Model Equations & Output

# 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-maximising 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

# Data Cleaning

**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

# 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 square (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 |

* 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 gasoline/fuel, car ownership & ease of drive, 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! But we restrict our analysis to demand function)*.

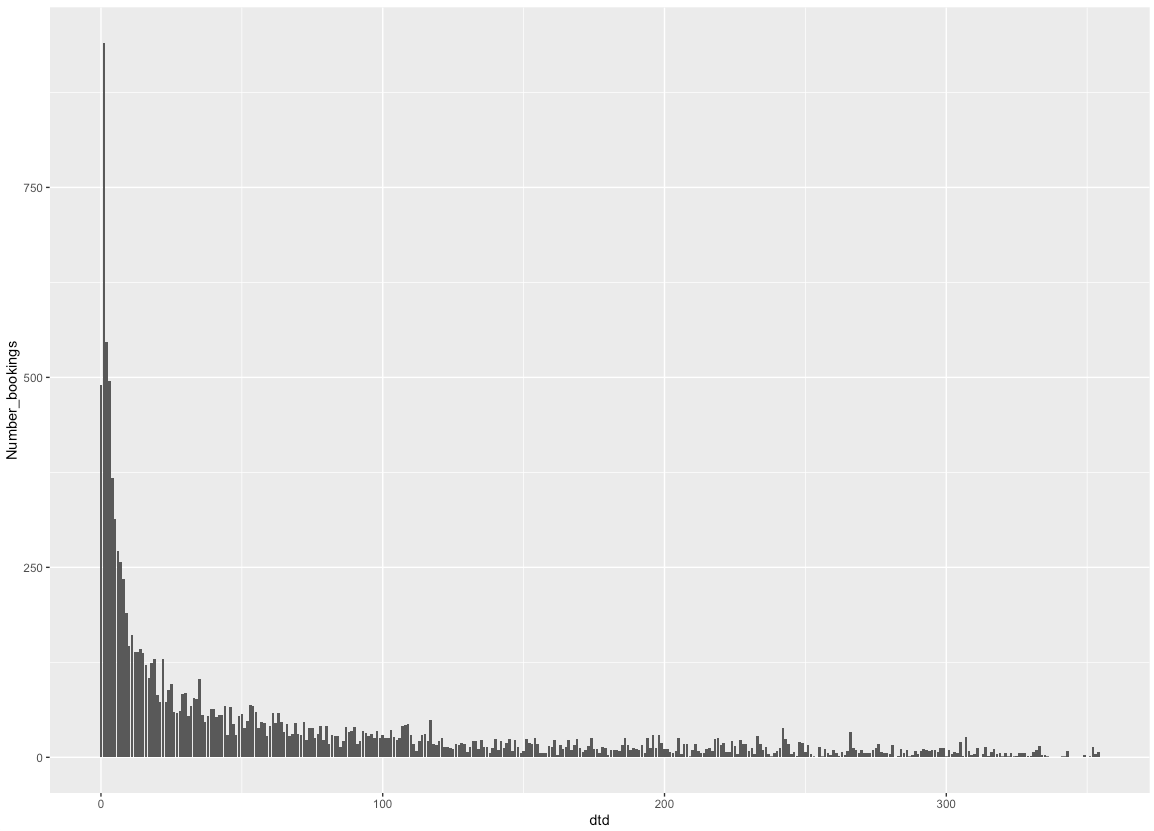
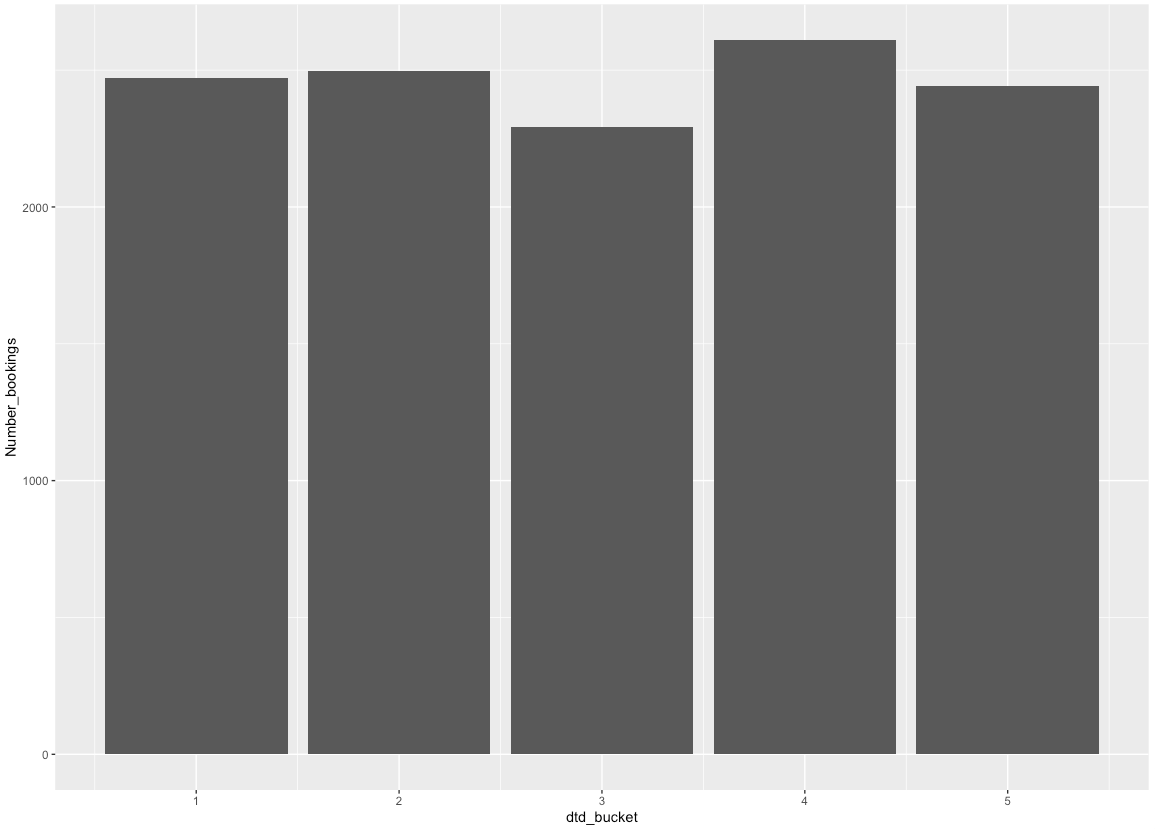
Based on the latter factors listed above **(in bold)** (some of which could be unpredictable), one would assume that the demand/supply and price relationship(s) adjust multiple times over different periods to achieve equilibrium, ***as these factors can cause the demand curve to shift*. As pointed above, 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 mechanic** 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:

\*Graphs not at identical scales

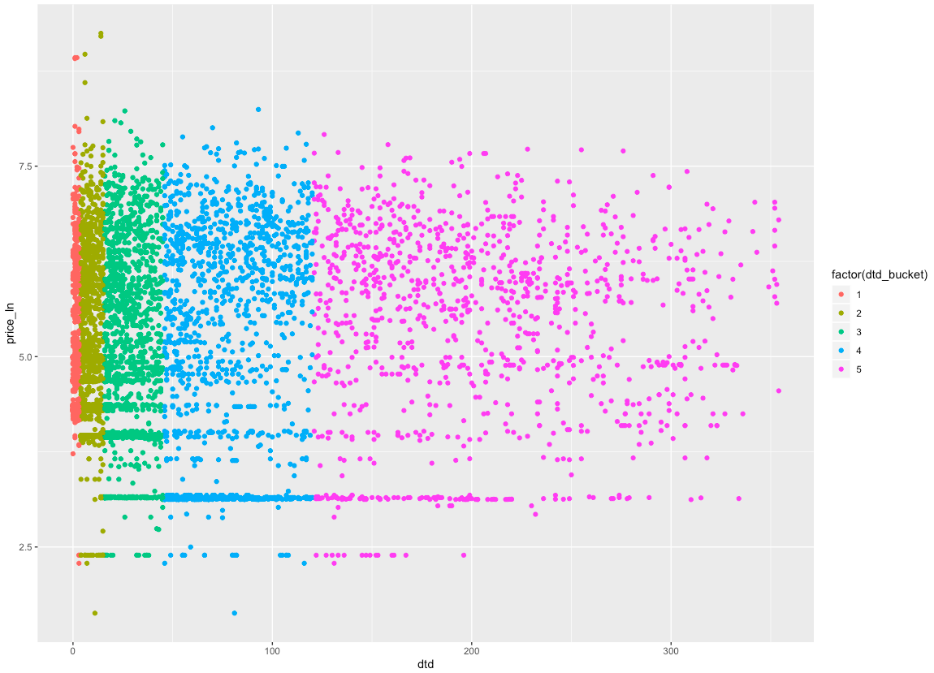
# Bookings

# Bookings

Date to departure buckets

Date to departure

To better understand this split, we see the view below:



Ln(Price)

Date to departure

# Equations based on clusters:

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

Failure of significance & IV tests

Elasticity closest to “overall” elasticity

Advance booking for normal passengers generally opens during this period

* 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 – Supporting Graphs

Table: 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*