

OTDK-dolgozat

2023

CORVINUS UNIVERSITY OF BUDAPEST
SCIENTIFIC STUDENTS' ASSOCIATIONS CONFERENCE

Dynamic Pricing in the Spanish High-Speed Rail
Industry

Vig Ádám

Közgazdasági elemző szak II. évfolyam

Velkey Artúr Ferenc

Közgazdasági elemző szak II. évfolyam

Konzulens: Dr. Hörcher Dániel

2022.04.19

Abstract

This paper examines the dynamic pricing scheme introduced in 2013 by the Spanish state-owned transport company, Renfe Operadora. Dynamic pricing could serve multiple purposes: It can be an effective tool for (1) maximizing profits through price discrimination or (2) maximizing the efficiency of capacity allocation under stochastic shocks in demand. Our paper tests the hypothesis that the objective of Renfe, a state-owned company, should be the maximization of social welfare instead of profit-oriented monopolistic price discrimination. The paper collects daily prices and seat availability for more than 1200 trains during a 30-day booking horizon and answers its research question using explanatory data analysis and linear probability models with unit fixed effects. Our findings suggest that during most of the 30 days of the investigated booking horizon, Renfe aimed to smoothen stochastic shocks in demand by adjusting prices only in response to deviations in capacity utilization. This suggests a welfare-maximizing focus instead of monopolistic price discrimination. However, the more substantial increases in prices in the last days of the booking horizon indicate profit-oriented price discrimination.

Keywords: high-speed rail, dynamic pricing

Table of Contents

I. Introduction	1
II. Literature review	2
<i>II. 1. High-speed rail.....</i>	<i>2</i>
<i>II. 2. Pricing strategies – Conventional and dynamic pricing</i>	<i>2</i>
III. Case of the Spanish high-speed railway.....	6
IV. Data.....	8
V. Exploratory Data Analysis.....	10
VI. Regression analysis	15
VII. Discussion	17
VIII. Conclusions.....	18
IX. References.....	20

I. Introduction

High-speed trains are a fast and efficient way of transportation that is becoming more and more popular nowadays. In this study, we analyze the Spanish high-speed rail company Renfe Operadora, more specifically, the main motivations behind its dynamic pricing strategy.

The first Spanish high-speed rail line was established in the 1990s; nowadays 300 high-speed trains are running every day, serving nearly 100,000 passengers (Hortelano et al., 2016, p. 92.). The dynamic pricing system was introduced in 2013, replacing the previous fixed price scheme, resulting in an overall reduction of mean prices across all ticket types. The company also experienced an increase in occupancy rates and revenues. (Hortelano et al., 2016) Dynamic pricing is known to be an effective tool for maximizing profits through price discrimination that we believe would be inadequate for a government-owned company whose main objective would be maximizing social welfare while minimizing its costs. Therefore, we examine the factors behind the existing dynamic pricing system and ask the following question: *What are the factors affecting dynamic pricing; and how do these factors contribute to changing prices of the Spanish high-speed railway company.*

First, we review the relevant literature on high-speed rails and dynamic pricing, focusing on different pricing strategies. We draw insights into dynamic pricing from the airline industry, which developed dynamic pricing much more intensely than the rail industry. Then, our article follows a case study design and performs a quantitative analysis on more than 1200 train price and occupancy data collected from February 2020 to March 2020. We observe the evolution of each train's price and seat occupancy during a 30-day booking horizon. We use exploratory data analysis and linear probability models with fixed effects to answer our research question.

Our findings suggest that on the 30 days of the investigated booking horizon, Renfe aimed to smoothen stochastic shocks in demand, by focusing on the remaining capacity of trains; however, in the last days before departure, prices tend to increase independently of trains capacity. This suggests a welfare-maximizing focus instead of monopolistic price discrimination. However, the more substantial increases in prices in the last days of the booking horizon indicate profit-oriented price discrimination.

II. Literature review

In the literature review, we aim to provide an overview of high-speed rail transportation, the firm strategy related to the pricing mechanisms of high-speed rail tickets, and introduce the dynamic pricing framework. In this section, we also introduce the center of our case study, the Spanish high-speed rail industry, and the properties of its current pricing strategy.

II. 1. High-speed rail

It is important to clarify that we are studying the pricing strategy of high-speed passenger rails. Pricing in the industry varies greatly from the freight and other purpose rail industries, and we do not include them in this research. We are also conducting this empirical study on the basis that most passenger rail companies at the national level are monopolistic and are often government-controlled and highly regulated firms.

According to the Environment and Energy Study Institute (Nunno, 2018), we understand that the high-speed railway (HSR) is a transport mode, which has gained increasing popularity throughout the world. In Europe, it has been rapidly spreading since the 1980s and is enjoying continuous investment (Nunno, 2018). One of the benefits of HSR is the lower environmental externality compared to other aviation (Nunno, 2018). The European HSR network stretches across the borders of many EU member states and is constantly expanding in connectivity across borders (Nunno, 2018).

HSR is a viable competitor in the transportation market. Based on current experience, we understand that HSR is more punctual and time conservative in many cases, depending on travel distance (Nunno, 2018). According to Lawrence (2019) and Hortelano et al. (2016), HSR is doing its best in travel time and competitiveness in the range of 200 to 900 km. In the lower range, it mostly competes with traditional rail, bus, and car, while in the higher end, it mostly competes with air travel (Hortelano et al., 2016).

II. 2. Pricing strategies – Conventional and dynamic pricing

Revenue management (RM) is an essential part of implementing a well-functioning pricing strategy. The focus of RM is to fundamentally understand the demand structure and the industry structure to design a pricing model. In revenue management, the main underlying aspects are understanding the demand structure and collecting sufficient data to devise an efficient algorithm to increase company revenue. (Talluri et al., 2006). Another aspect of pricing strategies must be mentioned: the aim of smoothening demand shocks and enabling passengers

with high price elasticity to spread out on different trains. We rely on Meissner (2010) to understand the difference and motivation between different pricing strategies.

In the following, we introduce the different pricing strategies based on revenue management and explain the case for railway ticket pricing. After the introduction of conventional pricing strategies, we examine the dynamic pricing strategy in the airline industry to grasp the essence of such a strategy. We are using the airline pricing scheme as our main basis for dynamic pricing because of the growing competition between railways and air in medium distances and because there is an abundance of research in recent years.

It is important to highlight, as we mentioned in the previous sections, that we are basing our research on the proposition that the firms in the high-speed railway industry are monopolistic and government-run. To investigate the phenomenon of dynamic pricing and the difference between the different motivations of a profit-maximizing or a welfare-maximizing firm, we rely on Meissner and Armstrong (2010) and in the case of airline dynamic pricing Williams (2022). When looking at conventional pricing strategies, we rely on the work of Daniel van Vuuren (2002).

In this section, we will focus on the conventional pricing strategies and introduce the positive effects of third-degree price discrimination, and then we will focus on the similarities between dynamic pricing and price discrimination, and lastly, on the benefits of dynamic pricing schemes in the railway industry. (Anderson & Renault, 2011)

When looking at the profit-maximizing angle, we use the different price discrimination options that van Vuuren (2002) introduced. According to van Vuuren and Stigler, price discrimination is defined as 'when two or more similar goods are sold at prices that are different ratios to marginal costs' (van Vuuren, 2002, p. 100-101.) In the following section we will explain three different discrimination mechanics: non-discriminatory pricing, second degree price discrimination, and third-degree price discrimination (van Vuuren, 2002).

Non-discriminatory pricing is the usual flat pricing framework for monopolistic firms that cannot use other methods in theory; it simply originates from the first-order condition of profit maximization. Second-degree price discrimination, in essence, is quantity-based pricing; the price of the product depends on the quantity that is consumed and sold. Third-degree price discrimination highlights that the firm is selling its product on different markets or to different separable groups of consumers, for example, discounting for students. (van Vuuren, 2002)

Following van Vuuren's work (2002), we bring in the aspect of welfare maximization, that is, efficient resource allocation, as one of the main motives for public service provision. The main requirement of the welfare optimization perspective is that prices should be equal to the marginal cost to avoid wasteful trips with a negative net welfare effect (Eliasson, 2021). However, this means that other costs, such as the different sunk and fixed costs of the firm, are not covered by the fare revenues. This brings the problem to a potentially large financial loss. Van Vuuren (2002) introduces the work of Braeutigam (1989), which indicates a solution related to the second-best case; ‘the constrained welfare-maximizing allocation and the socially optimal case of marginal cost pricing’ (van Vuuren, 2002, p. 101). This theory results in either a model that brings the public service to the break-even point or a state where the firm has a certain amount of profit. With the introduction of third-degree price discrimination, we can introduce different consumer groups. In different markets of consumers, we can use different prices, increasing the revenue of the firm itself.

Focusing on third-degree price discrimination, we discovered that differentiating across markets and consumers is noticeable in dynamic pricing. Distinguishing between certain consumer groups (e.g.: students) and different markets, such as the market for peak-hour and off-peak-hour trains, or the market for buying tickets in different timeframes from departure, we can assume that dynamic pricing is not explicitly a third-degree price discrimination, but it carries several similarities. As dynamic pricing tailors the price according to the current periods of demand and other factors (e.g.: last-minute buyers will have a more inelastic demand function.). In this case, by putting a time restriction on the price, we can segment the different consumers on the market, differentiating between consumers with more elastic and inelastic demand.

This essentially breaks down to the aspect that consumers who have a low price elasticity are inelastic in choosing their departure time (e.g.: they are aiming at a specific arrival time) and they will value the aspect that they are provided a ticket in a given period, while people that have a high price elasticity are elastic in choosing their departure time (e.g.: leisure travelers, in most cases pensioners) can change their departure time to a given period where the prices are lower. This phenomenon results in a demand smoothing effect which can be caused by price discrimination.

Van Vuuren (2002) in his study about railway pricing in the Netherlands discovered that a difference in departure time means a difference in consumer price elasticity between the different departure hours. In the case of morning trains in peak hours resulted in much lower

consumer price elasticity, meaning that a profit-maximizing perspective is not beneficial in the morning peak hours, in these cases, the firm should not raise prices. With this in mind, we are also considering peak hour and off-peak hour departure times in our research. In the following section, we aim to introduce the basic concepts of dynamic pricing strategies and their consequences.

To understand dynamic pricing, we rely heavily on Williams' (2022) work, which analyzed dynamic pricing strategies in the airline industry. To understand the basis of dynamic pricing, we used a paper that analyzes the airline industry, because dynamic pricing is more widely used by airline companies and because there is a lack of research focusing on dynamic pricing in European high-speed rail markets, and we found it easier to understand the core concepts and to apply it to high-speed trains.

Compared to our assumptions introduced at the end of the previous section, that third-degree price discrimination and dynamic pricing are related to each other, we also bring in intertemporal price discrimination as an additional basis for our assumptions. Intertemporal price discrimination is generally the practice of changing the price of a product during the period its sold (e.g.: in the case of DVD films, we can see a drop in price after a considerable time passed after the release), in our case, the price change of a railway ticket during its booking horizon, can be interpreted as temporal price discrimination, as well as dynamic pricing (Stokey, 1979). Intertemporal price discrimination in our case supports the argument that third degree and dynamic pricing can be related; in this case, intertemporal discrimination is reversed and follows a price increase, as we close in on the departure date of a said transport vehicle.

For an airline flight, prices depend on three key factors; capacity, filed fares, and inventory allocation for filed fares (Williams, 2022, p. 6). Filed fares are the prices for flights, and the third factor means the amount that the airline wants to sell at a filed fare. These three combined return a dynamic pricing structure based on two other methodological reflections. The first is that airline agencies develop filed fares (groups of products), that are dependent on the quality and timing of the purchase, e.g.: a seat quality is sold at a predetermined value until it reaches a date threshold. After it passes the previously determined date, its price increases. The second methodology that is combined with the filed-fares concept is a revenue management model, the essence of these models is explained in the previous sections of the study. (Williams, 2022)

Based on the introduced framework, we conclude that prices in dynamic pricing depend mainly on the previously determined file fares (price groups) and the current seat allocation (Williams, 2022).

Based on the literature review framed by Meissner and Armstrong (2010), an argument can be made that dynamic pricing could be an essential tool for increasing revenue. However, we cannot ignore the fact that pricing in the railway and all transport modes may have a motivation to smooth out demand. Smoothing demand is an essential way to reduce costs during peak hours of travel. Dynamic pricing and price discrimination can be an effective tool to shift consumers with a high price elasticity to other, not so high-density hours. (van Vuuren, 2020)

Next, we present Renfe's high-speed railway system and the main objectives of the company. Then we introduce our database and provide our descriptive and regression analysis.

III. Case of the Spanish high-speed railway

In the following section, we will concentrate on the subject of our study, the Spanish high-speed railway system, and industry. We will cover the short history of HSR development in Spain, the current system, its advantages and motivation, and the innovation strategy in pricing, starting in 2013.

Renfe Operadora is the Spanish public passenger rail transport operator company, and the company that owns and manages the HSR network is ADIF (Hortelano et al., 2016).

Figure 1. High-Speed Railway Corridors in Spain



Source: Hortelano et al., 2016, page 93

The high-speed railway has been the means of transportation that received the most investment in Spain, as it aggregated to more than 45 billion euros as of 2016 (Hortelano et al., 2016, 92). The HSR network in Spain includes several innovative milestones in the development of transportation, both for Spain and for the world. It is important due to the introduction of multimodal transportation systems that promote the usage of rail passenger transport above all else. According to Sánchez-Borràs et al. (2011), with this development strategy, Spain enabled itself to connect to international railway gauge standards and enabled incoming and outgoing train travel to Europe. With the focus on HSR, the Spanish network is now the leading HSR system in Europe and is second to the Chinese network in the world. (Hortelano et al., 2016).

The expansion of Spanish HSR networks had a great effect on its competition with airlines, resulting in an increase in the share of railway transportation of passengers to the extent where it was closing in on an even playing field against airlines (Hortelano et al., 2016). 'By 2012, 20 years after the inauguration of the first HSR line, more than 300 HS trains ran every day on

Spanish HSR tracks, served nearly 100,000 passengers, and reached 80 Spanish municipalities.' (Hortelano et al., 2016, page 92.).

As we are not focusing on nominal prices, only the underlying motivation of pricing strategies in the industry, we will not cover the exact price determination strategy, only the current system, and its motivation.

Before 2013, fixed prices were set for all HSR tickets and discount prices were also used; however, in 2013, the operator introduced more flexibility in prices, using a dynamic pricing mechanism, as well as an overall reduction of prices across all tickets. Hortelano et al. (2016) summarize the steps of the reduction and implementation of the program. The reduction of prices and the implementation of dynamic pricing resulted in the expected outcomes, the overall average occupancy increased by roughly 10%, and its revenue increased by 6.7% (Hortelano et al., 2016).

In the literature review section, our objective was to uncover certain aspects of the industry, pricing, and the change in strategy of the Renfe Operadora. Understanding the importance of future development and current usage of high-speed railways, we aimed to highlight why our research is relevant and interesting for the transportation industry. By introducing the conventional and dynamic pricing strategies, we can understand how a dynamic pricing strategy works and what its main motivation can be. In our case, the motivation is still unclear as, in theory, dynamic pricing can be welfare-enhancing and profit-maximizing in nature. For this, we use descriptive and regression analysis on a publicly available data set. The following sections explain the acquisition, feature engineering, and analysis of data.

IV. Data

For our quantitative analysis, we used a publicly available dataset that is available on Kaggle. The data was collected by scraping from the official Renfe site. The scraper robot collected all available information for all available trains, including price and available seats for various class and promotion categories (The Gurus, 2020). We make the cleaned dataset, cleaning code, and analysis code available in this OSF repository.

Although the data set is available from the beginning of 2019 until November 2020, the variable 'number of available seats' is only available from February 20, 2020, therefore we were able to analyze dynamic pricing only from that point. Another factor that forced us to narrow the timeframe is the Covid-2019 epidemic. There are two factors that we considered when dropping

data points after 31 March 2020. First, the epidemic caused serious lockdowns all around Europe and Spain, too, which would naturally decrease demand for high-speed trains; second, the dynamic pricing strategy was changed, and the company set fixed prices for its AVE type high-speed trains beginning at the end of March.

We concentrated on the economy class, although we have data available for the first-class class, too. Renfe Operadora sells various classes of tickets, and we chose Promo as the focus of our research because it is the most generic and follows the assumption that the consumer does not differentiate between seats when buying a ticket. There are more than 15 lines that are operated by Renfe; however, we only included the four biggest lines: Madrid-Barcelona; Madrid-Córdoba; Madrid-Zaragoza; Madrid-Sevilla.

Although we had data for in-day price and available seat changes, we aggregated these data to a daily level. We were able to calculate mean daily prices and price changes in both euros and percentages between days. Subtracting the maximum available seats and the minimum available seats per train and day, we created a variable that shows how many seats were sold on a day. Using the daily available seat variable and the maximum capacity of trains, we were able to calculate the load factor for each train and its change during the booking horizon, which describes the capacity utilization of a train.

Based on the relevant literature, we created a dummy variable that is supposed to indicate those trains that depart in the morning (7 am – 9 am) peak hours, because based on van Vuuren (2002) we suspect that peak and off-peak hour pricing strategies differ. We also extracted the weekday of departure as a categorical variable, which is a potential control variable (also used by Williams, 2022) and a continuous and a dummy variable that shows, for each train and day, how many seats were sold and, consequently, whether there were seats sold on the previous day.

V. Exploratory Data Analysis

We observed the price and availability of seats for 1,009 trains in the study period for 30 days before each train departure. Summary statistics for our main variables are given in Table 1. On average, a one-way ticket in our sample costs 62,91 euros, although fares vary substantially between routes because of different route lengths. The mean load factor is 39%; this is because the data cover 30 days before departure, however, the 95th percentile is only 68%. The mean load factor also varies between routes; on the day of departure the average utilization of a Madrid-Barcelona train is 68% but a Madrid-Cordoba train is only 50,03% – we suppose the same train sizes for each route based on what we experienced as the initial number of available seats in our data equalling 300. This variation in the mean load factor shows that Renfe differs in capacity optimization between routes, which could be caused by fixed train lengths combined with different amounts of demand.

Table 1.: Descriptive statistics for the data sample

Variable	Mean	Std. Dev.	Median	5th pctl	95th pctl
Fare (€)	62,91	19,84	58,00	37,00	96,00
Load Factor (%)	0,39	0,16	0,37	0,15	0,68
Daily Fare Change (€)	0,26	2,60	0,00	-0,01	4,29
Transactions Per Day	5,10	8,86	2,00	0,00	20,00
Unique Fares (per train)	4,19	2,11	4,00	1,00	8,00

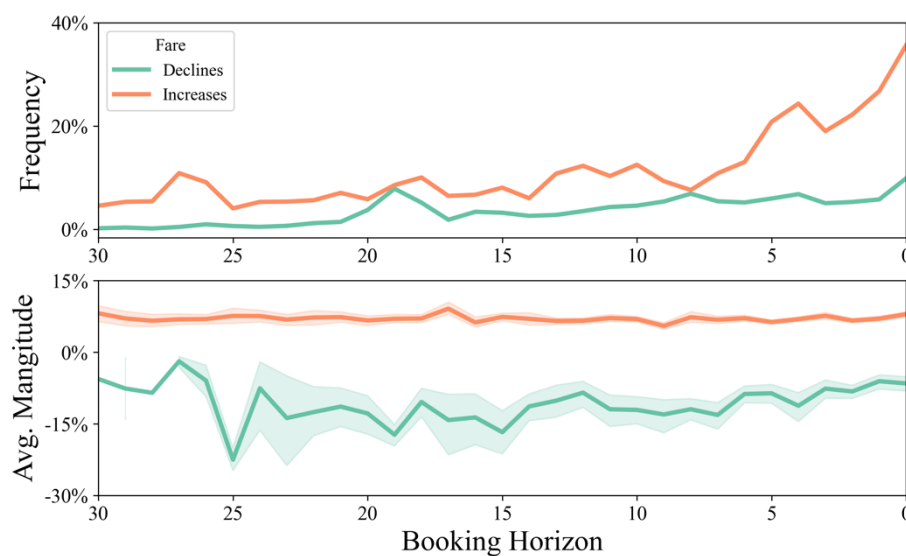
Note: Summary statistics for the 1209 trains tracked between 2020/02/21-2020/03/31. Each flight was tracked at least 15 days before departure. The total number of observations is 30336. The load factor is reported between zero and 100 at maximum capacity. The daily booking rate and daily fare change compare consecutive days. *Source:* own calculations based on data from renfe.com

The mean number of Transactions Per Day is 5,1 with a standard deviation of 8,86, the 5th percentile is 2, and the 95th percentile is 20. No sales per day or training were experienced for 28,6% of the data sample. This shows that there is a high variation in the daily demand for train tickets. Unique Fares (per train) show the number of unique fares for trains. The mean is 4,19, and the median is 4. Although there are four unique fares per train, for 12,6% of the trains, there were no price changes during the observed booking horizon and there was no price change for

84,86% of the trains per day observed. This shows that price changes are relatively rare; however, most of the trains experience it during the observed booking horizon.

Figure 2 shows the percentage of trains that see fares increase or decrease and the average magnitude of this fare change the day before departure. Price increase is always more possible than a decline: the average fare increase frequency is 11,28%, while the average fare decline frequency is only 3,59%, and also the probability of price change is increasing approaching the day of departure. There is a considerable growth in price increases in the last week before departure, which might be a sign of intertemporal price discrimination.

Figure 2: Frequency and magnitude of rate changes by day before departure



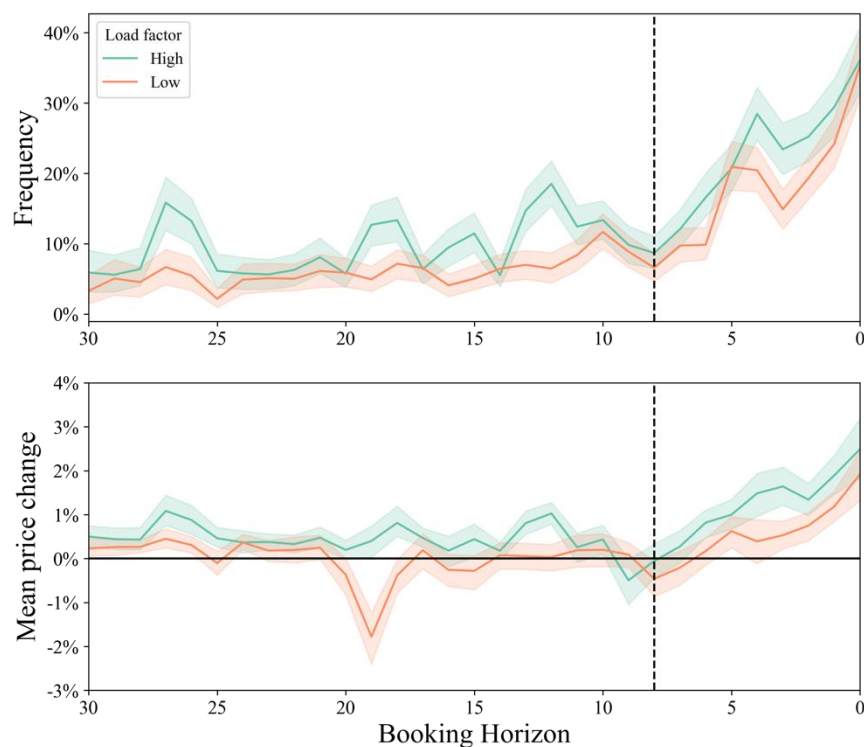
Note: The top panel shows the percentage of trains that see fares increase or decrease the day before departure. The lower panel shows the magnitude of the fare declines and increases by the day before departure. *Source:* own calculations based on data from renfe.com. Concept based on Williams (2022).

The average magnitude of the price change is also different when we consider the increase compared to the decline. The average magnitude of the price increase is 7,03% and is quite even during the booking horizon, while the average magnitude of the price decline is –10,95%. However, its consistency is less accurate than the magnitude of the price increase; this is due to fewer data points with declining prices. However, this magnitude of price decrease becomes smaller approaching the day of departure, the average magnitude of price decline is 8,83% in the last week before departure day.

In Figure 3, we show that the probability of an increase in the price of a train and the magnitude of the increase in the price are affected by two main factors. The first is the remaining days on the booking horizon. Apart from that, we calculated the median load factor for each route and

day subsample and cut the observations to create a categorical variable that shows whether we should consider a train more or less loaded compared to other trains on the same day and route. In the upper panel, until the 7th day before departure day, there are considerable increases in the proportion of trains that the price increased, but only if the train load factor was above the mean load factor on that day. This pattern is clear until the 10th day before departure, then the probability of price increase for higher and lower loaded trains comes close, and we cannot differentiate the probability significantly. This might be a sign of intertemporal price discrimination.

Figure 3: Proportion of Trains that are Price Increased and Mean Price Change



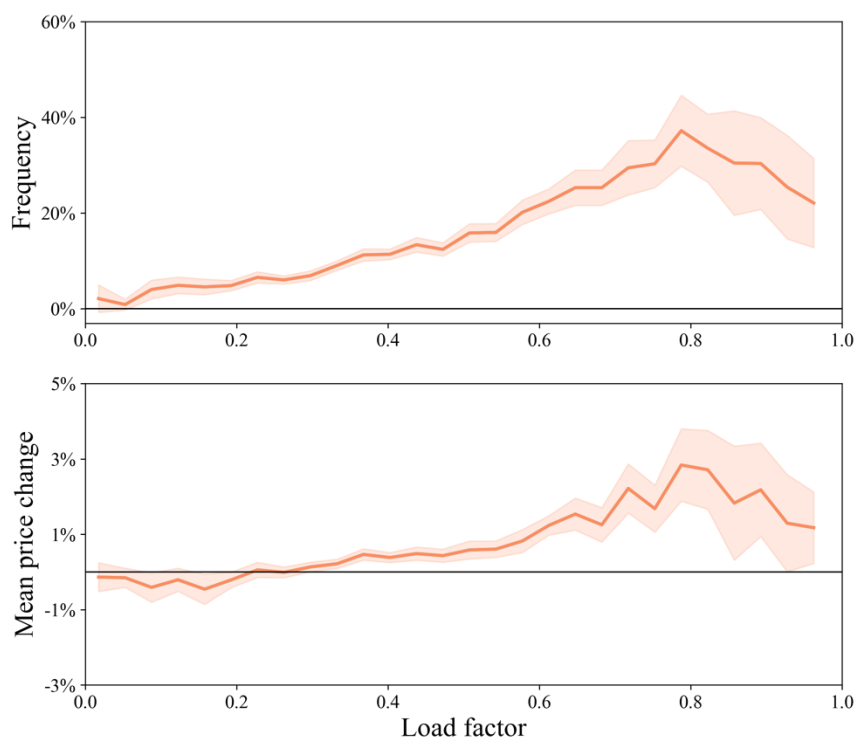
Note: The top panel shows the probability that the price of a train increases during the booking horizon. The high and low load factor is determined by the median load factor, calculated for each train-day sub-sample. The mean frequency of price increases is 8,0% before and 22,31% during the last week before departure day. The bottom panel shows the mean price increase. In the last week before departure day, there is a significant increase in the mean price change for trains with high and low load factor. 95% confidence intervals included. *Source:* own calculations based on data from renfe.com.

Entering the last week before departure day, the probability of price increase increases; on the day of departure, the probability of a train price increase is 35,6%. From the 7th day before departure, the price increase probability increases, but there is a considerable drop in this probability around 3 days before departure. The probability of price increase before jump on the seventh day is 8,0%, and it is 6,4% among those trains where the load factor is below the median. During the week before departure, this value is 22,3% on average, among those trains,

the load factor below the median is 19,7%. This shows that there is a huge increase in the probability of a price increase during the week before departure. Furthermore, we could suspect that this probability before the last week before departure is affected by the relative load factor, while during the last day it is affected only by the approaching departure.

We show that not only is the probability of a price increase higher during the last week before departure, but also the mean price changes on the bottom panel of Figure 3. The mean price change (in percentage) is around zero for trains with low load factor as with a significant 2% average price discount on the 19th day before departure, while trains with a high load factor experience significant price increases during the booking horizon. In the last week before departure day, there is a significant increase in the mean price change for both types of trains. This confirms our previous conclusion that there is a significant increase in prices during the last week independent of the utilization of the capacity of the trains.

Figure 4: Proportion of Trains that's Price Increased and Mean Price Change

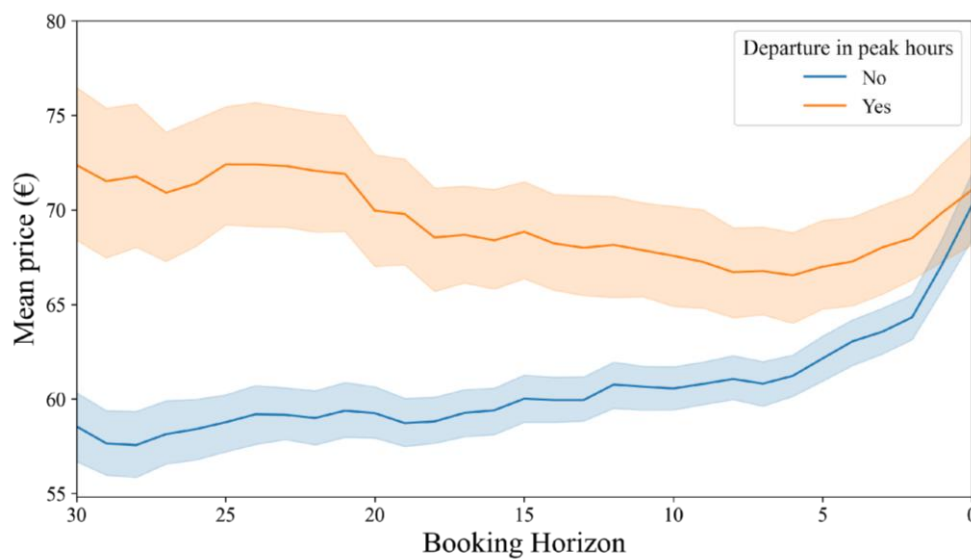


Note: Top panel shows the probability that a train's price increases while the train fills up, peaking around 40% at 80% capacity utilization. Bottom panel shows the mean price increase. 95% confidence intervals included. *Source:* own calculations based on data from renfe.com.

This hypothesis is further investigated in Figure 4, where we plot the frequency of price increase and the mean price change against the virtual load factor of the trains. The probability of price

increase increases as the train fills up with passengers, but this peaks around 80% capacity utilization. At its peak there is a 40% chance that a train's price would increase per day, however, we cannot accurately estimate the probability of price increase in the higher regions of the load factor, as we have fewer trains that reach this point. As we noted in Figure 2, the average magnitude of a price change is around 7%; this remains the same for trains with different load factors (not displayed here). Passengers could expect an average price increase of 3–4% per day as trains fill up and their total capacity is over 60%.

Figure 5: Mean price for peak-hour and off-peak trains



Note: The mean price of peak hour trains is 13,5 EUR more expensive than off-peak hour trains from 30 to 25 days on the booking horizon. This difference disappears during the last days of the booking horizon. 95% confidence intervals included. Source: own calculations based on data from renfe.com

According to van Vuuren (2012), it is important to differentiate between departure times based on peak travel hours. In trains with peak-hour departure times, we notice that the initial price is much higher (~72,4 EUR), however, there is a much lower interval in which the price moves. Also, it seems that peak fares are decreasing on average between 25 and 5 days before departure; therefore, we suspect that the increases in prices we observed in Figure 3 in this period are probably driven by off-peak fares. In off-peak hour trains, we can see that it starts from a lower price (~58,5 EUR), but in the end, the price is at the same level as it is in peak hour trains. van Vuuren's (2012) theory seems to hold for Renfe's case, as there is no price increase in the last couple of days.

In the next section, we examine the relationship between train capacity utilization, the booking horizon timeline, and price increase using linear probability models with unit-fixed effects.

VI. Regression analysis

We use OLS regression to investigate the effect of the capacity utilization of trains, the approaching departure day, and other factors on the probability of price increase. We use the fixed effect of the train to control for individual characteristics at the train level that could impact our coefficient estimates, thus identifying the effects of variation within the train of the variables. Therefore, we do not control for trains that depart during peak hours, because this is inherently controlled by the train-level fixed effects. Note that here we dropped those trains that depart with less than 50% of the seats utilized, because leaving 855 trains in our sample. This step was taken because we wanted to focus on trains that had a relatively high utilization rate of their capacity.

We estimate the following equation with OLS:

$$Y_{it} = \mu_i + \beta X_{it} + \epsilon_{it}$$

where Y_{it} is either a binary variable indicating price increase which results in a linear probability model, or the percentage of the price change. For each Y_{it} we estimated the equation with several dependent variables X_{it} such as the load factor of the train cut into five equal categories;

Table 2.: Fixed effect regressions on the probability of price increase

	Price increase dummy			
	(1)	(2)	(3)	(4)
Last 5 days before departure	0.176*** (0.009)			0.072*** (0.011)
Sales on previous day		0.073*** (0.005)		0.034*** (0.005)
Load factor 0.2–0.4			0.067*** (0.011)	0.052*** (0.011)
Load factor 0.4–0.6			0.174*** (0.013)	0.124*** (0.014)
Load factor 0.6–0.8			0.326*** (0.017)	0.234*** (0.020)
Load factor 0.8–1.0			0.360*** (0.027)	0.248*** (0.031)
Train fixed effect	Yes	Yes	Yes	Yes
Number of trains	855	855	855	855
Observations	20,792	20,792	20,779	20,779
R^2	0.119	0.078	0.129	0.136
Within R^2	0.054	0.01	0.064	0.072

Note: Clustered standard errors are in parentheses.

*p<0.1; **p<0.05; ***p<0.01

a dummy which describes whether we experienced positive sales on the previous day for a given train.

During the descriptive analysis, we explored the possibility that there might be different pricing strategies during this last period before departure than on the entire examined booking horizon of 30 days, so we created a dummy variable indicating this. β is a vector of our coefficients of interest, μ_i is train level fixed effect, and ϵ_{it} is the error term. We cluster our standard errors at the level of train id in each specification.

In Table 2, we show our results from the regressions with the price increase dummy as the dependent variable. In the first column we estimated that in the last 5 days before departure, there is an increased probability by 17% that a train's price would increase, which is in line with Figure 3., upper panel. In column (2) we estimated that when positive sales are experienced on the previous day for a train, that leads to the increased probability of price increase by 7.3% per day. We estimate the effect of the capacity utilization of the train (load factor) for five equally sized bins and the results are in line with what we saw in Figure 4, for example, the probability of an increase in the price is 36% higher *per day* when the load factor is between 80-100% compared to a poorly utilized train. The effect sizes are somewhat more moderate when including each variable of interest in one equation, but the results are substantially the same as when these factors are present simultaneously.

Table 3.: Fixed effect regressions on the changes in prices (%)

	Price change per day in %			
	(1)	(2)	(3)	(4)
Last 5 days before departure	1.280*** (0.086)			0.348*** (0.122)
Sales on previous day		0.588*** (0.001)		0.325*** (0.001)
Load factor 0.2–0.4			0.372** (0.149)	0.281* (0.148)
Load factor 0.4–0.6			0.920*** (0.160)	0.635*** (0.172)
Load factor 0.6–0.8			2.413*** (0.185)	1.912*** (0.227)
Load factor 0.8–1.0			3.177*** (0.342)	2.585*** (0.379)
Train fixed effect	Yes	Yes	Yes	Yes
Number of trains	855	855	855	855
Observations	20,792	20,792	20,779	20,779
R^2	0.048	0.037	0.055	0.057
Within R^2	0.016	0.003	0.023	0.024

Note: Clustered standard errors are in parentheses.

*p<0.1; **p<0.05; ***p<0.01

The results in Table 3, on the expected change in the prices are quite small; this is due to the average low probability of price increase. For all the variables of interest, we found a positive relationship between changes in the ticket price. The effect is greatest when a train is near full load, passengers should expect on average a 3,17% price increase *per day* for an almost fully loaded train. The coefficient estimates are again smaller for the specification with all the dependent variables included, but this is because the effect of the factors is present simultaneously and in the last 5 days before departure, there is an increase in the prices.

VII. Discussion

First, we showed that the operator is more likely to increase prices than decrease them during the booking horizon, with the average fare increase frequency being 11,28%, while the average fare decline frequency is only 3,59%. We also showed that the average magnitude of price increase is 7,03% while the average magnitude of price decline is higher, –10,95%.

Second, we show that the probability of a price increase is affected by the approaching departure day. There is a huge increase in the probability of a price increase during the week before departure day. We also showed that the capacity utilization of trains greatly affects both the probability of increase in prices and the expected increase in prices.

Third, in our regression analysis, we found that the main factor affecting both the probability of an increase in prices and the expected increase in prices is the capacity utilization of the train. In Table 2. column (4) we showed that during the last 5 days before there is a 7% higher chance of price increase per day while this effect is around 24% when considering the train's higher capacity utilization. This relationship was also present when we used the expected increase in prices instead of the probability of an increase in prices.

We tested the hypothesis that the objective of Renfe, a state-owned company, should be social welfare maximization instead of profit-oriented monopolistic price discrimination. We learned that on the 30 days of the investigated booking horizon, Renfe aimed to smoothen stochastic shocks in demand by focusing on the remaining capacity of trains, however, in the last days before departure, prices tend to increase independent of trains capacity. This suggests a welfare-maximizing focus instead of monopolistic price discrimination.

However, more substantial fare increases in the final days of the booking horizon indicate profit-oriented price discrimination. From a policy perspective, we think that a state-owned

company should not include any mechanism in its dynamic pricing strategy that could indicate profit-oriented price discrimination.

Both strategies can be validated by the main problems of the public transportation industry. First, demand smoothing results in fewer seat losses and capacity maximization on trains, resulting in fewer costs for the public service operator. Second, the profit-maximizing aspect, focusing on more inelastic passengers, does not result in a loss of demand, while it can enlarge income, resulting in a break-even state or profit.

VIII. Conclusions

In this research paper, we examine the high-speed rail industry in Spain. Focusing on Renfe Operadora, the Spanish public transport company's pricing strategy relating to the high-speed railway network in Spain. We aimed to identify the factors that affect dynamic pricing and decompose how these factors affect decision-making about changing ticket prices.

Our research is based on the literature on pricing systems in transport markets, introducing that the economic motivation of a service provider results in different pricing strategies. Applying this theory to our research of Renfe Operadora, we examined data on passenger train ticket prices and train specifications with both descriptive and regression analysis.

We used exploratory data analysis and OLS regression models with fixed effects to answer our research question. Our findings suggest that during the 30 days of the investigated booking horizon, Renfe aimed to smoothen stochastic shocks in demand, by focusing on increasing prices when the train capacity is already high; however, in the last days before departure, prices tend to increase independent of the train capacity.

This results in a mixed revenue management model in the case of Renfe Operadora. In the earlier days of the booking horizon, they apply a welfare-maximizing and cost-minimizing framework, that smooths out demand shocks and enables them to tailor demand at certain times in the booking horizon, resulting in cost decreasing and welfare enhancement. However, in the last days of the booking horizon, they use price discrimination and raise the prices when the passengers are more inelastic, resulting in a profit maximization framework.

To extend the research, we would need consumer-level data, in this way we could confirm the incentive of the consumer to buy tickets, this could help our understanding of the last 5 days in the booking horizon more deeply. Another new aspect could be the acquisition of competitor

airlines to investigate the competition between HSR and air travel, as well as uncover how the airline distorts demand in the industry.

IX. References

- Anderson, S. P., & Renault, R. (2011). Price discrimination. In: *A Handbook of Transport Economics*. Edward Elgar Publishing.
- Eliasson, J. (2021). Efficient transport pricing—why, what, and when? *Communications in Transportation Research*, 1, 100006.
- Hortelano, A. O., Guzman, A. F., Preston, J., & Vassallo, J. M. (2016). Price elasticity of demand on the high-speed rail lines of Spain: Impact of the new pricing scheme. *Transportation Research Record*, 2597(1), 90-98.
- Lawrence, M. (2019). Should countries invest in high-speed rail? September 08, 2019 Retrieved: 2020.12.19. from: <https://blogs.worldbank.org/transport/should-countries-invest-high-speed-rail>
- Meissner, J. & Armstrong, A (2010). Railway revenue management: Overview and models (operations research). Department of Management Science, Lancaster University Working Papers, (MRG/0019).
- Nunno, R. (2018). Fact Sheet: High Speed Rail Development Worldwide. The Environmental and Energy Study Institute, Retrieved: 2020.12.19. from: https://www.eesi.org/files/FactSheet_High_Speed_Rail_Worldwide.pdf
- Sánchez-Borràs, M., Robusté, F., & Criado, O. (2011). High-speed railways in Spain: Example of success? *Transportation research record*, 2261(1), 39-48.
- Stigler, G. (1987). J6, *The Theory of Price*. Macmillan, New York
- Stokey, N. L. (1979). Intertemporal price discrimination. *The Quarterly Journal of Economics*, Volume 93, Issue 3, August 1979, Pages 355–371
- Talluri, K. T., & Van Ryzin, G. J. (2006). *The theory and practice of revenue management* (Vol. 68). Springer Science & Business Media. Berlin
- The Gurus (2020). Spanish Rail Tickets Pricing – Renfe, Version 5. from: <https://www.kaggle.com/thegurusteam/spanish-high-speed-rail-system-ticket-pricing>. Downloaded: 2020.11.05
- Van Vuuren, D. (2002). Optimal pricing in railway passenger transport: theory and practice in The Netherlands. *Transport policy*, 9(2), 95-106.

Williams, K. (2022). The Welfare Effects of Dynamic Pricing: Evidence from Airline Markets *ECONOMETRICA*: MAR 2022, VOLUME 90, ISSUE 2;
<https://doi.org/10.3982/ECTA16180>