CORVINUS UNIVERSITY OF BUDAPEST

RAJK COLLEGE FOR ADVANCED STUDIES

Motivations Behind Dynamic Pricing in the Spanish High Speed Railway Industry

Authors: Ádám Vig, Artúr Velkey

Course: Transport Economics

Course Leader: Dániel Hörcher

12.20.2020

**Abstract**

This paper examines the motivations behind the dynamic pricing scheme introduced in 2013 by the Spanish state-owned transport company, Renfe Operadora. The new system resulted in an overall decrease in prices, increase in occupancy and revenues. The paper asks whether only smoothing stochastic shocks in demand or maximizing profits through price discrimination is a driver of the dynamic pricing system? Dynamic pricing could be an effective tool for maximizing profits through price discrimination, however a state-owned company’s main objective would be maximizing social welfare. The paper collects daily prices and seat availability for 1209 trains during a 30 day long booking horizon and answers its research question using descriptive statistics and logistic regressions. The results suggest that during the 30-day booking horizon price increase is affected by the relative capacity utilization of trains, the coefficient of load factor of trains is ~1,8% in the logistic regression on the probability of price increase after controlling for confounders. However, in the last 5 days before departure this effect disappears and the effect of approaching departure becomes significant and positive which allows for a conclusion that the company focuses on price increase and profit maximization in the last couple of days before departure.

**Table of Contents**

[**Introduction 1**](#_heading=h.30j0zll)

[**I.**](#_heading=h.1fob9te) **Literature review 2**

[*I.*](#_heading=h.3znysh7) *1. High-speed railway 2*

[*I.*](#_heading=h.2et92p0) *2. Pricing strategies 2*

[I.2.1. Conventional pricing 3](#_heading=h.tyjcwt)

[I.2.2. Dynamic pricing 4](#_heading=h.3dy6vkm)

[**II.**](#_heading=h.1t3h5sf) **Case of the Spanish high-speed railway 6**

[**III.**](#_heading=h.4d34og8) **Data 8**

[**IV.**](#_heading=h.2s8eyo1) **Descriptive Evidence 9**

[**V.**](#_heading=h.17dp8vu) **Regression analysis 13**

[**VI.**](#_heading=h.3rdcrjn) **Discussion 15**

[**VII.**](#_heading=h.26in1rg) **Conclusion 16**

[**VIII.**](#_heading=h.lnxbz9) **References 18**

# Introduction

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

The first Spanish railway tracks were established in the 1990’s, and nowadays there are 300 high speed trains 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, that resulted in an overall reduction of prices across all tickets. The company experienced increasing occupancy and revenues too. Dynamic pricing could be an effective tool for maximizing profits through price discrimination, that we believe would be inadequate for a government owner company, that’s main objective would be maximizing social welfare. Therefore, we examine the motivations behind the existing dynamic pricing system and ask the following question: whether only smoothing stochastic shocks in demand or maximizing profits through price discrimination is the driver of the dynamic pricing system used by the Spanish high-speed railway company, Renfe Operadora. (Hortelano et al, 2016)

First, we review relevant literature on high-speed railways and dynamic pricing focusing on different pricing strategies. We draw insights on dynamic pricing from the airline industry, that developed dynamic pricing much more intensely than the railway industry. Then our paper follows a case study design and execute quantitative analysis on 1209 trains price and occupancy data collected from 2020 February to 2020 March. We observed the evolution of each train’s price and seat occupancy during a 30-day booking horizon. We used descriptive analysis and logistic regression to answer our research question.

Our findings suggest that on the whole 30 days of the investigated booking horizon, Renfe aims to smoothen stochastic shocks in demand, by focusing on the remaining capacity of trains. In our view and according to our findings in the literature, this is a welfare optimizing point of view. However, when we are looking at the last 5 days of the observed 30 days, we can see that Renfe chooses price according to the same practice that airlines apply, resulting in a situation where in the last days the prices steeply go up independent of train’s capacity utilization.

# Literature review

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

## 1. High-speed railway

It’s important to mention and clarify that we are studying the pricing strategy of passenger high-speed railway. The pricing in the industry varies greatly from freight and other purpose railway industries, and we are not including them in this research. We are also doing this empirical study on the basis that most passenger railway companies on the national level are monopolistic in nature and are often government controlled, highly regulated firms.

According to the Environment and Energy Study Institute (Nunno, 2018) we can understand that high-speed railway (HSR) is a transport mode which has seen increasing popularity in the world. In Europe it has been spreading fast since the 1980s and is enjoying continuous investment. One of the benefits of HSR is the lower environmental externality than other means of transport. The European HSR network arcs through many EU member states borders and is constantly expanding in connectivity across borders. In the latter part of our review, we are going to examine the Spanish network and its pricing strategies.

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

## 2. Pricing strategies

Revenue Management (RM) is an essential part of implementing a well-working pricing strategy. The focus of RM is fundamentally understanding the demand structure and industry structure to devise a model for pricing. In revenue management the main underlying aspects are the understanding of the demand structure and the collection of sufficient data to devise an efficientalgorithm in order to increase firm revenue. (Talluri et al., 2006)

In the following, we introduce the different pricing strategies that are based upon revenue management and explain the case for railway ticket pricing. After the introduction to conventional pricing strategies, we examine the strategy of dynamic pricing in the airline industry to grasp the essence of such a strategy. We are using the airline pricing scheme as our main basis on dynamic pricing because of railways growing competition with air travel in medium distances and because it is a highly researched field.

### I.2.1. Conventional pricing

It is important to highlight, as we mentioned in the previous sections, that we are basing our research on the preposition that the firms in the high-speed railway industry are monopolistic and government run in nature. To investigate the situation of the Spanish HSR pricing strategy, we need to understand the basics of railway pricing strategy. In the following we are looking at the difference between welfare and profit maximization focused pricing strategies. It is essential to distinguish the two paradigms, in order to understand what the focus of the Renfe Operadoras motivation is. When looking at conventional pricing strategies we rely on the work of Daniel van Vuuren (2002).

When looking at the profit maximizing angle, we use the different price discrimination options that Vuuren (2002) introduced. According to 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” (Vuuren, 2002, p. 100-101.) In the following we will explain three different discrimination mechanics: nondiscriminatory pricing, second degree price discrimination, third degree price discrimination. (Vuuren, 2002).

Nondiscriminatory pricing is the usual 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 is dependent of the quantity it 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 the student discount. (Vuuren, 2002)

Focusing on third degree price discrimination, we discover 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 make the assumption that dynamic pricing is not explicitly a third-degree price discrimination, but it carries many similarities. Therefore, we can make the argument that dynamic pricing could be an essential tool for a profit maximizing strategy. (Vuuren, 2020)

To follow Vuuren’s (2002) work, we bring in the aspect of welfare maximization, as one of the main motives of government run firms. The main problem that a welfare optimization perspective faces is that prices should be equal to marginal costs, however this means that other costs, such as the different sunk and fixed costs of the firm are not in the equation. This brings the problem to a large amount of welfare loss. 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” (Vuuren, 2002, p. 101). This theory implies the profitability of the firm, either breaking even, or exceeding a certain amount of profit, the difference in the two is the value that the government needs to weigh and decide whether to impose restrictions on the firm’s strategy.

With the introduction of price discriminatory strategies, we can understand that the welfare maximizing standpoint changes as well. When we can use third degree price discrimination to use different prices at different markets and consumer groups, we can tailor prices that do not change the overall welfare of consumers, or at least it doesn’t lower them. Because of the above mentioned, dynamic pricing is also a great choice if we are going to focus on welfare enhancing.

Vuuren (2002) in his study about railway pricing in the Netherlands discovered that 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 its consequences.

### I.2.2. Dynamic pricing

To understand dynamic pricing, we rely heavily on Williams (2020) work, who analyzed dynamic pricing strategies in the airline industry. We used a paper that analyzes the airline industry because dynamic pricing is effectively used by airline companies and we found it easier to understand the core concepts and apply for high-speed trains too.

Compared to our assumptions introduced in 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 lowering prices well after the release of a product (e.g.: in the case of DVD films, we can see a drop in price after a considerable time passed after the release) (Stokey, 1979). Intertemporal price discrimination in our case supports the assumption that third degree and dynamic pricing is related, in this case intertemporal discrimination is reversed and follows an increase in price as we close in on the departure date of said transport vehicle.

For an airline flight, prices depend on three key factors; capacity, filed fares and inventory allocation for filed fares (Williams, 2020, p. 6). Filed fares are the prices for flights and the third factor means the amount that the airline wants to sell at filed fares. These three combined returns 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 quality and timing of purchase e.g.: a quality of seat is sold at a predetermined value until it reaches a date threshold. After it passes over the previously determined date, it rises in price. 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, 2020)

Based on the introduced framework, we conclude that prices in dynamic pricing depend on the previously determined filed fares (price groups) and the current seat allocation (Williams, 2020). In our research we are focusing on the importance of seat allocation and its effect on railway ticket prices.

Until now, we showed that there could be two main motivation in price discrimination and dynamic pricing: one that focuses on social welfare and another that focuses on company profits. We think that a government owned company should solely focus on social welfare maximization. We translate this conclusion to the following in our empirical analysis: we will investigate whether Renfe changes its prices based on a train’s capacity utilization, in a way that more utilized trains are more likely to get their price increased. Furthermore, we estimate motivation for profit maximization in a way that we investigate whether the company changes its prices approaching departure day independent of trains capacity utilization. This could lead us to a conclusion about revenue increasing motivations as it is increasing fares for last minute, urgent rides discriminating consumers irrespective to demand shocks.

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

# Case of the Spanish high-speed railway

In the following section we will now concentrate on the subject of our study, the Spanish high-speed railway system and industry. In this section 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 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*

A képen térkép látható

Automatikusan generált leírás

*Source: Hortelano et al, 2016, page 93*

High-speed railway has been the means of transportation that got the most investment in Spain, it aggregates to more than 45 billion euros as of 2016 (Hortelano et al, 2016). The HSR network in Spain bears innovative milestones for the development of transportation both for Spain and the world. It is important in the aspect of introduction of multimodal transportation systems, promoting the usage of railway passenger transportation above all else. According to Sánchez-Borràs et al, (2011) with this development strategy, Spain enabled itself to connect to international railway gage 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 its second to the Chinese. (Hortelano et al, 2016).

The Spanish HSR networks expansion resulted in a great effect in the competition with airlines, resulting in the increase of 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. In the following part, we are going through the strategy that was implemented in 2013 relating to the implementation of dynamic pricing in the network.

In the system before 2013, fixed prices were set to 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, Guzman, Preston, and Vassallo (2016) summarizes the steps of the reduction and implementation of the program, we base our knowledge on their study.

The reduction of prices and the implementation of dynamic pricing resulted in the expected outcomes, the overall average occupancy increased roughly ten percent and its revenue increased by 6.7% (Hortelano et al, 2016).

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

# Data

For our quantitative analysis we used a publicly available dataset that is accessible on Kaggle[[1]](#footnote-1). The data was collected by scraping from the official site of Renfe. The scraper 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 codes and code for analysis available in this OSF repository[[2]](#footnote-2).

Although the dataset is available from the beginning of 2019 until 2020 November, available seat variable is only available since 2020 February 20, therefore we were able to analyze dynamic pricing only from that. Another factor that forced us to narrow the timeframe is the Covid-2019 epidemic. There are two factors we considered when dropping data points after 2020 March, 31. First, the epidemic caused serious lockdowns in all around Europe and Spain too, that would naturally decrease demand for high-speed trains; second, dynamic pricing strategy was changed, and the company set fixed prices for its AVE type high speed trains beginning the end of March.

We concentrated on the economy class although we have data available for first class too. Renfe Operadora is selling various ticket classes, we chose Promo as the focus of our research because it's 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 this data to a daily level. We were able to calculate mean daily prices and price changes in both EUR and percentage between days. Subtracting maximum available seats and 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 seats 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 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 we suspect that peak and off-peak hour pricing strategy differs based on Vuuren 2002. We also extracted the weekday of departure as a categorical variable, which is a potential control variable (also used by Williams, 2020) 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. We also created a variable, that shows how many days are remaining until departure day and consequently created a variable opposite of this, that takes 30 at the day of departure and 0 thirty day before departure day. We did this on purpose, during the regression analysis we create an interaction variable among load factor and the approach of departure day.

# Descriptive Evidence

We observed price and seat availability for 1209 trains in this period for 30 days before each train’s departure. In total, we obtained 30336 variables. Summary statistics for our main variables are in *Table 1*. On average a one-way ticket in our sample costs 62,91 Euro, although fares between routes are varying substantially because of different route lengths. Mean Load Factor is 39% this is because data is covering 30 days before departure, however the 95th percentile is only 68%. Mean Load Factor also varies between routes, at the day of departure the average utilization of a Madrid-Barcelona train is 68% but a Madrid-Cordoba train is only 50,03% – we supposed same train sizes for each route based on our data. This variation in mean load factor shows that Renfe differs in capacity optimization between routes, that could be affected by fixed size wagons but different amounts of demand.

*Table 1.: Descriptive Statistics for the Data Sample*

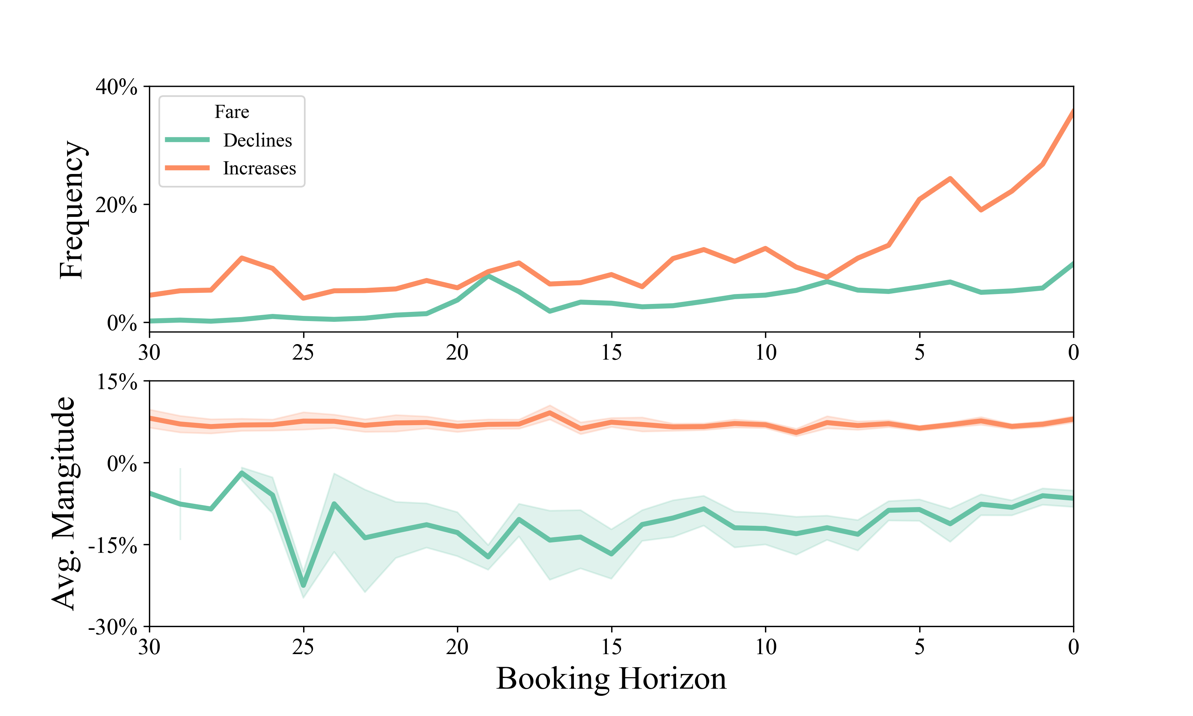
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Mean | Std. Dev. | Median | 5th pctile | 95th pctile |
| Fare (€) | 62,91 | 19,84 | 58,00 | 37,00 | 96,00 |
| Load Factor (%) | 0,39 | 0,16 | 0,37 | 0,15 | 0,68 |
| Daily Fare Change (€) | 0,26 | 2,60 | 0,00 | -0,01 | 4,29 |
| Daily Booking Rate | 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 tracked at least for 15 days before departure. Total number of observations is 30336. Load Factor is reported between zero and 100 at the maximum capacity. The daily booking rate and daily fare change compares consecutive days. *Source*: own calculations based on data from renfe.com

The mean Daily Booking Rate is 5,1 with a standard deviation of 8,86, the 5th percentile is 2 and the 95th percentile is 20. There were no sales experienced per day and train for 28,6% of the data sample. This shows that there is a high variation in daily demand for train tickets. Unique Fares (per train) shows the number of unique fares for trains. The mean is 4,19 and the median is 4. Although there are four unique fares per trains, for 12,6% of the trains, there were no price change 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 by day before departure. Price increase is always more possible than decline, the average fare increase frequency is 11,28%, while average fare decline frequency is only 3,59%, 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.

*Figure 2: Frequency and Magnitude of Fare Changes by Day Before Departure*

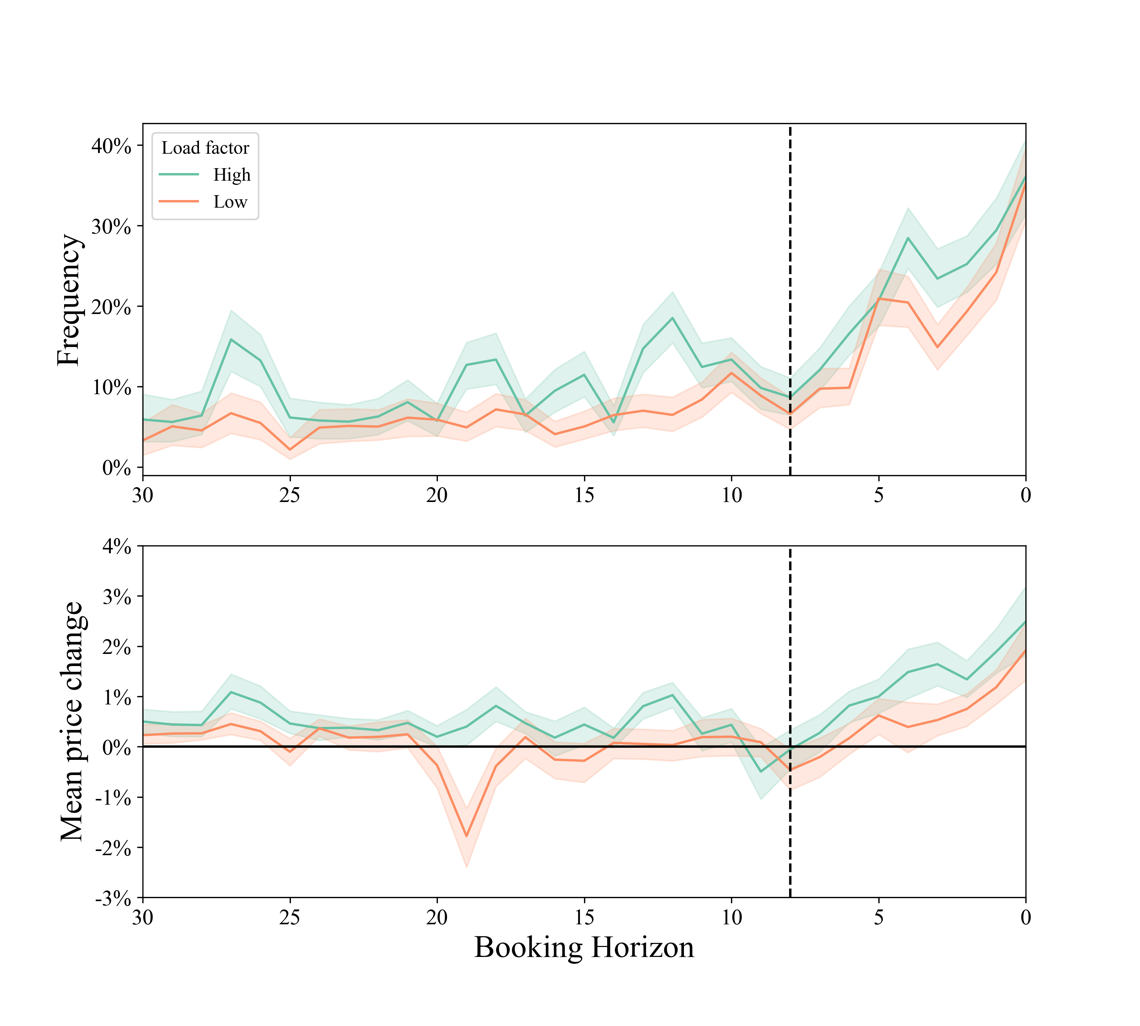
**

*Note*: The top panel shows the percentage of trains that see fares increase or decrease by day before departure. The lower panel plots the magnitude of the fare declines and increases by day before departure. *Source*: own calculations based on data from renfe.com. Concept based on Williams (2020).

The average magnitude of price change is also different when we consider increase compared to decline. Average magnitude of price increase is 7,03%, and it is quite even during the booking horizon, while average magnitude of price decline is -10,95%, however its consistency is less accurate than the magnitude of price increase – this is due to fewer data points with declining in price. However, this magnitude of price decrease is becoming 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 that a train’s price increases, and the magnitude of price increase is affected by two main factors. The first is obviously the remaining days in 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. On the top panel, until the 7th day before departure day, there are considerable jumps in the proportion of trains that’s price increased but only, if the train’s 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.

*Figure 3: Proportion of Trains that's Price Increased and Mean Price Change*

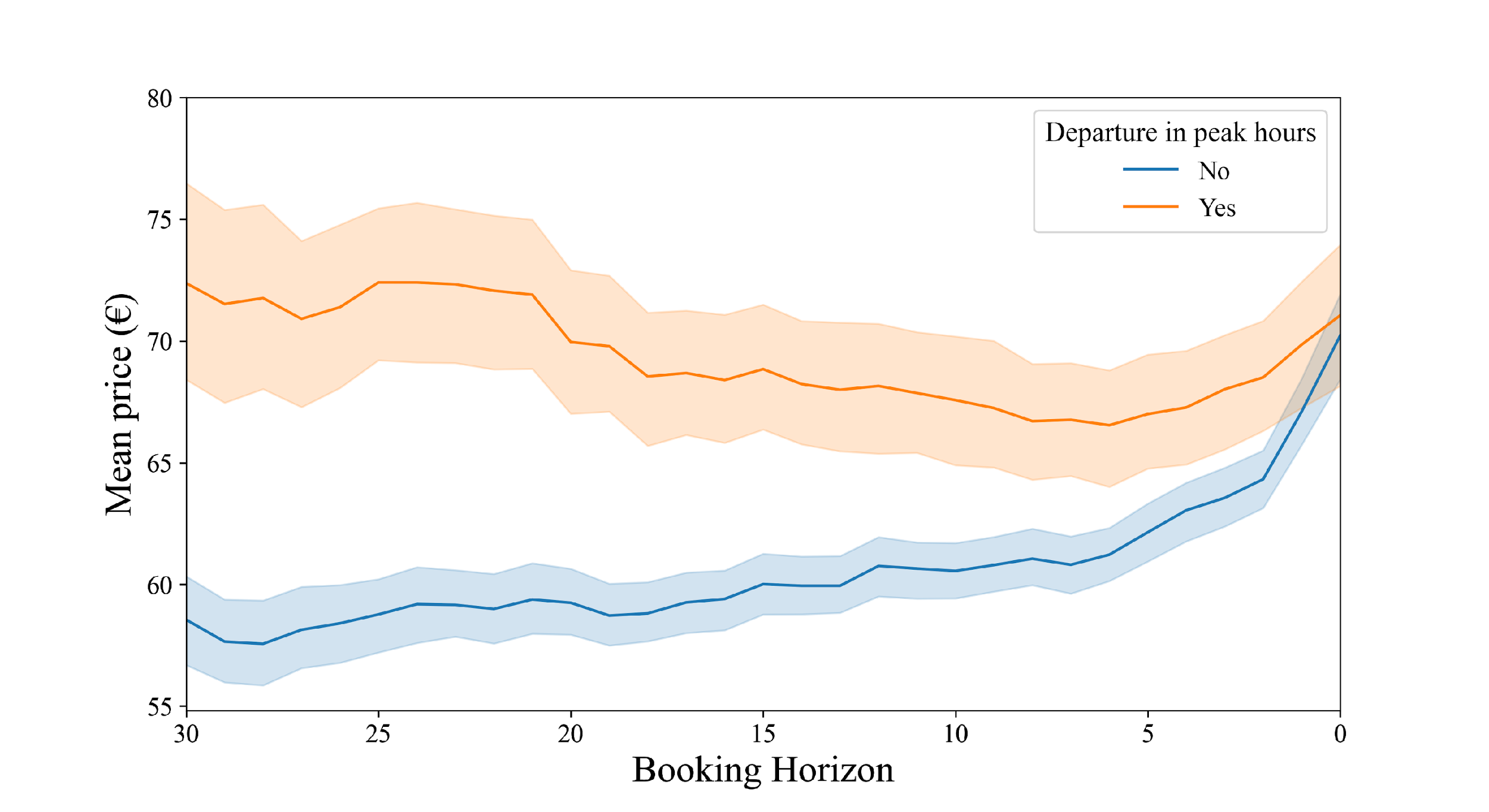


*Note*: Top panel shows the probability that a train’s price increases during the booking horizon. High and low load factor is determined by median load factor, calculated for each train-day subsample. Mean frequency of price increase is 8,0% before and 22,31% during the last week before departure day. Bottom panel shows the mean price increase. Entering the last week before departure day, there is a significant increase in mean price change for trains with high and low load factor. *Source*: own calculations based on data from renfe.com.

Entering the last week before departure day, the price increase probability jumps, at the day of departure the probability that a train’s price increased is 35,6%. From the 7th day before departure the price increase probability is increasing, but there is a considerable drop in this probability around 3 days before departure. Price increase probability before the jump on the 7th day is 8,0%, and it is 6,4% among those trains, that’s load factor is below median. This value during the week before departure is 22,3% on average, among those trains, that’s load factor is below median is 19,7%. This shows that there is a huge increase in the probability of price increase during the week before departure day. Furthermore, we could suspect that this probability before the last week before departure is affected by relative load factor, while during the last day it is affected only by the approaching departure.

We show that not only the probability of price increase is higher during the last week before departure, but the mean price change too on the bottom panel of *Figure 3*. The mean price change (in percentage) is around zero for trains with low load factor with a significant 2% average price discount on the 19th day before departure, while trains with high load factor experience significant price increases during the booking horizon. Entering the last week before departure day, there is a significant increase in 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 trains’ capacity utilization.

*Figure 4: Mean Price for Peak-Hour and Off Peak-Hour Trains*



*Note*: Mean price of peak hour departing trains is 60,7 EUR while for off-peak hour departing trains is 60,1 EUR. Furthermore, 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 *Source*: own calculations based on data from renfe.com

According to Vuuren (2012), differentiating between departure times based on peak travel hours is important. In our initial descriptive analysis and regressions, we used a dummy variable that cut the day in half, roughly around the median departing hour value. This was not efficient, and it did not show any promise according to the visualization. When using the framework of Vuuren (2012), we noticed a big difference between the peak and off-peak hour trains. In the 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, however 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 the peak hour trains. Concluding this, we decided to include this dummy variable in our logistic regression, and we state that the theory of Vuuren (2012) holds for Renfe’s case, as there is no price increase in the last couple days, thus in our opinion, it follows a welfare enhancing perspective, when pricing peak-hour departure trains.

We examine the relationship between train’s capacity utilization, the timeline of the booking horizon and price increase in the next section with logistic regression models.

# Regression analysis

We used logistic regression to investigate the effect of train’ capacity utilization and the approaching departure day on the probability of price increase. Price increase is a binary variable, we plotted its change during the booking horizon on the top plot of *Figure 3*. Among 30336 observations there were only 3529 day that a train’s price increased which is only 11,63% of the data sample. We examined our main variables' effect on the dependent variable in various settings. First, we estimated the effects on two samples from our available data: first we included all 30 days into the analysis, then we only estimated effects in the last 5 days before departure. During the descriptive analysis we explored that there might be different strategies in pricing during this last period before departure than on the whole examined booking horizon of 30 days.

First, we estimated only the effect of the *load factor* and the *approach of departure day* (day of the booking horizon) on the probability of price increase, then included the interaction of the two variables. We always included a categorical variable of the route in our models, as different routes could have different pricing strategies implemented. Next, we included various control variables: based on Williams (2020) we included the *weekday of departure* (0-6) as a categorical variable to capture the fluctuation in demand within weeks – for example weekdays and weekend could have different demand and also pricing conditions; based on Vuuren (2002) we included the dummy variable that categorized peak-hour and off peak-hour trains; finally, we included the dummy of whether there were *sales on the previous day*. We used clustered standard error to control for possible serial correlation in the data.

*Table 2: Logistic Regression Output Table. All 30 Days Included from the Sample*

A képen asztal látható

Automatikusan generált leírás

The output table for the regressions on all 30336 observations is in *Table 2*. In the first regression, out variables of interest are both significant and positive. The coefficient of the days of booking horizon is 0,036 that could be interpreted such that as departure day is approaching the probability of price increase is increasing by 3,6% each day. The coefficient of load factor shows that there is a 2,88% increase that prices will grow as a train’s capacity is increasing by 1%. When we include the interaction of these variables, its coefficient is also significant and positive even when we control for the weekday of departure. Controlling for peak hour departure and whether sales were experienced on the previous day, the effect of the approach of departure day disappears at a 5% significance level, while the effect of load factor remains significant although its magnitude decreased to 1,81%. The logit regression resulted in the same what we saw during the explanatory analysis, that peak-hour trains are significantly less likely to get their price increased and that sales experienced on the previous day has a significantly positive effect on current day increase.

Next, we estimated the same regressions but included observations only that were collected in the last 6 days before departure that included 5202 observations the estimated coefficients are completely different now, and after including our control variables the effect of load factor disappeared, it is not significant. However, the effect of the approaching booking horizon becomes significant at the 5% significance level even after including all control variables. The magnitude jumped to ~22% while when including the whole sample, it was only around 2% – after controlling, not even significant from 0.

*Table 3: Logistic Regression Output Table. Only the Last 5 Days until Departure Day Included from the Sample*

A képen asztal látható

Automatikusan generált leírás

# Discussion

To answer our research question and hypothesis, we used the previously mentioned regression models (see tables 2-3). In this section we are going to summarize the result relating to our hypothesis and the overall conclusion of our analysis.

Understanding the above mentioned two models, represented in table 2 and table 3 respectively, we see that the main difference between the last 30 days and 5 days is the significance of train’s capacity utilization. In our case, we can understand that in the whole 30-day period before the departure of the train, both the load factor, and the day of the booking horizon contributes to increasing the prices. In the last 5 days of the booking horizon, the load factor loses its importance, while the closing departure day still carries an effect. In this scenario, the price increase of the tickets depends on the fact that departure day is getting closer.

By interpreting the results relating to the research question, we get to the conclusion that there are different focuses in pricing, dependent on time. In the investigated 30-day booking horizon, Renfe Operadora aims to smoothen stochastic shocks in demand, by focusing on the remaining capacity of trains, in our view and according to our findings in the literature, this is a welfare optimizing point of view. However, when we are looking at the last 5 days of the observed 30 days, we can see that according to our regressions, Renfe Operadora chooses price according to the same practice that airlines apply. Resulting in a situation where in the last days the prices steeply go up. This combined with the lack of significance of the load factor and the interaction between the two, results in the conclusion that Renfe solely focuses on price increase and profit maximization in the last couple of days. These results accurately answer our research question and offer an accurate display of the pricing motivation of the HSR in Spain.

# Conclusion

In this research paper we examined the high-speed railway 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 uncover the motivation behind the dynamic pricing strategy, whether smoothing stochastic shocks in demand, or maximizing profit through price discrimination is the main driver of dynamic pricing in this case.

Our research is based on literature concerning the pricing systems and guidelines, introducing that a different motivation results in different pricing strategies. Applying this theory to our research of Renfe Operadora, we examined our acquired data of passenger railway ticket prices and train specifications with both descriptive and regression analysis. Both result in the findings that in the last 5 days before departure, Renfe Operadora switches its strategy to focus on profit maximization, while if we examine the last 30 days before departure, we get a result that means they focus on welfare maximization.

To extend the research, we would need consumer level data, this way we could confirm the consumers incentive in buying the 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.

# References

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>

Nunno, R. (2018). Fact Sheer: 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

Vuuren, D. (2002). Optimal pricing in railway passenger transport: theory and practice in The Netherlands. *Transport policy*, *9*(2), 95-106.

Williams, K. (2020). Dynamic airline pricing and seat availability. Cowles Foundation Discussion Paper No. 2103R, 28 May 2020

All data and code used in the paper is available here: https://osf.io/tzcj8/

1. https://www.kaggle.com/thegurusteam/spanish-high-speed-rail-system-ticket-pricing/ [↑](#footnote-ref-1)
2. https://osf.io/tzcj8/ [↑](#footnote-ref-2)