Table of Contents

Executive Summary

"Customer is King" seems to be affecting all markets including the rail transport industry. To maintain a competitive edge, the firm must understand defining characteristics in the demand curve in order to make optimal pricing decisions that are satisfactory to all stakeholders.

The emphasis of this report will be on the process behind deriving the demand function of train ticket sales over the past year (365 days). Given the limited nature of the provided dataset, industry understanding and exploratory data analysis offer insight into key data elements that are further incorporated into the model. **Our simplest (first) approach was to examine each variable and obtain a demand equation**, however, we tweaked the problem-solving approach given the **nuances of price sensitivity across different time frames. Findings from Two-Stage Least Squares confirm that there is unit elastic demand around 3 days prior to departure with lowering elasticity as we move away from departure date.** The chart below shows how price sensitivity varies across five segmented date groups:

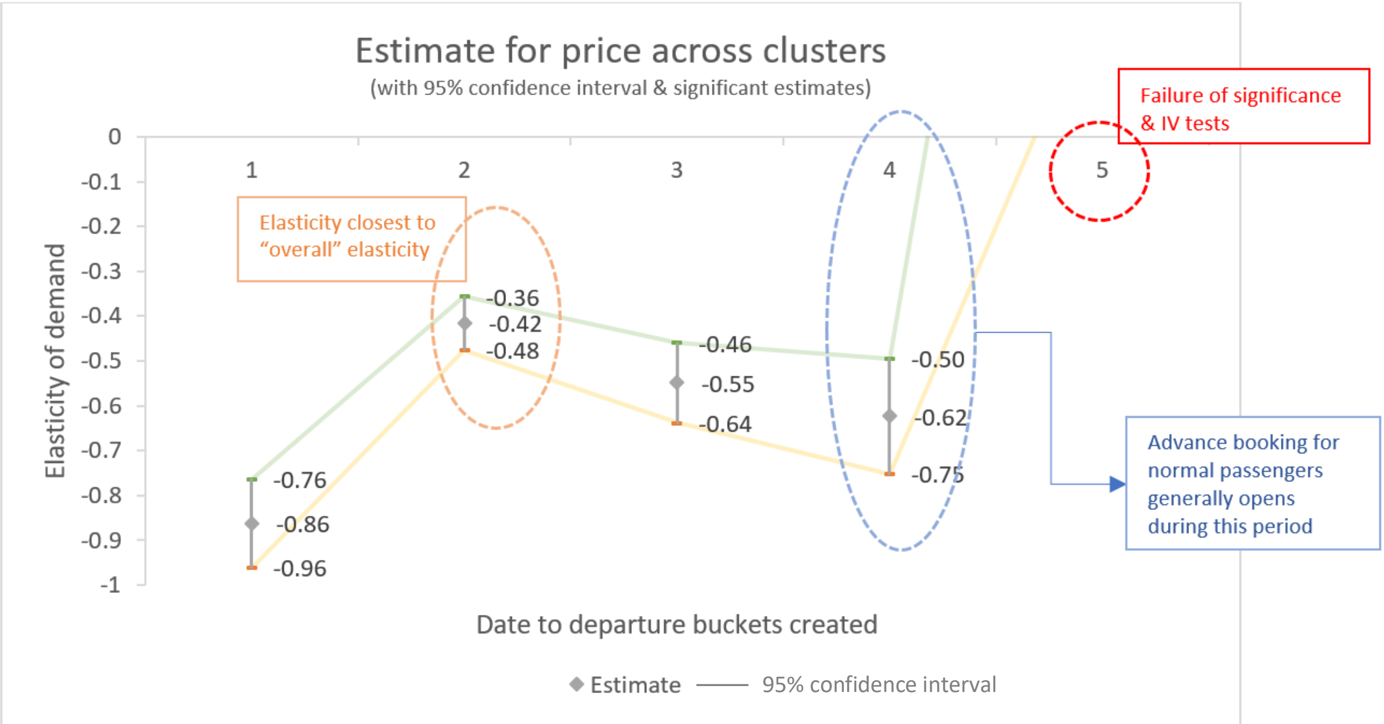


Fig 1

Industry Background

The rail transport industry is one of the most influential industries from the dawn of industrialization and has managed to achieve a degree of maturity, in markets around the world. Due to this, any pricing decision has to be a careful balancing act which requires a keen understanding of sensitivity of demand and motivation for the supplier. Some characteristic traits of a rail transport market are:

- **Enterprise Structure:** Characteristic of the operator (supplier) – public vs for-private profit organization, with the latter being more prone towards fluctuation in prices. A public monopoly, on the contrary, tends to seek cost-efficiencies, overall economic impact, job creation, etc. A market which allows private players **tends to lean towards monopolistic markets, due to high barriers of entry** – namely infrastructure costs, which is passed down to the customer.
- **Class of Ticket:** Quality of the coach and the services offered during/before/after the journey directly impact the ticket price.
- **Price Discrimination:** The transport industry largely relies on price discrimination to increase ridership and tap into consumer market surplus. They can be of various kinds – based on age, occupation, nature of travel, location etc.
- **Nature of Inventory:** The good being transacted, is the space in the train (limited good) – so even if demand increases, the supplier will not necessarily be able to capitalize on it to drive profitability; the inventory expires once the train departs. **Due to this time-sensitive nature, rail operators** use marginal pricing as a tool to adjust for demand.

Data: Transformations & Exploration

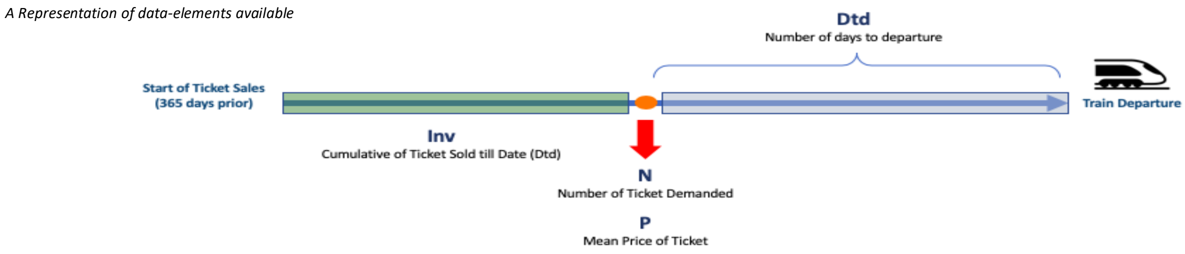


Fig 2

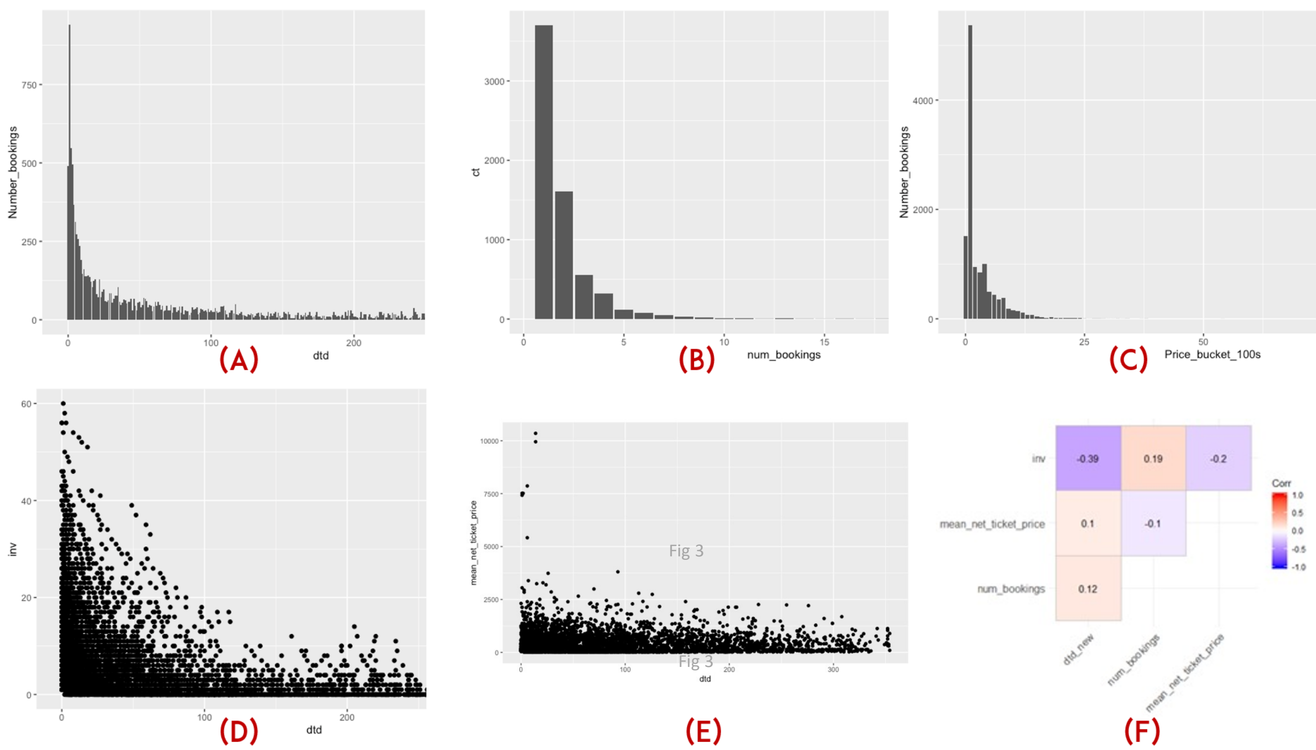
Data Cleaning

Due to insufficient context, **no outlier treatment was performed on mean price or cumulative tickets sold.** Few checks were performed to ensure integrity of data:

- *Integrity of date to departure:* Few days were found to have no bookings made on them. Since most of them were found to be 250 days before departure, no corrective action was deemed necessary.
- *Integrity of duplicate data:* Since our assumption clearly states that price-points are NOT an indicator of the type of journey undertaken, duplicate values are treated as valid data-points with differentiating factor between them missing.

Data Exploration

Variables were plotted to identify distinct patterns and relationships that are present in our limited dataset.



Some underlying trends were observed in the data. It is critical to understand them in order to build a robust model.

Please refer to Fig. 3 above for each of the points below:

A) The trend of number_bookings across dtd shows that the majority of the tickets are booked in the days/weeks/month prior to the departure of the train
B) Most of the bookings are made for a traveller travelling alone or with one companion
C) The high variation in prices cannot be currently attributed due to some data-elements not being present
D) Dtd vs Inv – As the date to departure approaches, the cumulative bookings increase
E) Dtd vs mean_price_tkt there is no real significant relationship in the data, when looked at in an aggregated view
F) Correlation Matrix for the data variables provided

Model Equations & Output

Demand & Supply Equation -

The demand & supply functions (in structural form) have been defined as below:

$$Demand = \alpha_1 + \beta_1 Price + \mu_1$$
$$Supply = \alpha_2 + \beta_2 Price + \gamma_2 Inv + \mu_2$$

The demand function, within the scope of this project, does not have any exogenous variables. The supply function has "cumulative bookings" Inv acting as an exogenous factor, since in a limited supply scenario:

$$cumulative\ bookings \sim Available\ seats\ left$$

Lastly, the principle of simultaneity in Supply & Demand function dictates the price to be endogenous in nature. Since there is an exogenous variable affecting supply, we will try to estimate the demand function by shifting supply curves.

To get the demand function, we need the true estimate of price. The reduced function for price is:
$Price = \pi_0 + \pi_1 Inv + \epsilon_1 \quad =>$ Where; $\mathcal{P}^* = \pi_0 + \pi_1 Inv$ is fed to the demand function.

Assumptions

Certain assumptions were made in our analysis, which have been listed down below:

- As highlighted in sections above, *the price of a ticket is based on enterprise structure, class of ticket & price discrimination*
    - Our assumption is that the data provided to us is from a private profit-maximizing firm operating in an open market I.e. market equilibrium dictates the price of tickets
    - The mean net price of ticket, in isolation, cannot be used to determine the characteristics of the ticket purchased
- For a given period of analysis, the price elasticity is constant

Model Build

With the demand & supply equations as specified in the equation above, where price (P) is treated as an endogenous variable:

- We run an **Ordinary Least Squares (OLS)** model, which should give us a *biased estimate for price* based on endogeneity by simultaneity between Supply & Demand curve
- To obtain the correct estimate of price, we then proceed with a **Two Stage Least Squares (2SLS)** model, using "*cumulative tickets sold* (*inv*)" acting as our instrumental variable to remove the collinearity between price & error in the demand function

Logarithmic functions have been applied to our demand & price variables to obtain the price elasticity for a linear demand function.

Below are the output parameters:

Parameter	Model	
	OLS	2SLS
Estimate for ln(Price)	-0.059	-0.405
P-value (of the estimate above)	< 2e-16	< 2e-16
Multiple R-Squared	0.016	-0.523
Weak Instruments	-	< 2e-16
Wu-Hausman	-	< 2e-16

Table 1

- We observe **a substantial difference between the true estimate & biased estimate** of price, with the estimates calculated with a very high degree of confidence (p-value < 0.05)
- The tests further prove that cumulative sales acts as a **significant IV**
- Although, even such a magnitude of difference, the true estimate places sales of these train tickets well within **inelastic range (-1 < μ < 1)** of price elasticity of demand

Does this provide a holistic view?

While our model above estimates the true value of price-estimate over the entire 1-year period, it is also true that there are various other external factors that impact demand/supply not captured in the current model. Fare elasticity to areas, different geographies, ticket classes, quality of service & reliability are only some of them. A key component of the demand, though, are certain factors that change over time. **Elasticity of other modes of transport, prices of fuel, weather** are factors that influence the decision to purchase a train ticket as they approach the travel date. *(Supply can also be adjusted to match demand, albeit at different price, by changing quota of seats within trains!)*

Based on the latter factors listed above **(in bold)** (some of which could be unpredictable), one would assume that the demand/supply and price relationship(s) adjust multiple times over different periods to achieve equilibrium, ***as these factors can cause the demand curve to shift*. As pointed above, they become increasingly prominent in influencing the purchase of a train ticket** & consequently the demand function as we approach departure date.
In an open market, this shift in demand curve would cause the supply curve to adjust & the equilibrium price remains constant. But because of limited supply of tickets available (*implying a restricted movement of the supply curve*), the equilibrium prices adjust upward/downward. Considering **this market mechanism** and **assuming market adjusts for the influence of external factors specified above,** we hope to capture the true estimate of price (& thus elasticity of demand) across multiple timeframes.

How have timeframes been decided?

The timeframes have been chosen in such a way, that they reflect equal number of bookings across them. This was done to ensure that there is no bias of bookings across a time period of consideration. Consequently, the below buckets were obtained:

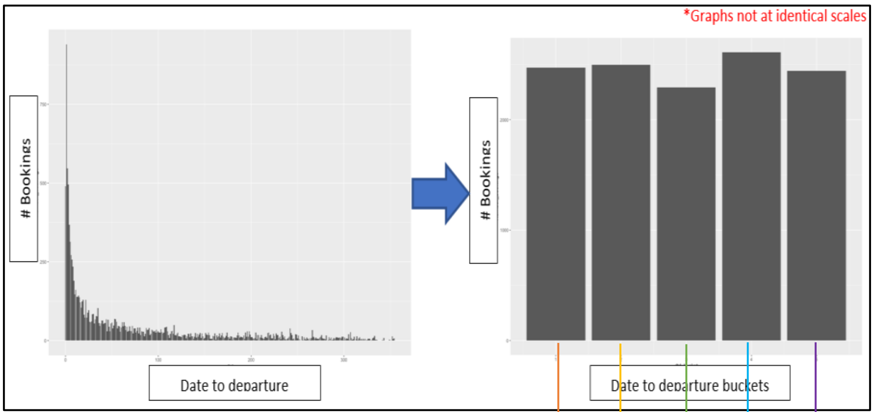


Fig 4

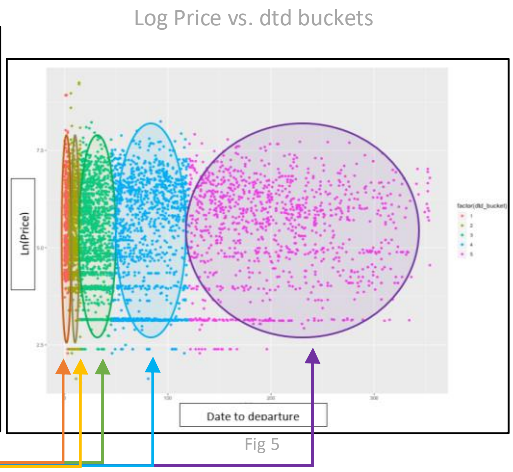


Fig 5

Equations based on clusters:

Estimating the demand function across clusters, we obtain the below estimate values for price when using a 2SLS equation with cumulative booking as IV

Table 2: Demand function parameters across time-based groups

Parameter	Model									
	Group 1 (<=3days)		Group 2 (4-15days)		Group 3 (16-45 days)		Group 4 (46-120 days)		Group 5 (> 120 days)	
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Elasticity*	-0.16	-0.86	-0.06	-0.42	-0.05	-0.55	-0.05	-0.62	-0.06	1.32
P-value (of the estimate above)	2.13e-14	<2e-16	3.47e-07	3.83e-12	5.77e-05	9.40e-10	8.15e-06	1.30e-06	8.32e-06	0.172
R-Squared*	0.049	-0.86	0.016	-0.57	0.01	-1.39	0.01	-2.27	0.01	-7.72
Weak Instruments	-	<2e-16	-	<2e-16	-	9.56e-14	-	3.23e-08	-	0.128
Wu-Hausman	-	<2e-16	-	2.7e-15	-	<2e-16	-	<2e-16	-	1.75e-05

**The values have been rounded off to 2 decimal places*

- We observe that the 1st bucket, created for **3 days before departure**, displays the highest elasticity of demand, albeit **close to unit elastic**. This behavior is expected, as we discussed in earlier sections about other factors playing a larger role close to departure date
- Groups 2, 3 & 4 have a sharp drop in elasticity from group 1 which gradually increases. It is important consider though, that this increase also accompanies an **increase in standard error** of the estimate; **which makes it more difficult to deduce the behavior of elasticity conclusively**
    - Our research has shown that bookings for normal passengers generally begin during the period of group 4 (45-120 days from departure). This could explain the increase in elasticity observed for that group
    - We also note that elasticity in Group 2 is the closest to the overall elasticity we have observed; further validating our assumption that elasticity needs to be observed across time
- Group 5, which has bookings from more than 4 months of departure has a positive coefficient; but with p-value of 0.172 (above our significance threshold) & failure of Weak instrument test, **the coefficient does not represent the true estimate with any degree of confidence**

Appendix

Reference(s):

1. http://eprints.whiterose.ac.uk/2034/1/ITS23_The_demand_for_public_transport_UPLOADABLE.pdf
The Demand for Public Transport: The Effects of Fares, Quality of Service, Income and Car Ownership *by*
Neil Paulley
Richard Balcombe
TRL, Crowthorne House, Nine Mile Ride, Wokingham, Berkshire, RG40 3GA
Roger Mackett
Helena Titheridge
Centre for Transport Studies, University College London, UK
John Preston
Transport Studies Unit, University of Oxford, UK
Mark Wardman
Jeremy Shires
Institute for Transport Studies, University of Leeds, UK
Peter White
Transport Studies Group, University of Westminster, UK
Published by University of Leeds & White Rose University Consortium

2. https://hbr.org/2015/08/a-refresher-on-price-elasticity
A refresher on Price Elasticity *by Amy Gallo*

3. Advanced Industrial Organisation | Identification of Demand Functions *by Måns Söderbom, University of Gothenburg*

4. https://www.citeman.com/2295-pricing-under-discrimination-monopoly.html
Pricing under discrimination Monopoly *by Iñaki Aguirre, Simon G. Cowan (2007)*

5. https://pdfs.semanticscholar.org/5ef0/79672f71ae5d853f1d322d1cbc9a090a5a74.pdf
Marta Jarocka , Urszula Ryciuk (2016) PRICING IN THE RAILWAY TRANSPORT