Pricing of the long-distance bus service in Europe: The case of Flixbus

Article *in* Economics of Transportation · September 2019

DOI: 10.1016/j.ecotra.2019.100120

3 authors, including:

Alberto A. Gaggero
University of Pavia

Branko Bubalo
University of Hamburg

Pricing of the long-distance bus service in Europe: the case of Flixbus*

Alberto A. Gaggero† — University of Pavia

Lukas Ogrzewalla — University of Hohenheim

Branko Bubalo — University of Hamburg

June 18, 2019

Abstract

We study empirically the price dynamics in the long-distance bus market using posted fares by Flixbus, the market leader in Europe. We find that, at a given point in time, the fare increases with the number of sold seats. This result largely explains, why the lowest available fare increases as the departure date approaches. No evidence is found in favor of intertemporal price discrimination, probably because of low consumer-heterogeneity throughout the entire booking period that characterizes the long-distance bus market.

JEL classification: L92; L11; D22, C13

Keywords: capacity effect, Flixbus, dynamic pricing, revenue management, temporal effect.

1 Introduction

A long-distance bus service is a public transport service that carries people by bus between cities. The buses that are used generally make a single stop at one location in (or nearby) a city and travel non-stop over longer distances. For this reason, *intercity* bus service is also used as a synonym of *long-distance* bus service.

*We are grateful to two anonymous reviewers, whose insightful comments contributed to improve this paper. We thank the participants to 2nd TEM Conference at Bard (AO). All errors are ours.

†Department of Economics and Management, University of Pavia, Via S. Felice 5, 27100 Pavia, Italy. Tel +39 0382 986226; fax +39 0382 986228; email: `alberto.gaggero@unipv.it`. *Corresponding author.*

In Europe the long-distance bus service sector has grown steadily during the recent years. One of the prominent European countries in this industry is Germany (Dürr and Hüschelrath, 2017). It is not a coincidence that Flixbus, the current market leader of long-distance bus service in Europe, was founded in Germany in 2011.

Following the market liberalization in January 2013, the German long-distance bus service sector showed an exponential growth in traffic.[1] The fast establishment of long-distance bus services in the market of passenger transportation is achieved by providing customers with low-cost travel options in a dispersed network of destinations. After a vibrant introduction phase characterized by tough pricing, fierce competition, mergers and acquisitions, the lead of the market is taken by Flixbus, whose market share in Germany is 95% in 2018 (Statista, 2019b).

From its very beginning, Flixbus rests on the increasing capabilities of digital business models. This is reflected by its web-based selling strategy in which customers book their bus journeys with their mobile phone application or on the company's web page. In this framework, Flixbus can instantaneously track the evolution of sales and can adjust fares over time. Bus travelers may therefore face different prices for the same journey depending on the day of their booking and the utilization of available seat capacity.

The reason for such possible price difference is a profit-maximizing strategy referred to as Revenue Management (RM). Such strategy was first implemented in the airline industry and it is now also adopted by railways, cruise lines, hotel firms, rental car firms, and also by long-distance bus service providers.

The airline RM literature suggests that during the booking period two effects could be in place: the *capacity effect* and the *temporal effect* (Alderighi et al., 2015). The *capacity effect* occurs when, at a given point in time, fares have a stepwise distribution over seats (Dana, 1999a). This means that the first seat on sale is cheaper, or at most equally priced, relative to the second seat on sale, which is cheaper, or at most equally priced, relative to the third seat on sale, and so on. If the capacity effect is in place, we observe an increasing pattern of the lowest available fare as we approach the departure date because the bus fills up day by day and only high-fare tickets remain on sale at the end of the booking period.

The *temporal effect* instead is driven by the time dimension and is captured by the day of the booking. The fare variation as the departure date approaches is explained with: *i)* the seat's decreasing option value, which pushes the fare down in order to spur the sale of the seat that otherwise run the risk to remain unsold (Alderighi et al.,

[1] 2.5 million passengers carried in the pre-liberalization year jump to 8.2 millions in 2013 and reach 23.2 millions in 2015 (Statista, 2019a).

2017); *ii)* the price discrimination strategy, which pushes the fare up, especially in the proximity to departure, in order to extract higher surplus from price-insensitive late bookers (Gaggero, 2010). This latter effect is also known as 'intertemporal price discrimination' (Gaggero and Piga, 2011).

In this work we investigate the existence of the above described effects in the long-distance bus market using primary fare data from Flixbus on selected one-way bus journeys taking place in one week of September 2017. Fares are collected daily, from 28 days until 1 day before departure, so that it is possible to observe their variation as the departure date approaches.

We apply panel fixed-effects techniques and correct for both endogeneity and sample selection. We find significant evidence in favor of the capacity effect, which largely explains the fare increase as the departure date approaches. This effect is, to some extent, mitigated by the seat's decreasing option value that comes into play during the last week before departure. No evidence is found in favor of intertemporal price discrimination, probably because of the very small number of price-insensitive late booking passengers in the long-distance bus market.

This paper continues as follows. The next section revises the relevant literature. Section 3 explains the process of data collection, followed by a descriptive analysis in Section 4. Section 5 conducts the econometric investigation with the description of the model, the discussion of the econometric issues and the illustration of the results. Finally, the concluding remarks are made in Section 6.

2 Literature review

Because RM was first applied by airline companies, the first papers studying RM pertain to the airline sector (Belobaba, 1987; Kimes, 1989). Subsequently RM has been extended to other sectors, such as: railways (Bharill and Rangaraj, 2008; Berto and Gliozzi, 2018), car rentals (Carroll and Grimes, 1995; Haensel et al., 2012), cruises (Maddah et al., 2010; Sun et al., 2011), hotels (Abrate et al., 2012; Guizzardi et al., 2019), and also long-distance bus service (Augustin et al., 2014; Grimaldi et al., 2017).

The analogies that can be found in the pricing behavior of all these industries are due to the fact that they all rely on the same RM techniques, which are not specifically developed for one sector.[2] Nevertheless, we can also find some differences

[2]RM is a pricing strategy that is applicable when the firm offers a limited capacity of a perishable service or good, the demand is uncertain, and customers are heterogenous in terms of their price-sensitivity (Weatherford and Bodily, 1992).

amongst industries that may affect the way in which the RM strategy is conducted. For example, the existence of intermediate stops in a long-distance bus journey makes this transport service more similar to rail transport or to the delivery of goods by truck rather than to air transport. One bus seat sold for the A-D journey with intermediate stops in B and C means one seat less on sale for the intermediate segments A-B, A-C, B-C and B-D. This feature makes the calculation of the shadow price of capacity more difficult, but the process of pricing a seat remains unchanged. A seat sold today cannot be sold at a higher price later on. For this reason, a fare set for today needs to account for the lost opportunity of selling the seat at a later date (Dana, 1999a).

Because a seat is a highly perishable product, its option value tends to zero as the departure comes closer. Therefore, the theoretical model of Talluri and Van Ryzin (2004) predicts a declining time-path of fares due to a falling option value. Strategic customers may, however, recognize the falling option value of seats and postpone their bookings in the hope of last-minute offers (Deneckere and Peck, 2012). In order to discourage such behavior, the company commits to fares increasing over time (Möller and Watanabe, 2010).

Alderighi et al. (2017) show that pricing in the airline industry hinges on a fare distribution which consists in pre-assigned seats to distinct fare classes, so-called 'fare buckets'. A bucket is described by a set of consecutive seats with the same fare tag. Based on the forecasts, the firm decides the set of fare classes and the number of seats contained in each bucket before the actual booking is opened. The resulting fare distribution determines the systematic fare changes as the airplane fills up. This effect in airline literature is sometimes referred to as 'systematic peak-load pricing' because in case of strong demand higher fares allow to shift some consumers from peak to off-peak times (Borenstein and Rose, 1994).

Dana (1999a) derives an optimal fare distribution from a theoretical point of view. Price dispersion is explained in a model with demand uncertainty, costly capacity (perishable assets), and price commitments. Before the actual demand is known, the firm is indifferent between selling a seat at a guaranteed low price or selling it at a higher price with uncertainty. The uncertainty emerges from the fact that the firm does not know whether it will face low or high demand during the booking period. The idea behind Dana's model is that the firm sets its optimal fares by adding a fixed markup to the cost of capacity. The cost of capacity can be best understood as the cost of 'reserving' seat inventory for the uncertain event of high demand. That is, the cost of capacity is inversely related with the chance of selling it. From this feature it follows that the optimal fare distribution shows an increasing profile as capacity (i.e. available aircraft seats) fills up. That means price dispersion cannot only be attributed to the need of segmenting customers, but also to the uncertain nature of demand. Alderighi et al. (2015) tackle this issue empirically in the airline industry and find support for fares increasing with a flight's load factor.

As discussed by Borenstein and Rose (1994), there also exists a 'stochastic peak-load pricing' component, which accounts for changes in the expected demand during the booking period. Therefore, there may be the need to adjust the fare distribution once the actual demand is revealed. Bilotkach et al. (2015), for example, show that RM interventions of low-cost carriers are more frequent when the actual demand deviates from the forecast.

To sum up, the RM literature shows that there can be two causes of price variation over time, loosely referred to as *dynamic pricing*. First, there is a systematic part of dynamic pricing accomplished by a predefined fare distribution, which allows to implement peak-load pricing to shift demand between peak and off-peak times; furthermore, the increasing stepwise slope of a fare distribution accounts for the shadow cost of capacity. Second, there is a stochastic part of dynamic pricing that deals with the adjustments of the fare distribution during the booking period to respond to changes in the expected demand.

The German Federal Office for Goods Transport indicates that bus service providers engage in dynamic pricing (Federal Office for Goods Transport, 2016, p. 34). In the review of the German long-distance bus market, it is observed that fares are the most expensive on the day before departure. Moreover, fares tend to be less expensive during off-peak times. It is therefore not so much a question of whether dynamic pricing is applied in the long-distance bus service industry, but rather to understand the prevailing effect that drives the evolution of bus fares during the booking period.

The literature on pricing of long-distance bus services is quite recent and therefore largely unexplored. Sampling observations for six European countries Fageda and Sansano (2018) find that long-distance bus fares are cheaper between high-income cities. This result, combined with the finding that higher bus frequency is observed between high-income cities, implies that low-income cities experience both higher prices and lower quality of service. In a fare comparison between a mature long-distance bus market (U.S.) and a young long-distance bus market (Germany), Augustin et al. (2014) find that distance and number of stops are significant determinants of the ticket price, while the bus frequency is only a relevant explanatory variable in the German market.

The work by Blayac and Bougette (2017) shows that long-distance bus operators adopt an aggressive pricing strategy when they start a new link between cities in order to induce demand for the new services, then, once the service becomes popular to more customers, they increase fares. Dürr et al. (2016) analyze the impact of competition on the fares of the German long-distance bus market using a sample of bus services operated during one week in November 2014 and one week in January 2015. In line with the predictions of economic theory they find that route-level average prices tend to decrease with competition, i.e. the number of active bus companies on a particular route. Their empirical analysis hinges on secondary fare data collected seven days before departure from an online search engine, thus little attention is given to the evolution of fares during the booking period.

An in-depth analysis on the evolution of long-distance bus fares over time lacks in the literature, which is largely devoted to airline pricing. The aim of the present paper is to fill this gap and expand the empirical research of RM to the long-distance bus service industry.

3 Sample collection and description

This paper is based on primary data directly retrieved from Flixbus' website. Our sample considers four domestic city pairs with both endpoints in Germany (Brunswick-Düsseldorf, Hanover-Bremen, Munich-Nuremberg and Tübingen-Freiburg), plus one international city pair linking Germany to Italy (Stuttgart-Milan). Each city pair may include more than one route.[3]

The database comprises information on the posted fares of one-way bus journeys scheduled in the week from September 4, 2017 to September 10, 2017. The fares are collected on a daily basis over a period of four weeks prior to the departure date. It is worth noticing that Flixbus offers only one ticket class, therefore all tickets carry the same characteristics.[4] This feature constitutes a main advantage of our dataset as it makes our data fit very well the settings described by various theoretical RM models (Dana, 1999a; McGill and van Ryzin, 1999; Talluri and Van Ryzin, 2004).

We acknowledge the limits of using a sample made by a relatively small number of bus connections, however, the selected city pairs represent the population of Flixbus routes very well, which are heterogeneous in terms of distance covered by the bus services, the frequency of departures, and the competitive environment.

[3]In our sample the city pair Munich-Nuremberg comprises the route Munich(Froettmaning)-Nuremberg and the route Munich(Central)-Nuremberg; the city pair Stuttgart-Milan includes Stuttgart(North-Milan) and Stuttgart(Airport)-Milan; the other city pairs are linked by only one route each.

[4]The collected fares include tax and handling fees. Furthermore, all tickets include hand luggage (max 7 kg) and a baggage allowance of 20 kg. However, all fares are net of add-ons and other fees, such as possible charges for the method of payment. Fares are for one-way bus journeys and quoted in euro.

Table 1: City pairs included in the analysis

City pair	Nbr. weekly services	Distance (km)	Travel time (hrs:min)	Alternative direct connection
Brunswick - Düsseldorf	17	354	5:20-5:40	Train
Hanover - Bremen	29	128	1:35	Train
Munich - Nuremberg	153	169	1:50-3:10	Train
Stuttgart - Milan	32	505	6:45-9:45	Airplane
Tübingen - Freiburg	37	159	2:25-4:04	None

NOTE. Distance is retrieved from Google Maps and measures the length of the shortest road itinerary connecting the city pair. Information on alternative travel options is retrieved from the online travel portal goeuro.de.

The data collection follows the approach adopted by Alderighi et al. (2017), in which the fare queries for a particular journey are iterated over an increasing number of seats.[5] By doing so, we are able to retrieve fares up to a maximum of 40 seats, the largest possible seat reservation for a single booking on Flixbus' website. This allows, at least partially, to determine the number of available seats on a given bus journey at a given day to departure.

Table 1 provides an overview of the sample by including for each city pair the distance, the bus travel time and the presence of alternative means of transportation. Because Flixbus has the dominant position on the long-distance bus market, the competitive environment for Flixbus is affected by the availability of close substitutes, i.e. direct connections offered by train and plane on the route (Federal Office for Goods Transport, 2016, p. 28).

When the long-distance bus services started, competition mainly took place on journeys to large German cities where bus service providers may expect high demand and revenues (Federal Office for Goods Transport, 2016, p. 1). In line with this presumption, the city pair Munich-Nuremberg exhibits by far the largest number of weekly connections in our sample (153). On the contrary, Flixbus offers only 17 weekly services on the low-demand market Brunswick-Düsseldorf. Most interestingly, also Deutsche Bahn (German Railways) have not identified a sufficiently large demand for this city pair, since they only operate few direct connections on the line Brunswick-Düsseldorf. These findings confirm the assessment of the Federal Office for Goods Transport arguing that long-distance bus service providers mainly compete for customers on highly attractive routes of long-distance rail services (Federal Office for Goods Transport, 2016, p. 50).

[5] Appendix A.1 describes in detail how the fare data are processed.

There is a moderate number of bus services (29) from Hanover to Bremen, even though both cities have more than half a million citizens. Most interestingly, a similar bus frequency (32) is observed between the much smaller towns of Tübingen and Freiburg: this result suggests that the supply of long-distance bus service is not necessarily based on the city's size. In fact, a large number of students live in Tübingen and Freiburg, whose high demand for affordable travel options induces Flixbus to provide more bus services on this city pair. Furthermore, Tübingen-Freiburg represents a particular case in the sample, since there is no direct railway service linking the two towns. For this reason, on this city pair Flixbus has a competitive advantage, not only in terms of a higher travel comfort offered by a direct link, but also in terms of a shorter travel time.[6]

Finally, on the city pair Stuttgart-Milan there is no direct connection by train, but easyJet serves this route on a daily basis. Because the business model of low-cost carriers is also based on the presence of a large portion of price-sensitive travelers, the competitor of Flixbus on this city pair is to be sought in aviation rather than in rail.

4 Descriptive analysis

4.1 The fare distribution

An important aspect of the RM strategy, which is the hallmark of airline pricing, is to define the different fare classes to which transport service providers assign a predetermined number of seats. This practice establishes a sequence of ascending fares with respect to the seats. The result is a fare distribution with the seats belonging to the same fare class gathered in groups called 'buckets'.

Figure 1 provides an example of a fare distribution set by Flixbus for the bus service Stuttgart(North)-Milan leaving Stuttgart at 10:40 and arriving in Milan at 19:05 on September 10, 2017. The fare distribution is depicted at different days to departure in order to gain a first insight on the temporal evolution of fares as the departure date approaches.

The top-left diagram shows the fare distribution 28 days prior to departure and depicts a situation in which there are at least 40 seats available, i.e. data are censored. One week later (top-right diagram) the seats on sale are 39, i.e. data are uncensored,

[6]Consider as a example the following two travel options for Tübingen-Freiburg with a similar departure time on weekdays: *i)* departure at 8:00 am, travel time of 2 hours and 54 minutes with two changes by train; *ii)* departure at 7:35 am, travel time of 2 hours and 25 minutes with a direct link by Flixbus.

and we are able to observe the entire fare distribution. The first 13 seats are on sale at €25.00 and represent the first bucket, then the second bucket is given by the 14^{th} up to the 30^{th} seat, priced €27.90. The 31^{st} up to the 35^{th} seat make the third bucket with the price of €32.90, followed by one-seat buckets, sequentially priced €37.90, €42.90, €49.90 and €59.90.

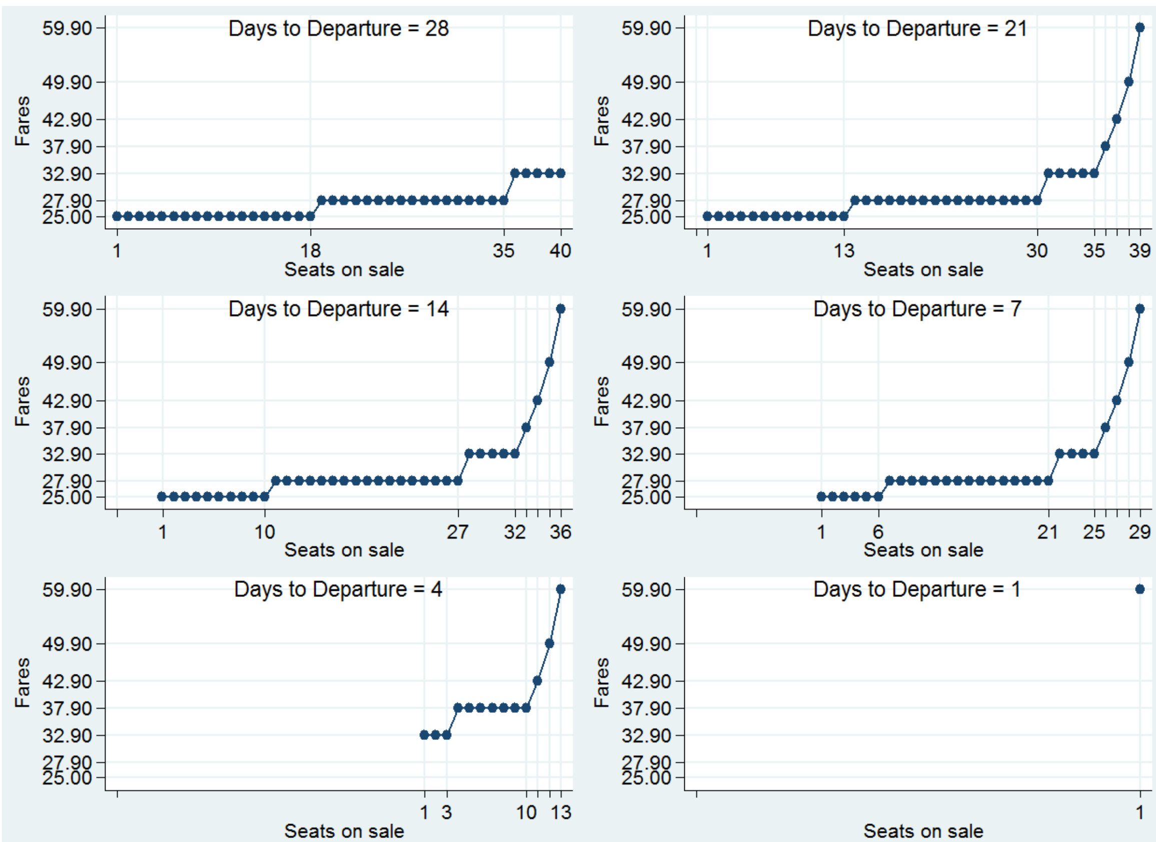

Figure 1: Fare distribution at different days to departure, Stuttgart(North)-Milan.

The lowest available fare is at the very left of the distribution and equals €25.00. From Flixbus' website, we retrieve the information that €25.00 is the starting fare for bus journeys from Stuttgart to Milan. This implies that we have observed the very first bucket, even though we could not estimate its initial size; the last bucket, which consists of one single seat in the present case; and all intermediate buckets. It is worth noticing that the top fare, i.e. the fare of the last bucket, is €59.90 which is 140 percent more expensive than the €25.00 starting fare. This will eventually translate into intertemporal price dispersion over the entire booking period (Gaggero and Piga, 2011).

In all diagrams the fare distribution is displayed by an increasing function of capacity utilization, which is in line with the empirical evidence of the airline industry (Alderighi et al., 2017). It should be pointed out, however, that while for airlines the size of each fare bucket is usually similar, in the case of Flixbus the seats included in lower-priced buckets often outnumber those ones contained in higher-priced buckets, as shown by the two top diagrams of the figure.

Dana (1999a) argues that the optimal fare distribution is increasing because the cost of seat inventory is inversely related with the probability of selling it. The rare occurrence of sold out events in our sample (less than 4%) suggests that the cost of reserving seats for high demand events is relatively high in the long-distance bus market. This possibility is reflected in a particularly steep slope of the fare distribution, especially for the last (high-priced) seats, which have the lowest chance of being sold. In other words, as sold out is very unlikely in the long-distance bus service industry, Flixbus spares only a very limited number of seats for price-insensitive, late-booking customers.

Figure 1 provides also a telling example of both the systematic and the stochastic elements of dynamic pricing discussed in Section 2. Broadly speaking, the systematic part is related to the movements along the curve, while the stochastic part becomes evident by a reshape of the fare distribution itself. Considering the fare distribution at 21 and at 14 days to departure, we observe that the size of the first bucket decreases by three seats, while the fare distribution itself remains unchanged. In this case, three seats have been sold at €25.00 as pre-specified in the fare distribution, with the consequence that the size of the first bucket shrinks from 13 to 10 seats.

Things look different if we compare the fare distributions 14 and 7 days before the departure date. The lowest available fare is still €25.00 seven days prior departure, however, the structure of the overall fare distribution has changed in comparison with the fare distribution observed one week earlier. In particular, the size of the second bucket, priced €27.90, and the size of the third bucket, priced €32.90, are respectively reduced from 17 to 15 seats and from 5 to 4 seats. Looking at the difference in the total number of available seats, we infer that 7 seats are sold within this week despite the size of the first bucket only shrinks by 4 seats, from 10 to 6. These numbers indicate a reallocation of seats from the third and the second buckets to the first bucket of the fare distribution.[7]

[7]Airline literature suggests that such rearrangements of the fare distribution are used as a strategy to strike a balance between two conflicting goals. Shifting seats from higher to lower buckets implies a reduction in the average fare of the overall distribution, and thereby enables to account

If on the one hand we observe fares being shifted downwards in the event that demand falls short of expectation, it seems conceivable that on the other hand fares are moved from lower to higher buckets when demand is higher than expected. We can find such a scenario comparing the fare distribution at 7 and at 4 days to departure. In this three-day period 16 seats are sold. Interestingly, the €37.90 bucket comprises only one seat 7 days prior departure, while it is enlarged to seven seats 3 days later. The sharp and possibly unexpected high demand may be the reason triggering this upward shift in fares.[8]

To sum up, the observed reallocation of fares points to active RM interventions to account for stochastic demand shocks. This suggests that the pricing strategy of Flixbus may encompass both a systematic component (increasing stepwise fare distribution) and a stochastic component (adjustment of the fare distribution due to unexpected changes in the demand).

4.2 The fare on the first seat on sale

As evinced from Figure 1, the fare on the first seat on sale, henceforth simply 'fare', is the lowest available fare. This initial fare varies over time for a particular trip until scheduled departure time approaches. The changes occur along a predefined path, i.e. along a stepwise fare distribution, which can also be altered over time.

If the predictions of the theoretical model of airline literature apply to Flixbus, one of the reasons explaining the increase of fare over time is because less seats are available on the bus. The monotonic increasing fare distribution showed in Figure 1 supports this view: as the departure date approaches, more seats are sold, and, as a consequence, the first seat on sale climbs up the fare distribution.

The main summary statistics of the fare are reported in Table 2. The results are sorted by city pairs and include the minimum, the maximum, the median and the mean of the fare; the table also includes the mean of the fare over distance (i.e. mean per 100km) to make the statistic comparable across city pairs of different length.

The table suggests a correlation between fare and travel distance to a certain

for a declining option value of the seats as departure approaches. These changes occur 'hidden in the background', i.e. they are not visible to customers. Therefore, customers may observe fares increasing over time, even though the average fare of the overall distribution effectively declines. By doing so, strategic customers, who anticipate the declining option value of a seat, are deterred from postponing their bookings in the hope of last-minute discounts (Alderighi et al., 2017).

[8]Figure 1 shows an example in which one day before departure there is just one seat left for sale and its fare corresponds to the maximum of the fare distribution, i.e. demand is high and the bus is about to sell out at the departure date. In many cases, demand is lower and the fare on the first seat on sale never reaches the level of the maximum fare.

Table 2: Summary statistics of the fare on first seat on sale

City pair	Min	Max	Median	Mean	Mean per 100km	Obs.
Brunswick - Düsseldorf	15.00	29.50	15.00	18.52	5.23	476
Hanover - Bremen	5.00	19.50	5.00	6.41	5.01	804
Munich - Nuremberg	5.00	19.90	7.90	9.10	5.38	4,284
Stuttgart - Milan	25.00	59.90	25.00	26.23	5.19	1,288
Tübingen - Freiburg	8.00	22.90	9.90	11.10	6.98	1,010
Overall	5.00	59.90	9.90	12.46	5.51	7,862

degree. However, a comparison of city pairs that are about the same length indicates that competition may also play a role in the shape of the fare. Even though the city pairs Munich-Nuremberg and Tübingen-Freiburg have almost the same distance (respectively 169km and 159km), the average fare of the latter turns out to be higher than the one of the former. Moreover, Tübingen-Freiburg displays the highest average fare per 100 kilometers among all city pairs (€6.98). This result is not surprising as Tübingen-Freiburg is the only city pair of the sample in which Flixbus faces weak competition.

The ratio of the maximum to the minimum value of the fare ranges from around two (Brunswick-Düsseldorf: €29.50/€15.00) to around four (Munich-Nuremberg: €19.90/€5.00); that is, the fare is found to be widely spread across the booking period. In the sample the relative difference between the maximum and the minimum value is more pronounced for short-haul routes (Hanover-Bremen, Tübingen-Freiburg, and Munich-Nuremberg).

Most importantly, for each city pair the minimum of the fare coincides with the starting bid, i.e., the lowest possible fare for a particular bus journey. This means that for each city pair we are able to collect the very first fare class of the fare distribution. A graphical representation of the evolution of the fare over time for each city pair is presented in Figure 2, which plots the fractional-polynomial prediction of the fare over the days to departure. Exception for the bottom-right diagram, which offers an overall comparison across city pairs dividing the fare by the travel distance, each diagram refers to a given city pair. The color and the pattern (i.e. dashed, dotted, etc.) of the curves are consistent within the entire figure.

The pattern of the fare is similar in all the city pairs: it stays quite stable up to seven days to departure, then it starts raising in the last week and finally grows aggressively during the last four days to departure. This means that the fare is described by an exponential function of time, where time is count top down by the days to departure.

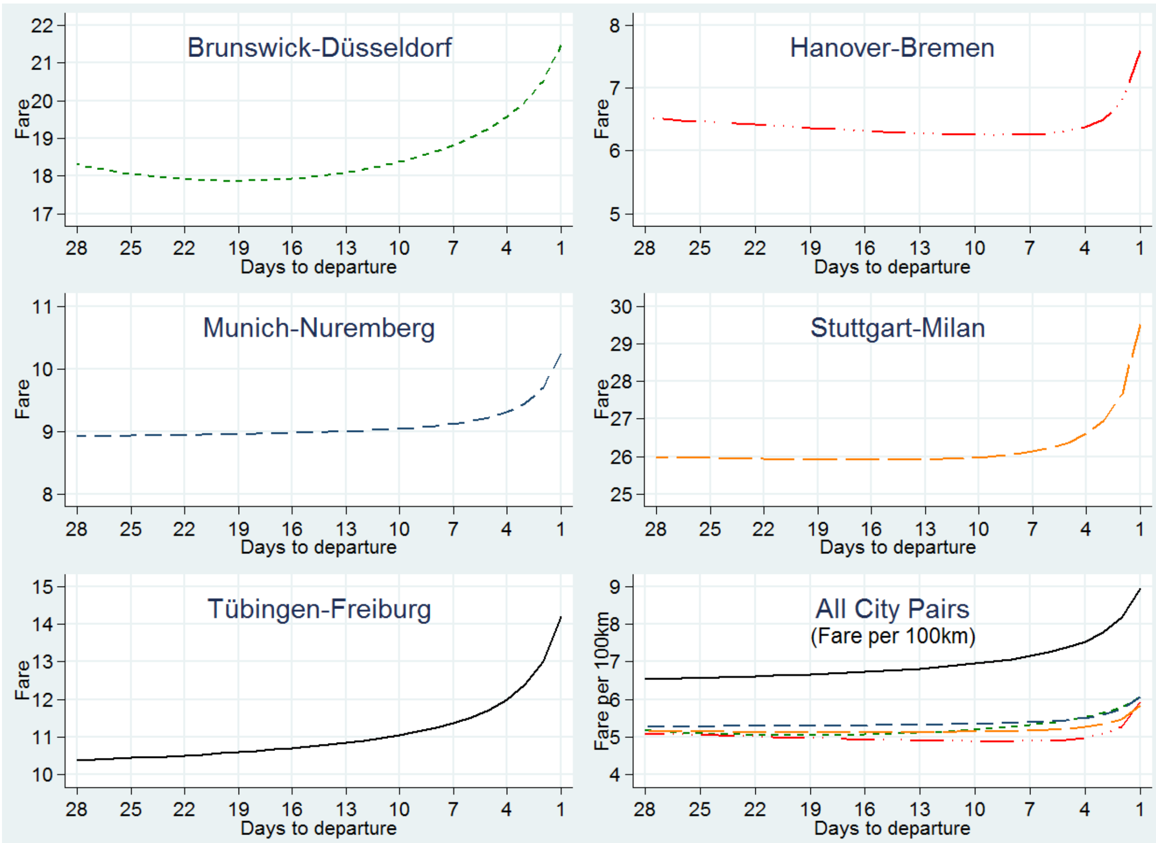

Figure 2: Fare and days to departure (fractional-polynomial prediction)

Economic theory suggests that the steep increase of the fare a few days before departure could be also explained by another effect than capacity. As departure date approaches more consumers resolve their demand uncertainty, and, hence, mobility service providers could rely on intertemporal price discrimination to segment customers differing with respect to the time they realize their need to travel which may reflect differences in their willingness to pay (Möller and Watanabe, 2010).

Empirical studies from the airline industry show that low-cost carriers employ intertemporal price changes to profitably exploit customers' heterogeneity (Gaggero, 2010; Alderighi et al., 2015, 2016). This strategy in particular allows to charge higher fares close to departure when price-insensitive business travelers usually turn up for booking. In contrast to the aviation industry, however, the demand for long-distance bus service largely consists of leisure passengers.[9] From this perspective, segmenting business and leisure travelers may appear less compelling to long-distance bus service providers. Nevertheless, as high-uncertainty, thus late booking, travelers in general exhibit a high valuation (Talluri and Van Ryzin, 2004), we cannot exclude that a fare increase in the last days before departure is also a sign that Flixbus intertemporally price discriminates.

Therefore, it remains unclear whether the increasing path of the fare is due to the lack of available seats (price goes up because the bus fills up) and/or to the mere proximity to departure (intertemporal price discrimination). The next section aims to clarify this issue with the aid of econometrics.

5 Econometric analysis

5.1 Model

In order to model the behavior of the fare on the first seat on sale we need to consider variables that can be related to the changes of fares. Findings from airline pricing suggest that the dynamic pattern of fares can be described by both a capacity component and a time component (Alderighi et al., 2017).

To investigate whether these forces are also into play in the long-distance bus service industry, we include two types of regressors. The regressor accounting for the capacity component is denoted by *AvailabeSeats*, which represents the number of unsold seats on a bus at given day of booking; the time dimension is identified by *DaysToDeparture*, which measures the temporal distance (in days) between the fare query and the departure date.

The pricing model we estimate to separate capacity-based from time-based effects is similar to the one proposed in Alderighi et al. (2015) and is summarized by the following equation:

$$Fare_{jt} = \beta_0 + \sum_{k=1}^{7} \beta_k DaysToDeparture_{kt} + \beta_8 AvailableSeats_{jt} + a_j + u_{jt} \qquad (1)$$

where the subscript j identifies a bus journey, which represents the panel individual, uniquely defined by the route, date and departure time; the day of query t

[9]The long-distance bus passengers traveling to visit friends & relative plus the passengers traveling for leisure and tourism are to about 83% of the total demand (IGES Institute, 2014).

represents the time dimension of the panel and spans from 1 to 28 days. The dependent variable $Fare$ is the log of the fare on the first seat on sale;[10] the independent variables are summarized by $AvailableSeats$ and $DaysToDeparture$; a_j represents the panel fixed-effects and u_{jt} the error term, assumed i.i.d. with zero mean.[11]

The $DaysToDeparture$ variable is decoded by a set of dummies which denote when a fare query is made. Our dummy categorization gathers a sequence of consecutive days to departure on a week spell, except for the last week which comprises five dummy variables: thus, we split $DaysToDeparture$ into a total of eight dummies denoting the fare query carried 1, 2, 3, 4-5, 6-7, 8-14, 15-21, 22-28 days to departure.[12] The baseline for evaluating fare changes over time is the point furthest away from departure, *DaysToDeparture 22-28*.

Two possible effects of $DaysToDeparture$ working in the opposite direction may exist in this context. First, we could observe a positive sign on the $DaysToDeparture$ dummies to indicate that fares increase with the departure date coming closer. If this result still holds when we control for capacity, it would lend support to intertemporal price discrimination.

Second, once we control for capacity, it is also possible to observe a negative sign on the $DaysToDeparture$ dummies. In a highly price-sensitive market, such as the long-distance bus market, it may be reasonable to lower prices if there is a considerably large number of unsold seats close to departure. This result would indicate the existence of a decreasing option value of the seats.

As far as the capacity effect is concerned, we expect a negative sign on the estimated coefficient of $AvailableSeats$. This result would confirm that the fare increases as the bus fills up, i.e. the fare increases with $AvailableSeats$ decreasing.

[10] As showed by Figure 2, the fare on the first seat on sale is an exponential function of the days to departure, hence taking the logarithm log-linearizes the variable.

[11] Note that frequency is not included in the model since the bus journeys of the present analysis are scheduled in the period September 4-10, 2017. In this respect, equation (1) may be considered a short-run model.

[12] In principle we could also use one dummy variable for each day to departure. However, there are two main reasons that speak against doing so. First, a check of the data suggests that there is only little variation of the fare on the first seat on sale if *DaysToDeparture* takes on higher values (i.e., the more time remains between booking and departure date). Hence, for high values of *DaysToDeparture*, a 'daily' dummy variable would add little information to our model. Second, the more dummies are used, the higher the number of parameters to be estimated in the regression, and therefore also uncertainty in the model as standard errors tend to increase with the number of regressors (Wooldridge, 2013, p. 93f.). These considerations lead to the above choice of the set of dummy variables.

5.2 Econometric issues

Before estimating the model there are two issues on *AvailableSeats* and one issue related to possible shocks hitting the demand of all the bus journeys of a given city pair that need to be considered.

Equation (1) comprises the panel fixed-effects term a_j, which captures journey-specific characteristics and time-invariant characteristics. Examples include bus-specific features such as departure time and coach model; route-specific features such as: competition and distance; city-specific features such as low/high income city and tourism/business destination.[13]

Among these, the competitive environment on a particular route is likely to affect the fare and be correlated with the independent variables, thereby giving rise to endogeneity concerns (e.g., a lower competition on a route impacts both the fare level and the number of sold seats, which eventually determine *AvailableSeats*).

The inclusion of the panel fixed-effects allows to fully control for all (i.e. observed and unobserved) panel time-invariant characteristics and tackles the omitted variable bias that would arise if time-invariant, journey-specific characteristics were ignored.

However, even with the use of panel fixed-effects, *AvailableSeats* may still be endogenous because the fare distribution could be subject to discretionary interventions by the RM analyst. For example, fares may be shifted downwards (upwards), if there are more (less) seats unsold than expected. For this reason, the correct specification of the model would require including the RM decision amongst the set of regressors. Because the logic triggering such RM behavior is unobservable, this effect falls into the regression error term u_{jt}. The endogeneity that arises here stems from omitting a variable that explains the link between *AvailableSeats* and discretionary RM interventions (Alderighi et al., 2015).

We tackle the endogeneity under omitted variable by means of an instrumental variable approach. A valid instruments z must satisfy two conditions. First, z must comply the instrument exogeneity, which states that z is exogenous in equation (1), i.e., $Cov(z, u) = 0$. Second, z must satisfy the instrument relevance, which implies that z is able to explain the variation in *AvailableSeats*, i.e., $Cov(z, AvailableSeats) \neq 0$.

Thus, we need to find at least one instrument which is correlated with *AvailableSeats*, but uncorrelated with the error term u_{jt}. In the spirit of Alderighi et al. (2015) and Bilotkach et al. (2015), we identify one instrument for the endogenous capacity component that can suit our model. The instrument is a dummy variable that

[13] Table A.3.1 in Appendix A.3 reports the pooled ordinary least squares estimation in which within-group time-invariant variables are included.

indicates whether a fare is collected during the weekend, $WeekendBook$. The underlying idea is that manual RM interventions are less likely during weekends, since limited RM staff is available; moreover, the booking behavior is also likely to be different during weekends.[14] This makes $WeekendBook$ in theory a valid instrument for $AvailableSeats$; formal tests in favor of the relevance and exgogeneity of $WeekendBook$, reported in Appendix A.2, confirm the instrument validity also in practice.

Besides endogeneity, another possible concern related to $AvailableSeats$ is that this variable is censored at 40 seats. This is due to the data collection procedure described in Section 3 and implies that $AvailableSeats = 40$ provides only the information that there are at least 40 seats available for booking.

Studies related to airline pricing tackle censoring of available seats by means of Tobit specifications (Alderighi et al., 2015) or Probit (Alderighi et al., 2017). The procedure consists in computing the regression residual (Tobit approach) or the inverse Mills' ratio (Probit approach) from an auxiliary regression in which the dependent variable is $AvailableSeats$. Then the main model is estimated using only the uncensored observations and including the Tobit residual or the inverse Mills' ratio among its regressors to correct for the sample selection (Heckman, 1979).

In general, however, we expect censoring at 40 seats to be a minor issue in our sample, since the difference between the censored and the actual value of $AvailableSeats$ is very likely to be smaller in buses than in airplanes: 40 seats represent only a fraction of an airplane's capacity, while they are relatively close to the maximum capacity of a bus.

Because the techniques correcting for censoring reduce the number of observations and because our sample comprises a rather small amount of uncensored observations, we deem that the costs of correcting for the sample selection outweigh the benefits. Nevertheless, the results correcting for sample selection are presented in the analysis. We run a model with Tobit specification following procedure 17.4 in Wooldridge (2002, p. 573ff.); among the regressors of the Tobit specification we include the day of the week of departure. Similar qualitative results are obtained when the correction for sample selection is implemented with Probit (not reported to save space but available upon request.)

[14] Alderighi et al. (2015) and Bilotkach et al. (2015) also use another instrument that captures the flight-specific slope of the fare distribution. For each day to departure this instrument is constructed using the lagged values of the preceding weeks for a particular day of departure. Because we have only bus journeys for a week in our sample, we cannot construct the second instrument (e.g., we observe the bus journeys leaving on Monday September 4th, 2017 and not the bus journeys leaving on the previous Mondays).

Finally, standard errors are clustered by city pairs, to allow the residuals of different bus journeys on the same city pair to be correlated. This choice of clusters takes care of possible shocks that are common to all journeys on a specific city pair and that may hit the demand (e.g. a famous festival taking place in one city, temporary road works delaying the bus journey, etc.).

5.3 Results

The results of our empirical analysis are summarized in Table 3. Five different models are reported. Models (1) and (2) stem from an ordinary least squares (OLS) regression. Model (3) is obtained from an instrumental variable (IV) approach in which $WeekendBook$ is used as instrument. Model (4) comes also from the same IV procedure, but it only considers uncensored observations in order to be comparable to Model (5), which originates from the aforementioned procedure of sample selection correction.

The coefficient of $AvailableSeats$ is negative and statistically significant in all estimations. This result supports the idea that the number of unsold seats is negatively related to the fare or, in other words, that the fare is set as an increasing function of capacity utilization.

The magnitude of this effect varies among models. The OLS estimation of Model (2) yields the lowest coefficient, where the sale of an additional seat is associated with an increase in fare by 1.3 percent on average.[15] Using the IV approach returns a slightly more pronounced capacity effect. In Model (3) an additional sold seat implies an increase of fare by 2.5 percent. At the mean value of the fare on the first seat on sale (€12.46, see Table 2) this effect quantifies in a price increase of around €0.31.[16] This finding indicates that the increase in fare, once a seat is sold, is moderate on average. Such a conclusion is consistent with the descriptive analysis of the previous section: because the majority of passengers are price-sensitive leisure travelers, they would probably refrain from traveling, should they see a sharp increase in fare. In this respect, the effect of capacity on fares must not be excessive.

The second effect on fare is the day of the booking, which is captured by seven $DaysToDeparture$ dummies. Model (1) shows that the fare on the first seat on sale follows an increasing temporal profile also in our sample, in line the pattern docu-

[15]Since equation (1) is a log-level model, the marginal effect of $AvailableSeats$ on the dependent variable is obtained as $\%\Delta Fare = 100\beta_1 \Delta AvailableSeats$, where $\%\Delta Fare$ is the percentage variation of the first seat on sale. Note that selling one seat means $\Delta AvailableSeats = -1$.

[16]The computation using the estimates from Model (4) and (5) yields very similar results, demonstrating our presumption that censoring is a minor issue in our sample.

Table 3: Results

	(1) OLS-FE	(2) OLS-FE	(3) IV-FE	(4) IV-FE	(5) IV-FE
Days to departure 1	0.145***	-0.053**	-0.245*	-0.395***	-0.598***
	(0.025)	(0.014)	(0.126)	(0.106)	(0.177)
Days to departure 2	0.080**	-0.046**	-0.170**	-0.297***	-0.471***
	(0.023)	(0.013)	(0.084)	(0.086)	(0.147)
Days to departure 3	0.063**	-0.026**	-0.113*	-0.215***	-0.381***
	(0.016)	(0.008)	(0.061)	(0.074)	(0.128)
Days to departure 4-5	0.036*	-0.017*	-0.068*	-0.141**	-0.287***
	(0.014)	(0.007)	(0.039)	(0.065)	(0.108)
Days to departure 6-7	0.024	-0.004	-0.031	-0.070	-0.202**
	(0.011)	(0.008)	(0.022)	(0.055)	(0.090)
Days to departure 8-14	0.012	0.001	-0.009	-0.014	-0.088
	(0.008)	(0.007)	(0.010)	(0.050)	(0.065)
Days to departure 15-21	0.001	-0.001	-0.004	0.004	-0.014***
	(0.005)	(0.005)	(0.003)	(0.006)	(0.004)
Available seats		-0.013***	-0.025***	-0.029***	-0.033***
		(0.001)	(0.010)	(0.004)	(0.005)
Tobit residual					0.021***
					(0.005)
R^2	0.108	0.262	0.112	0.263	0.408
Observations	7,862	7,862	7,862	1,849	1,849

NOTE. Dependent variable: natural logarithm of fare on the first seat on sale. Panel fixed-effects estimation. The statistical significance at 1%, at 5% and at 10% level of the estimated coefficients is denoted respectively by ***, ** and *. The standard errors are clustered by city pairs.

mented in Figure 2. However, this effect disappears once we control for capacity, in Models (2)-(5). More interestingly, the negative coefficients on the $DaysToDeparture$ dummies indicate that, keeping capacity constant, fares tend to decline as departure approaches. This finding supports the existence of a temporal effect on fares under the decreasing option value of seats rather than under intertemporal price discrimination.

The reason of this result may stem from the structure of the demand. As previously discussed, the demand for long-distance bus service consists of an overwhelming majority of leisure (price-sensitive) travelers and a negligible number of business (price-insensitive) travelers. For this reason, the intertemporal price discrimination strategy as it is adopted by airline carriers becomes unsuitable to the bus industry, where instead the decreasing option value of the seat prevails.

In other words, our estimates suggest that fares tend to be adjusted downwards, if there is a substantial number of unsold seats in the close proximity of departure. Möller and Watanabe (2010) state that such a conduct is optimal when capacity is relatively abundant and consumer valuation is relatively low. Both conditions appear to be applicable for the long-distance bus service industry.

Finally, the conclusions do not change qualitative using censored data, also with the correction of the sample selection. That is, our findings are confirmed with a different number of observations and with a different estimation approach.

To sum up, our empirical analysis suggests that bus fares are driven by the level of demand and capacity utilization in the first place. On the one hand, the positive relation between fares and capacity utilization reflects the cost of keeping seat inventory for uncertain events of high demand (Dana, 1999a); on the other hand, capacity-based fare increases may also aim at shifting demand from peak to off-peak departure times: travelers with a high valuation for departure during peak time will be charged more than those who are willing to go for off-peak departure times (Dana, 1999b).

6 Conclusion

In this paper we have studied empirically the price dynamics of the long-distance bus service industry. We have used fare data by Flixbus, the leading intercity bus company in Europe, collected from 28 days up to 1 day before departure on a daily basis. Although our sample comprises a relatively limited number of city pairs, they represent the portfolio of Flixbus routes which are heterogeneous in terms of distance, competition, per capita income at endpoints' cities, etc.

We have applied panel fixed-effects techniques and corrected for both endogeneity and sample selection. We have found that at a given point in time the fare is an increasing stepwise function of the number of sold seats. As we approach the departure date, the fare on the first seat on sale, i.e. the lowest available fare, increases because seats sell during the booking period.

Holding the number of available seats constant, the perishable nature of the product translates into a decreasing option value of each seat. This effect tends to push fares down as the departure date approaches in order to increase the chance of selling more seats. This is in line with the theoretical model of Sweeting (2012) that predicts a declining time-path due to a falling option value of each seat. We find statistically significant evidence supporting this effect during the last week before departure, when the pressure on the declining option values is more marked.

At last, but not least, we find no price effect referable to intertemporal price discrimination, that is, there is no evidence in favor of increasing fares in proximity of the departure, irrespective of the load factor, in order to extract more surplus from price-insensitive late bookers. This result may most likely depend on the composition of demand for long-distance bus service, which is quite homogeneous, consisting almost exclusively of price-sensitive travelers.

Our analysis has shown that the long-distance bus service industry adopts some of the revenue management practices which started in the airline industry. Despite the analogies in the revenue management techniques, these two industries show some differences. In the long-distance bus service industry, the seat availability may differ within the journey because the bus may make multiple stops before reaching its final destination. This feature does not characterize the airline industry, where the cases of an aircraft stopping over to board more passengers represent the exception rather than the rule. The passengers' heterogeneity is more evident in the airline industry, where more price-insensitive passengers are likely to show up near the departure date, than in the long-distance bus service industry, where passengers are predominantly price-sensitive irrespective of the day of the booking. These differences may explain why our results do not exactly overlap with the empirical findings in the airline sector (Alderighi et al., 2015, 2017).

Future research could expand the present analysis exploiting the entire fare distribution to gain more insights on the dynamic pricing in the industry. The study of Alderighi et al. (2017) provides a theoretical framework for this purpose. Furthermore, it would be interesting to investigate thoroughly how the fare distribution differs among routes with low and high competition, above all by other means of transport such as trains or airplanes (Fageda and Sansano, 2018; Knorr and Lueg-Arndt, 2016). From a theoretical point of view fares are expected to become more dispersed as firms face more competitive markets (Dana, 1999a).

Despite the long-distance bus service industry is very young in Europe, it is becoming an established mean of transport to link cities, also internationally. As the market grows, more sophisticated pricing techniques may be put in place.

Moreover, as the process of European integration continues, long-distance buses in Europe are becoming a closer substitute of trains or even airplanes (Alvarez-SanJaime et al., 2015; Capozza, 2016). Some complementarities between different means of transport may also emerge: for example, people may combine a point-to-point flight with a long-distance bus to their final destination, if the flight lands in another city. The interlink among transport industries, the implications of long-distance bus service on railway and air transportation, and their sustainability are bound to become an important topic of debate in the near future.

References

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

Alderighi, M., Gaggero, A. A., and Piga, C. A. (2017). The hidden side of dynamic pricing: evidence from the airline market. *MPRA Paper 76977, University Library of Munich, Germany.* (https://ssrn.com/abstract=3085585).

Alderighi, M., Nicolini, M., and Piga, C. A. (2015). Combined effects of capacity and time on fares: insights from the yield management of a low-cost airline. *The Review of Economics and Statistics*, 97(4):900–915.

Alderighi, M., Nicolini, M., and Piga, C. A. (2016). Targeting leisure and business passengers with unsegmented pricing. *Tourism Management*, 54:502–512.

Alvarez-SanJaime, O., Cantos-Sanchez, P., Moner-Colonques, R., and Sempere-Monerris, J. J. (2015). A model of internal and external competition in a high speed rail line. *Economics of Transportation*, 4(3):178–187.

Augustin, K., Gerike, R., Sanchez, M. J. M., and Ayala, C. (2014). Analysis of intercity bus markets on long distances in an established and a young market: the example of the U.S. and Germany. *Research in Transportation Economics*, 48:245–254.

Belobaba, P. P. (1987). Survey paper - Airline yield management an overview of seat inventory control. *Transportation Science*, 21(2):63–73.

Berto, A. and Gliozzi, S. (2018). Unconstraining the passenger demand for rail yield management at Trenitalia. *Electronic Notes in Discrete Mathematics*, 69:269–276.

Bharill, R. and Rangaraj, N. (2008). Revenue management in railway operations: a study of the Rajdhani Express, Indian railways. *Transportation Research Part A: Policy and Practice*, 42(9):1195–1207.

Bilotkach, V., Gaggero, A. A., and Piga, C. A. (2015). Airline pricing under different market conditions: evidence from European low-cost carriers. *Tourism Management*, 47:152–163.

Blayac, T. and Bougette, P. (2017). Should I go by bus? The liberalization of the long-distance bus industry in France. *Transport Policy*, 56:50–62.

Borenstein, S. and Rose, N. L. (1994). Competition and price dispersion in the U.S. airline industry. *The Journal of Political Economy*, 102(4):653–683.

Brueckner, J. K. and Spiller, P. T. (1994). Economies of traffic density in the deregulated airline industry. *The Journal of Law & Economics*, 37(2):379–415.

Capozza, C. (2016). The effect of rail travel time on airline fares: first evidence from the Italian passenger market. *Economics of Transportation*, 6:18–24.

Carroll, W. J. and Grimes, R. C. (1995). Evolutionary change in product management: experiences in the car rental industry. *INFORMS Journal on Applied Analytics*, 25(5):84–104.

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

Dana, J. D. (1999b). Using yield management to shift demand when the peak time is unknown. *The RAND Journal of Economics*, 30(3):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.

Dürr, N. S., Heim, S., and Hüschelrath, K. (2016). Deregulation, competition, and consolidation: the case of the German interurban bus industry. *Journal of Transport Economics and Policy*, 50(2):164–188.

Dürr, N. S. and Hüschelrath, K. (2017). Patterns of entry and exit in the deregulated German interurban bus industry. *Transport Policy*, 59:196–208.

Fageda, X. and Sansano, S. (2018). Factors influencing prices and frequencies in the interurban bus market: evidence from Europe. *Transportation Research Part A: Policy and Practice*, 111:266–276.

Federal Office for Goods Transport (2016). Marktbeobachtung Güterverkehr. Marktanalyse des Fernbuslinienverkehrs 2016.

Gaggero, A. A. (2010). *Airline Pricing and Competition: the J-curve of airline fares*. Lambert Academic Publishing: Saarbrücken.

Gaggero, A. A. and Piga, C. A. (2011). Airline market power and intertemporal price dispersion. *Journal of Industrial Economics*, 59(4):552–577.

Grimaldi, R., Augustin, K., and Beria, P. (2017). Intercity coach liberalisation. The cases of Germany and Italy. *Transportation Research Procedia*, 25:474–490.

Guizzardi, A., Pons, F. M. E., and Ranieri, E. (2019). Competition patterns, spatial and advance booking effects in the accommodation market online. *Tourism Management*, 71:476–489.

Haensel, A., Mederer, M., and H., S. (2012). Revenue management in the car rental industry: a stochastic programming approach. *Journal of Revenue and Pricing Management*, 11:99–108.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

IGES Institute (2014). Bahnkunden und Autofahrer lassen Fernbusmarkt wachsen.

Kimes, S. E. (1989). Yield management: a tool for capacity-considered service firms. *Journal of Operations Management*, 8(4):348 – 363.

Knorr, A. and Lueg-Arndt, A. (2016). Intercity bus deregulation in Germany – Intramodal and intermodal effects after two years. *Research in Transportation Economics*, 59:323–329.

Maddah, B., Moussawi-Haidar, L., El-Taha, M., and Rida, H. (2010). Dynamic cruise ship revenue management. *European Journal of Operational Research*, 207(1):445–455.

McGill, J. I. and van Ryzin, G. J. (1999). Revenue management: research overview and prospects. *Transportation Science*, 33(2):233–256.

Möller, M. and Watanabe, M. (2010). Advance purchase discounts versus clearance sales. *The Economic Journal*, 120(547):1125–1148.

Statista (2019a). Anzahl der Fahrgäste in Linienfernbussen in Deutschland von 2007 bis 2017 (in Millionen). (https://de.statista.com/statistik/daten/studie/263118/umfrage/fahrgaeste-der-fernbuslinien-in-deutschland/).

Statista (2019b). Marktanteile der größten Fernbusanbieter in Deutschland nach angebotenen Fahrplankilometern. (https://de.statista.com/statistik/daten/studie/263131/umfrage/marktanteile-der-fernbusanbieter-in-deutschland/).

Stavins, J. (2001). Price discrimination in the airline market: the effect of market concentration. *The Review of Economics and Statistics*, 83(1):200–202.

Sun, X., Jiao, Y., and Tian, P. (2011). Marketing research and revenue optimization for the cruise industry: a concise review. *International Journal of Hospitality Management*, 30(3):746–755.

Sweeting, A. (2012). Dynamic pricing behavior in perishable goods markets: evidence from secondary markets for major league baseball tickets. *Journal of Political Economy*, 120(6):1133–1172.

Talluri, K. T. and Van Ryzin, G. J. (2004). *The Theory and Practice of Revenue Management*. Springer Science & Business Media: Boston, MA.

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*, 40(5):831–844.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. MIT Press: Cambridge, MA.

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach*. South-Western: Mason, OH.

A Appendix

A.1 Data collection process

The data collection hinges on the following steps.
Step 1: Start a daily fare query to Flixbus' website for a bus journey that connects a given city pair on a specific departure date within the scheduled departure period September 4 to September 10, 2017. The query algorithm begins with a request for the fare of one seat, and then iterates with a request of the fares for 2, 3, up to 40 seats, which represent the largest party for a single booking allowed on Flixbus'

website. The bus journey is scheduled D days after the query date, with D taking on values within the range: 1, 2, up to 28 (e.g., $D = 1$ means there is one day separating the query date from the departure date).
Step 2: Two scenarios can be faced:
i) The website returns a valid total fare for the entire query requesting between 1 and 40 seats on a specific bus journey. In this case we can only infer that there are at least 40 seats available on this particular bus journey at D days before departure. We have no precise information on the bus' load factor because more than 40 seats could be on sale (i.e., the datum is censored).
ii) The website returns a valid total fare for $n < 40$ seats on a specific bus journey. In this case the datum is not censored and we can infer the exact number of seats remaining available for booking on the observed journey.
Step 3: As the website returns the total fare for the n seats requested, we obtain the fare of each single seat using the algorithm $p_n = P_n - P_{n-1}$ for $n = 2, 3, ..., 40$; where p_n is the bus fare of the n^{th} seat and P_n is the total fare returned for a query requesting n seats. The fare on the first seat is given by $p_1 = P_1$ and constitutes the initial condition to solve the algorithm recursively.

A.2 Instrument relevance and exogeneity

Model (6) in Table A.2.1 reports the first-stage estimates of Model (3) in Table 3. The dependent variable is $AvailableSeats$, hence the negative and highly statistically significant coefficients on the $DaysToDeparture$ dummies indicates that the number of available seats decreases as we approach the departure date, in line with the expectations. This effect is monotonic, as it must be, because more seats are sold and less are available on a bus as time goes by. Interestingly, the higher magnitude in absolute terms of the coefficients on *DaysToDeparture 6-7* up to *DaysToDeparture 1* indicates that the largest part of the booking occurs during the last week before departure, thereby confirming a documented feature of the long-distance bus industry.

Furthermore, and more importantly, the instrument $WeekendBook$ is found highly statistically significant, with a p-value below 1%. This finding points in favor of a strong correlation between $WeekendBook$ and $AvailableSeats$ and lends support to the instrument relevance condition of $WeekendBook$.

To verify the instrument exogeneity condition, we include $WeekendBook$ among the regressors of the FE-OLS estimation of the main model, i.e. Model (3) of Table 3. If $WeekendBook$ were statistically insignificant from zero, it would sustain the presumption that the variable is exogenous because $WeekendBook$ would have nothing to explain in the model. As Model (7) shows, $WeekendBook$ is highly statistically insignificant (with a p-value of about 30%), thereby supporting the exogeneity of $WeekendBook$.

It is also worth noticing that the inclusion of $WeekendBook$ in the regression of Model (7) does not affect the estimated coefficients of the other regressors, which are practically the equivalent, both in terms of magnitude and significance, to the ones reported in Model (3). This result shows that $WeekendBook$ has no explanatory power in equation (1) and corroborates the full exogeneity and, hence, validity of this variable as instrument for *AvailableSeats*.

A.3 Pooled regression

Table A.3.1 is the pooled OLS version of Table 3. The additional independent variables included in the pooled OLS regression are time-invariant within a given bus journey and aim to parallel the model estimated on similar data for airline fares (Stavins, 2001). Thus, we add the route distance in km, *Distance*; the geometric mean of population at the end points of the route, *Population*; the geometric mean of per capita income, $CapitaIncome$;[17] and a dummy variable equal to one if on the route there is no public-transport alternative (i.e., no direct air connection nor reasonable train service) to Flixbus, $FlixbusOnly$.

Distance is a proxy for costs, *Population* captures the dimension of the market, $CapitaIncome$ the wealth of the market, and finally *FlixbusOnly* aims to relate to route competition, since the absence of product substitutability may translate into stronger market power.[18]

The estimates on the variables of interest, $DaysToDeparture$ dummies and *AvailableSeats*, follow the exact same pattern they have shown in the columns of Table 3. As far as the new regressors are concerned, *Distance* is correctly positively signed, since longer bus journeys cost more. The negative coefficient on the *Population* variable shows that the gains from economies of density due to larger markets are passed through to consumers by means of lower fares, similarly to what observed in the airline industry (Brueckner and Spiller, 1994); however, the positive sign on $CapitaIncome$ variable indicates that wealthier routes pay a fare premium.

[17] Population and per capita income are respectively measured in hundred thousands and thousands. Data are obtained from the German Federal Statistical Office and the Land Statistical Offices except for Milan, whose population datum is collected from the Italian Statistical Office (ISAT), and per capita income from the Regional Statistical Yearbook of Lombardy (ASR Lombardia).

[18] We acknowledge the possible endogeneity of any (proxy of) competition variables included in a price equation. However, the endogeneity of *Flixbus only* is beyond the scope of these estimates, which should be treated with cautions in any case, since they ignore any panel fixed-effects.

Finally, the positive sign on the *FlixbusOnly* variable suggests that there is a positive correlation between fares and competition. A rough quantification of the competition effect on fares, i.e. neglecting the possible endogeneity of the *FlixbusOnly* variable, indicates that on routes where there is no public-transport alternative to Flixbus fares are about 31% more expensive than on routes where Flixbus faces some extent of competition.

Table A.2.1: Instrument relevance and exogeneity

	(6)	(7)
	OLS-FE	OLS-FE
Dependent variable	Available seats	log(Fare)
Days to departure 1	-15.367***	-0.052**
	(1.804)	(0.014)
Days to departure 2	-9.837***	-0.046**
	(1.279)	(0.013)
Days to departure 3	-6.924***	-0.026**
	(0.992)	(0.008)
Days to departure 4-5	-4.118***	-0.017*
	(0.584)	(0.007)
Days to departure 6-7	-2.142***	-0.004
	(0.277)	(0.008)
Days to departure 8-14	-0.800***	0.001
	(0.093)	(0.007)
Days to departure 15-21	-0.178**	-0.001
	(0.045)	(0.005)
Available seats		-0.013***
		(0.001)
Weekend booking	0.196***	-0.002
	(0.031)	(0.002)
R^2	0.586	0.262
Observations	7,862	7,862

NOTE. Panel fixed-effects estimation. The statistical significance at 1%, at 5% and at 10% level of the estimated coefficients is denoted respectively by ***, ** and *. The standard errors are clustered by city pairs.

Table A.3.1: Pooled ordinary least squares estimation

	(8) OLS-Pooled	(9) OLS-Pooled	(10) IV-Pooled	(11) IV-Pooled	(12) IV-Pooled
Days to departure 1	0.148***	-0.141**	-0.264**	-1.459*	-1.040***
	(0.026)	(0.042)	(0.112)	(0.873)	(0.203)
Days to departure 2	0.080**	-0.105**	-0.184**	-1.072*	-0.850***
	(0.023)	(0.030)	(0.074)	(0.592)	(0.170)
Days to departure 3	0.062**	-0.068**	-0.123**	-0.780**	-0.710***
	(0.015)	(0.023)	(0.054)	(0.391)	(0.155)
Days to departure 4-5	0.036*	-0.042**	-0.075**	-0.521**	-0.546***
	(0.013)	(0.013)	(0.035)	(0.246)	(0.124)
Days to departure 6-7	0.023*	-0.017*	-0.034*	-0.291**	-0.406***
	(0.011)	(0.006)	(0.020)	(0.120)	(0.101)
Days to departure 8-14	0.012	-0.003	-0.010	-0.190***	-0.238***
	(0.008)	(0.006)	(0.010)	(0.067)	(0.075)
Days to departure 15-21	0.001	-0.002	-0.003	-0.098	-0.094***
	(0.005)	(0.004)	(0.003)	(0.062)	(0.015)
Available seats		-0.019***	-0.027***	-0.101*	-0.046***
		(0.003)	(0.009)	(0.057)	(0.005)
Distance	0.004***	0.004***	0.004***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Population	-0.021***	-0.020***	-0.019***	-0.013	-0.022***
	(0.000)	(0.000)	(0.001)	(0.008)	(0.001)
Capita income	0.012***	0.011***	0.010***	0.003	0.013***
	(0.000)	(0.000)	(0.000)	(0.007)	(0.001)
Flixbus only	0.383***	0.334***	0.313***	-0.089	0.337***
	(0.000)	(0.009)	(0.023)	(0.289)	(0.021)
Tobit residual					0.028***
					(0.005)
R^2	0.760	0.778	0.775	0.506	0.742
Observations	7,862	7,862	7,862	1,881	1,881

NOTE. Dependent variable: natural logarithm of fare on the first seat on sale. Constant included but not reported. The statistical significance at 1%, at 5% and at 10% level of the estimated coefficients is denoted respectively by ***, ** and *. The standard errors are clustered by city pairs.