Development of a choice model for Railway ticket class selection

P.D.S.B.Chandrasena

(158320P)

Degree of Master of Science in Transportation

Department of Civil Engineering

University of Moratuwa
Sri Lanka

July 2020

Development of a choice model for Railway ticket class selection

P.D.S.B.Chandrasena

Index No 158320p

This is submitted in partial fulfillment of the requirement for the Master of Science in Transportation.

Department of Civil Engineering

University of Moratuwa

Sri Lanka

July 2020

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

The above candidate has carried out research for the Masters Dissertation under my supervision.

Name of the supervisor:

Signature of the supervisor: Date:

ABSTRACT

Sri Lanka Railway is a significant component of the public transport sector in the country with the passengers varying from office workers to tourists. The ability of the railway system to provide accessibility to rural areas and mobility in urban congested areas has resulted in railway system that was earlier established for goods transportation to be a more passenger centric transportation mode in the present.

The railway system consists of three classes namely 1st,2nd & 3rd class with the latter two being considered as non-reservation based or reservations at the counter being based on the availability and hence the focus area of the research.

Currently due to the non-availability of class wise passenger demand calculations has resulted in overcrowding in compartments with passenger complaints of no value for money paid which ultimately will result in passengers shifting to other modes. The research addresses the issue by identifying the behaviour of passengers in choosing the class.

Passenger data related to 333 railway station combinations was used to develop the choice model. Data preparation included calculating the passenger proportions for the two classes for each observed station combinations as the data was treated as aggregate.7 utility forms were considered based on various combinations of distance, price, connectivity strength & travel time attributes.

Based on the estimation results the utility function form with the best fit included price, connectivity strength and travel time as attributes. Travel time influences most towards 2nd class selection whereas connectivity strength or availability of directly connected trains with 3rd class compartments contributes most towards the 3rd class selection.

The developed class wise choice model has the ability to identify passenger behaviour with respect to selecting 2nd class or 3rd class. The model can be further developed by considering more attributes related to facilities of the compartments and also the choices can be expanded to incorporate 1st class also by extending towards reservation based data.

Keywords: Class wise choice model, Railway passenger demand, ALOGIT

ACKNOWLEDGMENT

Initially I have great pleasure in thanking the University of Moratuwa for giving me an opportunity to follow master's degree in Transportation. I am grateful to all the lecturers who guided me throughout my Masters degree programme; especially, Dr. Dimantha De Silva, Supervisor of my Research, who was of continuous guidance with worthwhile instructions throughout my research in making it a success.

I wish to extend my gratitude to all my lectures for valuable instructions given to me throughout my M.Sc. course and research period.

Furthermore, I would like to thank Mr. V.S Polwattage – Deputy Operating superintendent of Railways, Mr. Wijaya Samarasinghe Operating superintendent of Railways for helping me to follow lectures throughout my course in the post graduate programme and granting me the necessary provision to get relevant data required for my research.

Moreover, I am grateful to the higher management of my organization, Sri Lanka Railways for providing me with this valuable opportunity with financial aid to complete M.Sc. in Transportation.

P.D.S.B.Chandrasena
158320P
University of Moratuwa

Contents

LIST OF FIGURES **Page**

LIST OF TABLES Page

LIST OF ABBREVIATIONS

Abbreviation	Description
SLR	Sri Lanka Railway
GM	General Manager

LIST OF APPENDICES

1 INTRODUCTION

1.1 Sri Lanka Railway

1.1.1 History

Ceylon Government Railways (CGR) was established in 1850. Currently it is being functioned successfully as Sri Lanka Railways (SLR) providing unsurpassed service to mother Lanka as a major transport provider. The ministry for transport subject govern by department of Railways. The maid journey of Government Railway was started off from Colombo to Ambepussa with the construction of Fifty four kilometres long main Railway line. This momentous event was occurred on the 27th of December 1864. The official opening of this rail line for traffic was happened on 2nd of October 1865. This 54 kilometres line was vested with the public for transportation on 2nd of October 1865 marking a red letter day in the history of Sri Lanka. After that Sri Lanka Railways began to grow as a giant tree spreading its branches. Thereby, it was able to extend the total length of route up to 1460.91 kilometres as at 2018. Firstly, the main line was extended to Kandy, Nawalapitiya, Nanu-Oya, Bandarawela and Badulla respectively in 1867, 1874, 1885, 1894 and 1924. The other lines such as Mathale, Coast, Northern, Mannar, Kelani Valley, Puttalam, Batticaloa and Trincomalee are completed in the years 1880, 1895, 1905, 1914, 1919, 1926, 1928 respectively.

The underlying intention of the Sri Lanka Railway by transporting coffee and tea, cultivated in hill country to Colombo. It was the root cause, created the railway network in the country. Thereafter Sri Lanka Railways contributed to the establishment of agriculture and industrial areas and the corresponding economic and social development at that time. Rail transport in Sri Lanka initially played a prominent role in the growth of the export sector of tea, rubber and coconut which form the backbone of the economy in the country. Subsequently, it was popular as a mode of providing a valuable service to the public. However, how the prime aim of the Sri Lanka Railways is provision of an uninterrupted service to the commuters. Currently rail transport plays a vital role in reduction road congestion. Most of the people prefer to by train. Then rail line is used for the freight traffic too.

1.1.2 Organizations Structure and Work Force

General Manager (GM) is the Chief Executive Officer in Sri Lanka Railways. Sri Lanka Railways was administrated under many General Managers. Currently it is handled by the 28th incumbent. He is the person who is liable to the Ministry of Transport in respect of an uninterrupted passenger and freight traffic. To fulfil this giant task, he has an Executive staff. There are 03 Additional General Managers in order to carry out the functions in three different sectors namely Operation, Infrastructure and Administration. To perform these functions, Sri Lanka Railways is divided into 08 Sub Departments. They are as the follows.

- Commercial Superintend Sub Department
- Chief Mechanical Engineer's Sub Department
- Chief Engineer - Motive Power Sub Department
- Chief Engineer - Way & Works Sub Department
- Chief Accountant's Sub Department
- Superintend of Railway Stores Sub Department
- Chief Engineer - Signal & Telecommunication Sub Department
- Transportation Sub Department

The Heads of those Sub Departments are responsible for the functions pertaining to the Sub Department.

In addition to the afar 08 Sub Departments, it can be identified as five units within the Railways. There are as follows.

- Planning Unit
- Railway Protection Service Unit
- Sales & Marketing Unit
- Principal Costing officer's Unit Sri Lanka Railway German technical Training Centre

Department of Sri Lanka railways have an approved work force of 25000 out of which 18000 are presently employed. Because of the complexity of its operations, Sri Lanka Railway service is divided to several sub divisions. The office of General Manager Railways is the headquarters where main administrations are carried out. The repair work is the carried out at the Chief mechanical engineer′s sub department at Rathmalana. Currently maintenance work is done at Dematagoda Chief Engineer (Motive Power) sub department while work related to railway tracks are undertaken by Chief engineer (Way and Works) sub department. The railway signal system maintained by Chief Engineer (Signals and telecommunication) sub department. Commercial work is handled by the commercial superintendent sub department while financial work is taken care of by the Chief accountant′s sub department. All the procurements work are carried out by railway stores sub department and the transportation affairs are maintained by the transportation sub department. The safety of Railway passenger and Railway property is been provided by the Railway Security Service.

1.1.3 Railways Operations

The famous Sinhala Saying "Anguru kaka wathura bebee kolamba duwana yakada yaka″ (The iron demon who runs to Colombo eating coal and drinking water) proves its history. The locomotive or the engine of the train ran on steam and power in the past, and the Diesel Locomotives which a prime mover is a diesel engine are on railway at present. The general Electric built locomotives may be on railway in near future.

As the commercial Capital of the country thousands of people travel to the city of Colombo for their business transactions. The vast majority of services are provided by state owned Sri Lanka Railways as the premium national transportation mode to fulfil the needs of all types of passengers who choose and rely on Railways..

The present census revealed that more than three hundred thousand people has chosen the rail transport as their mode of travelling due to the convenience provided for passengers. The Ruhunu Kumari, Galu Kumari, Sagarika and Samudra Devi is operated from Matara and Galle while Sencadagala Manike travels between Kandy to Colombo. Muthu Kumari runs from Chilaw and the ever famous Mahawa Train runs from Maho Junction. Pode Manike and Udarata Manike are operated to Badulla and the Yarl Devi train travelled to Anuradapura, Kilinochchi and Jaffna. The train Meenagaya runs from Trincomalee and Batticaloa while numerous Night mails, Inter-cities and Power sets contributes to this task. If not for the Railway, thousands of man hours will be lost for the country, and as a developing country the drawbacks it creates may affect the future development of the country. People with no auto ownership would face huge difficulties in fulfilling their travelling needs.

The current railway system consists of total track length of 1460.91 km with 10 lines and 4 spur lines. The railway system also consists of 411 railway stations with 179 train stations, 162 sub stations and 70 unattended platforms (no officers and only with the facility of entrain and detrain).

The destination of the respective lines are as follows.

1. The main line = Badulla
2. Coastal line = Beliaththa
3. Kelani Valley line = Avissawella
4. Northern line = Kankasanthurai
5. Eastern line = Batticaloa
6. Chilaw line = Puttalam/Noornagar
7. Mannar line = Thalaimannar pear
8. Matale line = Matale
9. Trincomalee line = Trincomalee
10. Minhinthale Line = Mihinthale

Additionally, extension projects are underway to extend the coast line to Katharagama.

Line wise operational track lengths are given in Table 1.1

Table 1.1: Line wise operational track lengths

No.	Line Name	Length (km)	Length (km) Year-2018 Progress
1.	Main Line	290.49 km	
2.	Matale Line	33.75 km	
3.	Puttalam Line	133.86 km	4.392 km (SED-KTK) Double Line
	Puttalam Line – Airport Spur	2 km	
	Puttalam Line- Aruvakalu Spur	37 km	
4.	Northern Line: Polgahawela to Kankasanthurai (KKS)	339.35	
5.	Thalaimannar Line:Medawachchiya to Talaimannar Pier	106km	
6.	Batticaloa Line (MHO –BCO)	211.10 km	
7.	Trincomalee Line (GOA – TCO)	70.23 km	
8.	Coast Line (FOT –Beliatta)	158.73 km	7.399 km (KTS-PGS) Double Line
9.	Kelani Valley Line (MDA – AVS)	59.27 km	
10.	Mihintale Line (ANP – MHN)	15 km	
11.	Kolonnawa Spur	2 km	
12.	Harbour Spur	2 km	
	Total	**1460.91 km**	

Source: (2018 Sri Lanka Railway (SLR) Admin Report)

1.1.4 State of current railway passenger transport system

The railway system was mostly based on stakeholder requirements with factors such as geographical condition and technical potential. The originally established railway lines and stations being a major part of the current railway system is a significant aspect of the SLR.

Railway transportation has grown from goods transportation towards passenger transportation. Railway currently facilitates 3% of the total passenger transportation in the country. This is reflected in SLR statistics with passenger trains trips and good train trips having 95% and 5% respectively. Though the operational activities were affected by the trade union strikes in 2018, SLR managed to continue its passenger and freight transportation. Table 1.2 shows the comparison of no of train trips in year 2017 and 2018.

Table 1.2: No of Train Trips Run – Comparison 2017 & 2018

Description	2017	2018	2018 %	Inc/Dec	Inc/Dec %
No of Passenger Trains Trips Run	152,005	117,189	95%	(-34,816)	(- 22.90 %)
No of Goods Trains Trips Run	6,590	6,845	5%	255	03.87 %
Total	**158,595**	**124,034**	**100%**	**(-34,561)**	**(-21.79%)**

Source: (2018 Sri Lanka railway Admin Report)

It was at present that the importance of rail transport service was at a very high level. The following reasons have had a great impact in regards.

1.1.5 Reason for Passengers prefer Railway Transport & increase of the rail passenger demand

- When the Railway provides a good service that meets the preferences of commuters to encourage them to use rail. It is free from traffic conjunctions, parking problems and environmental problems. People who travel by train can freely end their journey without any stress. Somebody can get a solace by watching surroundings. It can gain an opportunity to take a rest.
- Travelling by train is more economical comparing with other modes of transportation. It saves both money and time.
- Travelling are free to book the service according to their budget. It is as the first class and the second class.
- Train is an Eco friendly mode of transportation in comparing with others like bus, aeroplane etc. What the train emits are not released directly to the air.
- Travelling by train is not only a fast journey but also a ride somebody can gain amazement. It is inspired by the beauty of the environment like waterfalls, lakes etc.
- People who travel by train automatically get opportunity to befriend with others. They easily become friends through chats. It builds new relationships and bonds. Train is a fair bridge of building new relationships among people.
- Train ride is not monotonous. Commuters can travel as they wish. They can move, chat, sleep without disturbances to others.
- Unlike airplanes, which travels from point A to point B with barely a glimpse of what′s in between, a train ride can be a destination in and of itself. Peering into the downtown of various stops, and lakes, mountains, caves, waterfalls or the permafrost landscape of destinations is something that cannot be done by a plane.

- People can catch attractive sceneries. They can forget sorrows in their hearts for a moment.
- This is very important for long distance travel. Travellers can gain the maximum enjoyment by providing facilities like private cabins.
- Travellers can free to carry wheelchairs, pets, music instruments etc.
- Train travellers can taste some countryside foods like mangoes, tamarind, corn etc. At the same time, they can have refreshment facilities in the train especially for long trips.
- To travel by train, it is not necessary to follow a long procedure. It is very simple and easy. It is not very difficult to travel by train.

1.1.6 Last Year's Passenger Demand Comparison

In 2018, it has been operated 7,709,520,908 passenger kilometres which was 7,495,063,930 passenger kilometres in the previous year and the number of passengers transported was 137,524,452 in 2018 which was 136,662,995 passengers in the previous year. Therefore, it is fantastic of increasing passenger kilometres by 2.86% and passengers by 0.63%.

Table 1.3: Passenger demand comparison

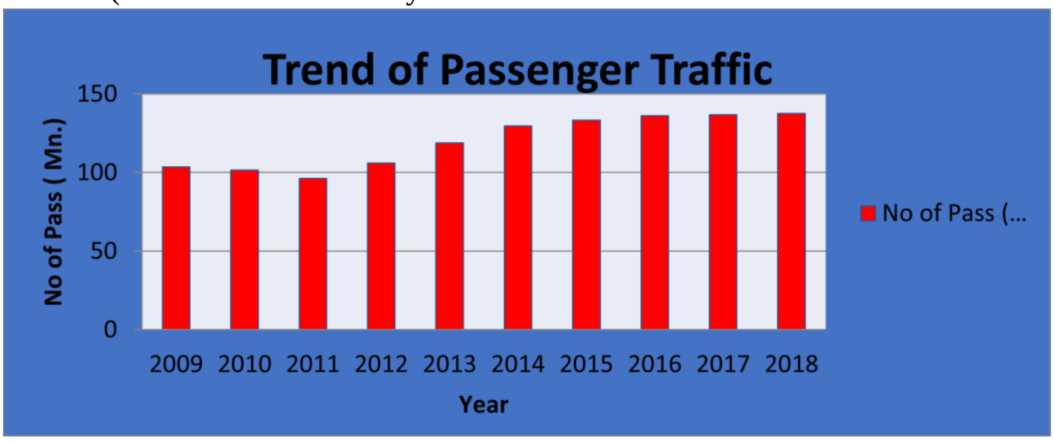
Description	2017	2018	Inc/Dec	%
No of Passengers	136,662,995	137,524,452	861,457	0.63%
Passenger kms	7,495,063,930	7,709,520,908	214,456,978	2.86%

Source: (2018 Sri Lanka railway Admin Report)

In year 2018 compared with year 2017, no of passenger demand increase slow down and increase percentage was 0.63 %. However, passenger kms increases near 3%. When railway passenger demand of the last 10 years shows gradually increase and in the last three years, it had come to peak level similar to the railway current infrastructure level.

Figure 1.1 trend of total no of passenger traffic for the 2009 to 2018.

Source: (2018 Sri Lanka railway Admin R



Considering the period from 2009 to 2015, passenger traffic is growing rapidly. With the end of the country's civil war in 2009, there was a high improvement in infrastructure development with the railway transportation. Preliminary initiatives for infrastructure development of the Sri Lanka Railway Department were carried out during the period from 2009 to 2011. During that time, passenger traffic showed a decline due to infrastructure repairs. But in the years that followed, (Year 2012 to 2015) there was a significant increase in rail traffic with the restoration of the Northern Railway line and the construction of new railway tracks like TML line. However, from the year 2016 onwards, railway passenger demand has been showing stability. This is due to the fact that the existing infrastructure of the Railway Department is not sufficient for increasing future passenger demand.

Figure 1.2 show the trend of total passenger kms traffic for the last ten years.

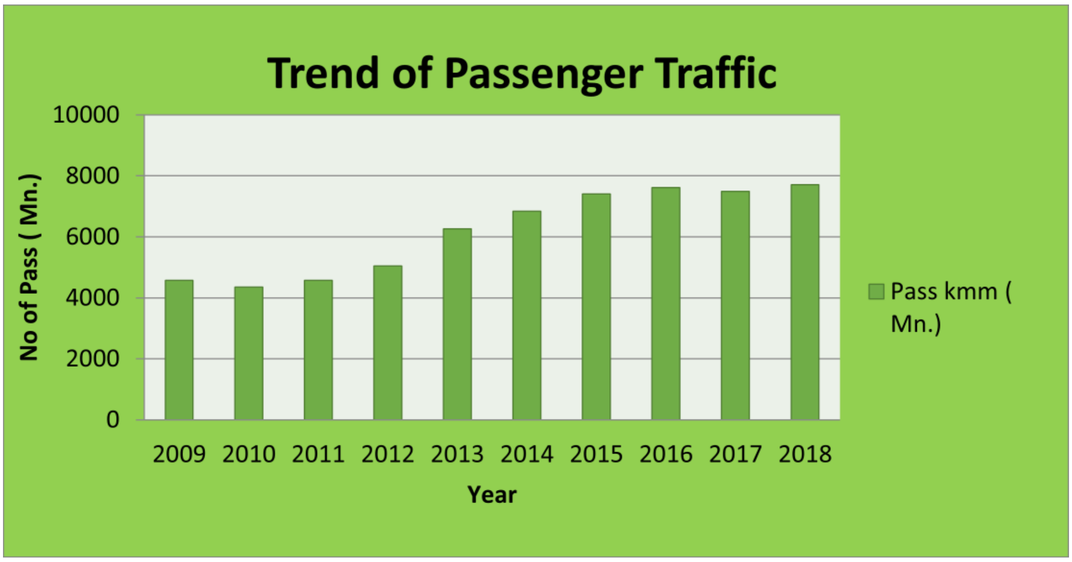


Figure 1.1: Passenger traffic trend – Passenger km

Source: (2018 Sri Lanka railway Admin Report)

Passenger kms chart is showed similar reflection of the no of passenger demand. Infrastructure development was showed a decrease in passenger traffic during the period from 2009 to 2011. But then came the period from 2012 to 2015 which saw a very rapid increase in passenger traffic. However, during the following years from 2015 to 2018, passenger traffic was stable.

1.1.7 Future Plans and Proposal

The Railway is now primary engaged in the transport of passengers, especially commuters to and from Colombo offering vital service and reducing road congestion and thus Railway Services play a pivotal in transportation of people and goods in the country. Over the past 150 years, Sri Lanka Railway service has provided a priceless service to the entire nation. The future generations expect the service to be more sophisticated and modern in the years to come.

The Sri Lanka Railway service has a focussed plan to make this expectation a reality. This plan comprises of train operated countries and modernized rail lines and many other novel devices. The total length of the railway lines in Sri Lanka is currently 1460.91 km. A further 732 km will be added to it in the future and as a result of it the total length of the railway lines in the country will be 2192.91 kms

Proposed new railway line:

- The New line from Kurunegala to Habarana (via Dambulla)
- The new line from Palavi Puttalam to Maho Junction
- The extension of Kelani Valley line from Padukka to Hambantota (Via Rathnapura Godakawela, Embilipitiya and Suriyawewa)
- The new line from Monaragala to Embilipitiya
- The new line from Anuradhapura to Kanthalai
- The rebuilding of the old line from Nanu-Oya to Nuwara Eliya
- The Light rail track connecting Anuradhapura, Trincomalee, Polonnaruwa and Dambulla which are due to be developed as Metros the future.
- The new line from Matara to Batticaloa (Via Hambantota, Monaragala, Ampara and Oluvil)

Among these proposed lines the new coastal line from Beliaththa to Hambanthota is significant in many ways. This line planned to build through the modern city of Hambantota which is comprised of Magam Ruhunupura port and Mattala international airport. Furthermore, because of the Kelani Valley lines that come to Hambantota across Ratnapura a new railway junction will be created in Hambanthota. It is lucrative many ways since this line runs across Ampara which is used from Dayata Kirula and empowered through Negenehira Navodaya. This will contribute towards to development of the Eastern tourism industry which is revolving Arugam-bay and Pasikuda. Feasibility studies have been carried out for most of these proposed lines and the necessary fund will be allocated in the future on a priority basis.

The modernization of the existing line is a vital part in this development process. Presently the trains are running at a speed of 80 kmph (Max speed) in the coastal line and all other lines will be upgraded to this standard in the future.

1.1.8 Important of the future rail passenger demand

Due to the importance of the railway system on both the socio and economic spheres of the country, the government in the last decade has made large investment on the train services. One important aspect that should be considered in these capital investment decisions is the passenger demand of train services. A proper assessment of the passenger demand will ensure that provided service will be efficient, productive and economically feasible. Currently passenger demand calculation for train operations in Sri Lanka is carried out in a very small scale by external institutes and hardly by the department.

One of the main features of public transport modes is the provision of categorised services. For example, passenger bus services in Sri Lanka are categorised as luxury, semi luxury and normal. But one of the key differences in air line and train services when compared with bus services is the categorisation in the same vehicle based on the different facilities provided. Railway service in Sri Lanka provides three classes namely 1st class, 2nd class and 3rd class based on the facilities provided with variation of train fare based on the class. The class segregation in railway services is important to attract different passenger types based on the facilities or comfort levels.

An important aspect of passenger demand calculation is determining the class wise passenger demand. It is useful when setting the number of train compartment of each class when introducing new train trips and when deciding on mixing classes in new train sets. Estimation and analysis of class wise passenger demand can also be utilized in manufacturing and importing train coaches.

1.2 Research problem

Non-availability of class wise passenger demand calculation has resulted in following issues in the current railway system.

1. Issue of 2nd class compartment being overcrowded and hence the passengers not being able to enjoy the facilities paid for.

2. Issues of one class being overcrowded while the other class is empty

3. Due to the above, issues of passengers shifting towards other transport modes.

Hence the need for identifying how railway passengers choose the class.

1.3 Research objectives

Based on the above research problem, the following research objective is derived.

- Understand the attributes that influences 2nd and 3rd class ticket purchase
- Development of a choice model railway ticket class selection.

2 LITERATURE REVIEW

2.1 Introduction

This chapter will review existing literature on areas related to the objective of development of a choice model for passengers' class preference.

2.2 Alternatives and choices

Existence of multiple alternatives results in a person having to make a choice. The set of all possible existing alternatives is defined as the universal alternatives set while the set of alternatives known to the decisionmaker is defined as the choice set.

(Fischhoff et al., 1977) and (Svenson, 1979) identified the following 4 categories of rules that can be used by the decisionmaker when in scenarios of multiple alternatives.

1. Difference between alternative values
2. Concern of level of satisfaction
3. Lexicographical rule
4. Utility maximization

The rule of utility maximization is recognized as suitable when modelling rational choice behavior of decisionmakers.

2.3 Passenger behavior choice models

(Shibata et al., 2010) used Multiple Indicator Multiple Cause model (MIMIC) to identify passenger behaviour in making train reservations. MIMIC model uses observed variable to predict latent variables and consists of structural and measurement equations as shown in figure 2.1.

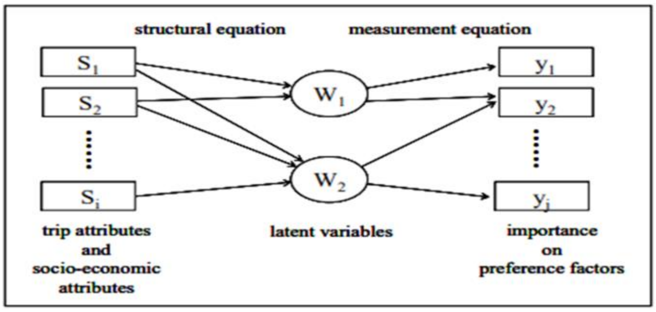


Figure 2.1: MIMIC model (Shibata et al., 2010)

Further the goodness of fit of the logit model has been improved by introducing latent variables.

(Hetrakul & Cirillo, 2014) used multinomial logit model and latent class model to develop a model for railway passenger seat booking behavior. The Multinomial logit model uses variables that represents passenger tradeoff behavior (i.e. fare and advanced booking) to develop utility function that calculates utility for each day. The latent class model uses individual specific variables that represents trip characteristics (i.e. time of day, day of week) to find utility separately for each class.

(Kim et al., 2017) also uses latent class modelling to analyse mode choice preference for freight transportation in New Zealand. Further latent class choice modelling is done using fixed parameters and random parameters with the model with fixed parameters providing a better overall model fit.

2.4 Logit choice modelling

(Lancaster, 1966) defined the utility function based on vector of attribute values for every alternative for every decisionmaker. (Ben-Akiva & Lerman, 1985) expands it defines the utility function by also considering vector of different characteristics of the decisionmaker (i.e. age, income, education background etc.) to incorporate variability in the population.

Utility consists of a measurable component and an error component. Probabilistic Choice Theory defines that a person's selected alternative can be identified by identifying the utility with the highest probability. The distribution of the error component will determine the probability expression for utility.

Error component has a normal distribution. But due to the complexity of the resulting probability expression (Probit function) the error component is assumed as a Weibull distribution. The resulting probability expression is known as the logit function and is as follows.

$$P[a', i] = \frac{e^{\lambda V(a', i)}}{\sum_{a \in A} e^{\lambda V(a, i)}}$$

Where:

P [a', i] : Probability of selection of alternative a' by individual i

V (a', i) : Utility of alternative a' by individual i

A : Set of all alternatives

λ : Dispersion parameter

Based on logit estimation results the T-ratio is calculated for each estimated parameter to check whether the parameter is significantly varies from 0. If the T-Ratio is greater than 1.96 the estimated parameter is significant.

$$t - ratio[k(1)] = \frac{k(1)}{SE_{K1}}$$

The T-Statistic is calculated between similar attributes that have similar estimates to check whether the parameters are significantly different. If the T-Statistic is greater than 1.96 the estimated parameters are significantly different.

$$t - stat\ [k(1), k(2)] = \frac{k(1) - k(2)}{\sqrt{{SE_{K1}}^2 + {SE_{K2}}^2 - 2r_{k1,k2}SE_{K1}SE_{K2}}}$$

3 DATA

3.1 Dataset

Monthly issued tickets from 333 railway stations for the following 2 types of ticket stocks were collected.

1. Pre-printed tickets (frequently issued)
2. Paper tickets (ad hoc issue due to pre-printed tickets not being available)

In addition, the data contains the ticket issued station and the destination, based on which an OD matrix was derived.

Data relating to May,2018 to September,2018 was collected. Based on 333 railway stations the dataset should have contained 110,889 combinations. But due to unavailability of records in some stations the dataset contains 97,020 combinations.

The following data was also available.

- Distance data for all railway stations from Colombo Fort railway station
- 2nd class & 3rd class composition for every train
- Train timetable

3.1.1 Total distance

Based on the available distance data, distance for every railway station combination was calculated. Distance (Dis) is a common attribute for both considered classes.

3.1.2 Availability

For each railway station combination, the relevant trains that connect the pair were identified. These trains were further categorized based on 2nd class & 3rd class composition to ultimately calculate available number of trains for 2nd class (Avi2nd) & for 3rd class (Avi3rd) for each station combination.

3.1.3 Travel time

Average speed of each train was calculated using distance data and train journey time obtained using the train timetable. Average train speed for each station combination was calculated based on the identified trains and class in 3.1.2 and the relevant average train speeds calculated. Based on the average speed of trains for each class and each station combination and station combination distance (calculated in3.1.1), travel time for 2nd class (Time2nd) and 3rd class (Time 3rd) was calculated. Waiting time adjustment was made based on train timetable.

Sample dataset is shown in table 3.1.

Table 3.1: Sample dataset

ID	Dis	Avi 2nd	2nd Price	Time 2nd	2nd pass	Dis	Avi 3rd	Time 3rd	3rd Price	3rd pass
1	0	0	0	0	0	0	0	0	0	0
2	1.768	10	20	0.680523	0	1.768	66	0.613799	10	5751

3	4.973	5	20	0.543629	0	4.973	53	0.545542	10	13015
4	6.776	3	20	0.564695	0	6.776	38	0.543728	10	7953

The dataset contains the following attributes.

ID : Journey identity number based on origin & destination stations

Dis : Distance of the journey in kilometres (x_{1i}, x_{1j})

Avi2nd : No. of directly connected trains with 2nd class compartments for the journey (x_{3i})

2ndPrice : 2nd class price for the journey in Rupees (x_{2i})

Time2nd : Travel time for available train with 2nd class compartments

$$S = \frac{\sum_{i=1}^{Avi2nd} s_i}{Avi2nd} \qquad \text{Time } 2^{nd} = \frac{Dis}{S}$$

where; s_i: average speed of i^{th} train mentioned at Avi2nd (x_{4i})

2ndpass : Average No. of 2nd class passengers who travelled the same journey

Avi3rd : No. of directly connected trains with 3rd class compartments for the journey (x_{3j})

3rdPrice : 3rd class price for journey in Rupees (x_{2j})

Time3rd : Travel time for available train with 3rd class compartments

$$S = \frac{\sum_{i=1}^{Avi3rd} s_i}{Avi3rd}; \qquad \text{Time } 3^{rd} = \frac{Dis}{S}$$

where;

s_i: average speed of i^{th} train mentioned at Avi3rd (x_{4j})

3rdpass : Average No of 3rd class passengers who travelled the same journey

1st class compartments were not considered due to it belonging to reservation based data and the focus of the research was on non-reservation passenger data.

3.2 Data analysis

3.2.1 Descriptive statistics

Table 3.2: Descriptive Statistics - Distance

Var	N	N*	Mean	SE Mean	St. Dev	Min	Q1	Median	Q3	Max
Dis	97020	0	191.02	0.376	117.05	0.000	96.00	178.00	275.00	572.00

The maximum distance between two stations is 572 km. The minimum is 0 due to non-availability of some data. The average distance between two stations is 191.

Table 3.3: Descriptive Statistics - 2nd class availability

Var	N	N *	Mean	SE Mean	St. Dev	Min	Q1	Median	Q3	Max
Avi 2nd	97020	0	0.3097	0.00333	1.0370	0.00000	0.00000	0.00000	0.00000	37.0000

Maximum availability which represents the maximum number of trains with 2nd class between two stations is 37. Minimum is 0 due to many station combinations not having trains with 2nd class compartments. The mean is a low value of 0.3 due to availability of 2nd class trains for large number of station combinations are 0 or very low.

Table 3.4 : Descriptive Statistics - 2nd class Price

Var	N	N*	Mean	SE Mean	St. Dev	Min	Q1	Median	Q3	Max
2nd Price	97016	4	265.09	0.458	142.53	0.000	160.00	260.00	360.00	800.00

Maximum price of 2nd class is Rs.800.00. Minimum of 0 is due to non-availability of data. Average of 2nd class price is Rs.265.

Variable	N	N*	Mean	SE Mean	St. Dev	Min	Q1	Median	Q3	Max
Time 2nd	97020	0	6.880	0.0851	26.506	0.000	0.000	0.000	0.000	328.9

Table 3.5 : Descriptive Statistics - 2nd class travel time

Time 2nd is the average travel time between two stations for trains with 2nd class compartments. Travel time data is available only at station combinations that are directly connected with trains with 2nd class compartments. The maximum value is 329 minutes (about 5 hours and 30minutes). Average of 6.8 minutes results from many data entries with zero, which occurs due to non-availability of directly connected trains.

Var	N	N*	Mean	SE Mean	St. Dev	Min	Q1	Median	Q3	Max
2nd Pass	97020	0	4.19	0.382	118.8	0.000	0.000	0.000	0.000	15422.0

Table 3.6 : Descriptive Statistics - 2nd class passengers

The number of passengers travelling with 2nd class tickets between station combinations is represented by the variable. The maximum number of 2nd class passengers between two stations is 15422. There are station combinations where no passengers used 2nd class. 4 passengers (2nd class) on average travel between two stations.

Table 3.7 : Descriptive Statistics - 3rd class availability

Var	N	N*	Mean	SE Mean	St. Dev	Min	Q1	Median	Q3	Max

Avi 3rd	97020	0	1.161	0.0143	4.454	0.000	0.000	0.000	0.000	140.000

Maximum availability which represents the maximum number of trains with 3rd class between two stations is 140. Minimum is 0 due to many station combinations not having trains with 3rd class compartments. The mean is a low value of 1.16 due to availability of 3rd class trains for large number of station combinations are 0 or very low.

Table 3.8 : Descriptive Statistics - 3rd class travel time

Var	N	N *	Mean	SE Mean	St. Dev	Min	Q1	Median	Q3	Max
Time 3rd	97020	0	7.458	0.0854	26.609	0.000	0.000	0.000	0.000	328.900

Time 3rd is the average travel time between two stations for trains with 3rd class compartments. Travel time data is available only at station combinations that are directly connected with trains with 3rd class compartments. The maximum value is 329 minutes (about 5 hours and 30minutes). Average of 7.4 minutes results from many data entries with zero, which occurs due to non-availability of directly connected trains. Any station in Sri Lanka is connected by train but not all of them are directly connected.

Table 3.9 : Descriptive Statistics - 3rd class passengers

Var	N	N *	Mean	SE Mean	St. Dev	Min	Q1	Median	Q3	Max
3rd Pass	97020	0	38.5	1.86	580.5	0.000	0.000	0.000	0.000	51238.0

The number of passengers travelling with 3rd class tickets between station combinations is represented by the variable. The maximum number of 3rd class passengers between two stations is 51328. There are station combinations where no passengers used 3rd class due to non-availability of directly connected trains. 38.5 passengers (3rd class) on average travel between two stations.

	2nd pass	3rd pass
Mean	4.073546	37.42816
Standard Error	0.371293	1.813928
Median	0	0
Mode	0	0
Standard Deviation	117.2328	572.7332
Sample Variance	13743.52	328023.4
Kurtosis	6823.066	2632.516
Skewness	71.08506	42.84011
Range	15422	51238

Minimum	0	0
Maximum	15422	51238
Sum	406104	3731326
Count	99693	99693

Table 3.10 : Descriptive Statistics - 3rd class Vs. 2nd class passengers

Accordingly table 3.10 2nd class number of passengers mean is 4.0 and 3rd class mean is 37.4. 3rd class passenger count distribute wide range as 51238 but 2nd only 15422.

3.2.2 Scatterplots

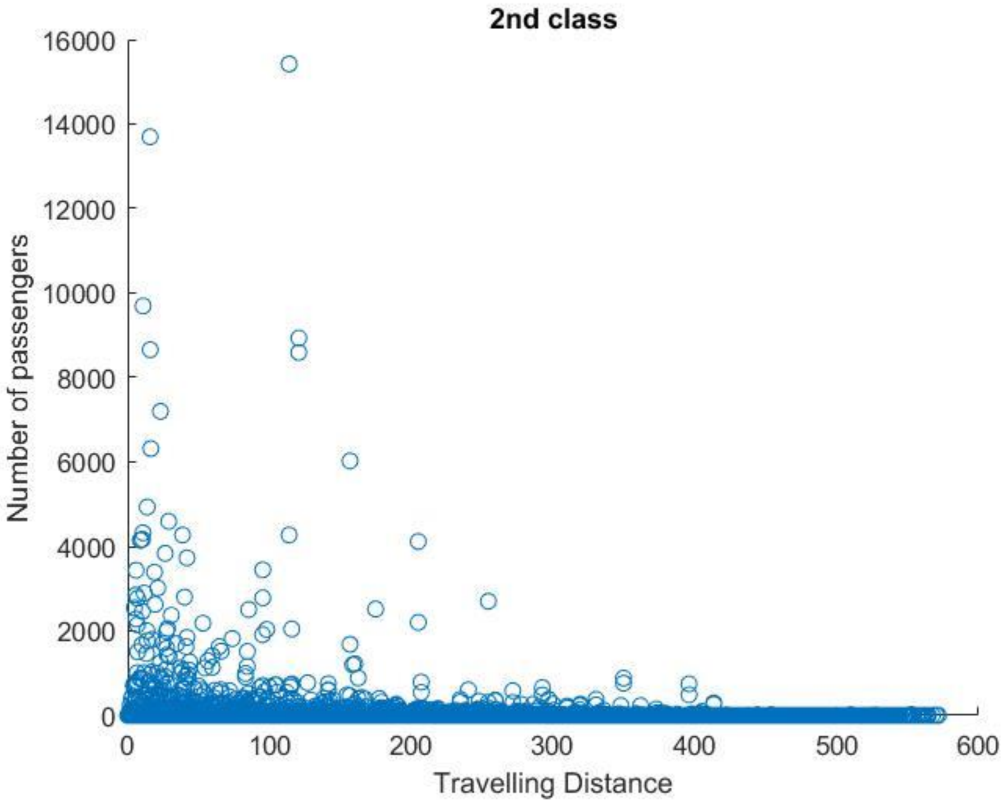


Figure 3.1 : Scatterplot of no. of 2nd class passengers vs. travelling distance

According to figure 3.1 there appears to be no linear relationship between the two variables. But it also shows data points skewed to 0 which implies that for most of station combinations, when travelling distance increases number of 2nd class passengers decreases.

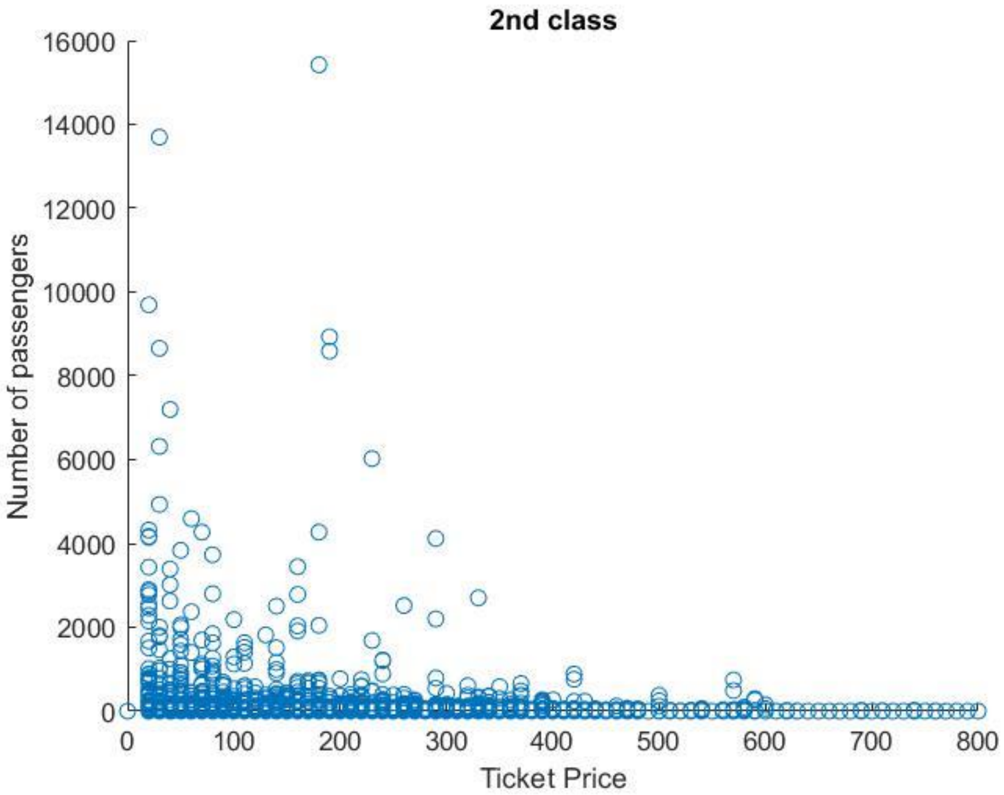


Figure 3.2: Scatterplot of no. of 2nd class passengers vs. travelling distance

According to figure 3.2 there appears to be no linear relationship between the two variables. But it also shows data points skewed to 0 which implies that for most of station combinations, when travelling distance increases number of 2nd class passengers decreases

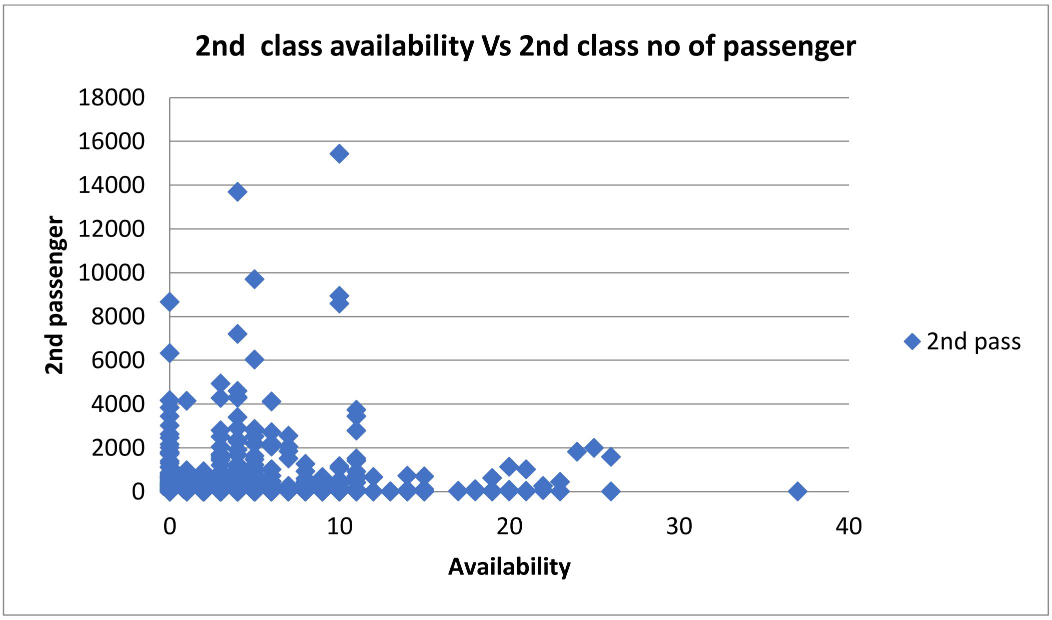


Figure 3.3: Scatterplot of no. of 2nd class passengers vs. availability

According to figure 3.3 there appears to be slightly linear relationship between the two variables. But it also shows data points skewed to 0 which implies that for most of station combinations.

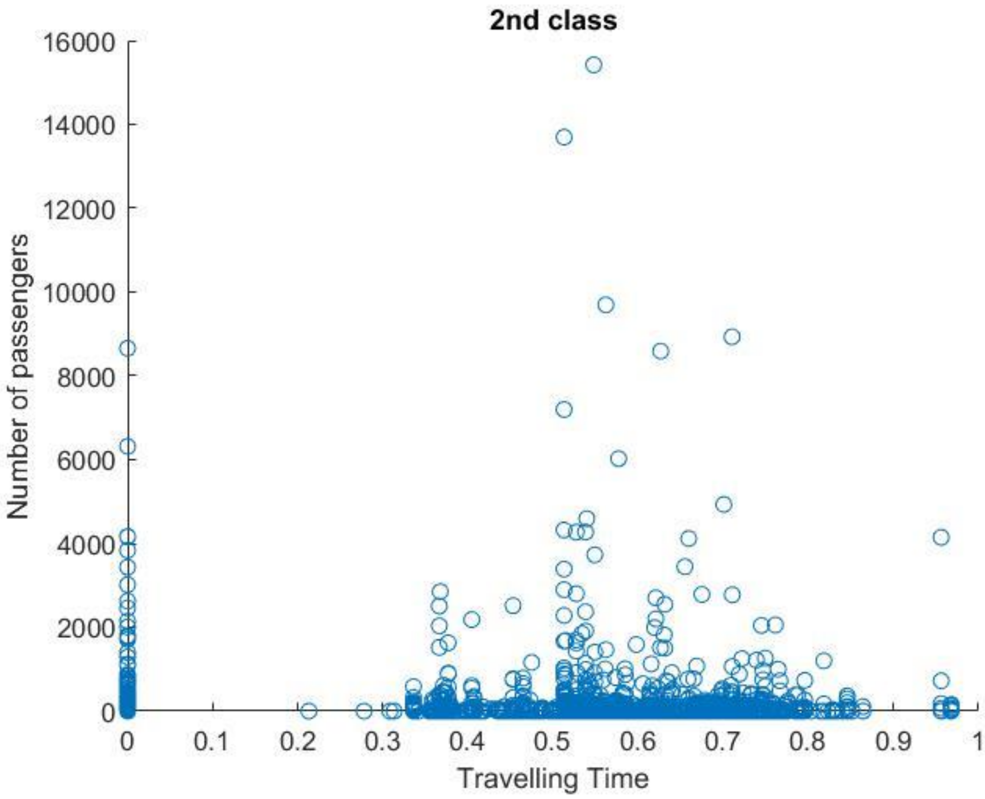


Figure 3.4: Scatterplot of no. of 2nd class passengers vs. travel time

According to figure 3.4 there appears to be no linear relationship between the two variables but it has slightly normal distribution. And also shows some of data points skewed to 0 which implies that for most of station combinations.

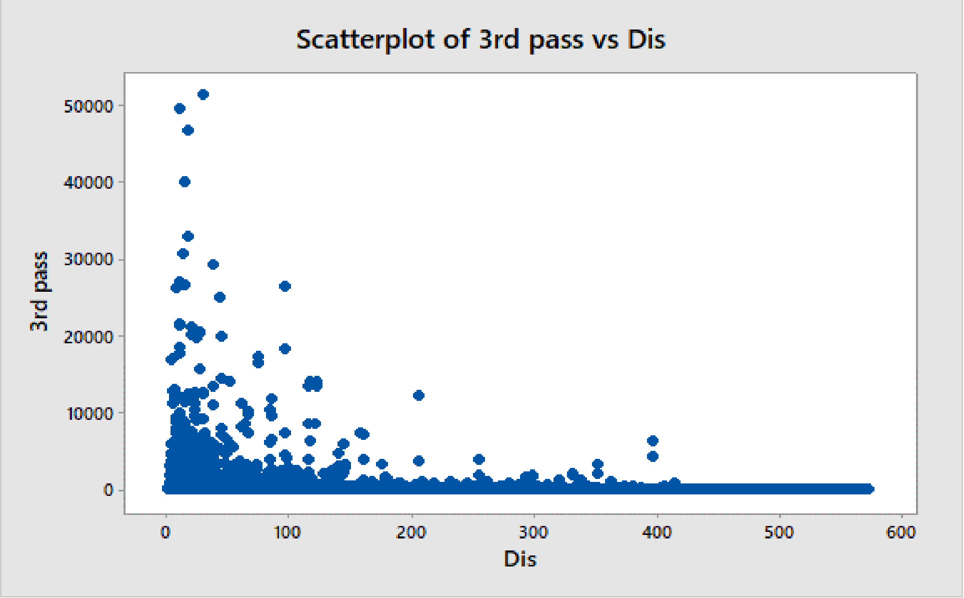


Figure 3.5: Scatterplot of no. of 3rd class passengers vs. travelling distance

According to figure 3.5 there appears to be no linear relationship between the two variables. But it also shows data points skewed to 0 which implies that for most of station combinations, when travelling distance increases number of 3rd class passengers decreases.

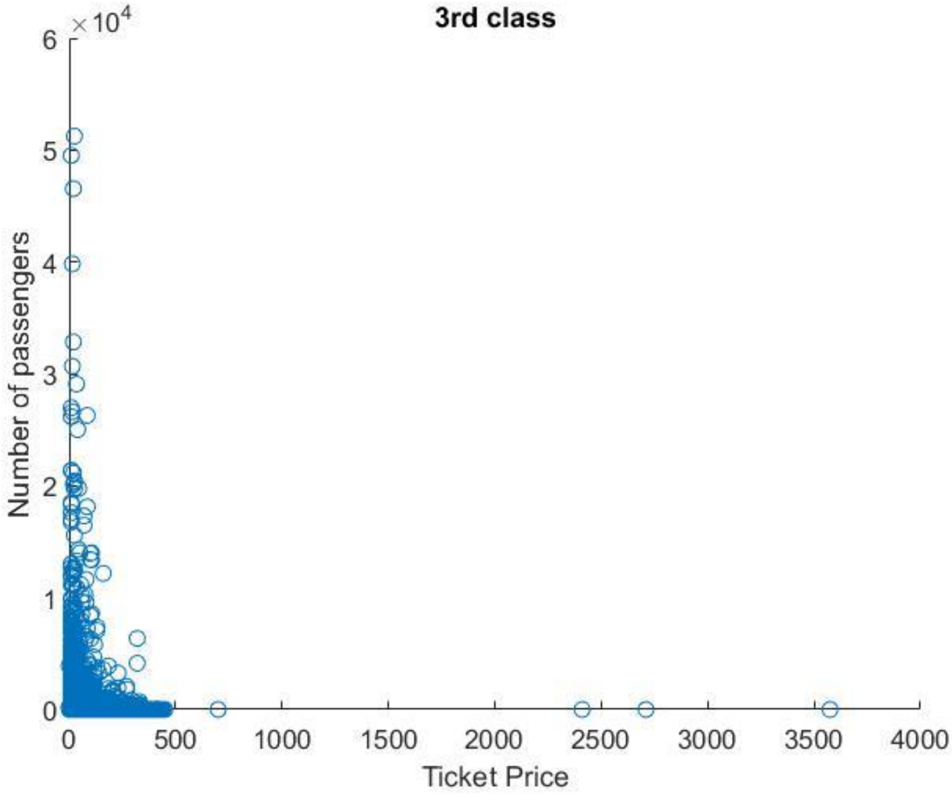


Figure 3.6: Scatterplot of no. of 2nd class passengers vs. tickets price

According to figure 3.6 there appears to be no linear relationship between the two variables. But it also shows some data points very skewed to 0 which implies that for most of station combinations

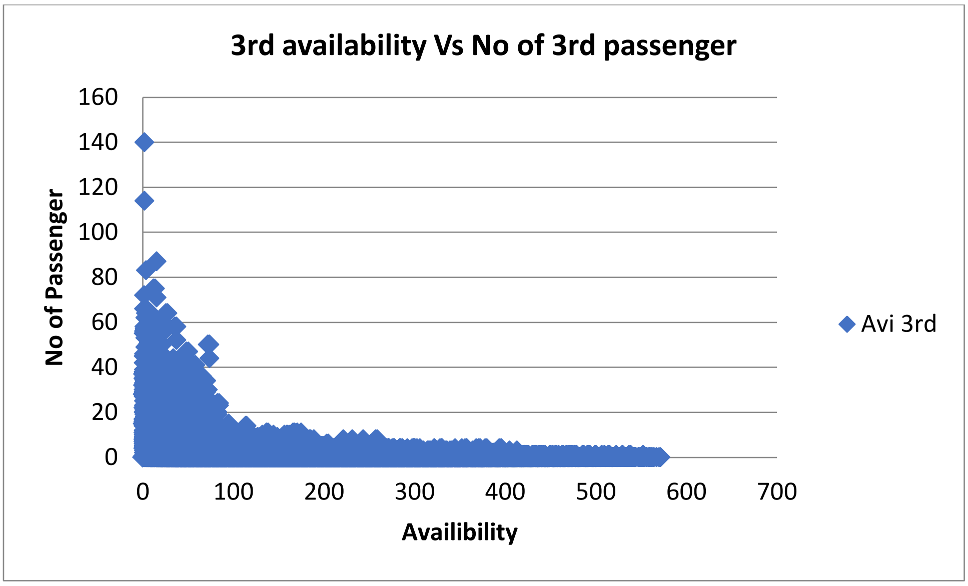


Figure 3.7: Scatterplot of no. of 2nd class passengers vs. availability.

According to figure 3.7 there appears to be no linear relationship between the two variables. But it also shows data points skewed to 0 with some distribution along to X axis which implies that for most of station combinations.

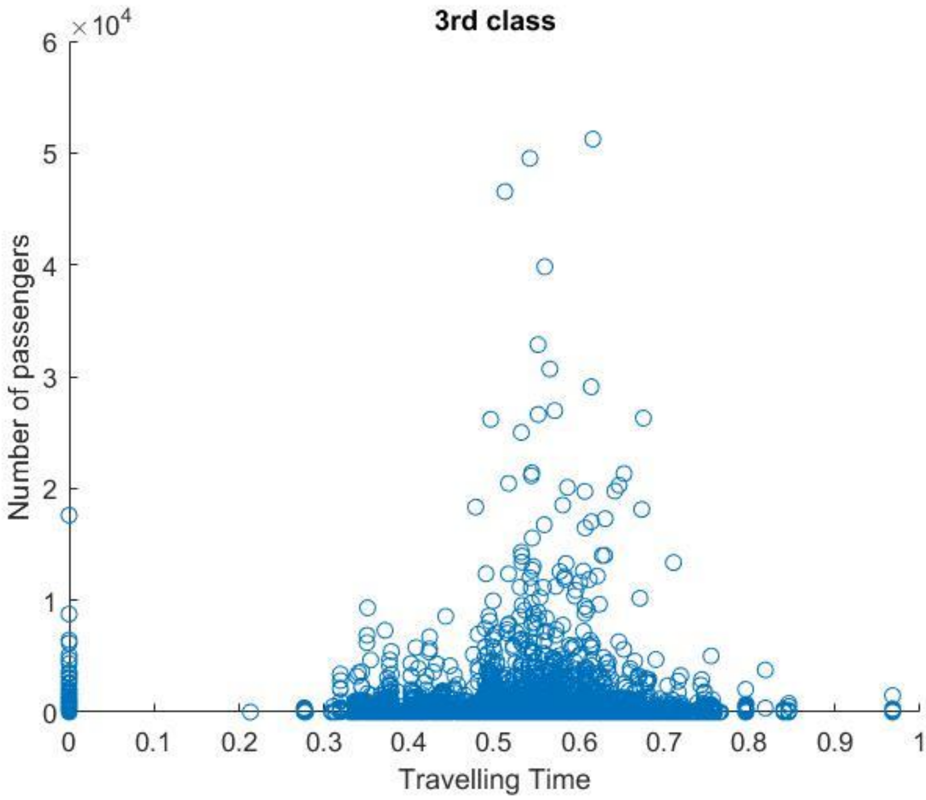


Figure 3.8: Scatterplot of no. of 3rd class passengers vs. travel time

According to figure 3.8 there appears to be no linear relationship between the two variables but it has slightly normal distribution. And also shows some of data points skewed to 0 which implies that for most of station combinations.

.

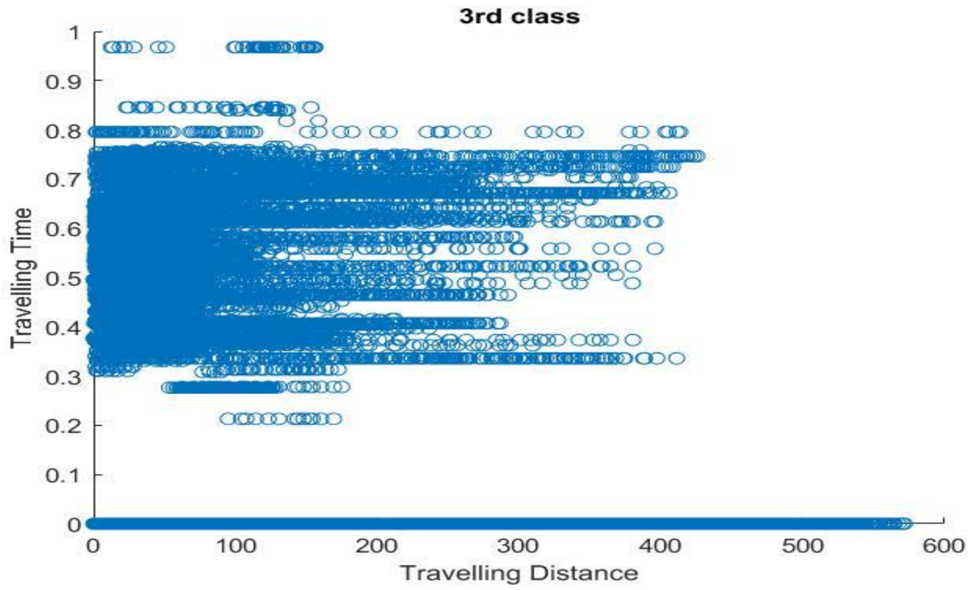


Figure 3.9: Scatterplot of 3rd travel time vs. distance

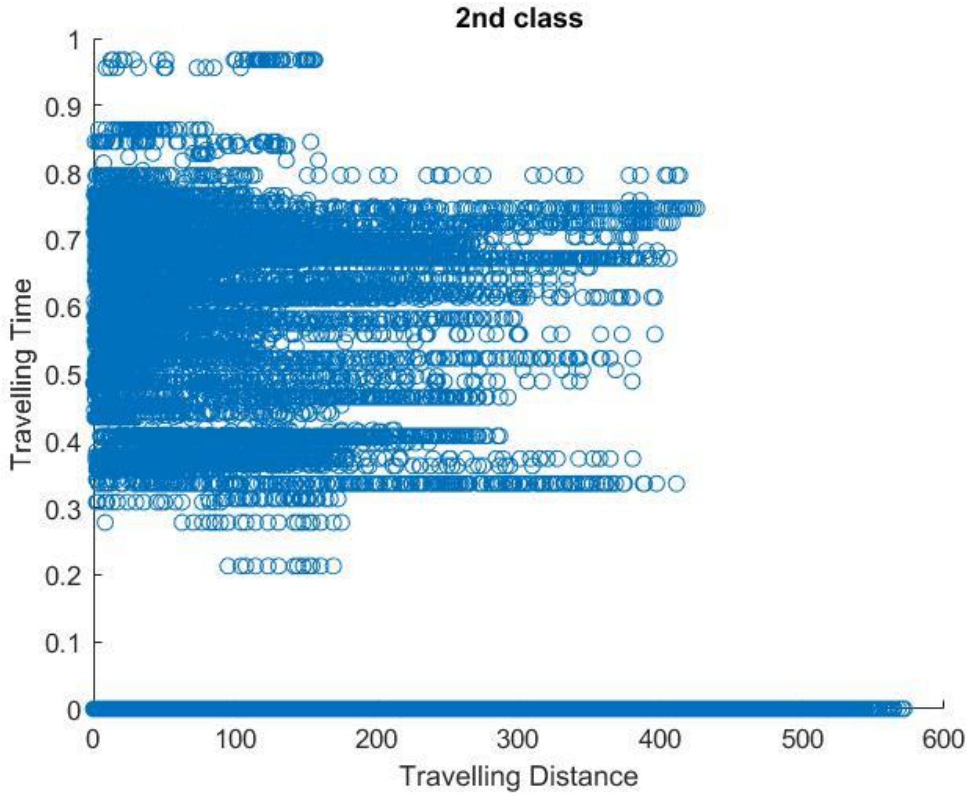


Figure 3.10: Scatterplot of 2^{nd} travel time vs. distance

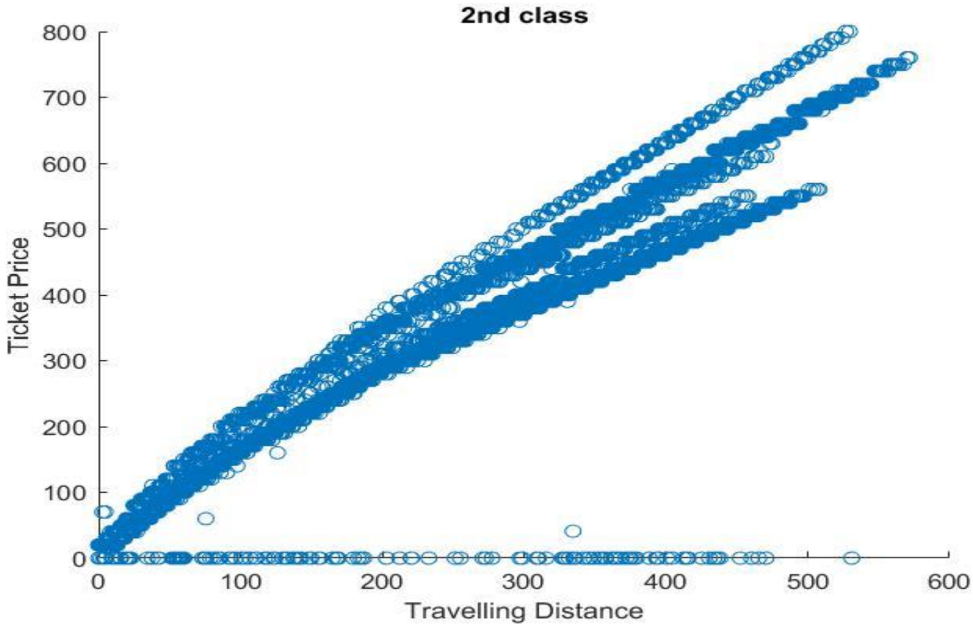


Figure 3.11: Scatterplot of no. of 2nd class tickets price vs. distance

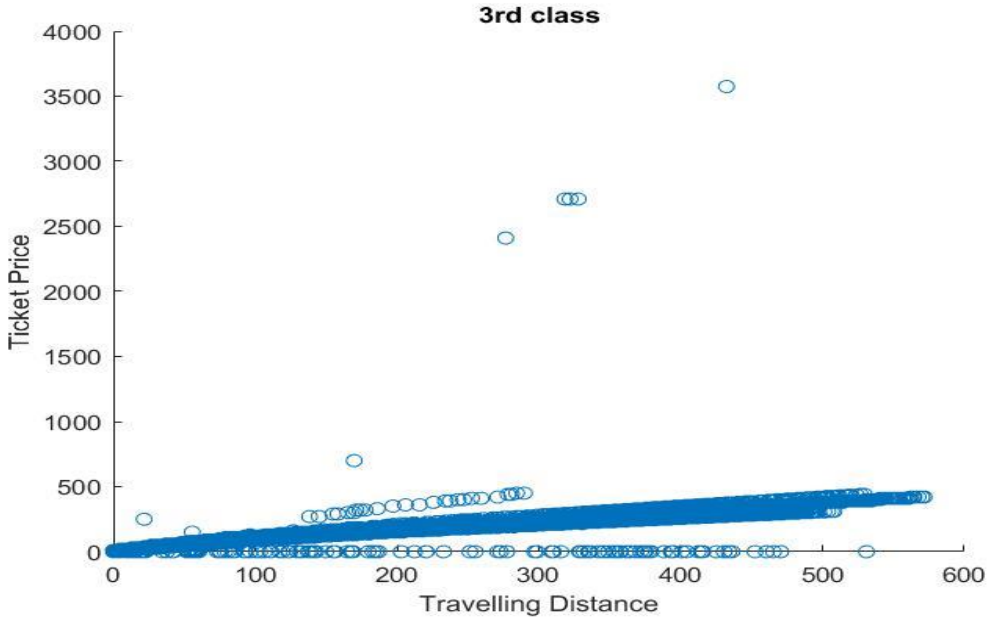


Figure 3.12: Scatterplot of no. of 3^{rd} class tickets price vs. distance

According to figure 3.11 and 3.12 there appears to be strong linear relationship between the two variables due to travel tickets price is a function of distance. But it shows some missing value data points.

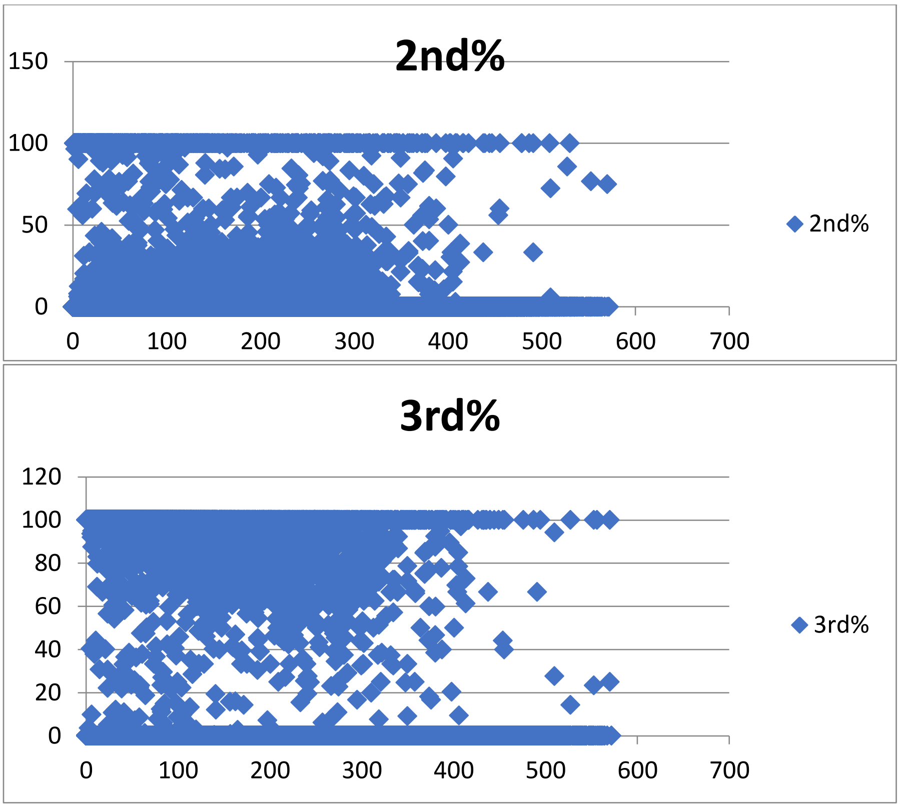


Figure 3.13: Scatterplot of percentage of no. of 3rd class & 2nd class passenger vs. distance

X axis of both charts is percentage of number of passenger traveling by each class by total passengers of the two stations. It seem 3rd class travellers percentage reaching to 100 when distant close to 0. But 2nd class percentage reaching 100 at about distant 100 to 200.

4 METHODOLOGY

4.1 Data preparation

Data is considered as aggregate when the observations consists of choices made by groups. In the considered data set the observation is defined as the station combination. Hence the passenger's choice of 2[nd] class or 3[rd] class falls under choices made by groups. Therefore, the proportion of allocation for each choice can be calculated as follows.

$Weight_i = 2ndpass_i + 3rdpass_i$

$Al2nd_i = 2ndpass_i / Weight_i$

$Al3rd_i = 3rdpass_i / Weight_i$

Where;

$Weight_i$ – Weight for station combination i

$2ndpass_i$ – Number of 2[nd] class passengers for station combination i

$3rdpass_i$ - Number of 3[rd] class passengers for station combination i

$Al2nd_i$ – Allocation of 2[nd] class passengers for station combination i

$Al3rd_i$ - Allocation of 3[rd] class passengers for station combination i

4.2 Utility function forms

The choice model will have the following two choices.

- Selection of 2[nd] class (U2)
- Selection of 3[rd] class (U3)

The following characteristics were considered in developing the choice model.

- Distance between stations (Dis2nd, Dis3rd)
- Alternative prices (price2nd, price3rd)
- Connectivity strength (Avai2nd, Avai3rd)
- Travel time (Time2nd, Time3rd)

By considering the above characteristics the following utility function forms.

1) Alternative prices only
2) Alternative prices + alternative specific constant
3) Alternative prices + connectivity strength
4) Alternative prices + connectivity strength + alternative specific constant
5) Alternative prices + connectivity strength + travel time
6) Alternative prices + connectivity strength + travel time + alternative specific constant
7) Alternative prices + connectivity strength + travel time + distance

4.3 Logit Estimation

For each of the utility function forms identified in 4.2 logit estimation was carried out using ALOGIT software.

Estimation using ALOGIT requires creation of the following two files.

1. Base data file (.dat) - consists of dataset
2. Control file (.bin) – controls needed for estimation

.dat file was created using the prepared data set.

In the .bin file the following controls were set.

- DATA line – data item of NCHOICES was set as 2 to reflect consideration of proportions of both classes
- Coefficients – for each utility form the relevant characteristics were defined
- Transformations – Transformation of ID & WEIGHT were defined
- Exclusions - Based on the utility form the relevant data exclusions were defined using 'IFEQ' function
- Utility Functions – For each utility form the relevant two utility functions were defined

The appropriate utility form was selected based on both model characteristics (i.e. Rho squared value) and statistics of estimated parameters (i.e. T-Ratio).

```
Railway Class choice - Function 07
---------------------------------

--- controls ---
DATA 0 2 5*0 2

END

--- coefficients ---

01 dist2nd
02 avai2nd
03 price2nd
04 time2nd
05 dist3rd
06 avai3rd
07 price3rd
08 time3rd


--- transformations ---

ID = d01
WEIGHT = d12


--- exclusions ---

  exclude = ifeq(d02,0) and ifeq(d03,0)

--- utility functions ---

u1 =  p01*d04
    + p03*d06
    + p02*d05
    + p04*d07

u2 =  p05*d08
    + p07*d11
    + p06*d09
    + p08*d10
```

Figure 4.1 : .bin file

Figure 0.1 : Results file (.out file)

```
Hague Consulting Group                                                   Page   1
ALOGIT Version 3.89 (321)                                        8:42:46 on 24 Feb 20

            Railway Class choice - Function 07 |

TEST: RAM requested 3000k.  Dynamic array size = 125000

Last input data item in transformations or utilities   12; maximum is  2000

Number of temporary data stores used    2; maximum is 12984

Number of transformation codes   23; maximum is 80000

PREPARE transformations
ID          = Data0001
WEIGHT      = Data0012
Temp0001    = IFEQ( Data0002  ,0          )
Temp0002    = IFEQ( Data0003  ,0          )
EXCLUDE     = AND ( Temp0001  ,Temp0002   )

Maximum Iterations 10
Convergence criterion is  .10E-01  Option 3

INFORMATION: no explicit specification - base file read with default format

ERROR: bad end-of-file in base data 6502

Report of user selections

 78859 Observations rejected because EXCLUDE    = 1.00

DATA INPUT COMPLETED

  from data file : data1.dat

Total observations read from file     :  89541

Observations rejected by user tests   :  78859

Observations rejected automatically   :      0

Observations accepted for processing :  10682

Sum of weights of observations :      4110862.00
```

5 RESULTS & DISCUSSION

5.1 Utility function form

Table 5.1 : Model results

Utility form	Variables	Model Estimation	ρ2(0)
1	price2nd, price3rd	Successful	0.2533
2	price2nd, price3rd, constant	Failed	-
3	price2nd, price3rd Avai3rd, Avai2nd	Successful	0.5049
4	Price2nd, price3rd Avai2nd, Avai3rd, constant	Failed	-
5	price2nd, price3rd Avai2nd, Avai3rd, Time2nd, Time3rd	Successful	0.5106
6	price2nd, price3rd Avai2nd, Avai3rd, Time2nd, Time3rd, constant	Failed	-
7	Dis2nd, Dis3rd, price2nd, price3rd Avai2nd, Avai3rd, Time2nd, Time3rd	Failed	-

When estimating some utility forms the model estimation was unsuccessful with the variables as a combination not correlating with the observations and hence the convergence criteria will not be met (matrix not invertible).

Utility form 1 only considers price as a factor (2 estimation parameters) and hence the obtained ρ2(0) value is low for model that estimates 2 parameters. Out of the 3 successful models utility forms 3 & 5 were the best fit models with ρ2(0) value being greater than 0.35.

Most appropriate utility form can be recognized as utility form 5 due to it containing additional attribute of time and hence will comparatively better explain the behaviour of passenger's choice compared with utility form 3.

The estimation results of other utility function forms are given in Appendix - A.

5.2 Utility function & estimation statistics

Table 5.2 : Estimation statistics

Attribute	Estimate	Standard Error	T-Ratio
avai2nd	-0.007789	0.000535	-14.6
price2nd	-0.08350	0.000364	-229.5

time2nd	0.01604	0.000280	57.3
avai3rd	0.06964	0.000125	557.7
price3rd	-0.1313	0.000668	-196.7
time3rd	-0.0009827	0.000364	-3.2

U2 = - 0.007789 * avai2nd - 0.0835 * price2nd + 0.01604 * time2nd

U3 = 0.0694 * avai3rd - 0.1313 * price3rd - 0.0009827 * time3rd

5.2.1 Price

Increase in price for both 2nd class and 3rd class will negatively impact its attractiveness to passenger and hence the estimated sign of negative is correct. Based on the estimated values it appears that increase in price in 3rd class has a higher negative impact when compared with 2nd class. This can be explained with consideration of type of passengers who use each class.3rd class passengers are relatively more price conscious and hence with an increment of price they will make decisions such as shifting towards other modes like bus and hence a higher negative impact on utility.

The T-Ratio for both price2nd andprice3rd are greater than 1.96 which implies that both the values are significantly different from zero and hence should be considered in the utility functions.

5.2.2 Travel time

Travel time and distance can be identified as attributes that can be used to represent an individual's train journey. With consideration for the fact that some train station combinations require usage of connected train trips, travel time is a better representative and hence is considered in development of the utility functions. Travel time considers both the journey time and the waiting time. The journey time and waiting time can vary between 2nd class and 3rd class for some station combinations due to availability of trains with each class.

Increase in travel time can result in passenger's seeking more comfort and hence 2nd class will become more attractive and 3rd class will be less. It is consistent with the signs for time2nd and time3rd being positive and negative respectively. Based on the magnitude of the impact, 2nd class has a more positive impact and 3rd class has a less negative impact with both the impacts can again be attributed to passenger types using each class.

The T-Ratio for both time2nd and time3rd are greater than 1.96 which implies that both the values are significantly different from zero and hence should be considered in the utility functions.

5.2.3 Availability

In the current rail system, there are trains with only 3rd class apartments (i.e. daily trains between urban locations) whereas trains with only 2nd class compartments/ no 3rd class compartments aren't present. Normally a train will have about 1:5 ratio between 2nd class and 3rd class compartments.

Increase in availability in 3rd class compartments will increase its attractiveness and hence the positive sign is consistent. In the context of the current railway system increase in 2nd class compartments will indirectly increase the availability of 3rd class compartments massively and hence 3rd class can be more comparatively attractive and more passengers opting for 3rd class .So in the current context of the railway system the negative sign for availability of 2nd class can be correct. Based on the magnitude of the impact, 2nd class has a less negative impact and 3rd class has a more positive impact with both the impacts can again be attributed to passenger types using each class.

The T-Ratio for both avai2nd and avai3rd are greater than 1.96 which implies that both the values are significantly different from zero and hence should be considered in the utility functions.

5.3 Applicability of the choice model

The created choice model is applicable in both existing and new operations to determine probabilities of passengers choosing 2nd class or 3rd class.

5.3.1 When introducing new trains

Determining the number of compartments and the class composition of compartments are important aspects that are to be considered when introducing a new train. The developed choice model can be used to calculate class wise passenger demands for each station combination on the route and based on the average values the total number of compartments and 2nd and 3rd class compartment composition can be determined.

5.3.2 Existing train operations

The passenger class wise choice model helps to identify the demand for each passenger class. The class composition of existing trains can be optimized based on results of the choice model by matching the class capacities of each train to its class demand. Hence the model provides a guide to decision making in scenarios such as whether to introduce 2nd class compartments to current 3rd class only train etc.

5.3.3 Existing train station operations

By determining the probabilities of 2nd class and 3rd class passenger demand the capacity decisions related to ticketing operations can be made for train stations. Hence the operations of the stations in terms of physical capacities (i.e. no. of ticketing counters for each class, required ticket stocks) and human resource capacities (i.e. time shifts of station staff) can be improved.

6 CONCLUSION & RECOMMENDATIONS

Non availability of class wise passenger demand calculation has resulted in issues such overcrowding in compartments and hence the passengers not getting value for money that will ultimately lead to passengers shifting to other modes. The research tries to address the issue by identifying how railway passengers choose the class.

The research scope is restricted to data relating to tickets issued over the counter and the reserved ticket data is not considered. Therefore, the two choices were 2nd class and 3rd class.

The observations were related to station combinations and the two choices for each combination was based on group choices and hence the data was treated as aggregate data. Initial data preparation included calculation of proportion of allocation for each of the two classes using weights. Distance, price, connectivity strength and travel time were considered as attributes and the most suitable utility function form was selected out of 7 utility forms. The estimation was done using ALOGIT software.

According to the figure 13 3rd class ticket users tends to travel short distance and also the 2nd class ticket users tends to travel longer. Long distance passengers seek more comfortability and willing to pay more. This is the reason for that.

The selected utility function consists of price, connectivity strength and travel time attributes with the estimated utility functions as follows.

U2 = - 0.007789 * avai2nd - 0.0835 * price2nd + 0.01604 * time2nd

U3 = 0.0694 * avai3rd - 0.1313 * price3rd - 0.0009827 * time3rd

Based on the estimated utility functions travel time influences the most in selection of 2nd class compartments. Connectivity strength or availability of directly connected trains with 3rd class compartments influences the most for passengers to choose 3rd class.

And also based on the estimated parameter values it can be observed that price has a higher impact on 3rd class passengers compared to 2nd class passengers. The developed choice model captures the impacts of train journey based on travel time instead of distance due to its more applicability in terms of existing train operations.

The developed passenger class choice model can be used to improve existing train & station operations as well as planning future train operations.

6.1 Future works

The developed choice model is restricted to 2nd and 3rd class choices and can be extended to consider 1st class passengers also with the scope extending to reserved tickets.

The choice model can be further improved by introducing attributes that reflect available facilities in each class directly rather than the indirect attributes that was considered in the research.

And also, consideration of express trains and normal trains separately will enable to identify how passenger choice behaviour changes with train types.

References

Ben-Akiva, M. E., & Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press.

Department of Sri Lanka Railway. (2017). *Sri Lanka Railway Administration Report - 2017*. Department of Sri Lanka Railway.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with Certainty: The Appropriateness of Extreme Confidence. *Journal of Experimental Psychology*, *3*, 552–564.

Hague Consulting Group. (1992). *ALOGIT Users' Guide—Version 3.2*.

Hetrakul, P., & Cirillo, C. (2014). A latent class choice based model system for railway optimal pricing and seat allocation. *Transportation Research Part E: Logistics and Transportation Review*, *61*, 68–83.

Kim, H.-C., Nicholson, A., & Kusumastuti, D. (2017). Analysing freight shippers' mode choice preference heterogeneity using latent class modelling. *Transportation Research Procedia*, *25*, 1109–1125.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, *74*(2), 132–157.

Shibata, M., Terabe, S., Professor, A., & Uchiyama, H. (2010). A Seat Class Choice Model on Intercity Rapid Train Passengers for Flexible Seat Class Assignment. *Journal of the Eastern Asia Society for Transportation Studies*.

Svenson, O. (1979). Process Descriptions of Decision Making. *Organizational Behavior and Human Performance*, *23*, 86–112.

Appendix A: Results – Other utility function forms

Utility Form 1: Price only

Attribute	Estimate	Standard Error	T-Ratio
price2nd	- 0.1246	0.000296	- 420.6
price3rd	- 0.1999	0.000544	- 367.2

Utility Form 3: Price & availability

Attribute	Estimate	Standard Error	T-Ratio
avai2nd	- 0.007264	0.000517	14.1
price2nd	- 0.08777	0.000358	- 245.0
avai3rd	0.07397	0.000124	597.3
price3rd	- 0.1498	0.000651	- 230.0