Erasmus University Rotterdam

Erasmus School of Economics

Master Thesis
Quantitative Marketing & Business Analytics

Exploring Consumer Preferences for Subscription Packages in Public Transport Using Smart Card Data

Supervisor
Dr. P. Bouman

Author
Kim de Groot
507070kg

Second Assessor
Dr. S. Sharif Azadeh

Abstract

In this paper, consumer preferences for subscription packages in public transport are explored using smart card data. Furthermore, the influence of travel demand on subscription choice is investigated. Additionally, we analyze the effect of the travel demand expenses and the differences between consumers on the decision process. These effects are modeled with six different logit models and a hierarchy in the decision process has been taken into account. Estimation is done by maximum likelihood. Results are based on four years of smart card data from MyOV of 4379 individuals that traveled with the Dutch Railways for more than a year. Overall, we find that travel demand, historical and realized, is a strong indicator of subscription choice and that first class travelers differ in these preferences. In addition, consumer have an economic incentive when choosing their subscription. However, when this subscription is expensive, consumers are not good at predicting their expected expenses. This research attempts to fill a gap in current literature regarding the non-linear pricing plans in public transport.

August 27, 2019

Acknowledgements

I would like to give my great appreciation to my thesis supervisor, dr. Paul Bouman, for his feedback and support while writing my thesis. Furthermore, I would like to thank dr. Shadi Sharif Azadeh, for assessing my thesis.

I would like to thank my colleagues at Lynxx for my time there, which I greatly enjoyed. Special thanks go out to Matthijs Bootsman as my supervisor at Lynxx, for his great support, patience and friendship while needed. Additional thanks go out to Paul Rooijmans and Sanneke Mulderink, for their industry knowledge and giving me the opportunity and support I needed to write my thesis.

Finally, I would like to thank my friends, family and fellow students for their continuous support and friendship throughout my studies.

Contents

List of Figures

List of Tables

1 Introduction

Mobility as we know it is about to change. Various mobility trends will largely determine the benefits and costs for business and society. The number of kilometers travelled per person is rising steadily and most of the traffic volume in the Netherlands is caused by cars. The resulting traffic jams, noise and air pollution are highly problematic and trigger a paradigm shift in favor of car-less mobility solutions. Usage of sharing options, such as public transport or ride sharing, is already spreading, especially among the younger generations. Other movements in mobility are already turning towards convenient individualization, but the desire for personalization has not been satisfied yet. The way we travel from point A to point B is changing, creating a new system of personal mobility. Impacted businesses should reconsider how they create value in this changing environment.

Public transport is an important tool in meeting the need for mobility in the Netherlands, since the road network is reaching the limits of its capacity, especially in the morning rush hour. Every day more than one million trips are made by train. The rail operator Dutch Railways (NS) holds the concession for the mainline rail network, making it the largest passenger carrier in the Netherlands (Eurlings 2010).

For mobility providers an interesting challenge is pricing of their service towards customers and providing platforms for ticketing. With the adoption of mobile ticketing in public transport, the smart card, the possibilities of adopting price and service differentiation improved. It follows that if service providers employ price differentiation and/or service expansion strategies they are more likely to have higher performance gains compared to the ones that use only the baseline strategy (Li, Van Heck, and Vervest 2009). Currently, public transport providers often sell their service through their own platform, often in the form of a subscription packages. Even though several subscriptions can be distinguished, there is still a considerable gain in personalization.

The present Mobility-as-a-Service (MaaS) trend offers the progression in the development of more personalized products or subscriptions, providing platform for ticketing, and dynamic pricing systems. The trend of MaaS raises the opportunity of changing the current way of ticketing and to design new products. Often in designing new products a conjoint analysis is conducted from stated preferences, but to obtain an understanding of the current pricing of mobility in general, the current market is informative. Looking at the present pricing in combination of current travel behaviour is informative on the established structure of the market and preferences of the consumers. In addition, smart cards provide the possibility to collect large amounts of information on travel behaviour of agents on an individual level.

With the use of descriptive analytics this paper provides insight into the decision process of individuals through realized behaviour and answers:

> **What are consumer preferences for subscription packages in public transport using smart card data?**

Currently, the Dutch Railways differentiate in price and offer various subscriptions. Subscriptions differ in the price of the service, the moment of usage and in other kinds of attributes, such as the quality of the service. Public Transport users decide upon their travel demand to subscribe to a package or to travel without a subscription and pay the full-tariff. In this paper travel demand is measured by the distance traveled by an individual. The exact definition of travel demand is formulated in Section 3.

The first question that arises is, if travel demand actually explains the subscription choice of consumers. Even though this is the service the NS provides to the consumers, it might be that other factors influence the subscription choice of consumers, such as convenience. In addition, since the subscription has to be chosen in advance of service usage, consumers might not always be able to pick the subscription that is best suited to them. They have knowledge of future information and plans to predict their travel demand and use this prediction to find a suitable subscription. Individuals might know in advance what their planning is going to be in the future and thus their travel demand. However, individuals could also weigh their past travel demand or previous travel patterns to make their future choices. For example, the past months an individual traveled twice a weekend over a certain distance and therefore a weekend subscription might be suitable. If travelers are able to perfectly predict their expected travel demand, they can choose the subscription that is most profitable. Yet, do the consumers have an economic incentive when choosing their subscriptions package and are they actually able to to decide on the correct package if they have an economic incentive. Finally, not all consumers similarly make choices, every individual is different and is different in their behaviour and therefore in making choices. In case of differences between individuals in preferences and taste, individuals can be separated into groups for different marketing strategies.

To this end, we formulate the following research questions

1: *Does travel demand explain the subscription choice of consumers and if so, do consumers look at previous or future patterns in travel behaviour?*

2: *When consumers choose their subscription do they take into account their expected travel expenses and thus have an economic incentive?*

3: *When making decisions on the subscription packages, do individuals differ in their preferences in travel demand?*

To answer these questions, we make use of smart card data from July 2015 and till April 2019. The data on the journeys of individuals are translated and aggregated to decision moments. In total the data contain information on 4379 individuals. A split in the data is made in first class and second class subscriptions and only non student and non business subscriptions are used. Additional information is gathered on the prices of the subscriptions. With discrete choice modelling several model formulation are constructed and estimated through maximum likelihood. The results are analyzed to discover consumer preferences for subscription packages in public transport.

From a business perspective, with the found consumer preferences, service providers can adapt and better fit the subscriptions to improve the situation of both the consumer and the service provider. As a result different marketing strategies are possible, namely targeting consumer groups with customized marketing strategies. This means, expanding and adjusting the assortment or in the more general case the service to be more aligned with individual preferences. Examples of improvements are different service attributes, higher/lower prices or more flexibility. Public transport providers can use this descriptive information to optimally design the subscription packages.

This paper academically contributes to the literature for two reasons. First, empirical evidence of the gains of fare management in combination with travel behaviour in public transport is limited, therefore, research on public transport pricing from the consumer point of view is a contribution to the current literature. Second, empirical research on the decision process specifically for subscription packages in public transport is limited.

The paper is structured as follows. In Section 2, a review of current literature concerning pricing in public transport is provided. Section 3 presents a description of the data and summary statistics regarding the attributes of interest. A conceptual framework is established in 4. In Section 5, background information on the used models is given. The model to answer the research questions is formulated in Section 6. The results are presented in Section 7 and afterwards Section 8 consists of conclusions drawn from the results and possibilities for further research and implications.

2 Literature review

2.1 Demand modelling and pricing in passenger transport

This empirical research is focused on demand modelling. Price is one of the attributes used to predict demand, see Section 4.3. Different from this study, most studies use price for revenue maximization problems and use price as an endogenous variable. This section gives a broad overview on the results of studies that have been done on pricing in public transport to better understand this topic. This information is then combined with the study in this paper, demand modelling.

The introduction of the smart card introduced the possibility to improve the analysis of travel behaviour. The literature states several benefits from the possibility to model travel behaviour of customers from smart card data, three benefits are pointed out. First, smart card data in comparison with existing data sources, obtain much larger samples, and enables to apply behaviour analyses over much longer periods (Bagchi and White 2004). Second, Cheung (2006) states that public transport authorities are required by the central government to facilitate competition and increase ridership. Smart card technology is seen as means to aid in the liberalization of the market and providing management information for planning and marketing. Finally, the smart card data can be linked back to the individual card and hence in many cases to a particular user if the smart card is personal and not anonymous. However, Bagchi and White (2004) comment on the still existing need for verificatory or complementary surveys, stated preferences, to obtain more information on the travelers ultimate origin and destination and purposes of the travel. Thus, smart card data, mostly revealed behaviour of travelers, do not fully explain travel behaviour. Pelletier, Trépanier, and Morency (2011) add to this, more research on user characterization and classification with having personal information on users should be done. In conclusion, smart card data are a strong data source in explaining travel behaviour, yet additional information is needed to fully clarify the behaviour.

In transport, customers are able to use any transport service without any contractual agreement. However, they can also enter into some form of pricing plan by applying for a type of period travel ticket, which offers enhanced potential for examining transport turnover (Lovelock, Vandermerwe, and Lewis 1996). The features of smart cards offer the possibility to adapt the fare in such a way that individuals pay the best fixed fee possible. Since start and end locations can be known, check in and out location, the price can be calculated by distance. This can be combined in individual subscription packages, which is done in The Netherlands with the “travel first, pay later” program (Pelletier, Trépanier, and Morency 2011). Link (2004) analyzed the yield-management scheme of the German Railways, but due to early amendment limited amount of information is available. She does find that all different types of rail customers have to be taken into account when designing a yield management scheme.

Different from this research, demand modelling, in the literature much research is done on assortment optimization. Alongside the pricing plans suitable for all type of rail customers, Link (2004) argues for empirical studies to determine whether yield-management or peak-load pricing is best for optimizing usage of capacity. Such an empirical study is done by Lovrić, Li, and Vervest (2013), they propose a sustainable perspective on revenue management that considers individual customers’ needs and requirements, environmental impacts, and public transportation operators financial viability. They find that a time-of-day differentiated pricing strategy is most effective from a sustainable perspective, in particular the Peak-Markup strategy which applies price markups during the morning and afternoon peak demand periods. This finding is in line with the research of Tirachini, Hensher, and Rose (2014) that explores the interplay between peak hours and fares in traffic congestion and bus crowding.

The potential of public transit smart card systems for commercial applications has not yet received intensive focus by researchers (Pelletier, Trépanier, and Morency 2011). In addition, Van Vuuren (2002) declares that the optimal pricing of passenger transport by railway companies is a topic that has not received much attention in the empirical literature. Most literature states that there is potential, but not how this potential can be explored in combination with the travel behaviour of consumers, thus demand modelling. Utsunomiya, Attanucci, and Wilson (2006) used the smart card data in combination with demographic profiles of the smart card users by route and by stations, to conduct a fare policy analysis which had the benefit to adjust the fares according to user needs. Another study to assortment optimization in the flexible mobility on demand system is done by Atasoy et al. (2015). With the use of simulation methods they observe that the trade-off between consumer surplus and operator's profit is critical and controlling for this trade-off improved results in terms of both profit and consumer surplus. However, in their study they did not use revealed or stated information to represent the needs of users.

In conclusion, pricing in public transport in revenue optimization problems contains potential for turnover increment for transport providers through the use of smart cards and different pricing options, such as the opportunity to find the best fare available per user, the option of not paying in advance, the possibility of non-personalized cards and Peak-Markup strategies to deal with capacity constraints. Utsunomiya, Attanucci, and Wilson (2006) argue that offering smart cards as a payment option raises the need of further tracking and analyzing of long-term individual travel behaviours. In addition, Pelletier, Trépanier, and Morency (2011) identify research issues on economic feasibility of the smart card, namely there is need to better assess the gains from fare management. In conclusion, empirical evidence of the gains of optimal pricing strategies is found. However, pricing combined with demand modelling in public transport is limited, therefore, empirical research on demand modelling with price as an attribute is a contribution to the current literature.

2.2 Choices and non-linear pricing methods

Even though there exists research on pricing and fare policy in public transport, research on decision making in pricing plans or subscription packages in public transport is limited. However, empirical studies on the decision processes of consumers of pricing plans in general in for example, the cell phone service industry, is reviewed in the literature. An overview of the findings of non-linear pricing plans in general is given in the following section.

The subscription packages are defined as long term package products, where consumers don't pay per usage, linear pricing, but pay a fixed fee independent of usage to obtain a different pricing scheme, non-linear pricing. Three different types of pricing plans can be explored in non-linear pricing, namely bucket, two-part and three-part pricing plans, see Figure 1[1]. With a three part pricing plan, a service provider charges a usage-independent fixed price, an allowance, and a marginal price per unit. The bucket pricing plans and two-part pricing plans are a special case of the three-part pricing plan, differing from the three-part pricing plan in the marginal usage-dependent price and allowance, respectively. In linear pricing there is only a usage-dependent price and not a usage-independent fee.

Since there is heterogeneity among consumers, most providers offer more than one pricing plan to make a distinction between consumers with different demands. Thus, consumers can choose without restriction among several combinations of fixed prices and allowances. Most empirical non-linear pricing studies have focused on consumer choice under two-part tariffs (Lambrecht, Seim,

[1] $n_{m,j}$ = consumption of a consumer m, when choosing pricing plan j

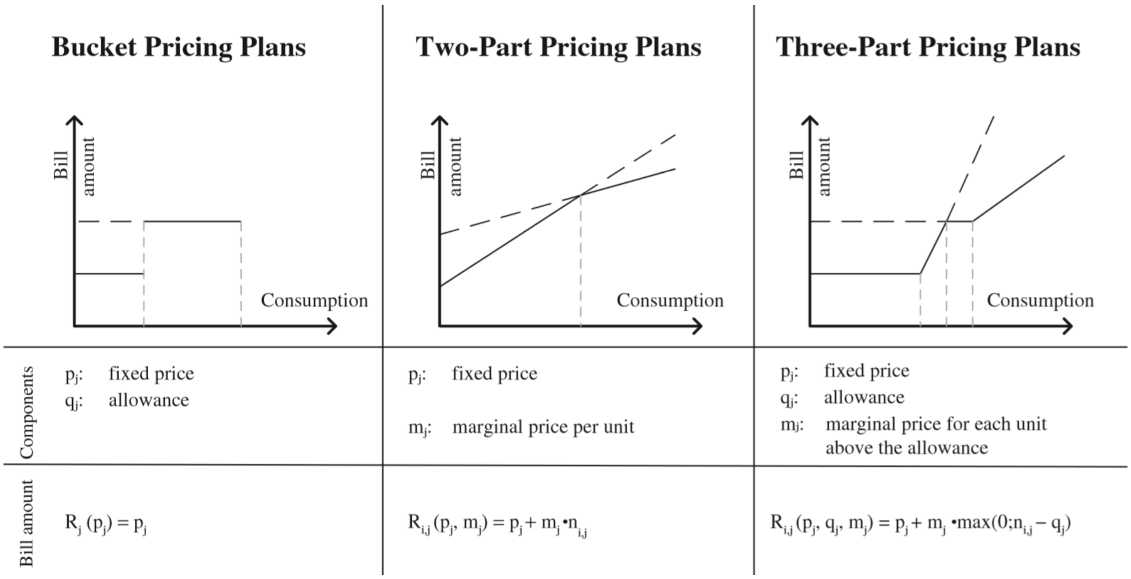


Figure 1: Visualization of non-linear pricing plans

and Skiera 2007), which resembles most of the Dutch Railways subscription packages. However, the Dutch Railways subscription packages differentiate in moment of travel during the day and the usage-dependent price differs per moment of travel. Therefore, the current subscription packages can differentiate between consumers with different travel preferences.

Next to non-linear pricing, service attributes also influence the decision of pricing plans, but how they do so depends on the particular service attribute and correctly modelling service attributes leads to more optimal prices and profits (Schlereth and Skiera 2012). For the Dutch Railways plans differ in pricing between moment of travel, routes, comfort, flexibility, but also differ in non-linear pricing methods.

Train, McFadden, and Ben-Akiva (1987) state that consumers seemingly make mistakes when they make plan choices, they often seem to choose plans that are not cost minimizing. One of the main reasons is that consumers have to decide on their plan before knowing their future usage. The determination of optimal plans requires knowledge about each consumer's simultaneous decision about service subscription, plan choice and consumption, which are interrelated and difficult to predict (e.g. Schlereth and Skiera 2012; Miravete 2003). In addition, Grubb and Osborne (2015) show that consumers are inattentive to past usage, consumers are overconfident, underestimating the variance of future usage. However, according to the theory of mental accounting, Prelec and Loewenstein (1998) state that a separated payment from consumption enables consumers to enjoy their consumption more.

The above research suggests that it is important to correctly account for uncertainty in future usage at the time of plan choice. Consumers can also account for these "mistakes", namely switching plans to correct for this "mistaken" plan choice. Iyengar, Ansari, and Gupta (2007) argue that during the decision process of a plan a consumer experiences two forms of uncertainty, namely the quality of the service provider and the uncertainty about their own usage of the service, which they capture while accounting for the non-linearity of pricing scheme. They find that consumer learning can result in a win-win situation for both consumer and firm with as key driver the retention rate with learning. Narayanan, Chintagunta, and Miravete (2007) find that consumers learn very rapidly if they have a measured plan but learn very slowly if they have a fixed plan. For example consumers switch less often from fixed plans to measured plans to gain from potential savings than vice versa. In addition, a fixed, known monthly price is often preferable for risk-averse consumers (Lambrecht and Skiera 2006).

In conclusion, it is important to differentiate between the various non-linear pricing plans and to take into account the interdependence between usage and price in modeling the decision process of consumers under different pricing plans. In addition, consumers often make mistakes when they make plan choices and do not always choose the plan that is cost minimizing.

2.3 Discrete choice model

In this section an introduction to the concept of discrete choice modeling is given. In Section 4 and 5 we elaborate on discrete choice modeling specific to the section.

Discrete choice studies have established themselves as an important tool for the analysis of individual decision making (Train 2009). They commonly focus on the simplified functional form for the way in which the observed sources of utility are represented. It is typically assumed that consumers act as if they are utility maximizers and linearly weighting the attributes. This simplified and appealing context is consistent with the Random Utility Maximization (RUM) model and is most commonly used in decision analyses, see Hess, Daly, and Batley (2018) for a comprehensive review of theory and implications.

Ben-Akiva and Lerman (1985) equate the term *rational behaviour* with behaviour consistent with utility maximization or based on the beliefs of an observer about what the outcome of a decision should be. Thus, as colloquially used, the notion of rationality is not a useful concept in describing individual behaviour. Nevertheless, there is ample evidence showing that in many contexts consumers judgments, preferences and behaviour are, taken at face value, systematically irrational (e.g. Tversky and Kahneman 1989). Rational relates to the model and not the the individual itself who argues to make rational choices. In this paper, *rational behaviour* of a consumer means that it is consistent and transitive, corresponding to the definition used by Ben-Akiva and Lerman (1985).

Hess, Daly, and Batley (2018) state in addition, in practice individuals have both differing taste and random behaviour. Balbontin, Hensher, and Collins (2019) divided heterogeneity into preference and process heterogeneity and showed with empirical evidence that combining both methods increases the model fit. Other key findings are the following: individuals' preferences and perceptions are unstable and depend on the context (Huber, Payne, and Puto 1982; Tversky and Simonson 1993), individuals are cognitively constrained (Jamasb and Pollitt 2005) and individuals tend to use different decision procedures and rules to deal with complex decision problems (Manzini and Mariotti 2007; Tversky and Kahneman 1974).

3 Data on travel behaviour

In this section, an overview is given of the journeys data from MyOV[2], the Dutch Railways, the preprocessing steps and the data description. The data preparation is done in *QlikView* and *R*

3.1 MyOV data

Since we would like to investigate the impact of travel behaviour on the subscription choice, revealed preference data on subscription choice per individual are needed. The available data contain information on the completed journeys on a sample of users.

The origin of the data is through a tool, MyOV. This tool is used to motivate users with a reward for traveling in less crowded trains. These rewards are given on trajectories crossing the station *Utrecht Central*. By assigning more points to lesser crowded train users, the tool attempts to spread out crowds during peak hour. After collecting the points, individuals could exchange them for free coffee or other vouchers. By participating, users allow for sharing their travel behaviour captured by the smart card before (1.5 years) and during the participation of MyOV. Through Translink[3], travel information is captured on check-in and check-out points, moment of travel, product used for traveling, etc, for public transport providers in the Netherlands.

The data obtained from TLS is not yet translated to a data set containing journeys. It contains only the smart card id and a check-in or check-out (*checks*) at station gate or pole. An algorithm to create journeys from all the checks obtained by an user exists and is used to construct journeys. The algorithm in short; the information on the checks are sorted by smart card id and the moment of the check, if two checks construct a realistic journey by one user, then they are linked as a journey (transformation from check-in and check-out to a single journey). A journey is realistic if a check-in is followed by a check-out and the distance traveled between the *checks* is done within a constrained time frame corresponding to the public transport route network in the Netherlands. After applying the algorithm, the obtained journey data set contains information on the check-in and check-out location, the moment of travel, the amount paid and most importantly the type of subscription package used for the travel.

From the constructed journeys with check-in and check-out location the total cost of the journey can be constructed. The Dutch Railways does not work with distance traveled in kilometers, but with distance units and the corresponding prices to these units[4]. From the check-in and check-out location the distance units can be linked to the journey and thus the total cost of the journey (with the given subscription). The distance in units and in kilometers are highly correlated, but not exactly the same, since some trajectories are differently priced. In this paper, we relate to the distance units as distance in kilometers.

One of the shortcomings of the data is that it contains journeys that have a missing check-in and/or check-out location and thus a missing distance. The reason can be no check-in or -out of the user, but also that the algorithm was not able to link two *checks* to a journey. These observations that are missing are removed from the journey data set, but this can only be done under the assumption that they are missing completely at random. It could be, for example, that an individual did not check-out on purpose, since it is cheaper to not check-out. Due to the large amount of data on the journeys and the aggregation of the journeys over time, we assume that these missing observations are negligible and can be removed from the data.

[2] The website to the tool: www.myov.nl

[3] Trans Link Systems B.V.: company in charge of processing the transactions

[4] An overview of the different tariffs and distance units is given on: www.ns.nl/klantenservice/betalen/tarieven-consumenten.html

Another shortcoming is that the MyOV tool is not a random sample of the public transport users and their subscriptions, since users had to subscribe themselves to the reward systems. Points could be earned on tracks from and to *Utrecht Central Station*, for which evidence is found in the data (mostly used check-in and check-out station). In addition, individuals that use the reward system are users that travel often during rush hour and travel frequently. The users in the journey data are thus not a completely random sample of public transport users, and not fully representative for travel behaviour of individuals. This can result in a bias in the coefficients and has to be taken into account throughout the analysis of this paper.

3.2 Additional information on subscriptions

Next to the journey data, information is needed on the subscription packages, how do they exactly differ and what do these different subscriptions cost. This type of information is not contained in the journey data, but can be found on the website of the Dutch Railways. The possible subscriptions, which vary in; moment of travel, type of discount, comfort level, duration and route traveled, differ in price. The prices for 2019 are collected and linked with the corresponding subscription in the journey data.

Besides, the journey data are over a time frame of 4 years and the prices have changed over these four years and between subscriptions. The collected prices are of 2019 and must be corrected for these changes. Since not all subscription prices can be fully traced back for all these years and the prices mostly change with the expected inflation (NS 2018), the expected inflation in the Netherlands[5] is used as an indicator for the price changes of the subscriptions over time. Next to the increase of subscription prices in line with inflation, there has been a change in taxes on the subscriptions in the Netherlands (NS 2018). This change is also taken into account in calculating the prices. Finally, certain subscriptions are sold with a discount, however, the size of the discount and when individuals bought a subscription with a discount cannot be traced back. Therefore, it is assumed that all individuals paid the cost for an non-discounted subscription package. Because the price every individual has paid for their subscription package is an indication and not the actual price, there is a possibility of bias in the results, which has to be taken account when interpreting the results.

3.3 Aggregation to decision moments

The MyOV travel data are available on the journey level of a card id, which contains information on every single trip made by a smart card over a certain amount of time. We assume that an individual has only one smart card and this card is directly linked to an user. Users aren't expected to make decisions on their subscription and cannot switch between subscriptions at the start of every journey. Based on the subscription package, which is monthly or yearly, travelers can switch between subscription packages. In addition, users do not necessarily have to decide on their subscription package at the beginning of the month, thus during the month users can change their subscription. Therefore in this paper *decision moments* are defined. A *decision moment* is "a specific moment in time where the individual is assumed to take a discrete choice on the way of payment for public transport provided by the Dutch Railways through the means of a subscription package", more elaboration on this definition in Section 4.1. In conclusion, not every journey is a decision moment and therefore, to correctly model these moments an aggregation step has to be made from single journeys to a summary of the single journeys for every individual.

[5]https://nl.inflation.eu/inflatiecijfers/nederland/historische-inflatie/cpi-inflatie-nederland.aspx

Two options are to summarize information of the single journeys to monthly or yearly travel behaviour corresponding to the duration of the subscriptions. However, there is a trade-off between aggregating the data and the information contained in the data. For example, aggregating to monthly data per individual loses information on the differences between the single journeys within that month and if the data are aggregated to yearly data more information will be lost. Another disadvantage is that, for most individuals the data contain information on only a couple of years (maximum of four years), but many subscriptions are yearly and individuals only make decisions once a year. The larger the time frame the data are aggregated to, the lower the number of observations over time per individual.

A lower-bound to define a decision moment could be monthly, since the shortest amount of time for a subscription is a month and an upper bound yearly, using the same train of thought on the longest subscription time. However, the majority of the subscriptions is yearly and most decisions are not made on a monthly level. In conclusion, in this paper it is assumed that every quarter of a year a user makes a decision on their subscription package. Thus, the journey data are aggregated to quarterly data, which is in between monthly and yearly aggregation. This way both monthly and yearly decision moments are captured and the loss of information is not too extreme.

After aggregating the data, an observation contains information on an individual at a quarter in time (for example, first quarter of 2017). With aggregating the data, different aggregating functions are used for the distance traveled, the amount paid and the subscription used. First, the journeys are split depending on the moment of the journey during the day. The moments used for this split correspond to the different moments subscription packages can differ in *off-peak*[6], *weekend* and *peak*. Note that for the Dutch Railways the moment of a journey is decided on the moment of check-in. The peak hours are defined as the check-in time between 6:30-9:00 and 16:00-18:30 on weekdays and outside these time frames on weekdays off-peak and in the weekend the moment is defined as weekend. Thus, for every individual per quarter we obtain the *off-peak*, *weekend*, *peak* and *total* distance traveled in kilometers. This moment split in the distance traveled leads to the subscription based kilometers, for example an individual with the WEEKEND FREE subscription has as subscription based distance the weekend kilometers and an individual with ALWAYS FREE subscription has total distance traveled as subscription based distance. An elaborate explanation of the different subscriptions is given in the next section and in Table 2.

Then, for every observation we sum over the kilometers traveled during *off-peak*, *weekend* and *peak*. The *Total Amount* paid and the total *Number of Travels* is not made moment dependent and is summed over the completed journeys that quarter of an user. The subscription package denoted for that observation is the "primary" subscription package, which is the package the highest distance is traveled with that quarter, more on this in Section 4.2.

Since the price of a trajectory subscription depends on the distance of the trajectory, an indicator for the chosen distance of the trajectory package is created. For every individual that has a trajectory subscription the average distance per journey over time is calculated. This average trajectory distance is linked with the corresponding trajectory price given by the Dutch Railways[7]. Again, these prices are corrected for inflation and taxes.

3.4 Data description

For this research, the quarterly aggregated journey data contain information on 4379 individuals that subscribed to the MyOV tool and traveled with the Dutch Railways. The first recorded journey

[6] the off-peak hours do not contain the weekend hours

[7] www.ns.nl/klantenservice/betalen/tarieven-consumenten.html

was July 2015 and the last was April 2019. With these variations over two dimensions the MyOV data are panel data. The data are not balanced, since the observations for every individual over time mostly cover a different time frame. On average every individual is observed over 10 quarters of time, or 2.5 years. This gives a pooled data set of 7538 observations, with for second and first class 5120 and 2418 observations, respectively. Note that individuals with 4 quarters or less are removed from the data, since lag functions of one to four quarters are calculated.

	Obs. (%)	Nr. of subscriptions 1 (%)	>1 (%)	Total dist. (km per Q)	Nr. travels (per Q)	Dist. per travel (km)
Other	56	68.7	32.3	1906	42	49
Student	38	68.1	31.9	1993	43	55
Business	6	99.9	0.1	3486	72	52
Total	100	70.0	30.0	2033	44	52

Table 1: Summary statistics per type of traveler

In Table 1 summary statistics per type of traveler are given. The different type of travelers are students, business and travelers other than business or student. The business travelers have subscriptions that are offered by businesses to their employees and the student subscriptions are a free subscription offered by the government to students to cover their travel expenses. The data set contains mostly other and student type of travelers and only few business travelers. The student and other travelers show similar travel behaviour. Of the individuals, 70% had only one subscription during the observation period and the other 30% have more than one subscription. On average per quarter, individuals travel in total 2032 kilometers of which, 61% in off-peak hours, 26% in the peak hours and 13% in the weekend (not shown in the Table). Per quarter, individuals travel on average 44 times, which is around 3.5 travels per week. Where travelers with business subscriptions travel on average 72 times and students 43 times, thus travelers with business subscriptions are the most frequent travelers.

A summary of the possible subscriptions, a short description and summary statistics are given in Table 2. A split is made in first and second class subscriptions. The motivation is given in Section 4. Possible subscriptions are the subscriptions of the Dutch Railways that last minimal one month and are available to all customers. There is also an option of not having a subscription, without a fixed fee, namely FULL TARIFF. The subscriptions in the table are the possible set of subscriptions in April 2019. Subscriptions that have changed in the past in name or characteristics are assigned to the subscription closest in characteristics to the given set of possible subscriptions.

	Subscription	Description	Second Class			First Class		
			Fixed Fee* (€)	Obs. (%)	Avg. Dist. (km/Q**)	Fixed Fee* (€)	Obs. (%)	Avg. Dist. (km/Q**)
1	WEEKEND FREE	Free travel during the weekend	34	4.3	2025	40	13.6	2469
2	OFF-PEAK DISCOUNT	40% discount during the off-peak hours	4.33	20.5	1048	4.33	3.4	908
3	OFF-PEAK FREE	Free travel during the off-peak hours	105	4.2	4140	133	24.4	5356
4	ALWAYS DISCOUNT YEARLY	40% discount during the off-peak hours, 20% discount during the peak hours, yearly	23	2.8	2544	23	1.4	2667
5	ALWAYS DISCOUNT MONTHLY	40% discount during the off-peak hours, 20% discount during the peak hours, monthly	28	0.6	3047	28	0.5	5077
6	ALWAYS FREE YEARLY	Free travels always, yearly duration	346	2.7	5987	583	0.4	5982
7	ALWAYS FREE MONTHLY	Free travels always, monthly duration	424	0.5	6680	728	-	-
8	TRAJECTORY FREE YEARLY	Free travels on a predefined trajectory, yearly	59-346***	5.5	2762	125-583***	1.4	2727
9	TRAJECTORY FREE MONTHLY	Free travels on a predefined trajectory, monthly	73-424***	1.4	2835	100-728***	-	-
10	FULL TARIFF	Always no discount	0	8.7	473	0	1.5	1010
		Total obs.		5120			2418	
11	BUSINESS OFF-PEAK DISCOUNT	For Businesses, 40% discount during the off-peak hours	4	1.0	2136	4	4.7	2670
12	BUSINESS TRAJECTORY FREE	For Businesses, free travels on a predefined trajectory	54-317***	1.5	3118	92-535***	1.9	3272
13	BUSINESS ALWAYS FREE	For Businesses, free travels always	317	1.8	4380	535	46.7	3674
		Total obs.		430			2771	
14	STUDENT WEEKEND	For students, free travels during the weekend	-	4.3	1526	-	-	-
15	STUDENT WEEK	For students, free travels during the week	-	40.2	2043	-	-	-
		Total obs.		4450			-	

* Fixed fee is monthly price in 2019

** Q: Quarter

*** Fixed fees of trajectory subscriptions depend on the chosen trajectory

Table 2: Subscription table

The focus in this paper is mostly on the subscriptions that are not business or student subscriptions, see Section 4.2 for an explanation. In the Table, the monthly fixed fee for 2019 is given, showing that first class subscriptions are more expensive than or equal to second class subscriptions. ALWAYS FREE MONTHLY is the most expensive and the OFF-PEAK DISCOUNT the least expensive of all the subscriptions. For second class travelers, these subscriptions are also the least and the most chosen, respectively. In addition, the more expensive the subscription, the higher the distance traveled per quarter. If individuals travel more per quarter, their travel expenses rise and therefore, the need for a subscription that saves on these expenses increases. Yet, the relation, the higher the price the lower the demand for the subscriptions, does not hold. Also three first class subscriptions that are offered by the Dutch Railways are not or little observed in the data, namely the monthly subscriptions, TRAJECTORY FREE MONTHLY, ALWAYS FREE MONTHLY and ALWAYS DISCOUNT MONTHLY. First class subscriptions are more expensive and therefore, the incentive to participate in the MyOV reward system could be less than for second class travelers. Note that for second class STUDENT WEEK and for first class BUSINESS ALWAYS FREE is most observed in the MyOV dataset, however as aforementioned this is not the focus of the paper.

The changes of travel behaviour over time is given in Figure 2. For every quarter of the observed time period the mean distance of off-peak, peak and weekend kilometers is given. The sum of the three is the mean distance traveled by individuals per quarter, thus the average total kilometers over time. From the figure one can see no changes in the average weekend and off-peak kilometers over time. However, the peak kilometers show a seasonal trend. In the second quarter of every year individuals travel less peak-kilometers, which can be explained by the summer period where many individuals go on a vacation and travel less in peak hours.

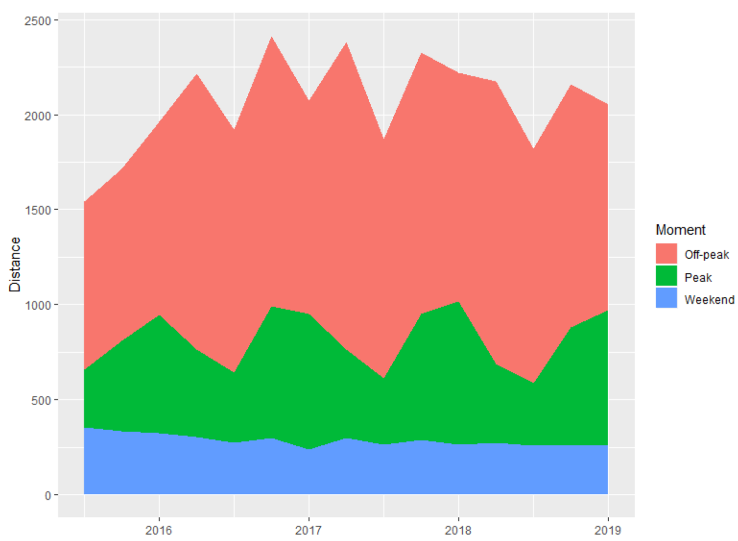


Figure 2: Time series of the avg. traveled distance (in km)

4 Conceptual framework

To answer the research questions, this paper makes use of a discrete choice model. In this Section the features common to discrete choice models are defined, namely the choice problem, the choice set and the choice attributes.

4.1 The choice problem

The objective of the problem is to explicitly model the choice behaviour. As defined in Section 3.3, a decision moment is: "a specific moment in time where the individual makes a discrete choice on the way of payment for public transport provided by the Dutch Railways through the means of a subscription package". Firstly, the choice of the individuals is discrete, since they choose one of the alternatives in the choice set. In addition, the individuals that have to make a choice on a subscription package is narrowed down to the group of travelers that is subscribed to MyOV for over a year. Furthermore, a decision is made at one moment in time in advance of the realized travel behaviour, thus the individual has to predict their own travel demand before the decision of the subscriptions has been made.

Before this decision moment, the individual goes through a decision process that leads to the final decision. McFadden (2001) has modelled one decision-making task visually, see Figure 3. He explains: "the decision process as a lifelong sequence with earlier information and choice operating through experience and memory to provide context for the current decision problem, and the results of this choice feeding forward to influence future decision problems". The heavy arrows in this figure from McFadden (2001) coincide with the economists' standard model of the choice process, a theory of rational choice in which individuals collect information on alternatives. The lighter arrows in the diagram correspond to psychological factors that enter decision-making. *Perception, preference* and *process* appear in both economic and psychological views of decision-making.

The heavy arrows in this decision process can be modeled by smart card data and are in line with the decision process individuals go through when making their choice of subscription package. However, not the full process can be modeled since not all information is available. Knowledge on past travel behaviour, through *Experience* and *Memory*, influences current *Perceptions* and knowledge on future expected travel behaviour affects these *Perceptions/Beliefs* through *Information.* In addition individuals differ in their *Preferences* which also influences the decision process, which can be explained by individual-specific differences. Nevertheless, for this empirical research there is no availability of stated preferences. Furthermore, there are choice set constraints, for example only students are able to choose a student subscription. The above explained sources of information enter the decision process which turns into a choice.

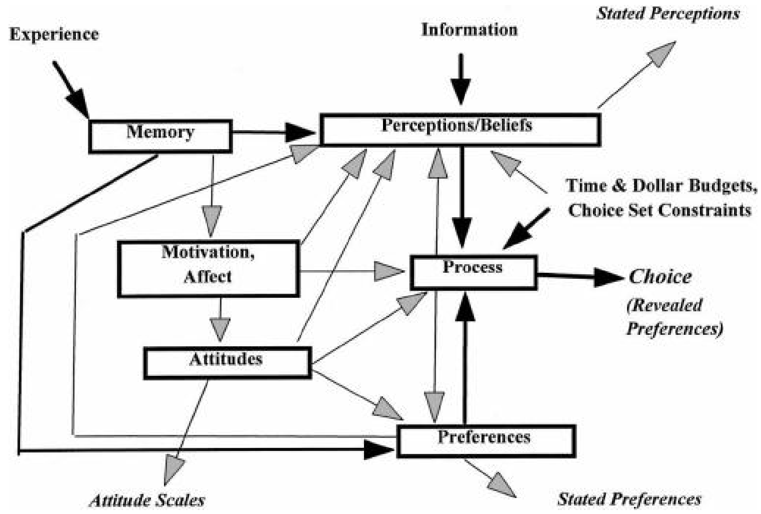


Figure 3: The choice process McFadden (2001)

4.2 The choice set

In Table 2 of the previous section, the set of possible subscription packages is given with a short description. As mentioned before, the packages differ in the moment of travel, type of discount, duration and trajectory. Since the research in this paper is on the subscription choice of travelers for the Dutch Railways, the focus is only on subscriptions of the Dutch Railways and no other public transport providers. In addition, the interest lies on more frequent travelers and no day trip travelers, thus journeys that are made with day tickets are excluded from the set of possible subscription packages. Nevertheless, if travelers continuously pay the full-tariff for their travels and did not pay a fixed fee, this is included in the choice set as *no subscription*.

The full set of subscription packages is split into two sets, namely a first and second class set. The decision to travel first or second class is not explained by travel behaviour, but mostly by other factors such as income (Brons et al. 2002). Because the decision process to choose between first and second class is a different process, two choice sets are defined. Hence, it is assumed that both decision processes are an independent process. Moreover, there is an availability constraint for the subscription packages. As aforementioned, three types of travelers can be distinguished, student, business and other. Yet, the business and student packages are not chosen by the traveler, but offered by the firm or government and the decision process on the type of package is different. If a business or student subscription is possible, the individual chooses one of these subscription and if none of the two are possible, the individual will choose from the *other* subscriptions. The focus of this paper is on the variable subscription packages consumers **choose** based on their expected travel behaviour and therefore the business and student subscription packages are excluded from the choice set. The choice set for first class subscriptions is limited to only yearly subscriptions, since there are no observations where a monthly subscription is chosen, see Table 2.

The sets of alternatives, the choice sets, exhibit the three characteristics given by Train (2009). First, the alternatives must be mutually exclusive from the decision maker's perspective. Second, the choice set must be exhaustive, in that all possible alternatives are included. The decision maker necessarily chooses one of the alternatives. Third, the number of alternatives must be finite, one can count the alternatives. The second and third restriction simply follows from the definition of the choice set above, however, the first restriction does not naturally hold.

The first restriction, the decision maker chooses only one alternative from the choice set, is not always true. The Dutch Railways offers the possibility to travelers to combine subscription packages. For example a student with a STUDENT WEEK and WEEKEND FREE subscription. However, from investigating the data it becomes clear that observations with double subscriptions are rare events, since the current subscription offering of the Dutch Railways is quite complete and seems to suffice the needs of travelers. Another possibility of a double subscription can be due to the aggregation to decision moments, see Section 3.3. There is a possibility to have several subscriptions in one quarter, since for example, an individual switched their subscription during that quarter. As a solution, the subscription is chosen that is mostly used by the individual, where mostly used is defined as the subscription the largest distance has been traveled with.

4.3 The choice attributes

To explain the decision between the choice alternatives, revealed or realized information on travel behaviour is used. Travel behaviour can be divided into different attributes which are explained below.

Realized travel demand

In Section 4.1 information is one of the factors influencing the *Perceptions/Beliefs*. The input of information in the decision process is the knowledge an individual has on their expected travel behaviour. A measure for the expected travel behaviour is the realized travel demand. At the moment of decision the individual is not aware of their realized travel demand and therefore this information is future knowledge. The realized travel demand is like a perfect prediction of their own travel demand. Two measures of realized travel demand variables are constructed, one is the total distance traveled that quarter and the other is total distance based on the type of subscription, split by the moment of travel (off-peak, peak, weekend). The split in moment is made, since the subscriptions differ in the moment of travel and are choice-specific.

Historic travel demand

The choice process of McFadden (2001) also points out the influence of *Experience* and *Memory*. This is modeled through past travel behaviour, hence the travel demand of previous quarters. The historic travel demand is measured in distance traveled in kilometers and again a measure of total historic distance and, subscription based historic distance is constructed. To measure historic travel demand, lag functions of variables are constructed. Several sizes of lags are used, namely of one quarter till four quarters (a year). The year is a maximum measure for historic demand, since it is not expected that decision-makers look back more than a year to make a subscription choice. The historic information of two quarters or more back is averaged out over the other historic lags. Due to the use of lags, the first four observations per individual are removed from the data set.

Realized savings

The economic value of subscription choice is modeled with the realized savings variable. This variable takes into account the fixed fee of the subscription, the costs of the realized travels with the chosen subscription and the full tariff that would have been paid for the travels if no subscription was chosen.

The calculation of the fixed fee of the subscription is given in Section 3.2 and is denoted by P_{it}, the price of choice alternative in the choice set for individual i at quarter t. Then the cost of the realized travels with the chosen subscription is given from the journey data and are aggregated to quarters, denoted by C_{it}. The full tariff, FT_{it}, that would have been paid for the travels if no subscription was chosen, is calculated through multiplying the price per kilometer by the total kilometers traveled that month. The price per kilometer is empirically calculated from the data, namely by calculating, for every year, the average price per kilometer paid by individuals that had no subscription (or FULL TARIFF subscription). This gives the following calculation of the realized savings, S_{it},

$$S_{it} = FT_{it} - (C_{it} + P_{it})$$

Thus, a positive result of the realized savings means that the chosen subscription has been economically a good choice. Individuals with the FULL TARIFF subscription did not realize any savings (S_{it}=0). However, the empirically calculated price is an indication of the true price per kilometer that could have been paid, since the Dutch Railways does not work with a fixed price per kilometer as explained in Section 3.

Individual characteristics

Differences between individuals in their travel behaviour can be obtained. Note that there are no demographics available. The average travel distance per travel per individual is informative on the length of the journeys the traveler makes. A traveler that travels long distances might go through a different decision process than short distance travelers.

5 Methodology

In this section the methodology is given. Note that bold symbols in the mathematical formulations refer to vectors and capital bold symbols refer to matrices.

The first part elaborates on choice probabilities. Then, the methodology is given behind the multinomial logit model (MNL). This model is used, since it is the most basic and widely used model in discrete choice modeling. However, one of the main shortcomings is the assumption of independence of irrelevant alternatives (IIA). To explore the hierarchy in choice two nested logit models are given, nested logit (NL) and cross nested logit (CNL), which are both part of the generalized extreme value models. The cross nested logit is given instead of the multilevel nested logit, due to it is more flexible structure and most multilevel nested logit models can be approximated by the cross nested logit. Next, the panel data structure of the data is taken into account in the mixed logit model (ML). The mixed logit model is a very flexible model, it accounts for preference heterogeneity and relaxes assumptions namely the IIA. The latent class model (LCM) also accounts for preference heterogeneity but for a discrete number of latent classes, whereas the ML accounts for preference heterogeneity for every individual separately. Even though the LCM is relatively more simple than the ML, this research estimates the ML. In this paper, we would like to investigate if there are differences across consumers preferences in general and not if they belong to a certain class, therefore the most flexible model, the ML, is estimated and given in this section.

5.1 Choice probabilities

The basic approach to the mathematical theories of individual preferences is that of micro-economic consumer theory. The theory provides the means to transform assumptions about desires into a demand function, that expresses the action of a consumer. This theory is based on “rational behaviour” and “rational behaviour” is what an observer beliefs about what the outcome of a decision should be. In discrete choice theory the same concept is applied, but with a discrete representation of the set of alternatives. Both hold the concept of the rational consumer; the only difference between choice theory and consumer theory will be that instead of deriving demand functions, it will work directly with utility functions. The random utility approach, formalized by Manski (1977), is in line with consumer theory, however, the utilities are unobserved and not known with certainty and they are therefore treated as random variables. Based on Random Utility Maximization the research question is mathematically formulated and answered.

The following choice situation is defined: a decision-maker, n, faces a set C_n with J subscription alternatives. There exists some vector of unobserved utility of choices, $(U_1, ..., U_J)$. In this paper, n refers to an individual and i, j refer to choice alternatives in the choice set. The probability of any alternative i being selected by person n from choice set C_n is given by:

$$P_n(i) = P(U_{in} \geq U_{jn}, \forall j \in C_n)$$

The probability that $U_{in} = U_{jn}$ for any i and j is here ignored in the choice set. Formally, if the distributions of U_{in} and U_{jn} can be characterized by a probability density function, $P(U_{in} = U_{jn}) = 0$. Note, in this formulation the choice set C can differ per individual n.

Reformulating the utility in an observed, V_{in} and unobserved, ϵ_{in}, component: $U_{in} = V_{in} + \epsilon_{in}$ gives

$$\begin{aligned}
P_n(i) &= P(U_{in} \geq U_{jn}, \forall j \in C_n, j \neq i) \\
&= P(V_{in} + \epsilon_{in} \geq V_{jn} + \epsilon_{jn}, \forall j \in C_n, j \neq i) \\
&= P(\epsilon_{jn} \leq V_{in} + V_{jn} + \epsilon_{in}, \forall j \in C_n, j \neq i)
\end{aligned}$$

the general derivation of any particular choice model given assumptions on the joint distribution of the disturbances.

The observed component V_{in} can be rewritten as a function of the vector of observed variables, $V(\mathbf{x}_{in})$. Assuming that the function, $V(\cdot)$ is specified by an additive model, thus linearly weighting variables gives:

$$V(\mathbf{x}_{in}) = \boldsymbol{\beta}'\mathbf{x}_{in}$$

here $\boldsymbol{\beta}$ is the vector of coefficients and $\mathbf{x_{in}}$ the vector of observed variables, both containing M components.

5.2 Multinomial Logit

The most widely used discrete choice model is the multinomial logit model. Two benefits of this formula are that it takes a closed form for the choice probabilities and it is easy interpretable. It's original formulation is due to Luce (1959). He used the assumptions about the characteristics of choice probabilities, namely the *independence from irrelevant alternatives* (IIA) property, which is discussed in Section 5.2.1. Marschak (1960) shows that the model is consistent with utility maximization.

To derive the logit model a specific distribution is used for the unobserved utility component. Using the similar choice situation as the one given above, the logit model is obtained by assuming that each unobserved component of utility has i.i.d. Extreme Value Type I distributed errors, also called the Gumbel distribution. The variance of this distribution is $\pi^2/6$ and normalizes the scale of utility. This handles two aspects of behavioural decision process, namely, *only differences in utility matter* and *the scale of utility is arbitrary.* The cumulative distribution for each unobserved component of utility is

$$F(\epsilon_{nj}) = e^{e^{\epsilon_{nj}}}$$

The difference between two extreme value variables is distributed logistic and composes to the logit choice probability. Hence, the probability of choosing alternative i is given by the Multinomial Logit formulation (McFadden 1974) is:

$$P_n(i) = \frac{e^{V_{in}}}{\sum_{j=1}^{J} e^{V_{jn}}}$$

This logit probability exhibits the properties of probability

$$0 \leq P_n(i) \leq 1, \text{ for all } i \in C_n$$

and

$$\sum_{i \in C_n} P_n(i) = 1$$

Rewriting the logit choice probabilty assuming that it is linear in the variables gives the following formulation:

$$P_n(i) = \frac{e^{\boldsymbol{\beta}'\mathbf{x}_{in}}}{\sum_{j=1}^{J} e^{\boldsymbol{\beta}'\mathbf{x}_{jn}}} \tag{1}$$

For estimation of Equation 1 two assumptions have to be made, namely, that the sample is exogenously drawn and that the explanatory variables are exogenous to the choice situation. A sample of N decision makers is obtained for estimation. As mentioned before as a benefit from the MNL, the logit probabilities take a closed form, which implies that estimation can be done by maximum-likelihood. Then, the probability of person n choosing the alternative that he actually chose can be written as

$$\prod_i (P_n(i)^{y_{ni}})$$

where $y_{ni} = 1$ if person n chose i and zero otherwise. From the assumption of the i.i.d. Gumbel distributed unobserved components follows that each individual's choice is independent of that of other individuals. Then, taking the log gives the log likelihood function

$$LL(\boldsymbol{\beta}) = \sum_{n=1}^{N} \sum_i y_{ni} \ln P_n(i) \tag{2}$$

where $\boldsymbol{\beta}$ is a vector containing the parameters of the model. The estimator is the value of β that maximizes this function. McFadden (1974) shows that $LL(\boldsymbol{\beta})$ is globally concave for linear-in-parameters utility.

5.2.1 IIA

As aforementioned, Equation 1 exhibits *independence form irrelevant alternatives.* The IIA property stated succinctly in words; if some alternatives are removed from a choice set, the relative choice probabilities from the reduced choice set are unchanged. The choice probabilities from a subset of alternatives is dependent only on the alternatives included in this subset and is independent of any other alternatives that may exist. Thus, for a specific individual the ratio of the choice probabilities of any two alternatives is entirely unaffected by the systematic utilities of any other alternatives. This can be easily shown for the MNL case as follows:

$$\frac{P_n(i)}{P_n(l)} = \frac{\frac{e^{V_{in}}}{\sum_{j=1}^{J} e^{V_{jn}}}}{\frac{e^{V_{ln}}}{\sum_{j=1}^{J} e^{V_{jn}}}} = e^{V_{in} - V_{ln}}$$

However, for the logit model this property can give rise to odd and erroneous predictions. One of the most used examples is the *red bus/blue bus paradox* where there is a hierarchy in choice and the IIA property does not hold, see Figure 4. Suppose an individual can choose between two options a car and blue bus with equal probability (probability of 0.5 each). Now another option is added, the red bus. Assume that individuals do not pay attention to the color of the bus they are traveling with. Thus, even though a third option is added, individuals are still expected to choose between car and bus with equal chance. Suppose an individual chooses between a car and a blue bus with equal probability. Then a third option is added, the red bus. Assuming that individuals that travel by bus do not care about the color of the bus, they are expected to choose between the bus and car still with equal probability. However, the IIA property implies that this equal probability does not hold, if both the ratio between car and blue bus and, the ratio between blue and red bus have to be maintained, the new probabilities must be 0.33 each. Thus the IIA does not successfully deal with the fact that the red and blue bus are very similar and share similar characteristics. In this specific choice problem, there is a hierarchy in choice, the individual first chooses between car or bus and if bus is chosen, then the color of the bus.

Even though the MNL seems unaffected from the IIA property it's problem lies in the assumptions of the MNL. The MNL model assumes that the disturbances are mutually independently distributed. However, in the case of the red buses and blue buses this is wholly implausible since both these alternatives have all the unobserved characteristics of buses. In fact, the disturbances of the red and blue bus modes are more reasonably assumed to be perfectly correlated.

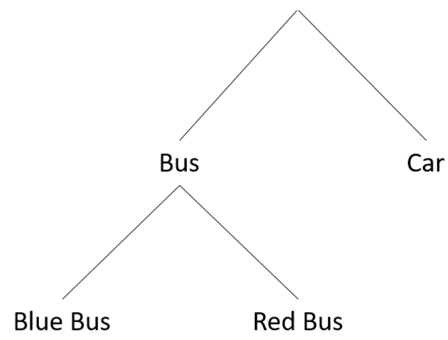


Figure 4: Example of a choice hierarchy

5.3 Generalized Extreme Value Models

In general it is not easy to model all kinds of correlation and to make sure the IIA does not hold. Hence, a more general model than standard logit is needed, such as Generalized Extreme Value model (GEV). GEV models combine different type of models that indicate all kind of possible substitution patterns. The main characteristic that classifies these models as GEV models is that the unobserved portions of utility for all alternatives are jointly distributed as a generalized extreme value distribution. This distribution allows for correlations over alternatives and is a generalization of the univariate extreme value distribution that is used for standard logit models. When all correlations are zero the GEV model becomes standard logit. Thus the class includes the standard logit model, but also a variety of other models. Similar to MNL model, GEV models have the benefit that the choice probabilities usually take a closed form. Thus, they can be estimated with maximum likelihood and without simulation.

McFadden (1978) has derived the generalized extreme value model. He shows that the choice model defined by equation 5.3 is consistent with random utility maximization. Three examples of GEV models: the MNL, the nested logit and the ordered generalized extreme value of Small (1981).

Let $G(y_1, y_2, ..., y_{J_n})$, for $y_1, y_2, ..., y_{J_n} \geq 0$ be a function with the properties given in the Appendix. If G satisfies the conditions and $G_i(y_1, y_2, ..., y_{J_n})$ denotes $\partial G / \partial y_i$, $i = 1, 2, ..., J_n$ then

$$P_n(i) = \frac{e^{V_{in}} G_i(e^{V_{1n}}, e^{V_{2n}}, ..., e^{V_{J_n n}})}{\mu G(e^{V_{1n}}, e^{V_{2n}}, ..., e^{V_{J_n n}})}$$

defines the GEV model. Properties of the GEV model are given in the Appendix.

5.3.1 Nested Logit

The most widely used member of the GEV family is called nested logit. A nested logit can be applied if the choice set can be divided into subsets, called nests. Thus, for example, with the red/blue bus paradox, the red and the blue bus can be partitioned into a nest. The subdivision should be done in such a way that for alternatives within a nest IIA holds and for alternatives between nest the IIA does not hold in general.

From Train (2009), let the set of alternatives j in the choice set be divided into K nonoverlapping nests denoted $B_1, B_2, ..., B_K$. The utility an individual n obtains from alternative j in nest B_k is denoted as $U_{nj} = V_{nj} + \epsilon_{nj}$, an observed and unobserved part. The nested logit model is obtained by assuming that the vector of unobserved utility, $\epsilon_n = (\epsilon_{n1}, ..., \epsilon_{nJ})$ has cumulative distribution

$$F(\epsilon_{nj}, \mu_k) = e^{(-\sum_{k=1}^{K}(\sum_{j\in B_k} e^{-\epsilon_{nj}\mu_k})^{(1/\mu_k)})} \tag{3}$$

with $\mu_k \geq 1$ a measure of correlation or degree of independence. If $\mu_k = 1$, it indicates that within a nest there is no correlation or complete independence. When $\mu_k = 1$ for all k, the nested logit reduces to the standard logit, since the GEV distribution becomes the product of independent extreme value terms.

As aforementioned, for the standard logit, each ϵ_{nj} is independent with a univariate extreme value distribution. However, for the nested logit, the ϵ_{nj}'s are correlated within nests and uncorrelated among nests. Stated differently: for any two alternatives, j and m in nest B_k, ϵ_{nj} is correlated with ϵ_{nm}, and for any two alternatives in different nests, the unobserved portion of utility is still uncorrelated: $\text{Cov}(\epsilon_{nj}, \epsilon_{nm}) = 0$ for any $j \in B_k$ and $m \in B_l$ with $l \neq k$.

The distribution given in Equation 3 leads to the following choice probability for alternative $i \in B_k$:

$$P_n(i) = \frac{e^{\mu_k V_{ni}}(\sum_{j\in B_k} e^{\mu_k V_{nj}})^{1/\mu_k - 1}}{\sum_{l=1}^{K}(\sum_{j\in B_l} e^{\mu_l V_{nj}^{(1/\mu_l)}})} \tag{4}$$

This expression is not very illuminating as a formula. However, the choice probabilities can be expressed in an alternative fashion through a decomposition into two logits. The choice probability of Equation 4 can be split in the product of a conditional and marginal probability; (1) the probability of choosing an alternative in nest B_k and (2) the probability of choosing the alternative i given that nest B_k is chosen. These two probabilities can be rewritten in the form of two separate logits, which can be algebraically solved into Equation 4 (see Train (2009) for the full proof).

Williams (1977) and Daly and Zachary (1975) proved that the nested logit model is consistent with utility maximization. Since the nested model takes on a closed form, the parameters can be estimated by maximum likelihood. Similar to the MNL, the choice probabilities of Equation 4 have to be substituted into the log-likelihood function which gives an explicit function of the parameters. Brownstone and Small (1989) shows that the values of the parameters that maximize this function are consistent and efficient (under moderately general conditions).

5.3.2 Cross Nested Logit

For the nested logit each alternative is a member of only one nest. However, this restriction of non-overlapping nests can be relaxed in the cross nested logit model (CNL) with overlapping nests.

Again, the nests of alternatives are described by $B_1, B_2, ..., B_K$. However, now each alternative can be a member of more than one nest and thus an alternative can be allocated among more than one nest. The allocation parameter α_{jk} assigns the alternative to the nests. This parameter can be seen as a percentage indicating to which extent the alternative, j is allocated to nest k. Since it is an percentage it is non-negative: $\alpha_{jk} \geq 0, \forall j, k$ and each alternative has to be fully allocated among the nests: $\sum_k \alpha_{jk} = 1, \forall j$. A zero value means that the alternative is not in the nest at all. The parameter μ_k is defined for each nests and its formulation is equal to the nested logit formulation. The probability that person n chooses alternative i is

$$P_n(i) = \frac{\sum_k (\alpha_{ik} e^{V_{ni}})^{\mu_k} (\sum_{j\in B_k} (\alpha_{jk} e^{V_{nj}})^{\mu_k})^{1/\mu_k - 1}}{\sum_{l=1}^{K}(\sum_{j\in B_l} (\alpha_{jl} e^{V_{nj}})^{\mu_l})^{1/\mu_l})} \tag{5}$$

The equation is similar to Equation 4 of the nested logit and if each alternative is only part of one nest, $\alpha_{jk} = 1$ for $j \in B_k$ and zero otherwise, the model becomes a nested logit model.

The parameters of the cross nested model can be estimated by standard maximum likelihood techniques. Substituting the choice probabilities of Equation 5 into the log-likelihood function gives an explicit function of the parameters in this model.

5.4 Mixed Logit Model

The mixed logit is a highly flexible model that can approximate any random utility model (McFadden and Train 2000). The specification is the same as for standard logit except that $\boldsymbol{\beta}$ varies over decision makers, rather than being fixed. It obviates limitations of standard logit by allowing for random taste variation, unrestricted substitution patterns, and correlation in unobserved factors over time (IIA). The simplest specification treats the coefficients that enter utility as varying over people, but being constant over choice situations for each person. The decision maker faces a choice among J alternatives at choice situation t, the utility of person n from alternative i is specified as

$$U(\mathbf{x}_{int}) = \boldsymbol{\beta}_n' \mathbf{x}_{int} + \epsilon_{int}$$

where $\mathbf{x}_{int}$ are observed variables that relate to the alternative, decision maker and decision moment, respectively. $\boldsymbol{\beta}_n$ is a vector of coefficients of these variables for person n representing that person's tastes. The unobserved part of the utility function ϵ_{int} is a random term that is i.i.d. Gumbel distributed over time, people and alternatives.

The coefficients vary over decision makers with density $f(\boldsymbol{\beta})$. Mixed logit is a mixture of the logit functions evaluated at different $\boldsymbol{\beta}$'s with $f(\boldsymbol{\beta})$ as the mixing distribution. The unobserved portion of utility is $\eta_{njt} = \boldsymbol{\beta}_n' \mathbf{x}_{jnt} + \epsilon_{int}$, which can be correlated over alternatives depending on the specification of $\mathbf{x}_{jnt}$. For the standard logit model, $\mathbf{x}_{jnt}$ is identically zero, so that there is no correlation in utility over alternatives. This lack of correlation gives rise to the IIA property and its restrictive substitution patterns.

Consider a sequence of alternatives, one for each time period $\mathbf{i} = i_1, ..., i_T$. The unconditional choice probability or the mixed logit probability is the product of logit formula's over time (since ϵ_{int} are independent over time) integrated over all possible variables of $\boldsymbol{\beta}_n$

$$P_n(i) = \int (\prod_{t=1}^{T_n} \frac{e^{\boldsymbol{\beta}' \mathbf{x}_{int}}}{\sum_j e^{\boldsymbol{\beta}' \mathbf{x}_{jnt}}}) f(\boldsymbol{\beta}) d\boldsymbol{\beta} \tag{6}$$

For estimation, a distribution is specified for the individual-specific coefficients and the parameters of this distribution are estimated. Because the log-likelihood does not take on a form that can be solved analytically, the mixed logit model is estimated using maximum simulated likelihood. Similar to the probability of one choice period, the probability is simulated. First, a draw of $\boldsymbol{\beta}$ is taken from its distribution. Then, the logit formula is calculated for each period, and finally the product of the obtained logits is taken. This process is repeated for R draws and the results are averaged over the draws by the following formula

$$LL(\boldsymbol{\theta}) = \sum_{n=1}^{N} \ln[\frac{1}{R} \sum_{r=1}^{R} (\prod_t P_n(y_{int})] \tag{7}$$

where $\boldsymbol{\theta}$ is the vector of coefficients to be estimated and $y_{nit} = 1$ if person n choses i and zero otherwise

6 Experimental design

In this section the experimental design of the empirical problem is given.

6.1 Model formulation

As explained in Section 4, the first and second class subscriptions are separated and two different choice sets are defined. In addition, the business and student subscriptions are excluded from the choice set. Since the monthly first class subscriptions are never chosen in the sample these are excluded from the choice set. For the first class travelers the choice set has seven possible subscriptions and for the second class travelers ten subscriptions. For both classes, the subscriptions are all available to everyone and thus the choice set is independent of the individuals, $C_n = C$. The first and second class choice problems only differ in their choice set and their utility functions and models are identical, therefore the formulation given below holds for both choice problems.

The observed component $V(\mathbf{x}_{in})$ is written as a function of the vector of explanatory variables. The explanatory variables are split in two groups $\mathbf{x}_{in} = (\mathbf{x}_{1,in}, \mathbf{x}_{2,in})$, $\mathbf{x}_{1,in}$ are the travel demand variables and $\mathbf{x}_{2,in}$ the intercept, realized savings and individual characteristics are the non-travel demand variables. As aforementioned, the observed utility is modeled by a linear relation of attributes and coefficients. The travel demand variables, $\mathbf{x}_{1,in}$, are log transformed and the other variables, $\mathbf{x}_{2,in}$, are scaled. Then, two different formulations will be used; (I) the realized and historic travel demand is choice-dependent and (II) the realized and historic travel demand is, the total travel demand, not choice-dependent. A general formulation of both model formulations are given below and a more elaborate formulation of both models are given in the Appendix.

Subscription based travel demand

For formulation (I), we use the subscription based travel demand, explained in 3.3. Since the subscriptions differ in the moment of travel, the subscription-specific travel demand for that certain choice is used. For example, for the OFF-PEAK FREE subscription only the travel demand in the off-peak moments[8] for which the free traveling applies to is used as explanatory variable.

For the choices $j = 1, ..., J$ the following observed component of the utility function is given

$$V(\mathbf{x}_{nj}) = \boldsymbol{\beta}_1' \mathbf{x}_{1,nj} + \boldsymbol{\beta}_{2,j}' \mathbf{x}_{2,n} \tag{8}$$

Since the explanatory variables $\mathbf{x}_{1,nj}$ (historic and realized travel demand) differ per choice alternative, a fixed coefficient, $\boldsymbol{\beta}_1$ across the choice alternatives is used. The explanatory variables $\mathbf{x}_{2,n}$ do not differ per choice and are therefore have a coefficient, $\boldsymbol{\beta}_{2,j}$ that is different for every subscription, j. In addition, only differences in utility are important and the scale of utility is not important, thus the coefficients, $\boldsymbol{\beta}_{2,j}$ for $j = 1, ..., J-1$ are relative to choice alternative J for which the coefficients are non-random and fixed to 0. Here choice alternative J is the non-subscription or FULL TARIFF option.

Total travel demand

For the total travel demand no split per moment is made, corresponding to formulation (II). The historic and realized travel demand is explained by the total distance traveled, hence the two explanatory variables do not differ per choice alternative.

[8] the off-peak hours are the off-peak hours during the week including the weekend

For the choices $j = 1, ..., J$ the following observed component of the utility function is given

$$V(\mathbf{x}_{nj}) = \boldsymbol{\beta}'_{1,j}\mathbf{x}_{1,n} + \boldsymbol{\beta}'_{2,j}\mathbf{x}_{2,n} \tag{9}$$

the explanatory variables $\mathbf{x}_{1,j}$ and $\mathbf{x}_{2,n}$ do not differ per choice alternative and choice specific coefficients are used. Again, utility is relative and thus, $\boldsymbol{\beta}_{1,j}, \boldsymbol{\beta}_{2,j}$ for $j = 1, ..., J-1$ are relative to choice alternative J similar to formulation (I).

As explained in Section 5 the unobserved component has an i.i.d. Gumbel distribution and the IIA property holds. In addition, for estimation it is assumed that each decision maker's choice is independent of that of other decision makers. In this case it is assumed that each *decision moment* is independent of other decision moments and thus every decision of a decision maker is independent of the previous and future decision.

6.1.1 Choice hierarchy

From inspecting the different subscription options, the assumption that the choice alternatives are independent is not probable. The subscription packages have similar characteristics and it can be argued that there is a hierarchy in choice. Two similar characteristics that can be observed are type of discount and the type of moment dependence. Relaxing this independence of choice alternatives assumption leads to a nested logit formulation. The different definitions used for the nested logit are given in Figures 5, 6, 7. The first layer are the nests and the second layer are the subscriptions. The grey shaded subscriptions are excluded from the first class choice set.

Moment dependent nests

The model with moment dependent nests is given in Figure 5. The five nests differ in the moment the subscription applies to, but also trajectory subscriptions are constructed as a nest and the full-tariff subscription is a nest. The FULL TARIFF subscription is a separate nest, since it is not a description and does not have similar characteristics with the other subscriptions. The "traject" nest has similar trajectory characteristic, and the subscriptions only differ in the duration of the subscription. The other nests, "weekend", "off-peak", "always" are constructed with similar reasoning. The subscriptions within the nests differ on subscription duration and type of discount.

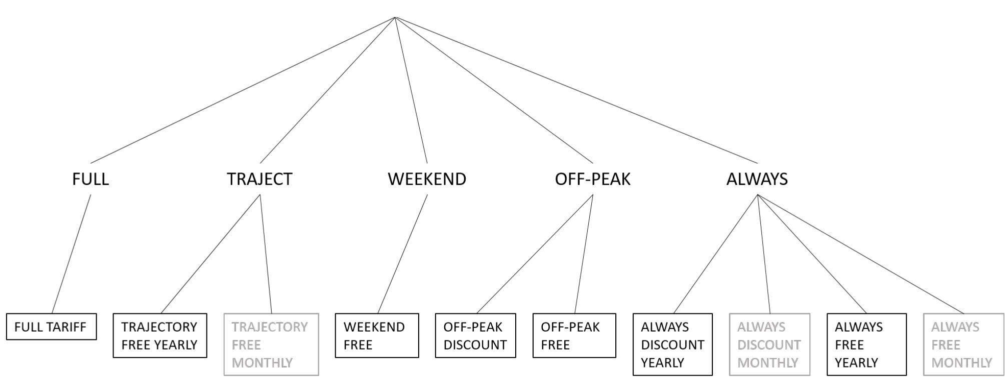


Figure 5: Nested logit model with type of moment nest

Type of discount nests

The model with discount dependent nests is given in Figure 6. The three nests constructed differ in the discount given, corresponding to the subscription, namely no discount, discount or full discount. Within the nests the subscriptions differ on subscription duration, trajectory dependence and moment dependence.

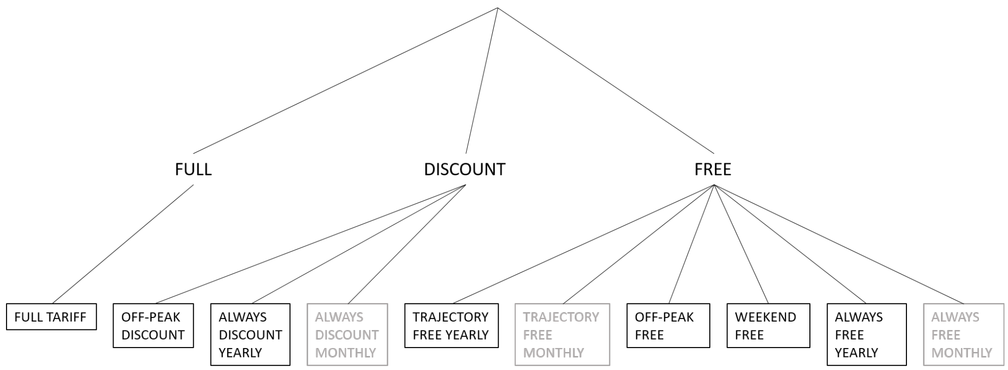


Figure 6: Nested logit model with type of discount nest

Cross nested logit

A cross nested logit model (CNL), with a mix of the characteristics of the two nested logit models given above, is constructed. The cross nested logit is a more flexible model than the standard nested logit and can capture more correlation patterns in this choice process. As an argument to use the cross nested logit instead of multi-level nested logit is that with the multi-level nested logit, the full correlation can only be accommodated along the highest dimension of nesting in the tree, an issue that was first discussed by Hess and Polak (2006). The CNL structure allows for the joint representation of correlation along the three choice levels, without the requirements of a multi-level nesting structure. In general, it is not necessary to use a multiple level model, as you can always use a cross nested model instead.

The moment dependent nests have been split into two groups, moment dependent, *NON-ALWAYS* and not moment dependent, *ALWAYS*. The other two nests separate free traveling subscriptions from non-free traveling subscriptions, *FREE* and *NON-FREE*, respectively. Every subscription can be allocated to two nests corresponding to the characteristics of the subscription. A full overview of the possible nest allocation of the choice set is given in Figure 7.

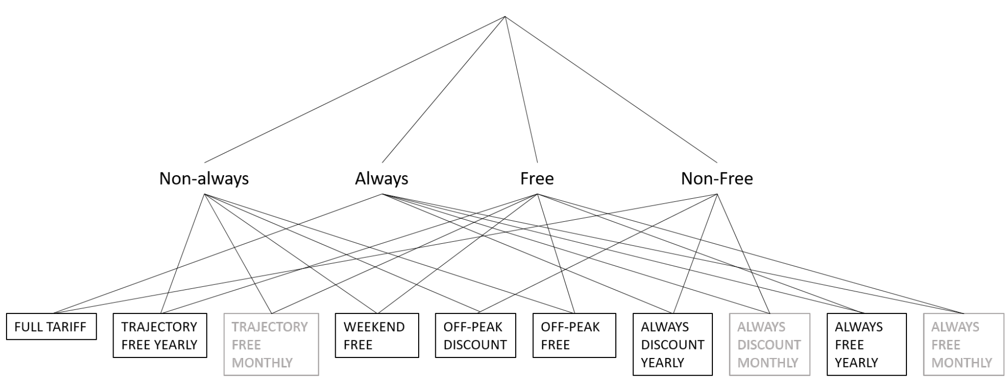


Figure 7: Cross nested logit model

6.1.2 Time dimension

In the setting of this research problem numerous choices made by each decision maker are observed. The smart card data, as mentioned in Section 3.4 is panel data. For the GEV models (MNL,NL and CNL) it is assumed that each *decision moment* and/or *decision maker* is independent between observations. If the unobserved factors that affect decision makers are independent over repeated choices, then logit can be used in the same way as purely cross-sectional data, meaning that panel data can be pooled for estimation. There are NT_n observations, for first class subscriptions 2418 observations and second class subscriptions 5120 observations.

However, the assumption that each decision moment and/or decision maker is independent between observations is a strong assumption. It makes sense to substantiate that, choices made by the same individual are trivially correlated, but two individuals making choices at the same time are non-trivially correlated.

For the GEV models all the observations are assumed to be independent, but for the mixed logit model it is possible that decisions made by the same individual have similar characteristics and are more dependent than decisions made by other individuals. However, the decisions made by the individual, thus the decision moments of that individual, are independent. Stating succinctly, the previous decision of an individual does not influence the future decision of that individual.

Mixed logit model formulation

The model formulation for the mixed logit is similar as in Equation 8, the differences are mentioned on. The time dimension, $t = 1, ..., T_n$ is individual specific and added. Then, for the choices $j = 1, ..., J$ the following observed component of the utility function is given

$$U(\mathbf{x}_{njt}) = \boldsymbol{\beta}'_{1,n}\mathbf{x}_{1,njt} + \boldsymbol{\beta}'_{2,j}\mathbf{x}_{2,nt} + \epsilon_{njt} \tag{10}$$

the random error term is assumed to be i.i.d. Gumbel distributed. The coefficients of the travel demand variables will be varying over individuals. As mentioned in Section 5, a distribution for $\boldsymbol{\beta}_1$ has to be specified. As in most applications (Revelt and Train 1998, Mehndiratta 1996 and Ben-Akiva and Bolduc 1996) $f(\boldsymbol{\beta})$ is specified to be normal: $\boldsymbol{\beta}_1 \sim N(\mathbf{b}, \mathbf{W})$ with the parameters $\mathbf{b}$ and $\mathbf{W}$ estimated.

6.2 Experimental setup

To answer the research questions, formulated in Section 1, the following experimental setup is constructed. Before answering *RQ1-RQ3*, the correct historic lag function has to be found to define the historic travel demand. Thus, formulation (I) is estimated with the MNL with a lag function for one, two, three and four quarters (Lag 1-4). The chosen lag function will subsequently be used as indicator for the historic travel demand. Formulation (I) and (II) are estimated with the MNL and analysed. Then, the expected hierarchy in choice is explored by estimating formulation (I) with the two nested logit models and the cross nested logit. Using these results and the results of estimating Equation 10 with the mixed logit the differences across consumers are explored to answer *RQ3*. *RQ1* is answered with the results of the MNL with formulation (I) and (II), and these results are compared to the results that model the hierarchy in choice. The second question, *RQ2*, is answered with two models that fit the decision process best. The coefficients of the realized savings variable of these two models are compared.

7 Empirical Results

In this section, we will elaborate on the results using the experimental setup given in the previous section. From the previous section two model formulations are given. Formulation (I) and (II) are estimated by the multinomial logit; and formulation (I) is estimated by the nested logit and cross nested logit. The mixed logit is estimated with Equation 10. Estimation of the models is done by a software packages made for choice models, Biogeme[9]. The first part of this section will elaborate on the travel demand as explanatory variable answering *RQ1*, the second part elaborates on the hierarchy in the choice process, the third part explores the differences across consumers and answers *RQ3* and the last part of this section compares the coefficients of the realized savings to answer *RQ2*.

The most important tables to answer the research questions are given in this section. The goodness of fit statistics given by Biogeme are given in Table 7. The formulas used to calculate the goodness of fit statistics are given in the Appendix. The standard error, $\hat{\sigma}_k$, is given in the parentheses[10].

7.1 Travel demand

The first part of this section elaborates on the results of formulation (I), subscription based travel demand, and selects the *historic* level of travel demand. The second part expands on the results of formulation (II), total travel demand, and the differences in travel demand between subscriptions. Both parts elaborate on the influence of travel demand on subscription choice and the differences between *realized* and *historic* travel demand. These results utilized to answer *RQ1*. Note, both models formulations are estimated by the multinomial logit model.

7.1.1 Subscription based demand

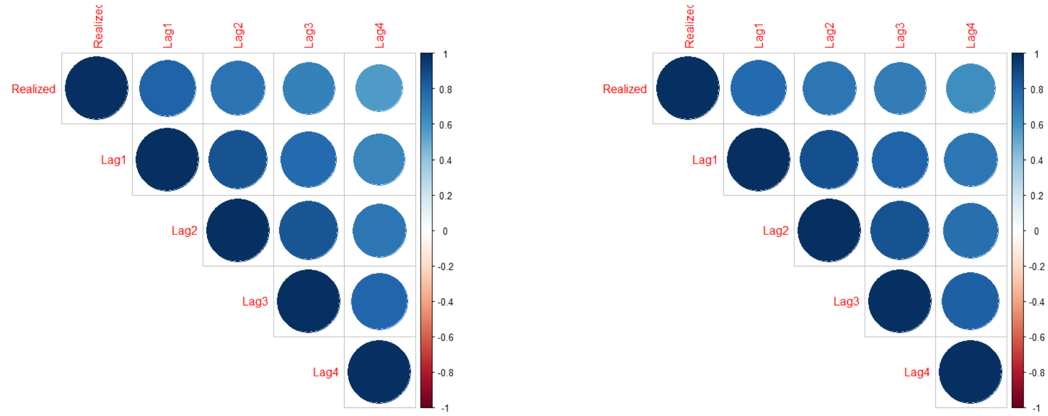
Section 4 elaborates on the choice attributes and the influence of experience and memory on the decision process. This process is translated in a lagged variable containing historic information. However, the most informative lag function, one, two three, or four quarters has to be found. Thus, the multinomial logit model is estimated for different levels of historic lags using subscription based travel demand.

In Figure 8 correlation plots are given for the *realized* and *historic* travel demand for the second and first class choice set. For estimation, subscription based travel demand is used, but for the correlation plot total travel demand is visualized as an indication. The correlation between the *realized* travel demand and the four different lag functions of *historic* travel demand is positive and very strong. Thus, travel demand is quite consistent over the lags and therefore over time. In addition, the higher the historic level of the lag, the lower the correlation. Stated differently, historic information on a year back is the least correlated with the *realized* travel demand. There are no differences in the correlation plots between first and second class thus the above holds for both choice sets.

The coefficients with their confidence interval for the *historic* and *realized* travel demand are plotted against the different levels of historic lag and are given in Figure 9. The y-axes with the coefficient size of both plots differ, the second class y-axis is smaller than the first class y-axis. The x-axes are equal and the higher the x-axis, the higher the level of historic lag. Also, the plot shows

[9]Biogeme is a open source Python package designed for the maximum likelihood estimation of parametric models in general, with a special emphasis on discrete choice models

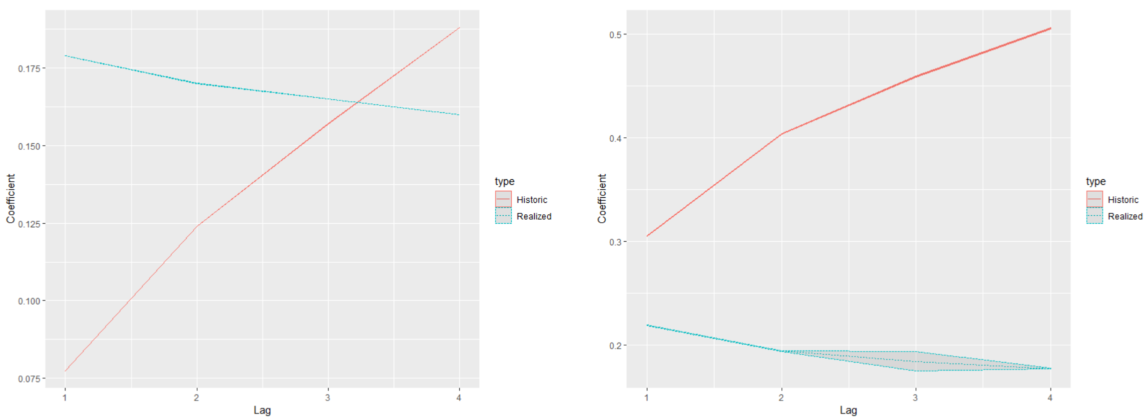
[10]The standard error is calculated as the square rooth of the kth diagonal entry of the Rao-Cramer bound



(a) Second class (b) First class

Figure 8: Correlation plot of realized and historic total travel demand

the confidence intervals of the coefficients, however, these intervals are tight and almost not visible in the graph, hence all the coefficients are significant at the 5% significance level. Thus, statistic evidence is found that travel demand does explain subscription choice.



(a) Second class (b) First class

Figure 9: Plot of realized and historic coefficient size (and confidence interval) for different lags

From Figure 9, it can be seen that there are clear differences between the sizes of the coefficients between the first and second class choice set. However, there are no large differences between the two choice sets for the *realized* travel demand. In addition, for both choice sets, the higher the historic level, the larger the influence of the *historic* travel demand and the smaller the influence of the *realized* travel demand. One explanation is that the higher the level of historic lag, the lower the correlation and the more variations are captured by both the *realized* and *historic* travel demand. Note that over the different lag functions the coefficients of the other attributes did not change.

In Table 8, 9 in the Appendix goodness of fit measures are given for the estimated models. The higher the level of historic lag, the higher the log likelihood and the rho-square bar, for both the first and second class models. Hence, using the arguments mentioned above we choose a lag function of one year. For further estimation, the level of historic information used is a year (Lag 4) for both the first and second class choice set.

Also, from Figure 9 it can be seen that, for the second class choice set, the *historic* travel demand coefficient ranges from 0.0774 to 0.188, which is a difference of 0.116 and for the first class the coefficient is from 0.305 to 0.506, a difference of 0.201. Thus, the influence of historic information and the influence of an increase of historic information is larger for first class travelers than for second class travelers, even though the correlation plot between first and second class doesn't show any differences. Also, for second class travelers the influence of historic information is smaller than realized information when less historic information is available, but when more than a year of historic information is available, the influence of the historic information is bigger than the information on *realized* travel demand.

As explained in Section 4.3, *realized* travel demand matches a perfect prediction of travel demand for individuals, however, in the decision problem it is the expected travel demand. For first class subscriptions, the influence of historic information is always larger than the influence of *realized* travel demand. The result that historic information on travel demand receives a higher weight than expected travel demand, can be explained by the fact that, individuals are more certain about their past behaviour than their expected behaviour. This effect shows to be stronger for first class travelers than for second class travelers.

7.1.2 Total demand

In this section the results of formulation (II) are discussed. Formulation (II) differs from formulation (I) in explanatory variable, see Section 6.1. The coefficients of total travel demand are choice-specific and therefore differ per subscription, thus the differences in travel demand between subscriptions can be evaluated. The results, the coefficients, standard error (in parentheses) and goodness of fit statistics, are given in Table 3. The coefficients are all with respect to the FULL TARIFF/no subscription package. First the results of the second class subscriptions are discussed and then the results of the first class subscriptions.

Second class

The results of the second class choice set show that the *realized* and *historic* travel demand coefficients of formulation (II) are significant and positive. Thus, an increase in total travel demand increases the probability of choosing a subscription other than the FULL TARIFF subscription. This is an expected result, since the more kilometers an individual travels the more useful a subscription becomes.

The coefficients of *realized* travel demand is the largest for the ALWAYS FREE MONTHLY subscription, 8.9, and the smallest for the OFF-PEAK DISCOUNT subscription, 0.276. This result corresponds to the subscriptions with the highest and lowest monthly fixed fee. For ALWAYS FREE YEARLY, *historic* travel demand has the largest influence, 4.6, on subscription choice, which is the second most expensive subscription. One could say that the higher the price of the subscription, the higher the influence of the total travel demand. Yet, the size of the coefficients are not sorted fully similar to the price of the subscription. In conclusion, if the subscription has a high fixed fee, travel demand (*historic* and *realized*) has a large influence on the subscription choice. However, this result is not the other way around for lower priced subscriptions.

The result of the previous section for formulation (I) shows a larger influence in *historic* demand than *realized* demand, but this difference in coefficient is small. Different from this result, the influence of *historic* travel demand is not consequently larger than *realized* travel demand. For about half of the subscriptions the *historic* travel demand has more influence on the decision.

		Second class	First class
Realized	ALWAYS DISCOUNT	0.692	0.726
	YEARLY	(0.103)	(0.22)
	ALWAYS DISCOUNT	1.97	
	MONTHLY	(0.307)	
	ALWAYS FREE	4.03	3.61
	YEARLY	(0.324)	(0.323)
	ALWAYS FREE	8.9	
	MONTHLY	(0.995)	
	OFF-PEAK DISCOUNT	0.276	0.0707
		(0.0373)	(0.113)
	OFF-PEAK FREE	0.3	0.162
		(0.085)	(0.113)
	TRAJECTORY FREE	1.19	2.3
	YEARLY	(0.113)	(0.355)
	TRAJECTORY FREE	0.756	
	MONTHLY	(0.236)	
	WEEKEND FREE	4.6	-0.0524
		(0.319)	(0.101)
Historic	ALWAYS DISCOUNT	1.19	0.513
	YEARLY	(0.379)	(0.197)
	ALWAYS DISCOUNT	0.756	
	MONTHLY	(0.0353)	
	ALWAYS FREE	4.6	3.36
	YEARLY	(0.121)	(0.341)
	ALWAYS FREE	2.35	
	MONTHLY	(0.12)	
	OFF-PEAK DISCOUNT	0.278	-0.117
		(0.149)	(0.101)
	OFF-PEAK FREE	1.51	1.97
		(0.0782)	(0.138)
	TRAJECTORY FREE	1.34	2.08
	YEARLY	(0.129)	(0.38)
	TRAJECTORY FREE	0.275	
	MONTHLY	(0.271)	
	WEEKEND FREE	0.764	0.429
		(0.336)	(0.106)
Parameters		50	40
Rho-square bar		0.365	0.482

All coefficients are significant at the 1% level

Table 3: Coefficient table of total travel demand for the MNL

These subscriptions (ALWAYS DISCOUNT YEARLY, ALWAYS FREE YEARLY, OFF-PEAK DISCOUNT, OFF-PEAK FREE and TRAJECTORY FREE YEARLY) are all the yearly subscriptions with an exception of the WEEKEND FREE subscription. Hence, for subscriptions with a duration of a year, information on travel demand dating a year back is more informative than expected travel demand that upcoming quarter. This is probable, since these subscriptions have a duration of a year, thus the decision for the upcoming quarter(s) also depends on past the quarter(s), depending on the start date of the subscription. If second class travelers have to make a long term decision, they rather look at their past behaviour than at their expected behaviour and, for a short term decision they rather look at their expected behaviour than at their past behaviour. As aforementioned, an exception on the above statement are individuals who choose the WEEKEND FREE subscription, they tend to look at their expected behaviour instead of their past behaviour, even though they make a long term decision. Since WEEKEND FREE is specific for the weekend, it could be that individuals differently consider their total travel demand, *realized* and *historic*.

First class

Different from the second class subscriptions, the travel demand coefficients of first class subscriptions are not all positive. An increase of *realized* travel demand does not increase the probability of choosing a WEEKEND FREE subscription compared to the FULL TARIFF subscription, which is an unexpected result. It could be that individuals are not very well at predicting their total expected travel demand for the WEEKEND FREE subscription and give more weight on their past travel demand. In addition, an increase of *historic* travel demand does not increase the probability of choosing an OFF-PEAK DISCOUNT subscription compared to a FULL TARIFF subscription. Again, for OFF-PEAK DISCOUNT subscriptions individuals could look more at their *realized* total travel demand than their *historic* total travel demand. For the other first class subscriptions an increase in total travel demand (*realized* and *historic*) increases the probability in choosing this subscription compared to the FULL TARIFF subscription, as expected.

Similar to the second class subscriptions the coefficient size of the most expensive subscription, ALWAYS FREE YEARLY, is the largest for total *realized* and *historic* travel demand, 3.61 and 3.36, respectively. Thus, for the subscription with the highest fixed fee, total travel demand has the highest influence on the probability of choosing this subscription. Note that this subscription also has the smallest intercept (see Table 9 in the Appendix). Similar to the second class subscriptions, the size of the coefficient are not sorted the same as the price of the subscription.

Even though the results of subscription based travel demand, formulation (I), demonstrate that *historic* travel demand has a larger influence on utility than *realized* travel demand, this result does not hold for total travel demand, formulation (II). Just as with the second class subscriptions the difference in impact of *realized* and *historic* travel demand differs per subscription. However, different from second class subscriptions no pattern can be explored in the type of subscription. Note that no monthly subscriptions are part of the first class subscription choice set.

Overall the impact of the total travel demand on subscription choice resembles that of the second class subscriptions, but the results are mixed.

7.2 Hierarchy in choice

As explained in Section 6.1.1, it is not likely that the choice alternatives are independent, since certain choice alternatives share similar characteristics. Therefore, problem formulation (I) is estimated for two nested formulations (moment and discount nests, Figure 5 and 6, respectively) and the cross nested formulation (Figure 7). The results of the estimated coefficients of the choice attributes for the first and second class choice set are given in the Appendix in Table 10 and 11. The estimated correlation coefficients of the models are presented in Table 4. Since some of the defined nests only contain one subscription package, the correlation coefficient is fixed to 1 (no correlation), as specified in the Table. Significance in the table indicates that the estimated correlation coefficient is significantly different from 1. As explained in Section 5.3.1 if one cannot reject the hypothesis that all correlation parameters are equal to 1, the nested logit reduces to the standard logit. Thus, if one of the correlation parameters is statistically different from 1, then statistical evidence is found for a nested model structure.

Second Class		Estimated				Fixed		
NL	Moment	Traject	Off-Peak	Always		Full	Weekend	
		1	1.61***	31.6		1	1	
		(-)	(0.0261)	(577)				
	Discount	Discount	Free			Full		
		14.7***	50.1***			1		
		(0.001)	(4.38)					
CNL	-	Non-always	Always	Free	Non-Free			
		1.08***	1	22.9***	22.2***			
		(0.185)	(-)	(2.36)	(1.68)			

First Class		Estimated				Fixed		
NL	Moment	Off-Peak	Always			Full	Traject	Weekend
		1	1			1	1	1
		(0.391)	(0.176)					
	Discount	Discount	Free			Full		
		1	2.15			1		
		(0.21)	(-)					
CNL	-	Non-always	Always	Free	Non-Free			
		4.1***	30.1***	24.3	43.7*			
		(0.0012)	(6.09)	(-)	(15.5)			

*p<0.1; **p<0.05; ***p<0.01; Standard error in parentheses; (-) std. error $> 2 * 10^6$

Table 4: Correlation coefficients of the NL and CNL

Second class

Inspecting the second class correlation coefficients of the three models, we find significant results for the correlation coefficients, which is evidence for a nested model structure. For the nested logit with "moment" nests there is a significant nest for off-peak subscriptions, namely OFF-PEAK DISCOUNT, OFF-PEAK FREE. However, no statistical evidence is found that the correlation coefficient of the trajectory and always nests are different from 1.

For the nested logit with the "discount" nests, the coefficients are significantly different from 1, at the 1% significance level. The subscriptions within the discount and free nest are correlated and therefore share similar characteristics. Hence, from the correlation coefficients, evidence is found for a nested structure for both "moment" and "discount" dependent nests.

Even though there is evidence for a hierarchy of choices within the decision process of subscription packages, the travel demand coefficients of the two nested logit models are not significant (Table 10 in the Appendix). This result is unexpected, since the travel demand coefficients are expected to be an important explanatory variable in the decision process. Although there is evidence for a nested structure in the discrete choice model, the non-significant travel demand coefficients are not reasonable and therefore we expect that these two nested logit models can be improved.

The correlation coefficients of the cross nested logit model also show evidence for a nested structure. As with the nested logit with "moment" nests, there is a partly significant result for correlation parameter of the moment dependent nests, namely for the non-always nest. The correlation parameters for the discount dependent nests are significant. Different from the two nested logit models the coefficients of the travel demand variables are positive and significant (Table 10), in line with the expectations. The size of the coefficients are similar to the multinomial logit coefficients (Table 8). The non-travel demand coefficients, $\boldsymbol{\beta}_2$, show significant and coherent results. Hence, the conclusions of the previous section remain.

In Table 5, the allocation parameters of the subscriptions in the predefined nests of the cross nested logit are given. For the second class subscriptions not all allocation parameters are significant at the 10% level or less. The parameters for the ALWAYS DISCOUNT YEARLY and OFF-PEAK DISCOUNT show that these subscriptions are cross allocated in two nests, which is evidence for a cross nested logit model. However, the allocation parameters of OFF-PEAK FREE and TRAJECTORY FREE MONTHLY give an indication of full allocation to one nest, the non-always and free nest, respectively. In addition, the two subscriptions that are similar in the moment and discount nest, but differ in the duration of the subscription, show mixed results. For the three cases where this occurs (ALWAYS DISCOUNT, ALWAYS FREE and TRAJECTORY FREE) only the ALWAYS FREE subscriptions show similar results. Overall the results of the allocation parameters of the cross nested logit model show mixed nest memberships of the second class subscriptions.

Table 7 shows goodness of fit measures of the nested logit models for moment and discount nests and for the cross nested logit. From the two nested logits, the model with the discount nests has the best fit, nevertheless, the cross nested logit model has better goodness of fit results than the nested logit models. The cross nested logit has a rho-square bar of 0.339, which is close to the rho-square bar of the multinomial logit of formulation (II), namely 0.365. Thus, a flexible nested structure (CNL) best explains subscription choices by travel demand, however, no clear allocation pattern across nests is observed.

First class

The first class correlation coefficients of the two nested logit models and the cross nested logit model are given in Table 4. The estimated coefficients of the nested logit don't show statistical evidence for a nested structure. The "moment" nests are not different from 1, where 1 indicates correlation. Inspecting the correlation parameters of the "discount" nested logit, there also is no evidence for shared characteristics within the nests. The estimated correlation coefficients are both 1 or not significantly different from 1.

In Table 11 in the Appendix the coefficients are given. Different from the second class nested models the travel demand coefficients are significant and similar to the multinomial logit formulation. This is in agreement with the theory that if correlation coefficients of the nested logit do not

Second class

	Non-Always	Always	Free	Non-free	Std. Error	Sign. level
ALWAYS DISCOUNT YEARLY		0.603		0.397	(0.244)	***
ALWAYS DISCOUNT MONTHLY		0.982		0.018	(-)	
ALWAYS FREE YEARLY		0.00249	0.99751		(0.00264)	
ALWAYS FREE MONTHLY		0.964	0.036		(-)	
OFF-PEAK DISCOUNT	0.553			0.447	(0.018)	***
OFF-PEAK FREE	1		0		(0.228)	***
TRAJECTORY FREE YEARLY	0.00976		0.99024		(0.0101)	
TRAJECTORY FREE MONTHLY	0.0991		0.9009		(0.00679)	***
WEEKEND FREE	0.0147		0.9853		(0.0156)	

First class

	Non-Always	Always	Free	Non-free	Std. Error	Sign. level
ALWAYS DISCOUNT YEARLY		0.847		0.153	(0.0233)	***
ALWAYS FREE YEARLY		0.788	0.212		(5.8)	***
OFF-PEAK DISCOUNT	0.86			0.14	(0.071)	***
OFF-PEAK FREE	0.999		0.001		(-)	
TRAJECTORY FREE YEARLY	0.378		0.622		(0.0367)	***
WEEKEND FREE	0.999		0.001		(-)	

*p<0.1; **p<0.05; ***p<0.01; (-) std. error $> 2 * 10^6$

Table 5: Allocation parameters for the CNL

show evidence for correlation the model approaches the multinomial logit model.

Different from the nested logit results, the cross nested logit does show significant results for a nested structure. The possibility for subscriptions to be allocated among nests gives significant results and no allocation among nests does not show significant results. Thus, the subscriptions have similar characteristics, but this is not explained through the moment and discount nested structures.

If we look at the allocation parameters in Table 5, most of the allocation parameters are significant at a 1% level. The subscriptions ALWAYS DISCOUNT, ALWAYS FREE and TRAJECTORY FREE have significant allocation parameters and are indeed subdivided among the nests. Even though both the ALWAYS FREE/DISCOUNT subscriptions are mostly allocated to the "always" nest, the subdivision between the type of discount is needed to find significant results, because in the "moment" nested logit model both these subscriptions are fully allocated to the "always" nests and no significant result is found there. Similar arguments can be used for the "non-free" nest containing two subscriptions that are also allocated over the "non-always" and "always" nests. Thus, for first class subscriptions the cross nested logit model demonstrates to be more appropriate than the nested logit models.

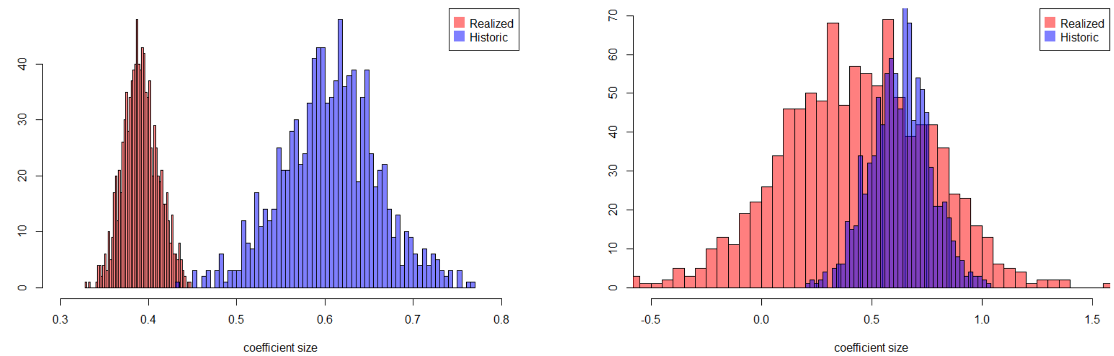
The coefficients of the cross nested logit (Table 11) are different in sign and size than the coefficients of the nested logit. Together with the evidence found for a cross nested logit structure, it is important to take into account the hierarchy of choices when modeling the subscription choice by travel demand. In addition, if we look at the goodness of fit of the cross nested logit, it has the highest log likelihood and rho-square bar compared to the multinomial logit and the nested logit. With a rho-squared bar of 0.515 the cross nested logit best explains the first class subscription choices of individuals.

7.3 Allowing for heterogeneity

The mixed logit model formulation allows for heterogeneity in the travel demand coefficients, see Section 6.1.2 for an elaborate explanation, and is utilized to answer *RQ3*. Stated in short, with the mixed logit formulation decisions made by the same individual with the same preferences and taste are modeled. If there are differences between the individuals in preferences and taste this is captured by the heterogeneous parameter. Appendix tables 8, 9 show the coefficients and Table 7 goodness of fit statistics of the ML model for second and first class, respectively. Further, with the mixed logit model the IIA property holds and no assumption has to be made on the independence of choice alternatives.

In Figure 10 the histograms of the individual-specific coefficients of the mixed logit model are given. This is an indication of the estimated parameters explaining the shape of the heterogeneous coefficients, **b** and **W** for the *historic* and *realized* travel demand. Since both coefficients are assumed to be distributed as the standard normal, the shape of the histograms is normal. For the second class subscriptions the mean of the distribution is 0.391 and 0.607 and the variance is 0.0193 and 0.0532 for *realized* and *historic* travel demand, respectively. From the figure, the influence of the *realized* travel demand is smaller, but there are less differences across individuals compared to the *historic* travel demand coefficients. Thus, individuals look more at their past travel demand than their expected travel demand, which is in line with the results found in the previous section. In conclusion, there are less differences between individuals in *realized* travel demand than in *historic* travel demand and the influence of *realized* travel demand is smaller than the *historic* travel demand.

For first class subscriptions the mean of the distributions is 0.422 and 0.634 and the variance 0.331



(a) Second class

(b) First class

Figure 10: Histograms of individual-specific coefficients of ML model

Note that the scale of the coefficient size is different for first and second class subscriptions

and 0.132 for *realized* and *historic* travel demand, respectively. The mean of the distribution is similar for second and first class subscriptions. Nevertheless, the variance is larger for the first class subscriptions than for the second class subscriptions. First class individuals show more differences across individuals regarding the influence of travel demand on subscription choice. These differences are larger for *realized* than *historic* travel demand. In addition, note that the *realized* travel demand can be negative and larger than the *historic* travel demand for certain individuals. Thus, there are more differences across individuals that travel first class than individuals that travel second class. An explanation could be that first class travelers have less awareness on their past and expected travel demand, since they are expected to have a stronger financial position (Section 4.2).

The goodness of fit of the mixed logit is worse than the multinomial logit (formulation (I)) and the cross nested logit if we look at the log likelihood and the rho-square bar for both second and first class choice sets. Despite the fact that the mixed logit is a more flexible model and that for first class subscriptions there is evidence for differences across individuals, the fit did not increase. One of the possible explanations is that there is a large sample of individuals and a small number of observation over time, which could lead to a worse fit of the mixed logit.

7.4 Comparison of realized savings

In this Section, we will investigate the economic incentive of individuals by inspecting the coefficients of the realized savings variable, for a more comprehensive explanation see Section 4.3. The higher the realized savings variable, the better the choice economically. One of the research questions, *RQ2* about the economic incentive of individuals when choosing their subscriptions, will be answered in this section. Note that the realized savings variable is the savings after a decision has been made on the subscription, thus this variable is also an indication if individuals can predict their savings from traveling.

Table 6 contains the coefficients for the realized saving variable for the cross nested logit and the mixed logit. Only the cross nested logit and mixed logit results are given, because in the previous section evidence is found that the choice alternatives are not independent and the multinomial logit does assume IIA. The coefficients are all with respect to the FULL TARIFF/no subscription package. In the table, the subscriptions are sorted from the subscriptions with the lowest fixed fee (per month) to the highest. Note though, the TRAJECTORY subscriptions can be less expensive

	Second class		First class	
	CNL	ML	CNL	ML
OFF-PEAK DISCOUNT	2.14	-0.348	1.38	0.722
ALWAYS DISCOUNT YEARLY	4.55	2.8	0.462	1.8
ALWAYS DISCOUNT MONTHLY	5.9	3.61		
WEEKEND FREE	4.28	3.67	2.14^{+}	2.92
OFF-PEAK FREE	6.47	4.79	2.04^{+}	3
TRAJECTORY FREE YEARLY	4.39	2.87	0.851	1.06^{+}
TRAJECTORY FREE MONTHLY	-0.771^{+}	-0.438^{+}		
ALWAYS FREE YEARLY	4.28	3.62	0.606	1.98
ALWAYS FREE MONTHLY	-1.53	-1.24		

$^{+}$ not significant at the 5% significance level

Table 6: Coefficient table of the realized savings attribute for the CNL and ML

than the OFF-PEAK FREE subscriptions, since it depends on the trajectory the subscription applies to. All the coefficients in the table are significant at the 5% significance level, unless stated otherwise. First the second class choice set will be discussed and then the first class choice set.

For the second class choice set the coefficients are similar in sign, except two subscriptions. The positive coefficients indicate, the higher the realized savings, the higher the observed utility of the choice alternative and the higher the probability of choosing this subscription with respect to the FULL TARIFF subscription. This relation indicates that individuals have an economic incentive and that they are able to predict their expected savings. Even though the literature stated that individuals are not always able to predict their expected expenses, in this specific empirical problem this does not hold (Section 2). One significant exception on this relation is the most expensive subscription, namely ALWAYS FREE MONTHLY. Since the subscription is monthly, we supposed that individuals were better in predicting their expected savings compared to a yearly subscription. Nevertheless, we explain this negative sign to individuals not being able to correctly predict their savings.

Next to the economic incentive of individuals, when choosing their subscriptions, we supposed that the higher the price of the subscription, the higher the risk of choosing this subscription. However, this increased risk does not indicate a higher expected return or savings from travelling. One of the reasons could be the difference in the moment dependence of the subscriptions. Or the differences between type of customers, where similar customers choose similar subscriptions.

In Table 6, all the first class subscriptions have positive coefficients, note that some subscriptions are not significant at the 5% level. With respect to the FULL TARIFF subscription, where realized savings is 0, the higher the predicted savings the higher probability of choosing a subscription other

Second class	MNL (I)	MNL (II)	NL Moment	NL Discount	CNL	ML
Number of parameters	32	50	35	34	47	33
Sample size	5120	5120	5120	5120	5120	3864
Log likelihood	-9047.21	-7740.62	-9140.61	-8839.47	-8071.99	-8879.95
Rho-square bar	0.26	0.365	0.255	0.277	0.339	0.246
AIC	18158	15581	18286	17747	16238	17826
BIC	18367	15908	18317	17969	16545	18032

First class	MNL (I)	MNL (II)	NL Moment	NL Discount	CNL	ML
Number of parameters	26	40	28	28	39	27
Sample size	2418	2418	2418	2418	2418	521
Log likelihood	-2857.42	-2710.85	-2857.42	-2815.31	-2538.89	-2311.73
Rho-square bar	0.457	0.482	0.457	0.465	0.515	0.41
AIC	5767	5501	5771	5687	5156	4677
BIC	5917	5733	5933	5848	5382	4792

Table 7: Goodness of fit statistics

than the FULL TARIFF subscription. For first class individuals, there is an economic incentive when choosing a subscription. This statistical evidence is stronger for first class travelers than for second class travelers. However, just as with the second class subscriptions, it is not that the higher the price of the subscription, the higher the weight on the expected savings. Again, the same reasoning as for the second class subscriptions can explain this dissimilarity, namely moment dependence of the subscriptions or differences in type of customers.

8 Conclusion

In this paper, we have aimed at understanding consumer preferences for subscription packages in public transport using smart card data. By use of different logit models, the influence of travel behaviour on the decision process of subscription packages of the Dutch Railways has been assessed. To answer the research questions data from the MyOV tool is used. Travel behaviour is measured in terms of historic and realized distance traveled at moment of decision. In addition, the economic incentive of individuals is assessed by the savings of the subscription package. Finally, in one of the models unobserved differences between consumers were analyzed.

First of all, travel demand, historical and realized, have shown to be a strong indicator of subscription choice. For first and second class subscriptions statistical evidence is fount that, consumers use past experience of the entire previous year. In addition, an increase in total travel demand increases the probability of choosing a subscription other than no subscription. For subscriptions with a duration of a year, information on travel demand dating a year back is has more explanatory power than expected travel demand that upcoming quarter. If travelers have to make a long term decision, they rather look at their past behaviour than at their expected behaviour and, for a short term decision they rather look at their expected behaviour than at their past behaviour. This result holds for both second and first class travelers, however, for first class travelers this result is less strong.

In addition, a hierarchy in choice is found for second and first class travelers when they make a decision. However, no clear pattern in similar characteristics between the subscription options can be established. Features that are type of moment or discount dependent give an indication for the structure of the choice hierarchy but, mixing these two features in the cross nested logit results in the best fit. In this research a flexible model construction is used for estimation, but in the future more investigation can be done in the exact structure of the choice hierarchy.

When allowing for differences in heterogeneity across consumers it is found that there are differences in preferences. Before estimation a split has been made between the first and second class travelers. We found that first class travelers have more differences in taste and preferences, compared to second class travelers. For second class travelers, especially historic information is differently weighted across individuals, but for first class travelers, mostly realized travel demand has a different influence on subscriptions choice across individuals. In further studies the demographics of customers can be added to characterize the differences in taste and preferences to further complete the investigations.

Finally, the results indicate that there is indeed an economic incentive when individuals choose their subscription package and that individuals are able to predict their expected savings. The latter contradicts the findings from the literature, which stated that individuals are often not able to predict their expected savings. However, when the subscription becomes expensive individuals are not very well at predicting their expected savings. No pattern can be found between the price of the subscription and the economic incentive, thus, for further research it will be interesting to explore this. In addition, for second class travelers the amount of savings is more important than for first class travelers.

The results have different managerial implications. First, the results indicate the use of different marketing strategies for first class travelers than second class travelers, since there are more differences between these travelers. Though, as aforementioned, more investigation is needed in the exact differences between individuals. Second, the expensive subscriptions and the WEEKEND FREE subscription can be more exploited, since consumers are not able to predict their expected savings for these subscriptions. A suggestion is to offer bucket and three-part tariff pricing plans next to the current two-part pricing plans. As stated in the literature, offering several pricing plans gives consumers a wider range options and also has shown to lead to an improved situation for both the supplier and consumer. However, for the other subscriptions, different from earlier findings on non-linear pricing plans, we find that in public transport consumers with two-part pricing plans consumers are actually able to predict their expected savings. Where consumers seemingly make mistakes when they make plan choices, in public transport they do have an economic incentive and seem to be cost minimizing.

Besides providing general insights in consumer preferences for public transport subscriptions, this paper brings up more possible extensions and suggestions for further research. We assume in this paper that travel behaviour influences the subscription choice of the consumer. However, this process can also be vice versa, the subscription choice could also influence the travel behaviour. For example, an individual with a free travel subscription can change his/her behaviour since he/she has can travel freely rather than an individual with a subscription with a discounted travel subscription. Further, the data from MyOV aren't a random sample of the population. Thus, further examination can be done to investigate how this biases or doesn't bias the results. For example by studying less frequent travelers and focusing on other stations than Utrecht Central.

References

Akaike, Htrotugu (1973). “Maximum likelihood identification of Gaussian autoregressive moving average models”. In: *Biometrika* 60.2, pp. 255–265.

Atasoy, Bilge et al. (2015). “The concept and impact analysis of a flexible mobility on demand system”. In: *Transportation Research Part C: Emerging Technologies* 56, pp. 373–392.

Bagchi, Mousumi and Peter White (2004). “What role for smart-card data from bus systems?” In: *Municipal Engineer* 157.1, pp. 39–46.

Balbontin, Camila, David A Hensher, and Andrew T Collins (2019). “How to better represent preferences in choice models: the contributions to preference heterogeneity attributable to the presence of process heterogeneity”. In: *Transportation Research Part B: Methodological* 122, pp. 218–248.

Ben-Akiva, Moshe and Denis Bolduc (1996). *Multinomial probit with a logit kernel and a general parametric specification of the covariance structure*. Departement d’economique, Universite laval with Department of Civil.

Ben-Akiva, Moshe and Steven R Lerman (1985). *Discrete choice analysis: theory and application to travel demand*. Vol. 9. MIT press.

Brons, Martijn et al. (2002). “Price elasticities of demand for passenger air travel: a meta-analysis”. In: *Journal of Air Transport Management* 8.3, pp. 165–175.

Brownstone, David and Kenneth A Small (1989). “Efficient estimation of nested logit models”. In: *Journal of Business & Economic Statistics* 7.1, pp. 67–74.

Cheung, Francis (2006). “Implementation of Nationwide Public Transport Smart Card in the Netherlands: Cost–Benefit Analysis”. In: *Transportation research record* 1971.1, pp. 127–132.

Daly, Andrew and Stanley Zachary (1975). “Commuters’ Value Of Time”. In:

Eurlings, Camiel (2010). “Breaking new ground in building a stronger public transport system”. In: *Minister of Transport, Public Works and Water Management*.

Grubb, Michael D and Matthew Osborne (2015). “Cellular service demand: Biased beliefs, learning, and bill shock”. In: *American Economic Review* 105.1, pp. 234–71.

Hess, Stephane, Andrew Daly, and Richard Batley (2018). “Revisiting consistency with random utility maximisation: theory and implications for practical work”. In: *Theory and Decision* 84.2, pp. 181–204.

Hess, Stephane and John W Polak (2006). “Exploring the potential for cross-nesting structures in airport-choice analysis: a case-study of the Greater London area”. In: *Transportation Research Part E: Logistics and Transportation Review* 42.2, pp. 63–81.

Huber, Joel, John W Payne, and Christopher Puto (1982). “Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis”. In: *Journal of consumer research* 9.1, pp. 90–98.

Iyengar, Raghuram, Asim Ansari, and Sunil Gupta (2007). “A model of consumer learning for service quality and usage”. In: *Journal of Marketing Research* 44.4, pp. 529–544.

Jamasb, Tooraj and Michael Pollitt (2005). “Electricity market reform in the European Union: review of progress toward liberalization & integration”. In: *The Energy Journal*, pp. 11–41.

Lambrecht, Anja, Katja Seim, and Bernd Skiera (2007). “Does uncertainty matter? Consumer behavior under three-part tariffs”. In: *Marketing Science* 26.5, pp. 698–710.

Lambrecht, Anja and Bernd Skiera (2006). “Paying too much and being happy about it: Existence, causes, and consequences of tariff-choice biases”. In: *Journal of marketing Research* 43.2, pp. 212–223.

Li, Ting, Eric Van Heck, and Peter Vervest (2009). "Information capability and value creation strategy: advancing revenue management through mobile ticketing technologies". In: *European Journal of Information Systems* 18.1, pp. 38–51.

Link, Heike (2004). "Pep–a yield-management scheme for rail passenger fares in Germany". In: *Japan Railway & Transport Review* 38.3, pp. 50–55.

Lovelock, Christopher, Sandra Vandermerwe, and Barbara Lewis (1996). "Service marketing: A European perspective". In: *Prenice Hall*.

Lovrić, Milan, Ting Li, and Peter Vervest (2013). "Sustainable revenue management: A smart card enabled agent-based modeling approach". In: *Decision Support Systems* 54.4, pp. 1587–1601.

Luce, R Duncan (1959). "On the possible psychophysical laws." In: *Psychological review* 66.2, p. 81.

Manski, Charles F (1977). "The structure of random utility models". In: *Theory and decision* 8.3, pp. 229–254.

Manzini, Paola and Marco Mariotti (2007). "Sequentially rationalizable choice". In: *American Economic Review* 97.5, pp. 1824–1839.

Marschak, Marianne (1960). "A method for evaluating child-parent interaction under controlled conditions". In: *The Journal of genetic psychology* 97.1, pp. 3–22.

McFadden, Daniel (1974). "The measurement of urban travel demand". In: *Journal of public economics* 3.4, pp. 303–328.

— (1978). "Modeling the choice of residential location". In: *Transportation Research Record* 673.

— (2001). "Economic choices". In: *American economic review* 91.3, pp. 351–378.

McFadden, Daniel and Kenneth Train (2000). "Mixed MNL models for discrete response". In: *Journal of applied Econometrics* 15.5, pp. 447–470.

Mehndiratta, Shomik Raj (1996). "Time-Of-Day Effects in Inter-City Business Travel". In:

Miravete, Eugenio J (2003). "Choosing the wrong calling plan? Ignorance and learning". In: *american economic Review* 93.1, pp. 297–310.

Narayanan, Sridhar, Pradeep K Chintagunta, and Eugenio J Miravete (2007). "The role of self selection, usage uncertainty and learning in the demand for local telephone service". In: *Quantitative Marketing and economics* 5.1, pp. 1–34.

NS (2018). *Tarieven consumenten 2019*. URL: https://www.ns.nl/klantenservice/betalen/tarieven-consumenten.html.

Pelletier, Marie-Pier, Martin Trépanier, and Catherine Morency (2011). "Smart card data use in public transit: A literature review". In: *Transportation Research Part C: Emerging Technologies* 19.4, pp. 557–568.

Prelec, Drazen and George Loewenstein (1998). "The red and the black: Mental accounting of savings and debt". In: *Marketing science* 17.1, pp. 4–28.

Revelt, David and Kenneth Train (1998). "Mixed logit with repeated choices: households' choices of appliance efficiency level". In: *Review of economics and statistics* 80.4, pp. 647–657.

Schlereth, Christian and Bernd Skiera (2012). "Measurement of consumer preferences for bucket pricing plans with different service attributes". In: *International Journal of Research in Marketing* 29.2, pp. 167–180.

Schwarz, Gideon et al. (1978). "Estimating the dimension of a model". In: *The annals of statistics* 6.2, pp. 461–464.

Small, Kenneth A (1981). *Ordered logit: A discrete choice model with proximate covariance among alternatives*. Econometric Research Program, Princeton University.

Tirachini, Alejandro, David A Hensher, and John M Rose (2014). "Multimodal pricing and optimal design of urban public transport: The interplay between traffic congestion and bus crowding". In: *Transportation research part b: methodological* 61, pp. 33–54.

Train, Kenneth (2009). *Discrete choice methods with simulation*. Cambridge university press.

Train, Kenneth, Daniel McFadden, and Moshe Ben-Akiva (1987). “The demand for local telephone service: A fully discrete model of residential calling patterns and service choices”. In: *The RAND Journal of Economics*, pp. 109–123.

Tversky, Amos and Daniel Kahneman (1974). “Judgment under uncertainty: Heuristics and biases”. In: *science* 185.4157, pp. 1124–1131.

— (1989). “Rational choice and the framing of decisions”. In: *Multiple Criteria Decision Making and Risk Analysis Using Microcomputers*. Springer, pp. 81–126.

Tversky, Amos and Itamar Simonson (1993). “Context-dependent preferences”. In: *Management science* 39.10, pp. 1179–1189.

Utsunomiya, Mariko, John Attanucci, and Nigel Wilson (2006). “Potential uses of transit smart card registration and transaction data to improve transit planning”. In: *Transportation research record* 1971.1, pp. 118–126.

Van Vuuren, Daniel (2002). “Optimal pricing in railway passenger transport: theory and practice in The Netherlands”. In: *Transport policy* 9.2, pp. 95–106.

Williams, Huw (1977). “On the formation of travel demand models and economic evaluation measures of user benefit”. In: *Environment and planning A* 9.3, pp. 285–344.

9 Appendix

Generalized Extreme Value Models

Let $G(y_1, y_2, ..., y_{J_n})$, for $y_1, y_2, ..., y_{J_n} \geq 0$, be a function with the following properties:

1. G is non-negative
2. G is homogeneous of degree $\mu > 0$; that is $G(\alpha y_1, \alpha y_2, ..., \alpha y_{J_n}) = \alpha^\mu G(y_1, y_2, ..., y_{J_n})$
3. $\lim_{y_i \to \infty} G(y_1, y_2, ..., y_{J_n}) = \infty$, for $i = 1, 2, ..., J_n$
4. The $l$the partial derivative of G with respect to any combination of l distinct y_i's, $i = 1, ..., J_n$ is non-negative if l is odd, and non-positive if l is even

If G satisfies these conditions and $G_i(y_1, y_2, ..., y_{J_n})$ denotes $\partial G/\partial y_i$, $i = 1, ..., J_n$, then

$$P_n(i) = \frac{e^{V_{in}} G_i(e^{V_{1n}}, e^{V_{2n}}, ..., e^{V_{J_n n}})}{\mu G(e^{V_{1n}}, e^{V_{2n}}, ..., e^{V_{J_n n}})}$$

defines the GEV model.

Utility functions

The two model formulations, subscription based and total travel demand, given in Section 6 are defined by Equation 8 and Equation 9, respectively. Below the general equations given in the main text are given with the fully written out formulation.

In the fully written out formulation numbers are used for the subscriptions. The following $j = 1, ..., J$ corresponds to the following subscriptions:

1. WEEKEND FREE
2. OFF-PEAK FREE
3. OFF-PEAK DISCOUNT
4. ALWAYS DISCOUNT MONTHLY
5. ALWAYS DISCOUNT YEARLY
6. ALWAYS FREE MONTHLY
7. ALWAYS FREE YEARLY
8. TRAJECTORY FREE MONTHLY
9. TRAJECTORY FREE YEARLY
10. FULL TARIFF

Note that the monthly subscriptions are only used for the second class choice set, thus for the first class choice set $J = 7$ and the second class choice set $J = 10$.

Formulation (I)

The general formulation:

$$V(\mathbf{x}_{nj}) = \boldsymbol{\beta}_1' \mathbf{x}_{1,nj} + \boldsymbol{\beta}_{2,j}' \mathbf{x}_{2,n}$$

The fully written out formulation[11]:

$$\begin{aligned}
V_{1,n} &= \alpha_1 + \beta_{hist.} * (x_{lag_km_weekend,n}) + \beta_{real.} * (x_{km_weekend,n}) + \beta_{1,saved} * x_{saved,n} + \beta_{1,dist/travel} * x_{dist./travel,n} \\
V_{2,n} &= \alpha_2 + \beta_{hist.} * (x_{lag_km_weekend,n} + x_{lag_km_offpeak,n}) + \beta_{real.} * (x_{km_weekend,n} + x_{km_offpeak,n}) \\
&\quad + \beta_{2,saved} * x_{saved,n} + \beta_{2,dist/travel} * x_{dist./travel,n} \\
V_{3,n} &= \alpha_3 + \beta_{hist.} * (x_{lag_km_weekend,n} + x_{lag_km_offpeak,n}) + \beta_{real.} * (x_{km_weekend,n} + x_{km_offpeak,n}) \\
&\quad + \beta_{3,saved} * x_{saved,n} + \beta_{3,dist/travel} * x_{dist./travel,n} \\
V_{4,n} &= \alpha_4 + \beta_{hist.} * (x_{lag_km_total}) + \beta_{real.} * (x_{km_total}) + \beta_{4,saved} * x_{saved,n} + \beta_{4,dist/travel} * x_{dist./travel,n} \\
V_{5,n} &= \alpha_5 + \beta_{hist.} * (x_{lag_km_total}) + \beta_{real.} * (x_{km_total}) + \beta_{5,saved} * x_{saved,n} + \beta_{5,dist/travel} * x_{dist./travel,n} \\
V_{6,n} &= \alpha_6 + \beta_{hist.} * (x_{lag_km_total}) + \beta_{real.} * (x_{km_total}) + \beta_{6,saved} * x_{saved,n} + \beta_{6,dist/travel} * x_{dist./travel,n} \\
V_{7,n} &= \alpha_7 + \beta_{hist.} * (x_{lag_km_total}) + \beta_{real.} * (x_{km_total}) + \beta_{7,saved} * x_{saved,n} + \beta_{7,dist/travel} * x_{dist./travel,n} \\
V_{8,n} &= \alpha_8 + \beta_{hist.} * (x_{lag_km_traject,n}) + \beta_{real.} * (x_{km_traject,n}) + \beta_{8,saved} * x_{saved,n} + \beta_{8,dist/travel} * x_{dist./travel,n} \\
V_{9,n} &= \alpha_9 + \beta_{hist.} * (x_{lag_km_traject,n}) + \beta_{real.} * (x_{km_traject,n}) + \beta_{9,saved} * x_{saved,n} + \beta_{9,dist/travel} * x_{dist./travel,n} \\
V_{10,n} &= \beta_{hist.} * (x_{lag_km_total,n}) + \beta_{real.} * (x_{km_total,n})
\end{aligned}$$

[11]Note: $x_{lag\_km\_total} = x_{lag\_km\_weekend,n} + x_{lag\_km\_offpeak,n} + x_{lag\_km\_peak,n}$ and $x_{km\_total} = x_{km\_weekend,n} + x_{km\_offpeak,n} + x_{km\_peak,n}$

Formulation (II)

The general formulation:

$$V(\mathbf{x}_{nj}) = \boldsymbol{\beta}'_{1,j}\mathbf{x}_{1,n} + \boldsymbol{\beta}'_{2,j}\mathbf{x}_{2,n}$$

The fully written out formulation:

$$V_{1,n} = \alpha_1 + \beta_{1,hist.} * (x_{lag\_km\_total}) + \beta_{1,real.} * (x_{km\_total}) + \beta_{1,saved} * x_{saved,n} + \beta_{1,dist/travel} * x_{dist./travel,n}$$
$$V_{2,n} = \alpha_2 + \beta_{2,hist.} * (x_{lag\_km\_total}) + \beta_{2,real.} * (x_{km\_total}) + \beta_{2,saved} * x_{saved,n} + \beta_{2,dist/travel} * x_{dist./travel,n}$$
$$V_{3,n} = \alpha_3 + \beta_{3,hist.} * (x_{lag\_km\_total}) + \beta_{3,real.} * (x_{km\_total}) + \beta_{3,saved} * x_{saved,n} + \beta_{3,dist/travel} * x_{dist./travel,n}$$
$$V_{4,n} = \alpha_4 + \beta_{4,hist.} * (x_{lag\_km\_total}) + \beta_{4,real.} * (x_{km\_total}) + \beta_{4,saved} * x_{saved,n} + \beta_{4,dist/travel} * x_{dist./travel,n}$$
$$V_{5,n} = \alpha_5 + \beta_{5,hist.} * (x_{lag\_km\_total}) + \beta_{5,real.} * (x_{km\_total}) + \beta_{5,saved} * x_{saved,n} + \beta_{5,dist/travel} * x_{dist./travel,n}$$
$$V_{6,n} = \alpha_6 + \beta_{6,hist.} * (x_{lag\_km\_total}) + \beta_{6,real.} * (x_{km\_total}) + \beta_{6,saved} * x_{saved,n} + \beta_{6,dist/travel} * x_{dist./travel,n}$$
$$V_{7,n} = \alpha_7 + \beta_{7,hist.} * (x_{lag\_km\_total}) + \beta_{7,real.} * (x_{km\_total}) + \beta_{7,saved} * x_{saved,n} + \beta_{7,dist/travel} * x_{dist./travel,n}$$
$$V_{8,n} = \alpha_8 + \beta_{8,hist.} * (x_{lag\_km\_total}) + \beta_{8,real.} * (x_{km\_total}) + \beta_{8,saved} * x_{saved,n} + \beta_{8,dist/travel} * x_{dist./travel,n}$$
$$V_{9,n} = \alpha_9 + \beta_{9,hist.} * (x_{lag\_km\_total}) + \beta_{9,real.} * (x_{km\_total}) + \beta_{9,saved} * x_{saved,n} + \beta_{9,dist/travel} * x_{dist./travel,n}$$

Note that $V_{10,n}$ is not added since the estimated coefficients are with respect to the FULL TARIFF subscription and the utility function is fixed.

Goodness of fit statistics

The goodness of fit measures are calculated and given by the estimation package, Biogeme.

Rho-squared bar

$$\bar{\rho}^2 = 1 - \frac{\mathcal{L}(\hat{\theta}) - K}{\mathcal{L}(0)}$$

The Akaike Information Criteria (AIC) (Akaike 1973)

$$AIC = 2K - 2\mathcal{L}(\hat{\theta})$$

The Bayesian Information Criteria (BIC) (Schwarz et al. 1978).

$$BIC = -2\mathcal{L}(\hat{\theta}) + KN$$

where $\mathcal{L}(0)$ the null model, $\mathcal{L}(\hat{\theta})$ the estimated model, K number of parameters, N the sample size

MNL (Second Class)

			(I)				(II)
	Lags		LAG1	LAG2	LAG3	LAG4	LAG4
$\boldsymbol{\beta}_1$	Realized		0.179***	0.17***	0.165***	0.16***	
			(0.00377)	(0.00378)	(0.00379)	(0.00379)	
		ALWAYS DISCOUNT					0.692***
		YEARLY					(0.103)
		ALWAYS DISCOUNT					1.97***
		MONTHLY					(0.307)
		ALWAYS FREE					4.03***
		YEARLY					(0.324)
		ALWAYS FREE					8.9***
		MONTHLY					(0.995)
		OFF-PEAK DISCOUNT					0.276***
							(0.0373)
		OFF-PEAK FREE					0.3***
							(0.085)
		TRAJECTORY FREE					1.19***
		YEARLY					(0.113)
		TRAJECTORY FREE					0.756***
		MONTHLY					(0.236)
		WEEKEND FREE					4.6***
							(0.319)
	Historic		0.0774***	0.124***	0.157***	0.188***	
			(0.00417)	(0.00525)	(0.00605)	(0.00664)	
		ALWAYS DISCOUNT					1.19***
		YEARLY					(0.379)
		ALWAYS DISCOUNT					0.756***
		MONTHLY					(0.0353)
		ALWAYS FREE					4.6***
		YEARLY					(0.121)
		ALWAYS FREE					2.35***
		MONTHLY					(0.12)
		OFF-PEAK DISCOUNT					0.278***
							(0.149)
		OFF-PEAK FREE					1.51***
							(0.0782)
		TRAJECTORY FREE					1.34***
		YEARLY					(0.129)
		TRAJECTORY FREE					0.275***
		MONTHLY					(0.271)
		WEEKEND FREE					0.764***
							(0.336)
$\boldsymbol{\beta}_2$	Dist / travel	ALWAYS DISCOUNT	0.759**	0.753**	0.747**	0.747**	-0.423***
		YEARLY	(0.0866)	(0.0864)	(0.0863)	(0.0862)	(0.336)
		ALWAYS DISCOUNT	1.01**	1**	0.995**	0.979**	0.385***
		MONTHLY	(0.152)	(0.152)	(0.151)	(0.152)	(0.537)
		ALWAYS FREE	2.23***	2.22***	2.21***	2.21***	2.72***
		YEARLY	(0.068)	(0.0679)	(0.0678)	(0.0678)	(0.416)
		ALWAYS FREE	2.44***	2.41***	2.42***	2.42***	5.63***
		MONTHLY	(0.102)	(0.102)	(0.102)	(0.102)	(0.917)
		OFF-PEAK DISCOUNT	1.09***	1.05***	1.03***	1***	0.386***
			(0.0553)	(0.0553)	(0.0552)	(0.0552)	(0.152)

			(I)				(II)
	Lags		LAG1	LAG2	LAG3	LAG4	LAG4
$\boldsymbol{\beta}_2$	Dist / travel	OFF-PEAK FREE	1.35***	1.33***	1.3***	1.29***	0.46***
			(0.0805)	(0.0805)	(0.0804)	(0.0804)	(0.262)
		TRAJECTORY FREE	0.0382	0.011	-0.00445	-0.0192	-4.52***
		YEARLY	(0.0793)	(0.0793)	(0.0793)	(0.0794)	(0.46)
		TRAJECTORY FREE	-0.105	-0.138	-0.157	-0.181	-6.88***
		MONTHLY	(0.13)	(0.13)	(0.13)	(0.13)	(0.949)
		WEEKEND FREE	2.09***	2.06***	2.04***	2.01***	1.22***
			(0.0636)	(0.0635)	(0.0635)	(0.0634)	(0.187)
	Saved	ALWAYS DISCOUNT	3.73***	3.74***	3.75***	3.76***	0.426***
		YEARLY	(0.119)	(0.12)	(0.12)	(0.12)	(0.536)
		ALWAYS DISCOUNT	4.12***	4.13***	4.14***	4.16***	-0.915***
		MONTHLY	(0.194)	(0.194)	(0.194)	(0.194)	(0.835)
		ALWAYS FREE	4.82***	4.82***	4.83***	4.84***	-4.29***
		YEARLY	(0.102)	(0.102)	(0.102)	(0.102)	(0.607)
		ALWAYS FREE	3***	3***	3.02***	3.03***	-11.9***
		MONTHLY	(0.242)	(0.242)	(0.242)	(0.242)	(1.3)
		OFF-PEAK DISCOUNT	0.93***	0.927***	0.937***	0.949***	1.86***
			(0.0753)	(0.0754)	(0.0755)	(0.0756)	(0.396)
		OFF-PEAK FREE	5.95***	5.95***	5.95***	5.96***	4.15***
			(0.0962)	(0.0963)	(0.0964)	(0.0965)	(0.464)
		TRAJECTORY FREE	2.56***	2.52***	2.5***	2.48***	-3.29***
		YEARLY	(0.115)	(0.115)	(0.115)	(0.115)	(0.487)
		TRAJECTORY FREE	-0.975***	-1.01***	-1.03***	-1.06***	-7.28***
		MONTHLY	(0.13)	(0.13)	(0.13)	(0.13)	(0.685)
		WEEKEND FREE	4.68***	4.72***	4.76***	4.79***	3.03***
			(0.0975)	(0.0977)	(0.0979)	(0.0981)	(0.467)
	Intercept	ALWAYS DISCOUNT	-1.68***	-1.68***	-1.67***	-1.67***	-13.6***
		YEARLY	(0.044)	(0.0439)	(0.0438)	(0.0438)	(0.831)
		ALWAYS DISCOUNT	-3.42***	-3.41***	-3.41***	-3.41***	-21.4***
		MONTHLY	(0.0857)	(0.0856)	(0.0855)	(0.0856)	(2.08)
		ALWAYS FREE	-2.8***	-2.79***	-2.79***	-2.79***	-70***
		YEARLY	(0.0465)	(0.0464)	(0.0464)	(0.0464)	(3.68)
		ALWAYS FREE	-4.38***	-4.37***	-4.37***	-4.37***	-95.1***
		MONTHLY	(0.0923)	(0.0922)	(0.0922)	(0.0922)	(9.47)
		OFF-PEAK DISCOUNT	-0.325***	-0.503***	-0.636***	-0.757***	-2.54***
			(0.0291)	(0.0314)	(0.0334)	(0.0351)	(0.222)
		OFF-PEAK FREE	-3.16***	-3.34***	-3.48***	-3.6***	-14.2***
			(0.0485)	(0.05)	(0.0514)	(0.0526)	(0.922)
		TRAJECTORY FREE	0.38***	0.343***	0.326***	0.306***	-21.3***
		YEARLY	(0.0367)	(0.0367)	(0.0367)	(0.0368)	(0.964)
		TRAJECTORY FREE	-0.959***	-0.995***	-1.01***	-1.03***	-23.9***
		MONTHLY	(0.0559)	(0.0559)	(0.056)	(0.056)	(1.85)
		WEEKEND FREE	-1.47***	-1.4***	-1.34***	-1.29***	-8.67***
			(0.0402)	(0.0404)	(0.0407)	(0.0409)	(0.549)
Number of estimated parameters			32	32	32	32	50
Sample size			5120	5120	5120	5120	5120
Log likelihood			-9082.054	-9062.496	-9058.021	-9047.207	-7740.616
Rho-square bar			0.258	0.259	0.26	0.26	0.365
AIC			18228	18189	18180	18158	15581
BIC			18437	18398	18389	18367	15908

*p<0.1; **p<0.05; ***p<0.01; Standard error in parentheses

Table 8: Coefficient table of MNL (Second class)

MNL (First Class)

			(I)				(II)
	Lags		LAG1	LAG2	LAG3	LAG4	LAG4
$\boldsymbol{\beta}_1$	Realized		0.219***	0.194***	0.184***	0.177***	
			(0.0202)	(0.0203)	(0.459)	(0.0206)	
		ALWAYS DISCOUNT					0.726***
		YEARLY					(0.22)
		ALWAYS FREE					3.61***
		YEARLY					(0.323)
		OFF-PEAK DISCOUNT					0.0707***
							(0.113)
		OFF-PEAK FREE					0.162***
							(0.113)
		TRAJECTORY FREE					2.3***
		YEARLY					(0.355)
		WEEKEND FREE					-0.0524***
							(0.101)
	Historic		0.305***	0.404***	0.459***	0.506***	
			(0.0212)	(0.026)	(0.0293)	(0.0318)	
		ALWAYS DISCOUNT					0.513***
		YEARLY					(0.197)
		ALWAYS FREE					3.36***
		YEARLY					(0.341)
		OFF-PEAK DISCOUNT					-0.117***
							(0.101)
		OFF-PEAK FREE					1.97***
							(0.138)
		TRAJECTORY FREE					2.08***
		YEARLY					(0.38)
		WEEKEND FREE					0.429***
							(0.106)
$\boldsymbol{\beta}_2$	Dist. / travel	ALWAYS DISCOUNT	-0.0392	-0.0655	-0.0706	-0.0634	-0.929*
		YEARLY	(0.505)	(0.505)	(0.505)	(0.506)	(0.51)
		ALWAYS FREE	0.304	0.267	0.259	0.263	-1.77***
		YEARLY	(0.429)	(0.429)	(0.429)	(0.43)	(0.543)
		OFF-PEAK DISCOUNT	0.293	0.204	0.168	0.156	-0.147
			(0.432)	(0.432)	(0.432)	(0.432)	(0.402)
		OFF-PEAK FREE	0.513	0.904	0.391	0.37	-1.23***
			(0.384)	(0.371)	(0.384)	(0.385)	(0.374)
		TRAJECTORY FREE	-3.07***	-3.24***	-3.33***	-3.39***	-10.9***
		YEARLY	(0.736)	(0.745)	(0.751)	(0.756)	(1.37)
		WEEKEND FREE	1.04***	0.984**	0.956**	0.946**	0.202
			(0.38)	(0.38)	(0.381)	(0.381)	(0.357)
	Saved	ALWAYS DISCOUNT	1.78***	1.78***	1.78***	1.76***	1.63***
		YEARLY	(0.432)	(0.433)	(0.432)	(0.431)	(0.616)
		ALWAYS FREE	1.99***	2***	1.99***	1.96***	-0.32
		YEARLY	(0.393)	(0.394)	(0.394)	(0.393)	(0.555)
		OFF-PEAK DISCOUNT	0.622	0.681	0.684	0.667	1.71***
			(0.394)	(0.395)	(0.395)	(0.395)	(0.608)
		OFF-PEAK FREE	3.06***	3.12***	3.13***	3.12***	3.37***
			(0.374)	(0.375)	(0.374)	(0.374)	(0.526)
		TRAJECTORY FREE	2.23***	2.21***	2.2***	2.17***	0.0461
		YEARLY	(0.496)	(0.498)	(0.497)	(0.495)	(0.59)
		WEEKEND FREE	2.76***	2.81***	2.86***	2.86***	3.38***
			(0.376)	(0.376)	(0.376)	(0.376)	(0.526)

			(I)				(II)
	Lags		LAG1	LAG2	LAG3	LAG4	LAG4
	Intercept	ALWAYS DISCOUNT	-0.245	-0.232	-0.228	-0.228	-8.06***
		YEARLY	(0.353)	(0.352)	(0.352)	(0.352)	(1.6)
		ALWAYS FREE	0.116	0.137	0.143	0.148	-52.2***
		YEARLY	(0.307)	(0.307)	(0.306)	(0.307)	(3.65)
		OFF-PEAK DISCOUNT	-0.721**	-1.16***	-1.43***	-1.66***	1.14*
			(0.305)	(0.311)	(0.316)	(0.321)	(0.676)
		OFF-PEAK FREE	-0.334***	-0.777***	-1.05***	-1.28***	-13.7***
			(0.279)	(0.286)	(0.292)	(0.297)	(1.07)
		TRAJECTORY FREE	2.33***	2.27***	2.25***	2.24***	-27.7***
		YEARLY	(0.341)	(0.342)	(0.343)	(0.344)	(3.26)
		WEEKEND FREE	2.3***	2.41***	2.45***	2.5***	-1.2*
			(0.271)	(0.272)	(0.273)	(0.273)	(0.699)
Number of estimated parameters			26	26	26	26	40
Sample size			2418	2418	2418	2418	2418
Log likelihood			-2894.481	-2868.513	-2864.224	-2857.421	-2710.854
Rho-square bar			0.45	0.455	0.456	0.457	0.482
AIC			5841	5789	5780	5767	5501
BIC			5991	5940	5931	5917	5733

*p<0.1; **p<0.05; ***p<0.01; Standard error in parentheses

Table 9: Coefficient table of MNL (First class)

NL,CNL,ML (Second Class)(I)

			NL Moment	NL Discount	CNL	ML
β_1	Realized		0.156	0.00964	0.0597***	0.391***
			(-)	(-)	(0.007)	(0.0199)
	Historic		0.185	0.0111	0.12***	0.607***
			(-)	(-)	(0.0142)	(0.0445)
β_2	Dist / travel	ALWAYS DISCOUNT	1.6	1.21***	-0.829***	0.983***
		YEARLY	(-)	(0.0008)	(0.293)	(0.313)
		ALWAYS DISCOUNT	1.63	1.23***	-0.554	1.4***
		MONTHLY	(-)	(0.0000)	(0.503)	(0.485)
		ALWAYS FREE	1.77	1.56	0.607***	2.78***
		YEARLY	(-)	(-)	(0.0975)	(0.243)
		ALWAYS FREE	1.79	1.57	1.28***	3.05***
		MONTHLY	(-)	(-)	(0.31)	(0.358)
		OFF-PEAK DISCOUNT	1.09***	1.28***	-0.0164	2.14***
			(0.00683)	(0.0000)	(0.0275)	(0.284)
		OFF-PEAK FREE	1.25***	1.54	0.337	2.3***
			(0.00897)	(-)	(0.209)	(0.339)
		TRAJECTORY FREE	0.0184*	1.45	0.259**	0.823***
		YEARLY	(0.011)	(-)	(0.109)	(0.27)
		TRAJECTORY FREE	-0.104*	1.44	-1.38***	0.441
		MONTHLY	(0.054)	(-)	(0.484)	(0.46)
		WEEKEND FREE	2.09***	1.55	0.63***	3.09***
			(0.00477)	(-)	(0.098)	(0.263)
	Saved	ALWAYS DISCOUNT	3.78	2.13***	4.55***	2.8***
		YEARLY	(-)	(0.0008)	(0.387)	(0.381)
		ALWAYS DISCOUNT	3.79	2.24***	5.9***	3.61***
		MONTHLY	(-)	(0.0001)	(0.461)	(0.526)
		ALWAYS FREE	3.78	3.26	4.28***	3.62***
		YEARLY	(-)	(-)	(0.265)	(0.317)
		ALWAYS FREE	3.76	3.23	-1.53***	-1.24**
		MONTHLY	(-)	(-)	(0.549)	(0.506)
		OFF-PEAK DISCOUNT	1.12***	1.72***	2.14***	-0.348
			(0.0231)	(0.0000)	(0.201)	(0.365)
		OFF-PEAK FREE	4.6***	3.27	6.47***	4.79***
			(0.00724)	(-)	(0.524)	(0.394)
		TRAJECTORY FREE	2.27***	3.27	4.39***	2.87***
		YEARLY	(0.0264)	(-)	(0.268)	(0.337)
		TRAJECTORY FREE	-1.02	3.25	-0.771	-0.438
		MONTHLY	(-)	(-)	(0.54)	(0.43)
		WEEKEND FREE	4.23***	3.26	4.28***	3.67***
			(0.0187)	(-)	(0.265)	(0.35)
	Intercept	ALWAYS DISCOUNT	-1.34	0.681***	-1.42***	-1.72***
		YEARLY	(-)	(0.0039)	(0.426)	(0.15)
		ALWAYS DISCOUNT	-1.41	0.567***	-3.54	-3.36***
		MONTHLY	(-)	(0.0001)	(0.201)	(0.26)
		ALWAYS FREE	-1.44	-0.252	-1.12***	-2.85***
		YEARLY	(-)	(-)	(0.0686)	(0.152)
		ALWAYS FREE	-1.51	-0.297	-4.57***	-4.65***
		MONTHLY	(-)	(534)	(0.288)	(0.312)
		OFF-PEAK DISCOUNT	-0.74***	0.76***	-0.059	-3.65***
			(0.00359)	(0.0000)	(0.0747)	(0.212)
		OFF-PEAK FREE	-2.56***	-0.335	-3.37***	-6.54***
			(0.0134)	(-)	(0.566)	(0.249)
		TRAJECTORY FREE	0.275	-0.131	-0.592***	1.14***
		YEARLY	(-)	(-)	(0.0813)	(0.134)
		TRAJECTORY FREE	-1.05	-0.157	-1.63***	-0.223
		MONTHLY	(-)	(-)	(0.354)	(0.195)
		WEEKEND FREE	-1.28***	-0.188	-0.787***	-1.99***
			(0.0181)	(-)	(0.0746)	(0.175)
Number of estimated parameters			35	34	47	33
Sample size			5120	5120	5120	3864
Log likelihood			-9140.61	-8839.465	-8071.99	-8879.95
Rho-square bar			0.255	0.277	0.339	0.246

*p<0.1; **p<0.05; ***p<0.01; std. error in parentheses, (-) std. error $> 2 * 10^6$

Table 10: Coefficient table of NL/CNL/ML (Second class)

NL,CNL,ML (First Class)(I)

			NL Moment	NL Discount	CNL	ML
$\boldsymbol{\beta}_1$	Realized		0.177***	0.115***	0.129***	0.422***
			(0.0206)	(0.00655)	(0.129)	(0.0436)
	Historic		0.506***	0.302***	0.251**	0.634***
			(0.0318)	(0.0501)	(0.251)	(0.0662)
$\boldsymbol{\beta}_1$	Dist / travel	ALWAYS DISCOUNT	-0.0714	-0.0157	0.00958	-0.176
		YEARLY	(0.503)	(0.488)	(0.00958)	(0.518)
		ALWAYS FREE	0.253	0.0488	-0.122**	0.146
		YEARLY	(0.43)	(0.383)	(-0.122)	(0.439)
		OFF-PEAK DISCOUNT	0.146	0.115	-0.164***	0.683
			(0.437)	(0.416)	(-0.164)	(0.518)
		OFF-PEAK FREE	0.361	0.414	-0.0328	1.18**
			(0.385)	(0.366)	(-0.0328)	(0.471)
		TRAJECTORY FREE	-3.39***	-1.85***	-0.863***	-2.79***
		YEARLY	(0.756)	(0.468)	(-0.863)	(0.883)
		WEEKEND FREE	0.938**	0.663	0.188	0.801**
			(0.381)	(0.363)	(0.188)	(0.407)
	Saved	ALWAYS DISCOUNT	1.77***	1.74***	0.462***	1.8***
		YEARLY	(0.409)	(0.42)	(0.462)	(0.442)
		ALWAYS FREE	1.98***	2.36***	0.606***	1.98***
		YEARLY	(0.413)	(0.36)	(0.606)	(0.406)
		OFF-PEAK DISCOUNT	0.68	0.571	1.38***	0.722*
			(0.745)	(0.389)	(1.38)	(0.449)
		OFF-PEAK FREE	3.13***	2.86***	2.04	3***
			(0.438)	(0.363)	(2.04)	(0.402)
		TRAJECTORY FREE	2.18***	2.58***	0.851***	1.06
		YEARLY	(0.508)	(0.379)	(0.851)	(0.682)
		WEEKEND FREE	2.87***	2.78***	2.14	2.92***
			(0.408)	(0.363)	(2.14)	(0.403)
	Intercept	ALWAYS DISCOUNT	-0.223	-0.258	0.785***	-0.165
		YEARLY	(0.192)	(0.0984)	(0.785)	(0.354)
		ALWAYS FREE	0.155	1.24	0.842***	0.224
		YEARLY	(0.24)	(0.25)	(0.842)	(0.309)
		OFF-PEAK DISCOUNT	-1.66***	-0.671	-0.45***	-2.39***
			(0.552)	(-)	(-0.45)	(0.416)
		OFF-PEAK FREE	-1.27***	0.0857	-0.431	-2.09***
			(0.341)	(0.132)	(-0.431)	(0.393)
		TRAJECTORY FREE	2.24***	2.49	0.888***	1.9***
		YEARLY	(0.344)	(0.28)	(0.888)	(0.409)
		WEEKEND FREE	2.5 ***	2.51	1.62	1.84***
			(0.275)	(0.253)	(1.62)	(0.296)
Number of estimated parameters			28	28	39	27
Sample size			2418	2418	2418	521
Log likelihood			-2857.42	-2815.307	-2538.89	-2311.73
Rho-square bar			0.457	0.465	0.515	0.41

*p<0.1; **p<0.05; ***p<0.01; std. error in parentheses, (-) std. error $> 2 * 10^6$

Table 11: Coefficient table of NL/CNL/ML (First class)