business model) in different geographical areas.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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.

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

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.

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

Escobari, D. (2012). Dynamic Pricing, Advance Sales and Aggregate Demand Learning in Airlines. *Journal of Industrial Economics*, 60(4):697–724.

Escobari, D. and 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.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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*, 59(4):552–577.

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

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

Gallego, G. and van Ryzin, G. J. (1994). Optimal Dynamic Pricing of Inventories with Stochastic Demand over Finite Horizons. *Management Science*, 40:999–1020.

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

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.

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

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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.

Piga, C. A. and Bachis, E. (2007). Pricing strategies by European traditional and low cost airlines: or, when is it the best time to book on line? In Lee, D., editor, *Advances in Airline Economics. The Economics of Airline Institutions, Operations and Marketing*, pages 319–344. Elsevier, Amsterdam, Holland.

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.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

Table 1: Distribution of $TopFare$, by route, when $Seats < 50$

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	27.1

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

Table 6: Tobit and First Stage estimates (dependent variable: *Sold Seats*)

	Tobit		First stage	
Lag Mean Slope	2.536	(0.072)***	2.388	(0.005)***
Booking Day1	63.752	(0.697)***	61.354	(0.112)***
Booking Day4	58.949	(0.705)***	56.374	(0.110)***
Booking Day7	54.357	(0.713)***	52.006	(0.110)***
Booking Day10	49.909	(0.707)***	47.345	(0.109)***
Booking Day14	44.182	(0.706)***	41.966	(0.106)***
Booking Day21	34.468	(0.695)***	32.538	(0.103)***
Booking Day28	25.293	(0.694)***	23.756	(0.101)***
Booking Day35	17.162	(0.700)***	16.005	(0.099)***
Booking Day42	10.144	(0.696)***	9.429	(0.090)***
Booking Day49	5.395	(0.698)***	5.039	(0.087)***
Booking Day56	2.754	(0.658)***	2.537	(0.080)***
Booking Day63	1.651	(0.627)***	1.529	(0.077)***
N. UK airports serving arrival	-1.138	(0.185)***		
Tobit residual			0.925	(0.001)***
Booking Day is in Holiday period			-0.186	(0.025)***
Constant	110.826	(5.209)***		
DUMMIES:				
Month booking	No		Yes	
Week	Yes		No	
Route	Yes		No	
DOW Booking	Yes		No	
Time Departure	Yes		No	
Number of obs.	511,226		100,031	

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

Pseudo R2	0.1621	
Centered R2		0.9731
Test excluded instruments:		$F(2,4490) = 1.0e{+}05^{***}$
Underidentification K-P LM Test		$\chi^2(2)=1.0e{+}05^{***}$
Anderson-Rubin Wald test		$F(2,4490)= 908.97^{***}$
Anderson-Rubin Wald test		$\chi^2(2)=1818.79^{***}$
Stock-Wright LM S statistic		$\chi^2(2)=726.44^{***}$

Note: *Lag Mean Slope* is the mean obtained by taking the 7, 14 and 21 days lagged (L) values of a template's *slope*. See the main text for details on its construction. Significant at *10%, ** 5%, and *** 1%.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

Table 7: Pricing equation results using the full sample and different estimation methods (dependent variable: *LnFare1*)

	IVFE		FE-OLS	
Sold Seats	0.0311	(0.001)***	0.0343	(0.001)***
Booking Day1	0.4121	(0.053)***	0.2248	(0.054)***
Booking Day4	0.1213	(0.051)**	-0.0542	(0.053)
Booking Day7	-0.0962	(0.049)**	-0.2560	(0.050)***
Booking Day10	-0.1205	(0.047)*	-0.2631	(0.049)***
Booking Day14	-0.2589	(0.044)***	-0.3815	(0.047)***
Booking Day21	-0.2062	(0.042)***	-0.2963	(0.044)***
Booking Day28	-0.1316	(0.039)***	-0.1948	(0.042)***
Booking Day35	-0.0804	(0.038)**	-0.1210	(0.040)***
Booking Day42	-0.0710	(0.037)*	-0.0944	(0.041)**
Booking Day48	-0.0399	(0.038)	-0.0524	(0.040)
Booking Day56	-0.0129	(0.038)	-0.0190	(0.042)
Booking Day63	-0.0009	(0.037)	-0.0046	(0.036)
Tobit residual	-0.0005	(0.0004)	-0.0025	(0.0004)***
DUMMIES:				
Month booking	YES		YES	
Number of obs.	100,031		100,031	
Centered R2	0.5680		0.5683	
Excluded instruments:	2			
Underidentification K-P LM Test	$\chi^2(2) = 1151.62^{***}$			
Hansen J statistic	$\chi^2(1) = 2.158$			

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.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

Significant at *10%, ** 5%, and *** 1%. K-P=Kleibergen-Paap.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

Table 8: Pricing equation results interacting with dummies for periods near the departure (dependent variable: *LnFare1*)

	Interaction with dummy for 7 days before dep.		Interaction with dummy for 10 days before dep.		Interaction with dummy for 14 days before dep.	
Sold Seats	0.0314	(0.001)***	0.0314	(0.002)***	0.0295	(0.003)***
Sold Seats*booking period	-0.0010	(0.002)	-0.0006	(0.003)	0.0024	(0.005)
Booking Day1	0.4155	(0.053)***	0.4120	(0.053)***	0.4204	(0.053)***
Booking Day4	0.1218	(0.051)**	0.1197	(0.051)**	0.1333	(0.053)**
Booking Day7	-0.0991	(0.049)**	-0.0995	(0.050)**	-0.0797	(0.055)
Booking Day10	-0.1258	(0.048)***	-0.1253	(0.051)**	-0.1002	(0.058)*
Booking Day14	-0.2639	(0.045)***	-0.2633	(0.048)***	-0.2340	(0.063)***
Booking Day21	-0.2102	(0.042)***	-0.2098	(0.044)***	-0.1910	(0.050)***
Booking Day28	-0.1343	(0.039)***	-0.1340	(0.041)***	-0.1211	(0.044)***
Booking Day35	-0.0820	(0.038)**	-0.0818	(0.039)**	-0.0745	(0.039)*
Booking Day42	-0.0721	(0.037)**	-0.0719	(0.037)**	-0.0673	(0.037)*
Booking Day48	-0.0404	(0.038)	-0.0404	(0.038)	-0.0378	(0.038)
Booking Day56	-0.0133	(0.038)	-0.0132	(0.038)	-0.0116	(0.038)
Booking Day63	-0.0012	(0.037)	-0.0012	(0.037)	0.0005	(0.037)

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

Tobit residual	-0.0006 (0.0004)	-0.0005 (0.0004)	-0.0004 (0.0004)
DUMMIES:			
Month booking	YES	YES	YES
Number of obs.	100,031	100,031	100,031
Centered R2	0.5673	0.5676	0.5693
Excluded instruments:	2	2	2
Underidentification K-P LM Test	$\chi^2(2)$ =363.314***	$\chi^2(2)$=248.265**	$\chi^2(2)$=88.711**
Hansen J statistic	$\chi^2(2)$=2.330	$\chi^2(2)$=2.212	$\chi^2(2)$=2.080

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. Significant at *10%, ** 5%, and *** 1%. K-P=Kleibergen-Paap.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

Table 9: Pricing equation results in flights that had high vs low volatility (dependent variable: *LnFare1*)

	High volatility		Low volatility	
Sold Seats	0.0292	(0.001)***	0.0323	(0.001)***
Booking Day1	0.5844	(0.080)***	0.2907	(0.067)***
Booking Day4	0.2545	(0.077)***	0.0273	(0.065)
Booking Day7	0.0118	(0.074)	-0.1741	(0.062)***
Booking Day10	-0.0180	(0.071)	-0.1968	(0.060)***
Booking Day14	-0.1721	(0.068)**	-0.3267	(0.057)***
Booking Day21	-0.0865	(0.063)	-0.2950	(0.053)***
Booking Day28	-0.0530	(0.058)	-0.1972	(0.049)***
Booking Day35	-0.0224	(0.058)	-0.1340	(0.047)***
Booking Day42	0.0170	(0.057)	-0.1488	(0.048)***
Booking Day48	0.0547	(0.060)	-0.1229	(0.045)***
Booking Day56	0.0366	(0.061)	-0.0563	(0.053)
Booking Day63	0.0056	(0.065)	-0.0116	(0.048)
Tobit residual	0.0003	(0.001)	-0.0008	(0.001)
DUMMIES:				
Month booking	YES		YES	
Number of obs.	40,728		59,303	
Centered R2	0.5586		0.5768	
Excluded instruments:	2		2	
Underidentification K-P LM Test	$\chi^2(2)$ =616.293***		$\chi^2(2)$=764.002**	
Hansen J statistic	$\chi^2(2)$=1.906		$\chi^2(2)$=0.455	

Note: The two samples are built by selecting those flight codes which in a given month had a standard deviation of *Sold seats* respectively larger and smaller than the sample

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

one. Bootstrap Standard Errors (SE) are reported in parenthesis, clustered by route and week. 250 repetitions. Significant at *10%, ** 5%, and *** 1%. K-P=Kleibergen-Paap.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

Table 10: Pricing equation results by time of the day and season (dependent variable: *LnFare1*)

	Morning		Evening		Winter		Summer	
Sold Seats	0.0296	(0.001)***	0.0347	(0.001)***	0.0275	(0.001)***	0.0324	(0.001)***
Booking Day1	0.6579	(0.093)***	0.2777	(0.085)***	0.8413	(0.153)***	0.1038	(0.060)*
Booking Day4	0.3540	(0.09)***	0.0235	(0.081)	0.4924	(0.152)***	-0.1568	(0.057)***
Booking Day7	0.0875	(0.087)	-0.1423	(0.076)*	0.1928	(0.149)	-0.3362	(0.054)***
Booking Day10	0.0502	(0.083)	-0.1545	(0.074)**	0.1095	(0.146)	-0.3331	(0.052)***
Booking Day14	-0.1275	(0.079)	-0.2509	(0.070)***	-0.0606	(0.144)	-0.4620	(0.049)***
Booking Day21	-0.1161	(0.075)	-0.1980	(0.064)***	-0.1071	(0.141)	-0.3779	(0.046)***
Booking Day28	-0.0574	(0.074)	-0.1305	(0.060)**	-0.0354	(0.139)	-0.3023	(0.042)***
Booking Day35	0.0119	(0.073)	-0.0791	(0.057)	0.0027	(0.136)	-0.2379	(0.041)***
Booking Day42	0.0075	(0.070)	-0.1151	(0.057)**	-0.1146	(0.142)	-0.1803	(0.038)***
Booking Day48	0.0086	(0.075)	-0.0081	(0.053)	-0.1478	(0.142)	-0.1269	(0.040)***
Booking Day56	0.0527	(0.075)	-0.0199	(0.059)	-0.2182	(0.139)	-0.0657	(0.039)*
Booking Day63	0.0499	(0.080)	-0.0101	(0.054)	-0.0749	(0.135)	-0.0348	(0.039)
Tobit residual	0.0022	(0.001)***	-0.0034	(0.001)***	0.0010	(0.001)	-0.0008	(0.001)
DUMMIES:								

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

Month booking	YES	YES	YES	YES
Number of obs.	38,885	27,193	31,859	68,172
Centered R2	0.5865	0.5735	0.5596	0.5791
Excluded instruments:	2	2	2	2
Underidentification K-P LM Test	$\chi^2(2)$=572.6***	$\chi^2(2)$=527.1***	$\chi^2(2)$ =399.5***	$\chi^2(2)$=784.8***
Hansen J statistic	$\chi^2(2)$=0.072	$\chi^2(2)$=0.440	$\chi^2(2)$=0.505	$\chi^2(2)$=0.002

Note: Morning=6am-11am; Evening=4pm-10.15pm; Winter=November-March; Summer=April-October. Bootstrap Standard Errors (SE) are reported in parenthesis, clustered by route and week. 250 repetitions. Significant at *10%, ** 5%, and *** 1%. K-P=Kleibergen-Paap.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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	0.0332	(0.001)***	0.0285	(0.001)***	0.0332	(0.001)***	0.0296	(0.001)***
Booking Day1	0.4235	(0.085)***	0.4528	(0.069)***	0.3355	(0.087)***	0.4596	(0.068)***
Booking Day4	0.0974	(0.082)	0.1651	(0.066)**	0.0687	(0.085)	0.1513	(0.065)**
Booking Day7	-0.1728	(0.080)**	-0.0140	(0.063)	-0.1380	(0.082)*	-0.0744	(0.063)
Booking Day10	-0.2134	(0.078)***	-0.0322	(0.059)	-0.1518	(0.078)*	-0.1071	(0.060)*
Booking Day14	-0.4085	(0.076)***	-0.1321	(0.055)**	-0.2827	(0.075)***	-0.2505	(0.057)***
Booking Day21	-0.3638	(0.071)***	-0.0922	(0.051)*	-0.2028	(0.070)***	-0.2176	(0.053)***
Booking Day28	-0.2630	(0.070)***	-0.0433	(0.046)	-0.1154	(0.066)*	-0.1521	(0.050)***
Booking Day35	-0.1094	(0.068)	-0.0552	(0.041)	-0.0661	(0.063)	-0.0944	(0.047)**
Booking Day42	-0.0725	(0.067)	-0.0770	(0.040)*	-0.0398	(0.064)	-0.0979	(0.047)**
Booking Day48	-0.0792	(0.070)	-0.0109	(0.040)	-0.0242	(0.062)	-0.0552	(0.045)
Booking Day56	-0.0697	(0.084)	0.0016	(0.041)	-0.0723	(0.072)	0.0214	(0.045)
Booking Day63	-0.0249	(0.073)	0.0032	(0.041)	0.0246	(0.068)	-0.0239	(0.045)
Tobit residual	0.0001	(0.001)	-0.0015	(0.001)**	-0.0013	(0.001)*	0.0001	(0.001)
DUMMIES:								

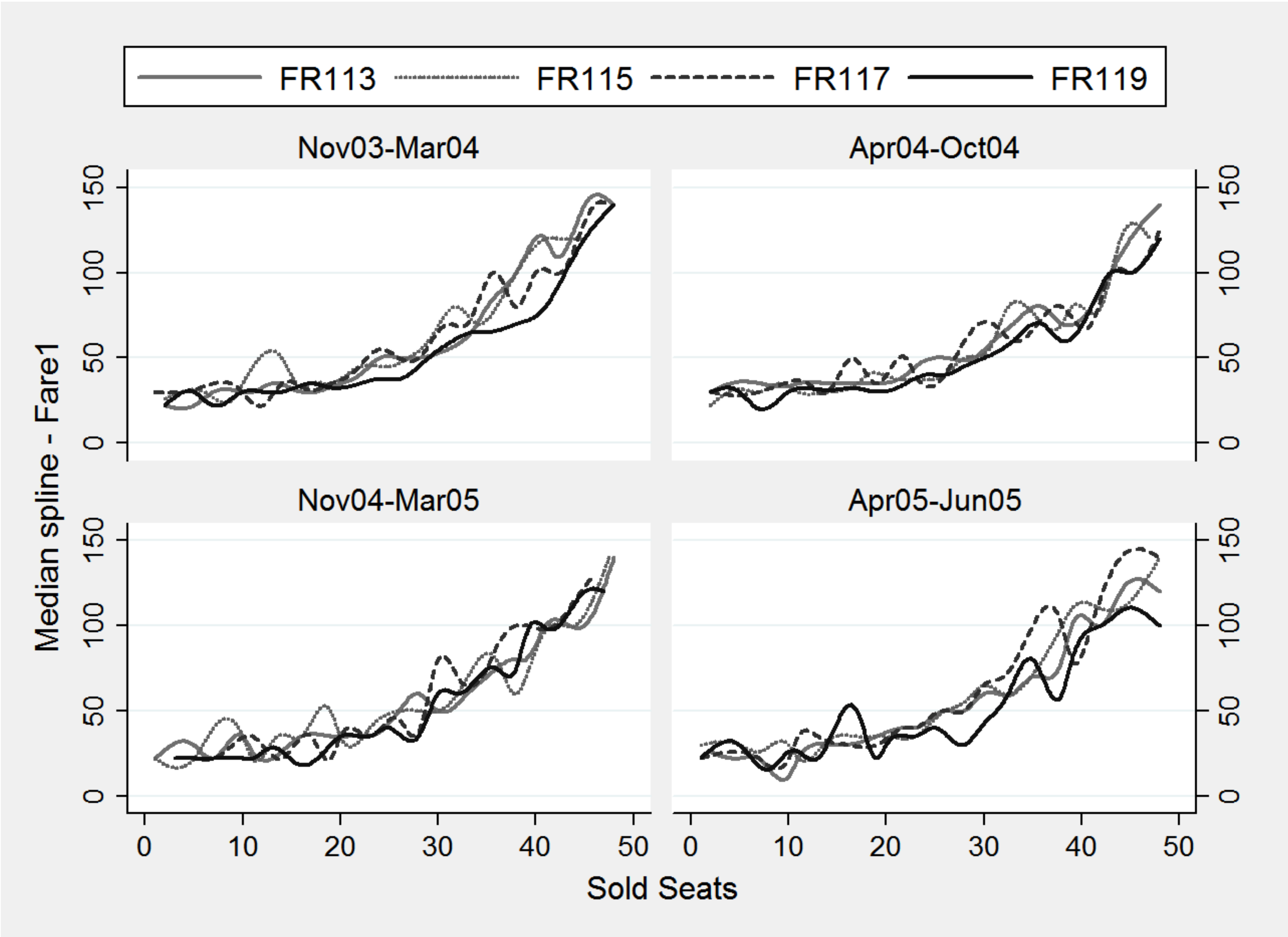
Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

Month booking	YES	YES	YES	YES
Number of obs.	38,332	52,523	41,536	58,495
Centered R2	0.6459	0.5115	0.5499	0.5825
Excluded instruments:	2	2	2	2
Underidentification K-P LM Test	$\chi^2(2) = 457.4^{***}$	$\chi^2(2) = 575.4^{***}$	$\chi^2(2) = 621.4^{***}$	$\chi^2(2) = 584.1^{***}$
Hansen J statistic	$\chi^2(1) = 2.141$	$\chi^2(1) = 0.322$	$\chi^2(1) = 0.698$	$\chi^2(1) = 1.254$

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. Significant at *10%, ** 5%, and *** 1%. K-P=Kleibergen-Paap.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

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 legend

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00451

Figure 2: Nonparametric fit between percentage deviation from mean $Fare1$ and percentage deviation from mean occupancy