Mode choice modeling including new forms of mobility

Nicolas Schoenn-Anchling

Travail de fin d'études

pour le diplôme d'ingénieur de l'École nationale des travaux publics de l'État

Année 2022-2023

Voie d'approfondissement :
Transport – Ingénierie des mobilités

Soutenu le 4 septembre 2023

Devant le jury composé de :

- Président du Jury : Christine BUISSON
- Maître de TFE : Silvia Francesca VAROTTO
- Expert : Guilhem MARIOTTE

Par

Nicolas Schoenn-Anchling

Modélisation du choix modal comprenant les nouvelles formes de mobilités

Organisme d'accueil
LICIT-ECO7

Notice analytique

AUTEUR			
Nom	Schoenn-Anchling		
Prénom	Nicolas		
ORGANISME D'ACCUEIL			
Nom de l'organisme et Localité	LICIT-ECO7 (3 rue Maurice Audin 69120 Vaulx-en-Velin)		
Nom du Tuteur	Varotto Silvia Francesca		
ANALYSE DU TFE			
Titre (français)	Modélisation du choix modal comprenant les nouvelles formes de mobilités		
Titre (anglais)	Mode choice modeling including new forms of mobility		
Résumé (français)	Ce travail a pour objectif de modéliser le choix modal pour un trajet aller de manière à déceler les facteurs clés des nouvelles formes de mobilités tout en rendant compte de la corrélation qui existe entre ce choix et les informations que l'on possède sur le trajet retour. Pour ce faire, a été conduit un sondage préférences déclarées pour les membres de l'ENTPE. L'analyse des données récoltées et les estimations de modèles logit ont permis de montrer que le temps hors véhicules et le temps de recherche d'un parking au retour sont des facteurs essentiels dans le choix de mode de transport tout comme la météo. Les différents modèles développés ont aussi permis de mettre en valeur l'existence d'une hétérogénéité dans les préférences modales entre différents groupes de personnes qui ne pouvaient être décrit à partir des variables socio-économiques récoltés.		
Résumé (anglais)	This work aims to model mode choice for an inward trip in order to understand the key factors of new forms of mobility while accounting for the correlation that exists between this choice and the information that we have on the return trip. To do this, a stated preference experiment was conducted for ENTPE members. The analysis of the collected data and the estimations of the logit models made it possible to show that the out-of-vehicle time and the parking search time on the way back have a negative impact on the shares of the new mobilities. The models also showed that the weather had a negative impact on bike and bike-sharing system. The various models developed have also made it possible to highlight the existence of heterogeneity in mode preferences between different groups of people that could not be described from the socio-economic variables collected.		
Mots-clés (français, 5 maxi)	Choix modale, Préférences déclarées, modèles de régression logistique, Prévision de la demande		
Mots-clés (anglais, 5 maxi)	Mode choice, Stated preferences, Logistic regression models, Forecasting		
Termes géo-graphique (français)	Vaulx-en-Velin, Métropole de Lyon, ENTPE		
COLLATION			
	Nb de pages	Nb d'annexes (nb de pages)	Nb de réf. biblio.
	61	5	58

Déclaration de travail personnel

Je déclare que ce rapport constitue l'aboutissement d'un travail personnel et ne peut être suspecté de plagiat.

Le travail présenté distingue explicitement ce que j'ai produit de ce que j'ai emprunté à d'autres. A ce titre, les citations sont clairement identifiables et les sources (écrits, images) qui ont alimenté ma réflexion sont référencées.

Remerciements

Je souhaiterais commencer par remercier Silvia Francesca VAROTTO, ma tutrice de stage, pour m'avoir aidé, conseillé et guidé tout au long de cette aventure. Le sujet que nous avons construit m'a apporté nombre de connaissances et de compétences que j'aurai l'occasion d'employer au cours de la thèse à venir mais aussi dans mon cursus professionnel.

Ensuite j'aimerais remercier toutes les personnes du LICIT-ECO7 pour leur accueil chaleureux.

Puis je voudrais remercier toutes les personnes qui ont relu ce rapport et m'ont aidé à le parfaire.

Enfin, je finirai en remerciant Flore, Vincent, Matthias et Madeline pour les grands débats du midi.

Sommaire

Liste des illustrations

Liste des tables

Liste des abréviations

ASC: alternative specific constant

ATA model: model with all trip attributes

Avail: number of spot available at arrival for bike sharing-system

BSS: bike sharing-system

CT: time spent in congestion the day before

ECP model: model with error component panel

ESS: e-scooter sharing system

ESSPT: e-scooter combined with transit

MNL: multinominal logit model

ML: mixed logit model

NbT: Number of transfer

OT: out-of-vehicle time

PC: private car

PB: personal bike

PST: parking search time on the way back

PT: public transportation

RI: rain on the way in

RP: revealed preferences

SEV model: model with socio-economic variables

SP: stated preferences

TT: travel time

UF: utility function

WA model: model with the weather attributes

WR: probability of rain on the way back

1. Introduction

In recent years, new forms of mobility have emerged in urban environments. These new forms are the result of technical innovations, the use of certain traditional modes of transport which have been redesigned or with the emergence of new technologies and what came out of it. These changes have brought modifications to the original transport systems, either by adding new services that can results in the saturation of that system or by transforming people behavior within that system. To observe these changes and predict them in order to ensure the resilience of transport networks, experts in the field have developed transport models. These models can help decision-makers to design policies in the context of urban planning.

Transports models have two main components: supply and demand. In transport planning it is common to use the so-called four steps model that was first introduced by the Chicago Area Transportation Study in the 50's (Hensher and Button, 2007). Historically, this is a static macroscopic trip-based model. The term static refers to the demand side of the model and it means that the model can give a picture of the level of service of a network at a certain time t because the OD matrices are supposed constant. Macroscopic refers to the scale of the model and means that the data is aggregated which shortens computation time, and hence facilitates the analysis of large-scale networks. Lastly, trip-based refers to the demand component, and it means that all simulated trips are supposed to be independent from one another. To summarize, this model can give a complete vision of a large-scale network at a certain time t with a simplified modelling of congestion phenomenon and users behavior. However, as mentioned before, the urban environment tends to be more complex than when this model was developed. Congestion phenomenon are more frequent and people's behavior becomes more complex because of the wide range of available services. Hence, as the four-steps model is static it does not capture the propagation of congestion phenomenon in time and space and the correlation between trips within a day, other models have been developed to better represent urban environments.

To refine the modeling of the level of service, dynamic traffic models have been developed (Johari et al., 2021). These models acknowledge the impact of the state of the network at time t-1 on the level of service of the network at time t. Hence, they better represent the state of the network because they capture congestion propagation in time and space. Nonetheless, for many years those models only existed at the micro or mesoscopic scale making them time consuming when implemented for large-scale networks. Recently, the LICIT-Eco7 lab developed a dynamic traffic simulation platform, MnMs (Becarie et al., 2022), that works at the macroscopic level allowing the analysis of large-scale networks such as urban environments. It captures the spread of congestion phenomenon in time and space but also the multimodal compound of such transport systems. To do so the model uses the macroscopic fundamental diagram (Daganzo, 2007) to which has been added a third dimension to acknowledge multimodal traffic (Geroliminis et al., 2014).

Alongside the efforts made to better represent the supply, researchers developed tools to enrich the estimation of the demand. To achieve that, activity-based models have been developed. Those models consider that the trip demand is not inherent, but it is the result of the need to perform activities along the day. Therefore, trips are not independent as they are connected through a schedule. Moreover, if trip-based models often consider that user always choose the shortest path, activity-based models use more parameters to better capture the decision mechanism of users as the cost of the trip or the comfort of the chosen mode. Furthermore, those models, because they better represent the process of choosing a certain trip and allow the forecast of more complex scenarios by playing with more variables, are more suited to analyze transport policies than trip-based model (Zargayouna, 2021).

Dynamic traffic simulations and activity-based modeling have been investigated rather separately for a long time. Research have been led to integrate one with another, but some gaps are still to be field. Section 1.1 presents some of those integrated models, section 1.2 highlights the limitations in the research field and section 1.3 depicts the problem of the current study.

1.1. Integration of activity-based modeling and dynamic simulations

In this section I review different techniques that have been used to integrate activity-based modeling with traffic simulation. Section 1.1.1 presents the sequential integration of schedule choice and route choice. Section 1.1.2

introduce the use of a supernetwork. Section 1.1.3 is about departure time module. Section 1.1.4 introduces several equilibria that have been developed to solve those models. Finally, section 1.1.5 depicts how to extend those models.

1.1.1. Sequential integration of schedule choice and route choice

The first one, and the most direct one, consists in having a module that create schedules for people from disaggregated data (Bekhor et al., 2011; Halat et al., 2017; Javanmardi, 2012; Lin et al., 2008; Xu et al., 2017b, 2017a). Those schedules are then the core that generates trip demand, meaning that the route choice and the traffic simulation are done in a second module. Listed in Table 1 are articles in which the authors used that technic and some relevant characteristics of the model they developed.

Article	Route choice	Convergence criteria	Simulation's scale
(Lin et al., 2008)	Least Generalized cost path	OD matrices or travel times	Mesoscopic
(Bekhor et al., 2011)	Shortest Path	Not Explicit	Mesoscopic
(Javanmardi, 2012)	Shortest Path	Level of service and Input ABM	Microscopic
(Halat et al., 2017)	Hyperpath	Penalty inconsistency schedule and Generalized path cost	Micro/Mesoscopic
(Xu et al., 2017a)	Hyperpath	Generalized Cost Path	Mesoscopic
(Xu et al., 2017b)	Hyperpath	Utility	Mesoscopic

Table 1. Schedule module and Route Choice Module Integration

The fact that there are two units in the computation process means that they are two criteria for the convergence of the model or that there are two convergences in the process. The route choice can be represented through several calculations : shortest path (which is not the real process when looking at the data) (Bekhor et al., 2011; Javanmardi, 2012), least generalized cost path (travel time, gas price, toll, comfort…) (Halat et al., 2016; Lin et al., 2008; Zhang et al., 2018) or using the concept of hyperpath (least generalized perceived cost[1]) (Halat et al., 2017; Verbas et al., 2016; Xu et al., 2017b, 2017a). The advantage of such models is that most of them are easy to manipulate, therefore they can model complex behavior such as the correlation between modes within a day, or the interaction between several household members. Nevertheless, they have limitations. The first one would be that in order to keep such level of detail ones need a lot of disaggregate data and a lot of information about the characteristics of the population, those data not always exist depending on the area we are working with because of anonymization policies. The second would be that because the route choice and the traffic simulation are done in a second phase, there can be inconsistencies in the produced schedules. Specifically, in each iteration we first fix travel time to choose a schedule and then we fix the schedules in order to choose a route, hence being late or early at an activity is seen as an inconsistency because it does not follow the chosen schedule. Thirdly, in order to reduce those inconsistencies and because of the convergence issues of the two modules, the models would need a lot of iterations making them computationally burdensome, and in the end some inconsistencies are never

[1] perceived cost means that when ones computes it, they make an assumption on which factors the users will consider as part of the cost. For example, most individuals do not consider insurance fee when looking at the cost of a trip even if it is part of the cost of the car alternative.

solved. Fourthly, this type of integrations has been only implemented in microscopic and mesoscopic simulations to keep a high-medium level of detail and because aggregating the data to have a macroscopic simulation would increase the computation time.

1.1.2. Supernetwork to jointly choose schedule and route

The second method is based on the way one represents the transportation system in the network. Specifically, the system is converted into a network in which the nodes represent activities, and the links represent either a route (if connecting two nodes) or the duration of an activity (if connecting one node with itself). Each user chooses a set of activities in a global set of all feasible sets of activities. Each link generates either a utility, if related to an activity, or a disutility, if related to a trip. Therefore, when a user chooses a certain route, it naturally produces a schedule to which a utility is related. The aim of the model is to maximize this utility. References can be found in Table 2.

Article	Traffic Assignment	Convergence criteria	Added Value
(Ramadurai and Ukkusuri, 2010)	Dynamic	Utility	Dynamic Supernetwork
(Fu and Lam, 2018)	Static	Utility	Joint utility (homogenous household)
(Vo et al., 2020)	Dynamic (static equivalent)	Travel Time	Joint utility (heterogenous household)
(Vo et al., 2021)	Dynamic (static equivalent)	Travel Time	Mixed equilibrium (individual, household)

Table 2. Supernetwork Integration

One advantage of those types of models is that users only choose a set of activities, time is flexible, meaning that inconsistencies as described in section 1.4.1 does not exist. The preferred arrival time and duration of an activity are modelled by the utility related to the performance of that activity. This modelling is called supernetwork (Ramadurai and Ukkusuri, 2010). This natural integration allows to acknowledge both the utility of performing an activity and the disutility of traveling at the same time which includes exogenous factors such as gas price and tolls (Vo et al., 2020). Again, the simplicity of such models allows the modeling of correlation between modes within a day and leverage between household utility and individual needs (Fu and Lam, 2018; Vo et al., 2021, 2020). However, because of the fact that each activity has to be represented with a node and that each activity is linked to another through several links the level of knowledge needed to represent the network is huge and the data might not be available. It also induces an important computation time. To reduce the computation burden, researchers converted the dynamic problem into a static equivalent which does not correctly model the congestion phenomenon. Moreover, this kind of network hasn't been implemented in large-scale networks, most recent studies have only been implemented for the Sioux Falls network (Fig 1). Lastly, only few modes are acknowledged in those supernetwork (car, bus and metro).

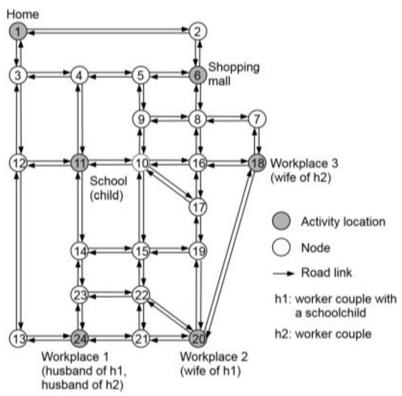


Fig 1. Sioux Fall Network from (source : Vo et al., 2021, 2020)

1.1.3. Departure time module within the dynamic assignment

The last method uses a departure time module that is directly integrated in the simulation (Cantelmo et al., 2018). This module can model departure time choice, route choice and mode choice at the same time. Both the utility of performing an activity and the disutility of traveling are acknowledged in a unique utility function which is the core of the module allowing a unique convergence criterion. In addition, the utility of the activity can model complex phenomena such as clock-based (changing over time) and duration-based (dependent of how long one performs the activity) utilities (Cantelmo and Viti, 2019). Literature can be found in Table 3.

Article	Activity utility	Travel disutility	Decision variables
(Cantelmo et al., 2018)	Not Explicit	Travel time and being early or late	Departure Time, mode choice and route choice
(Cantelmo and Viti, 2019)	Clock-based and Duration-based	Travel time and being early or late	Departure Time, mode choice and route choice

Table 3. Departure Time Module

Here schedule conflicts can arise but are seen as an additional disutility for being early or late rather than an inconsistency as discussed in section 1.4.1. Another advantage is the fact that those models have been developed for macroscopic simulations making them more efficient for urban environments. The integration between those modules uses the extension of the bottleneck proposed in Li et al., 2014. The issue with those integrations is that only few decision variables are in the models (departure time, route, mode) and the disutility of travelling only includes travel time and being late or early. Also, heterogeneity between users is not acknowledge, parameters for both the utility function of an activity and the disutility of travel are only trip/purpose dependent. Moreover, if the correlation between trips is correctly captured thanks to the activity duration, there can be inconsistencies in the mode choice within a day and there is no representation of the household members' interactions. Finally, exogenous factors such as the weather are not acknowledged in this model.

1.1.4. Developed equilibria

About the equilibria formulated in those integrations, in the first two types of integration it is mainly fixed-point problem (Flötteröd et al., 2012; Halat et al., 2017; Javanmardi, 2012; Verbas et al., 2016; Xu et al., 2017b) or variational inequality (Fu and Lam, 2014; Halat et al., 2016; Lin et al., 2008; Ramadurai and Ukkusuri, 2010; Vo et al., 2021, 2020), which might be converted into a fixed-point problem thanks to a GAP function. Those equilibria mainly focus on satisfying Wardrop's first principle (Wardrop and Whitehead, 1952). Concerning the third type the formulation is a complementarity problem (Cantelmo and Viti, 2019) which aims to satisfy the Nash equilibrium (Nash, 1950). Comparison of those equilibria are available in Table 4.

Article	Equilibrium	Solver	Type
(Flötteröd et al., 2012)	Fixed-point	Not Explicit	Route
(Javanmardi, 2012)	Fixed-point	Not Explicit	User
(Verbas et al., 2016)	Fixed-point	Not Explicit	Wardrop's 1st Principle
(Halat et al., 2017)	Fixed-point	Method of Successive Average	Micro/Mesoscopic
(Xu et al., 2017b)	Fixed-point	Method of Successive Average	Mesoscopic
(Lin et al., 2008)	Variational Inequality	Method of Successive Average	Wardrop's 1st Principle – User

(Ramadurai and Ukkusuri, 2010)	Variational Inequality	Gradient Projection Algorithm	Wardrop 1st Principle – User
(Fu and Lam, 2014)	Variational Inequality	Not Discussed	User
(Halat et al., 2016)	Variational Inequality	Quasi-Newton Algorithm and Method of Successive Average	Wardrop 1st Principle
(Vo et al., 2020)	Variational Inequality	Diagonalization and Method of Successive Average	Household Optimum and Household System Optimum
(Vo et al., 2021)	Variational Inequality	Diagonalization	Mixed Equilibrium (Individual, Household)
(Cantelmo and Viti, 2019)	Complementarity problem	Method of Successive Average	Nash Equilibrium

Table 4. Equilibria

For solving those equilibria, the most common solver is the method of successive average (Cantelmo et al., 2018; Cantelmo and Viti, 2019; Halat et al., 2017, 2016; Lin et al., 2008; Vo et al., 2020; Xu et al., 2017b). It might be paired with a quasi-Newton algorithm (Halat et al., 2016), a diagonalization (Vo et al., 2021, 2020) or a gradient projection algorithm (Cantelmo et al., 2018) which can also be used alone (Ramadurai and Ukkusuri, 2010). Others use systematic relaxation algorithm (Bekhor et al., 2011).

1.1.5. Extending the models

As mentioned earlier, transport models aim to ensure the resilience of transportation systems, assist decision-makers in creating new public policies, or even help them for urban planning. To achieve this, these models must be robust and therefore be regularly updated, for example, by incorporating the new forms of mobility we have already discussed. In section 1.2.5.1, I provide more detailed information about these new mobility forms. Section 1.2.5.2 introduces the existence of two types of data and explains how combining the two allows for the inclusion of these new ways of getting around.

1.1.5.1. Extension of dataset and reduction of data needs

The new forms of mobility are diverse. As previously mentioned, they can manifest either as new modes of transportation or as new uses of traditional modes of transportation. Research has been conducted to understand the impact of these forms on the current modal split to see their share but also to understand from which conventional modes sharing system users are coming from (Manout et al., 2023). Others worked on understanding how those new services are used (Reck et al., 2021). In the meantime, some studies have been led to highlight the key factors that are influencing the utility functions of those alternative in mode choice modeling (Baek et al., 2021; Krauss et al., 2022)). When used in the context of mode choice models, shared mobility systems are often considered modes used for short distances or as secondary options for first or last-mile connectivity. However, shared bicycles are used for a multitude of distances, serving as a possible alternative to more traditional modes even for medium-distance trips (Reck et al., 2021). As for electric scooters, on their own, they may not be a competitive alternative for such distances (Reck et al., 2021). Nevertheless, with the emergence of Mobility as a Service platforms, electric scooters from shared services could very well be used for medium-distance trips (Meng et al., 2020). However, to my knowledge, no study has been conducted in this direction.

1.1.5.2. Revealed-preferences and stated-preferences

Revealed-preferences data refers to dataset that include data about trips that have been done. Surveys to obtain such data often contain a set of questions to report the trips a household did during a determined period (in general households are asked to report the trips they made the day before the survey). The survey often includes questions about the purpose of the trip and socio-economic characteristics of each member of the household.

Stated-preferences data refers to dataset that include preference over several alternatives that don't necessarily exist. Surveys to obtain such data often contain a set of different situations and propose different alternatives in each situation (e.g., a person is proposed different transportation option to make a trip and depending on the information they have on each option they must choose which one suits them best).

The studies mentioned before include revealed-preferences dataset as inputs to estimate the current demand on the current network. However, activity-based modeling can also be used to estimate the impact of a new service on the network. Because revealed-preferences dataset don't include information about a service that does not exist yet, stated-preferences are preferred to estimate new modal shares. The issue with stated preferences is that they are biased. Studies show that there is a difference between stated-preferences and revealed-preferences. To overcome this issue it is common to use both dataset and to link them using a scale parameter for their standard deviations (Dissanayake and Morikawa, 2010; Dosman and Adamowicz, 2006; Kim et al., 2022; van Essen et al., 2020; Weis et al., 2021). In most recent studies, researchers proposed to build the stated-preferences survey based on the revealed-preferences answers of people to facilitate the response and to reduce bias from the beginning (Rudloff and Straub, 2021; Weis et al., 2021). However, some recent studies tried to use revealed preferences to estimate the new modal shares (De Clercq et al., 2022). To do so, they first estimate the modal share for the alternatives that are present in the current revealed preference dataset and then they define the attributes for the new modes. Next, they compute how close the new modes of transportation are to the current modes. After, they use this closeness to define scale parameters and compute the new utility functions using those scale parameters. The issue with that methodology is that it hasn't been compared with a classic methodology using stated preferences data so it's hard to tell if the estimates are reasonable. In conclusion, to the state-of-the-art, the use of stated preferences data is still the best way to implement new modes in transport models.

1.2. Research gap

Although some major advances have been made in the integration of activity-based model and dynamic traffic simulation some limitations remain:

- **Research Gap 1:** Bike sharing system and e-scooter sharing system are not included together with conventional modes in mode choice modeling for medium-trip distance.
- **Research Gap 2:** Few studies acknowledge the existence of new mobilities in the mode choice process.
- **Research Gap 3:** Models that have been implemented in macroscopic simulation do not include mode consistencies within a day.
- **Research Gap 4:** If some studies solved modes consistencies, to my knowledge, none focused one the correlations between mode choice within a day.
- **Research Gap 5:** Integrated models at a macroscopic scale only acknowledge a few decision variables (departure time, route choice and mode choice).
- **Research Gap 6:** Exogenous factors, such as the weather or the weekday, are rarely acknowledged in the models; only gas price and tolls are considered in supernetworks and schedule module.
- **Research Gap 7:** Because of their complexity, recent models are still computationally burdensome, which might discourage policymakers that manage large-scale network from using them.

1.3. Research objective

This study aims to address research gap 1. The idea is that in order to integrate an activity-based model with dynamic traffic simulation we first need to have a robust mode choice model in which new mobility is included. I will benefit from this work to also address research gap 2, research gap 4 and research gap 6 by adding these factors in the choice model for future integration with traffic simulation. This means that the research objective is to model mode choice with new mobilities for medium-distance trip while acknowledging potential obstacles and constraints from the return trip. The model will be used to observe changes in the modal split and to see what are the key factors that could support an increase in the shares of emerging forms of mobility. To do so I intend to create a model by using an SP-experiment which will help by bringing new forms of mobilities and new variables that aims to represent obstacles and correlation related to each alternative. The SP-study will be restricted to home-work mid distance trip for people that work or study at the ENTPE. Mid-distance means that all origins will be in the Grand Lyon area and the only destination is the ENTPE. Hence the questions I will try to answer are:

Research question 1: Which factors influence the mode choice with emerging forms of mobility?

Research question 2: Do the information that one possesses on the return trip influence the inward mode choice?

The methodology for this study is as follows:

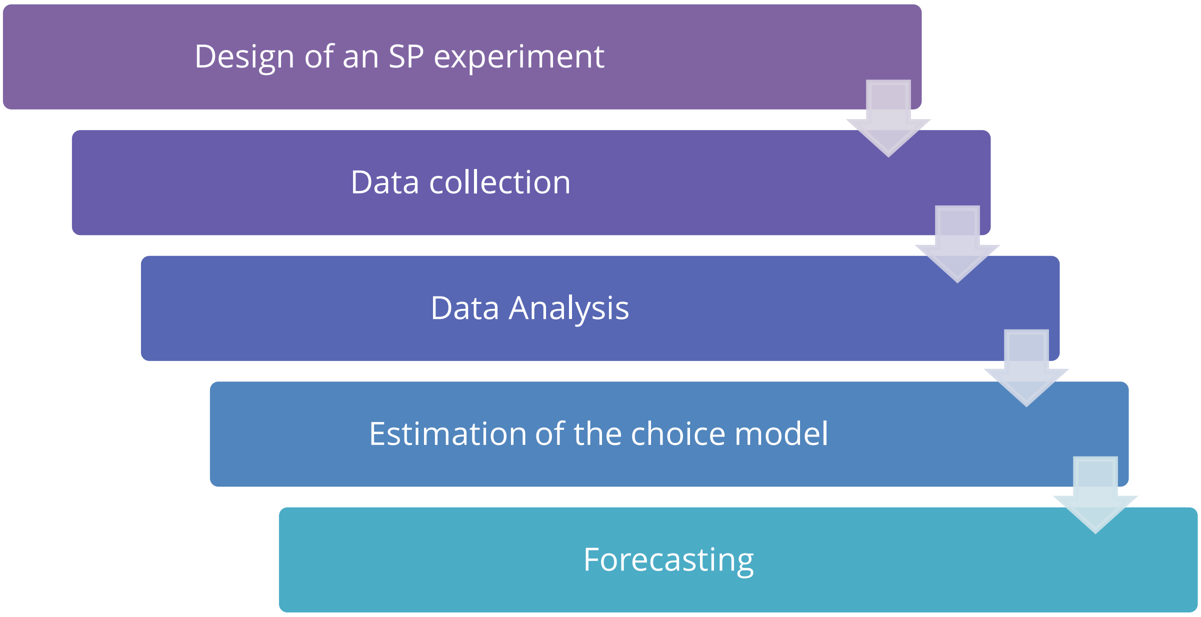


Fig 2. Methodology for the present study

The remaining of this report is organized as follows: Section 2 details the methodology that allowed me to design an efficient SP-experiment. Section 3 analyses the collected answers. Section 4 focuses on the mode choice model and the estimation of their parameters. Section 5 presents some scenarios to understand how new mobility could be more attractive. Finally, section 6 presents my conclusion on this study, its limitations and future work.

2. Material and methods

The population whose mobility habits I wish to model is the population of the Lyon metropolitan area. The most recent data concerning this population are those from the 2015 household travel survey. Unfortunately, this "revealed preferences" data does not contain all the forms of mobility I wish to represent, in particular, it does not include self-service electric scooters. What's more, this data is 8 years old, and the habits of the people of Lyon have probably changed since then.

One solution would be to collect a stated preferences (SP) survey containing the alternatives I want to model on a smaller sample such as the ENTPE population. Section 2.1. presents the trips chosen for the questionnaire. Section 2.2. introduce the different attributes that will be used to describe these trips. Section 2.3 details the socio-economic variables that will be collected. Finally, section 2.4. describe the method and tools used for the design of the survey.

2.1. Origin, destination and modes

To carry out this survey, I need to describe the situations I am going to present in it. First, I chose the trip I wanted to represent in our questions. I chose the Home-ENTPE trip. To facilitate the respondents thought process, I chose to design the survey using realistic trips. I used Google Maps to obtain different levels for each attribute. To ensure a certain variability in attribute levels, I chose 3 possible home locations:

Villeurbanne (53 rue Anatole France, 69100 Villeurbanne)

Lyon - Centre-Ville (69-59 Bd des Canuts, 69004 Lyon)

Banlieue (36 Rue Romain Rolland Bis, 69200 Vénissieux)

In order to obtain levels as close to reality as possible, I took the attributes of the home - ENTPE routes for an arrival at 9 am on a Thursday morning according to Google Maps forecasts. Once I had the trips, I had to choose which transport alternatives would be available for each trip.

In Lyon, the available modes are walking, private car, public transportation, personal bicycle (electric or not), personal scooter (electric or not), other private micro-mobility (roller skates, skateboard...), bike sharing-system (Vélo'v) (electric or not), e-scooter sharing system (Dott or Tier) and car sharing system.

Because of the distance between origin and destination (3.97 km, 7.37 km and 9.37 km as the crow flies), walking and e-scooter sharing-system have been excluded from the options as stand-alone modes. In order to include them, they are integrated with the public transport mode, giving us the first two alternatives in the survey: Walking/Public Transport/Walking (PT) and shared e-scooter/Public Transport/Walking (ESSPT). To simplify the questionnaire, I assumed that e-scooters were available at each origin, which is not the case (Dott, 2023; Tier, 2023). I have ruled out the possibility of completing the trip by shared e-scooter, given the short distance between the nearest stop and ENTPE (450m).

The other three alternatives are the private car (PC), the personal bicycle (PB) and the bike sharing-system (BSS). The combination of bike sharing-system and public transport is not considered, as the distance between the starting point and the nearest metro station, or the distance between the starting point and the start of the second section of the public transport route (around 1.5 km), is too short for this option to be relevant (Reck et al., 2021).

I have chosen to exclude the car sharing-system alternative. First, because I need to minimize the number of alternatives in order to optimize the number of responses required. Second, I decided to exclude this one, on the assumption that this service is used more when there are loads to carry or strong time constraints, and therefore it is used on an occasional basis in the context of a home - ENTPE trip.

2.2. Attributes of the alternatives

Once the alternatives had been chosen, I had to determine the attributes (i.e., characteristics) of each alternative. Based on previous studies, I assumed that mode choice is influenced by in-vehicle travel time[2] (TT), out-of-vehicle time (walking and waiting combined) (OT), number of transfers (NbT) and cost[1].

In addition, I investigated potential obstacles to mobility services: Availability at Arrival (Avail) for the BSS option and Rain on Arrival (RA) for the PT option. To select these variables, I drew on empirical research conducted on obstacles to shared mobility (Shui and Szeto, 2020). I considered three other variables to understand whether the information about the return trip can influence the mode chosen for the inward trip. These variables were the following ones: weather forecast for the return trip (WR), parking search time for the return trip (PST), and additional time spent in traffic jams by car during the return trip the day before (CT).

All times are expressed in minutes and the costs in euros. The number of changes is an integer number, and availability at arrival is expressed as the number of spots remaining at the "VAULX- Hôtel de Ville" Velo'v station at the time the decision is made. Rain during the inward journey is a binary variable equal to one when it rains, weather forecast on the return trip is an integer between 0 and 100 expressing the percentage of the rain probability on the return trip, and extra time due to traffic jams the day before is a number between 0 and TT*0.5.

To define the levels of these attributes, I mostly used Google Maps, following the methodology described earlier. For the ESSPT alternative, I had to build several alternatives in order to have variability. For the departure from Vénissieux, there are 2 possible stops to start the PT trip. To calculate 2 alternatives, I took the distance between these 2 stops and the origin. For Lyon and Villeurbanne, I took the distance to the stop closest to the origin and the distance to the first transfer (Hôtel de Ville for Lyon and Laurent-Bonnevay for Villeurbanne) to calculate our 2 alternatives. I then calculated the travel time using a commercial speed of 15km/h (Manout et al., 2023), to which I added the characteristics of the public transport routes departing from each of these points.

Costs were calculated using the fares available on the transport providers' websites (Dott, 2023; JCDecaux and Lyon Métropole, 2023; SYTRAL, 2023; Tier, 2023). In order to have several levels, I considered prices with and without subscriptions, using the corresponding fares which are also available on the websites of these services[3]. For cars, I took the distance available on Google Maps, the average consumption of diesel (resp. petrol) for French vehicles (Statista Research Department, 2023) and multiplied it by the average price of diesel (resp. petrol) (Global Petrol Prices, 2023) to obtain two price levels for each origin. Availability at arrival varies between 0 and 30, 30 being the maximum number of free spots at the station. The possible parking search times are 0, 6 and 12 min. 0 corresponding to having a parking at home or an available spot at the nearest Vélo'V station for the BSS option, and 6 and 12 min corresponding to the search time according to the density of the zone considered.

Attributs	Units	Levels		
RI	Binary	0 1		
WR	Binary	0 10 50 90		
$Origin$	City	$Villeurbanne$	$Lyon\ city\ center$	$Lyon\ suburbs$
TT_{PV}	[min]	16 26 35	20 30 40	18 29 40
TT_{PT}	[min]	13 16 22	19 25 31	39 52 65
TT_{PB}	[min]	20 22 24	34 36 38	45 48 51
TT_{BSS}	[min]	18 20 22	30 34 38	43 46 49

[2] For the ESSPT option, travel time and cost are divided by mode (i.e. one TT and one cost for ESS and another for TC, TT_{ESSPT} may differ from TT_{PT}).

[3] Only extra cost was considered and not subscription fares

TT_{ESS}	[min]	2	12	1	8	4	8
TT_{ESSPT}	[min]	13 16 22	8 10 12	30 35 40	19 23 27	61 63 65	39 41 44
$COST_{PV}$	[€]	0.74 0.86		1.25 1.46		1.42 1.66	
$COST_{PT}$	[€]	0 2					
$COST_{BSS}$	[€]	0 1.8		0 1.8 0.2 2 0.4 2.2		0.65 2.45 0.8 2.6 0.95 2.75	
$COST_{ESS}$	[€]	0.5 1.5	3 4	0.25 1.25	2 3	1 2	2 3
PST_{PV}	[min]	0 6 12					
PST_{PB}	[min]	0 6 12					
PST_{BSS}	[min]	0 6 12					
OT_{PT}	[min]	2 11 20					
OT_{BSS}	[min]	2 11 20					
OT_{ESSPT}	[min]	2 11 20					
NbT_{PT}	Number	0 1 2					
NbT_{ESSPT}	Number	0 1 2					
CT_{VP}	[min]	0 0.1 0.3 0.5					
$Avail_{BSS}$	Number	0 3 12 30					

Table 5. Summary table of the trip-related attributes and their levels

2.3. Socio-economic characteristics of the respondents

Collecting socio-economic data from respondents is important for several reasons. Firstly, these data enable us to check that our sample is representative of the population surveyed, and in the event that it is not, to correct the results using weights. Secondly, these characteristics enable the study oof differences in the decision-making mechanisms of different groups in the sample. To allow a direct comparison with the EMD 2015 survey, I have used the socio-economic variables they used. Thus, I collected information regarding age, last diploma obtained[4], employment[3], possession of a driving license, vehicle ownership, type of dwelling occupation, number of adults and minors in the household, respondent's main modes of transport[3], respondent's secondary modes of transport[3] if any, area of residence[5] and possession of a transport season ticket.

	-1	Missing value
Socio-economic variable		
ID	Respondent’s ID	
GENDER	Respondent’s Gender	
	0	Man
	1	Woman
	2	Non-Binary
AGE	Respondent’s age (years)	
EDUC_LEV	Last diploma	
	0	BAC

[4] For these questions, the possible answers have been adapted to the ENTPE population.

[5] For zone of residence, we have aggregated the EMD 2015 zones of residence by arrondissement for Lyon and by commune for the others.

	1	DUT/BTS
	2	LICENCE
	3	MASTER/Engineer/MAITRISE
	4	PHD/AGGREG
	5	HDR
JOB	occupation	
	0	Student
	1	Student civil servant
	2	Technician and other
	3	Administrative profession
	4	Scientific profession
DRIVER_LICENSE	Owning of driver's license	
	0	No
	1	Yes
NB_CAR	Number of cars in the household	
NB_BIKE	Number of bike in the household	
RESIDENCY_TYPE	Housing type	
	0	Owner
	1	Rental: Social Housing/Student's residence
	2	Rental: Other
	3	Housed for free
NB_ADULTS	Number of adults in the household	
NB_MINOR	Number of children in the household	
NB_2RM	Number of motorcycles in the household	
NB_SCOOTER	Number of scooters in the household	
NB_PROVEH	Number of professional vehicles in the household	
MM	Main modes	
	1	Is in the combination
	0	Is not
MM_W	Walk	
MM_C	Car	
MM_TC	Transit	
MM_PMM	Private micro-mobility (Bike/Scooter)	
MM_SS	Sharing-system	
MM_O	Other modes	
SM	Secondary modes	
	1	Is in the combination
	0	Is not
SM_W	Walk	
SM_C	Car	
SM_TC	Transit	
SM_PMM	Private micro-mobility (Bike/Scooter)	
SM_SS	Sharing-system	
SM_O	Other modes	
DISTANCE	Estimation of the distance between the house and the ENTPE	
TYPE	Residency zone type	
	1	City
	2	Suburb

	3	Countryside
SUB	Subscription	
	0	Don't have it
	1	Have it
SUB_TC	Transit	
SUB_BSS	Bike sharing-system	
SUB_CSS	Car sharing-system	
SUB_ESS	E-scooter sharing-system	
SUB_O	Other modes	

Table 6. List and description of the socio-economic variables

2.4. Survey design

Once the attributes and socio-economic variables had been determined, the choice situations had to be created. My aim was to obtain the maximum amount of information with the minimum number of answers. To do this, I decided to adopt an efficient design for the survey, and to find an efficient design I used Ngene software (ChoiceMetrics, n.d.). Ngene is a software package which, starting from a model specification, selects the attribute levels of each alternative in such a way as to minimize the chosen criterion. In this case, I chose the "S-optimality" criterion (Eq. 1) which is a necessary condition to have significant parameters (unfortunately it is not sufficient).

$$S - optimality = \max_{k \in K}(\left(\frac{se_1(\beta_k)t_\alpha}{\beta_k}\right)^2)$$

Eq. 1 Equation of the S-optimality used in Ngene (Rose and Bliemer, 2013)

Where K is the set of parameters' names, $se_{1,k}$ is the asymptotic standard error of parameter k with a sample size of 1, X is the vector of attributes, β is the vector of parameters, t_α is the t-value corresponding to the $(1 - \alpha)$ confidence interval.

To operate, Ngene needs the type of model, the number of choice situations to be generated, the number of situation blocks, the model's utility functions, the possible levels for each attribute, and an estimate of the model's parameters, called the priors.

Regarding the type of model, I will use logit models. Those models provide a framework that aims at reproducing people's choice mechanics. To do so logit models use utility functions (Eq. 2) which are supposed to describe the different factors that influence those behaviors using explanatory variables (Ben-Akiva and Lerman, 1985). In our case the possible explanatory variables are the trip attributes, as described in section 2.2., and the socio-economic variables, as described in section 2.3. Because there is multiple alternatives the logit model used is a multinominal logit model (MNL).

$$\forall i \in I, \forall j \in J, \forall k \in K \;\; U_{ijk} = ASC_j + \beta_j * X_{ijk} + \epsilon_{ijk}$$

Eq. 2 Utility function for logit models

where I is the set of individuals, J is the set of alternatives, K is the set of choice situations, U_{ijk} is the utility function of individual i for alternative j and choice situation k, ASC_j is the alternative specific constant of j, β_j is the vector of parameters to be estimated, X_{ijk} is the vector of explanatory variables of individual i and alternative j in choice situation k and ϵ_{ijk} is the logistic-distributed error term.

Once the parameters are estimated, it is possible to compute the probability of choosing a certain alternative (Eq. 3).

$$\forall i \in I, \forall j \in J, \forall k \in K \;\; P_{ij} = \frac{e^{ASC_j + \beta_j * X_{ijk}}}{\sum_{m \in J} e^{ASC_m + \beta_m * X_{imk}}}$$

Eq. 3 Formula for computing the probability of choosing an alternative with a logit model where I is the set of individuals, J is the set of alternatives, P_{ij} is the probability that individual i choose alternative j, ASC_j is the alternative specific constant of j, β_j is the vector of parameters to be estimated, X_{ij} is the vector of explanatory variables of individual i and alternative j.

There is a possibility that when choosing the explanatory variables, I didn't think of every factor that actually influence people's choice and so that it exists a correlation between repeated observations over time that cannot be captured with an MNL. Hopefully, advanced choice modeling techniques such as the mixed-logit model can capture these correlations using individual specific error (Train, 2009). In classic logit models the preference of individuals regarding the different alternatives is captured by the alternative specific constant (ASC) which is the same for every individual. In the mixed logit model with individual specific error, the error components aim at allowing the ASC to be individual specific meaning that each individual can have their own mode preference. When using that choice model, I will use Eq. 4.

$$\forall i \in I, \forall j \in J, \forall k \in K \;\; U_{ijk} = ASC_j + \beta_j * X_{ijk} + \sum_{m \in J} \gamma_{jm} * \sigma_{mi} + \epsilon_{ijk}$$

Eq. 4. Utility function used for capturing individual's heterogeneity

where I is the set of individuals, J is the set of alternatives, U_{ij} is the utility function of individual i for alternative j, ASC_j is the alternative specific constant of j, β_j is the vector of parameters to be estimated, X_{ij} is the vector of explanatory variables of individual i and alternative j for choice situation k, γ_{jm} is the parameter of alternative j that captures the impact of the error term σ_{mi}, σ_{mi} is the individual specific error components of alternative m that aims at representing the distribution of individual preferences for that alternative and ϵ_{ijk} is the logistic-distributed error term.

In Ngene, we first estimated a MNL and then the mixed-logit model. However, the number of necessary responses estimated by Ngene for the ML was greater than the number of people at ENTPE. I therefore considered the MNL specification for the design of the study. I chose to generate 30 choice situations (minimum possible for the specified model) and to have 5 blocks of 6 questions. Ngene does not allow for socio-economic variables. Utility functions (Eq. 5) are available below.

$$U_{Ngene_{PV}} = \beta_{TT\,PV} * TT_{PV} + \beta_{cost} * COST_{PV} + \beta_{PST\,PV} * PST_{PV} + \beta_{RI\,PV} * RI * TT_{PV} + \beta_{CT\,PV} * CT_{PV} * TT_{PV}$$

$$U_{Ngene_{PT}} = ASC_{PT} + \beta_{TT\,PT} * TT_{PT} + \beta_{cost} * COST_{PT} + \beta_{OT\,PT} * OT_{PT} + \beta_{NbT\,PT} * NbT_{PT} + \beta_{RI\,PT} * RI * TT_{PT} + \beta_{WR\,PT} * WR$$

$$U_{Ngene_{PB}} = ASC_{PB} + \beta_{TT\,PB} * TT_{PB} + \beta_{PST\,PB} * PST_{PB} + \beta_{RI\,PB} * RI * TT_{PB} + \beta_{WR\,PB} * WR_{PB}$$

$$\begin{aligned} U_{Ngene_{BSS}} = {} & ASC_{BSS} + \beta_{TT\,BSS} * TT_{VLS} + \beta_{cost} * COST_{BSS} + \beta_{OT\,BSS} * OT_{BSS} + \beta_{PST\,BSS} * PST_{BSS} + \beta_{Avail\,BSS} * Avail_{BSS} \\ & + \beta_{RI\,BSS} * RI * TT_{BSS} + \beta_{WR\,BSS} * WR \end{aligned}$$

$$\begin{aligned} U_{Ngene_{ESSPT}} = {} & ASC_{ESSPT} + \beta_{TT\,ESS} * TT_{ESS} + \beta_{TT\,ESSPT} * TT_{ESSPT} + \beta_{cost} * (COST_{ESS} + COST_{TC}) + \beta_{OT\,ESSPT} * OT_{ESSPT} \\ & + \beta_{NbT\,TC} * NbT_{ESSPT} + \beta_{RI\,ESS} * RI * TT_{ESS} + \beta_{RI\,TC} * RI * TT_{ESSPT} + \beta_{WR\,ESSPT} * WR \end{aligned}$$

Eq. 5 Utility functions used for the design in NGENE

To construct the priors, I had the choice between using assumed value or a distribution with a MonteCarlo simulation but to keep it simple I have chosen to use the values found in the literature. The issue is that it was quite hard finding one that could be use. As mentioned in section 1.1.5.1. most studies conducted on BSS and ESS consider them as a full option for a short distance trip or as a first or last mile option and I couldn't find a study in which one of them was combined with PT (Baek et al., 2021; Krauss et al., 2022; Reck et al., 2022, 2021). I finally chose to use the values of (Krauss et al., 2022) (except for RI, WR and CT) because it was the closest to my own in terms of alternatives and in variables. However, there are still some differences with the current study. First, their survey was conducted on a different population. Here I focused on members of the ENTPE so mostly students, researchers, and people from administration, in their study they interrogated people from several cities of Germany and did not focus on a specific part of the population. Then, their experiment was divided in two, one for short distance (0.5km to 4km) and another for medium distance (2 to 20km), population was randomly divided in two so that each groups answer only for short distance or medium distance. Finally, they have more alternatives; for short distance they had e-scooter sharing-system, bike sharing-system, walk and private car; for medium distance they had car sharing-system, ride pooling, transit and private car, meaning that it also did not have a private bike option in both cases. Also, I used their parking search time estimate even though it was for the inward trip and mine is for the return trip because I couldn't find any studies that integrates variables of the return trip in the inward trip utility function. RI was taken from (Reck et al., 2022) but I transformed distance*precipitation in TT*precipitation. For WR and CT I just assumed an estimate based on the other because I could not find any studies with variables that were like those ones. All priors can be found in Table 7.

Parameter	Prior	S-optimality	Parameter	Prior	S-optimality
ASC_{PT}	-1.205	NA	$\beta_{PST\,PB}$	-0.03	32.15
ASC_{PB}	-1.1	NA	$\beta_{PST\,BSS}$	-0.03	32.15
ASC_{BSS}	-1.199	NA	$\beta_{OT\,PT}$	-0.035	27.58
ASC_{ESSPT}	-1.39	NA	$\beta_{OT\,BSS}$	-0.035	29.38
$\beta_{TT\,PC}$	-0.057	5.69	$\beta_{OT\,ESSPT}$	-0.035	32.08
$\beta_{TT\,PT}$	-0.065	3.25	$\beta_{NbT\,PT}$	-0.298	26.90
$\beta_{TT\,PB}$	-0.09	4.81	$\beta_{CT\,PC}$	-0.04	2.86
$\beta_{TT\,BSS}$	-0.09	5.04	$\beta_{RI\,PC}$	0.022	32.13
$\beta_{TT\,ESS}$	-0.116	24.13	$\beta_{RI\,PT}$	0.022	21.75
$\beta_{TT\,ESSPT}$	-0.057	6.81	$\beta_{RI\,PB}$	-0.04	11.73
β_{cost}	-1.886	0.74	$\beta_{RI\,BSS}$	-0.02	27.81
$\beta_{PST\,PC}$	-0.03	32.13	$\beta_{RI\,ESS}$	-0.371	9.38
$\beta_{WR\,BSS}$	-0.02	5.06	$\beta_{WR\,PT}$	-0.01	11.09
$\beta_{WR\,ESSPT}$	-0.01	20.09	$\beta_{WR\,PB}$	-0.04	2.10
$\beta_{Avail\,BSS}$	0.06	5.22			

Table 7. Summary table of parameters, their retained estimates and the S-optimality criterion related to each

The highest S-optimality criterion is 32.15, which means I need a minimum of 33 respondents per block, or 165 respondents in total. The complete table of choice situations is available in Appendix 1. For the distribution of the survey, I used the Qualtrics Metrics tool ("Qualtrics," n.d.). The questionnaire comprises two parts. The first collects the respondent's socio-economic data. The second is the SP manual users, and the SP experiment. An example of a choice situation is available in Appendix 2.

In order to obtain a homogeneous number of responses per block, a counter was set up to generate an equal number of blocks across the entire sample. To force respondents to read the question, and thus avoid them clicking in the same spot throughout the survey (in the event that they always wish to use the same mode of transport), the order of the mobility options was randomized for each choice situation. However, for reasons of readability and due to the limitations of the software used, the order of questions within each block could not be randomized. Thus, there may be a bias due to the user's fatigue when completing the survey.

The questionnaire was sent to all members of the ENTPE via internal e-mail addresses. To reduce the number of responses required, I decided to have each participant complete 2 blocks of 6 questions. In total, 398 respondents participated in the survey. 292 responses were 100% complete, and 33 contained only the person's socio-economic information (these will not be considered as they contain no information on the SP experiment), giving a response rate of around 33%, which is close to the average response rate for surveys sent by e-mail (Aubagna, 2021). Finally, each block obtained at least 143 responses (complete or incomplete), which is well above the minimum requirement of 33 (+300%).

3. Results and statistics

Once the data has been cleaned up, we obtained 3,619 observations from 332 separate respondents. In this section, we present the descriptive statistics of the survey. Section 3.1. describes explains mobility habits of ENTPE members for the home-work trip. Section 3.2. compares mobility habits and survey answers to show rationality in people's answers. Section 3.3. presents the descriptive statistics regarding the socio-economic variables. Finally, section 3.4. highlights the fact the sample I obtain was not representative and shows how adjustments were made.

3.1. Mobility habits for the Home – ENTPE trips

In the socio-economic parts of the survey question 36 was: "In your everyday life when you need to go from your home to the ENTPE which combination of modes do you mostly take" and question 37 was about the second mostly taken combination. I thought it could be interesting to look at those data to see if people kept on with their habits or if they tried new things during the experiment. Results are displayed in Fig 3.

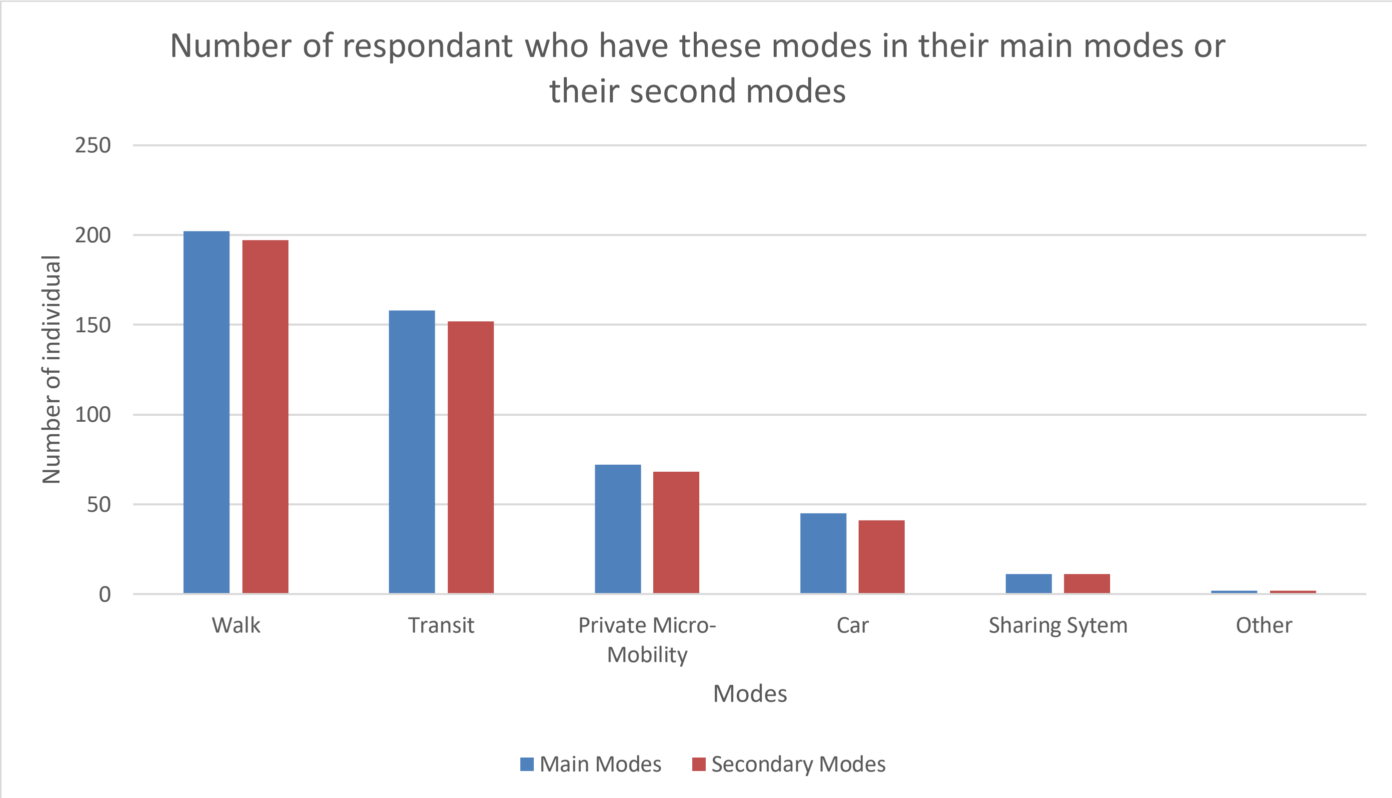


Fig 3. Main modes and secondary modes of surveyed individuals

First, what we see is that people that do have a secondary combination of modes have something that is really similar to their main one. Then, we note that there is a strong tendency to include walk to the combination. The main reason of that is that a lot of people leave near the ENTPE (see Fig 4). Another reason is that when taking another mode such as transit or sharing-system you must combine it with walk in order to get in the network. Then we have transit which again makes sense because there is a bus stop which is 500m away from the school. Moreover, the two other main residency area are Villeurbanne (3.6 km in Fig 4) and Lyon (6.9 km in Fig 4) which are well connected through the TCL network. After is private micro-mobility which can be explained by the fact that there are secured infrastructure for bike on almost all trip from Villeurbanne or Lyon, also there is a lot of parking infrastructure in the campus to encourage members to use their bike. Next, we have car which can be justified by the fact that most answers are from students (see Table 10) which mostly don't own a car compared to staff members (see Fig 5 and 6). Then come sharing-systems which tells that today for a home – work trip this option is not really considered. However, if we look at the number of subscriptions per modes (see Fig 7) we see that almost 1/3 of the respondents actually have a pass for bike sharing-system meaning that they might use them for other purposes. So, we can expect that in the SP-experiment there is a pattern due to people's practice in their everyday lives. New forms of mobility might not have a strong share in the developed model because of the fact that we are investigating a home – work trip but hopefully I'll be able to capture why with the new variables and hence create scenarios to find policies that might change that.

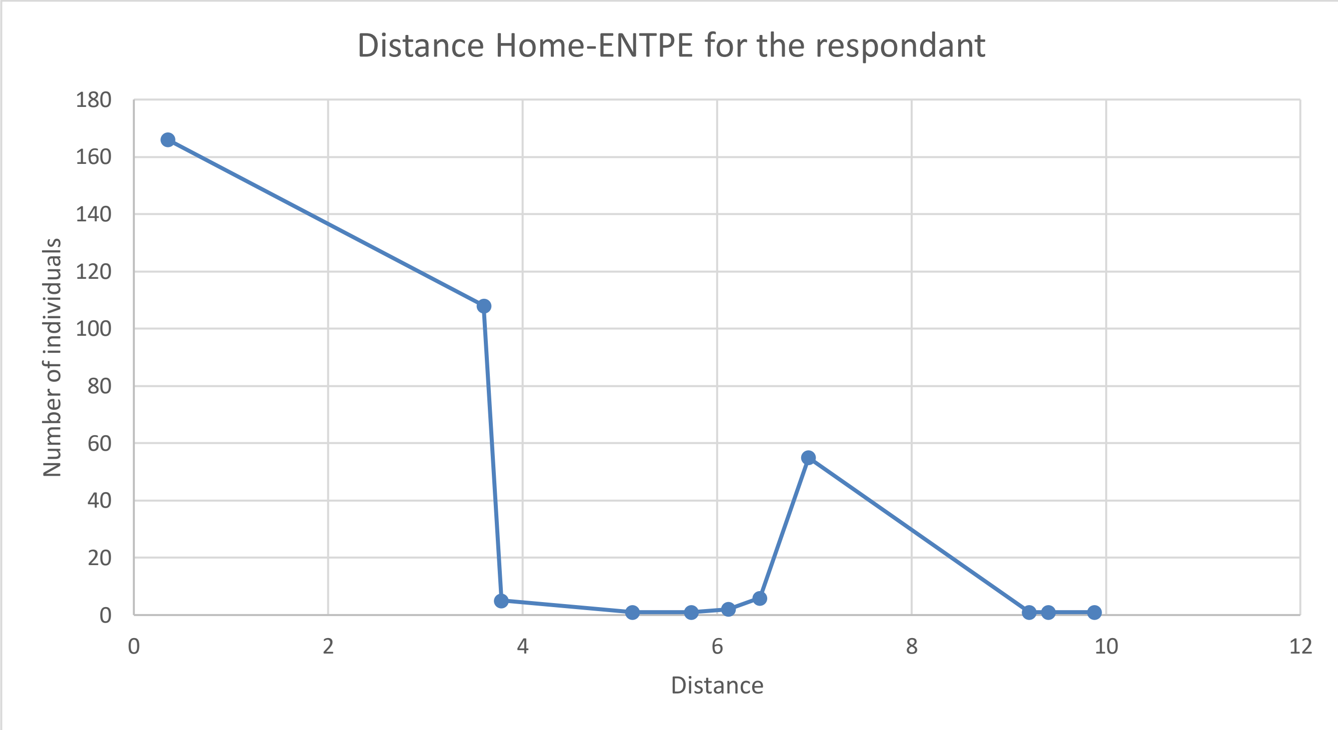


Fig 4. Distance Home-ENTPE for the people who answered the survey[6]

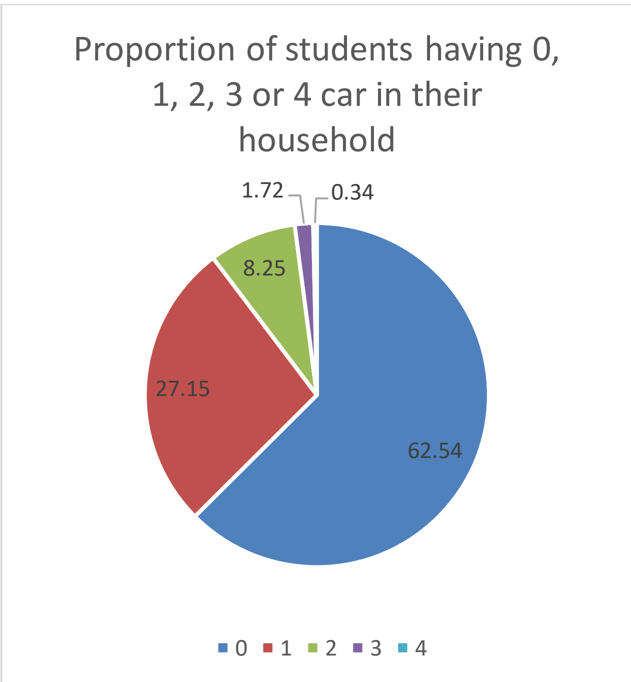


Fig 5. Proportion of student and civil-servant student per NB_CAR

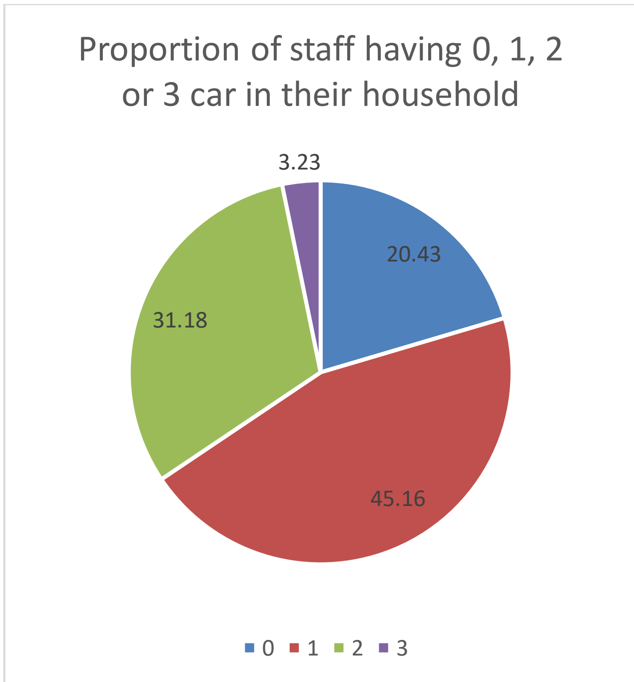


Fig 6. Proportion of staff per NB_CAR

[6] I stopped at 10km to make the graph readable but in reality, 18 other respondents live between 10km and 100km away from the school.

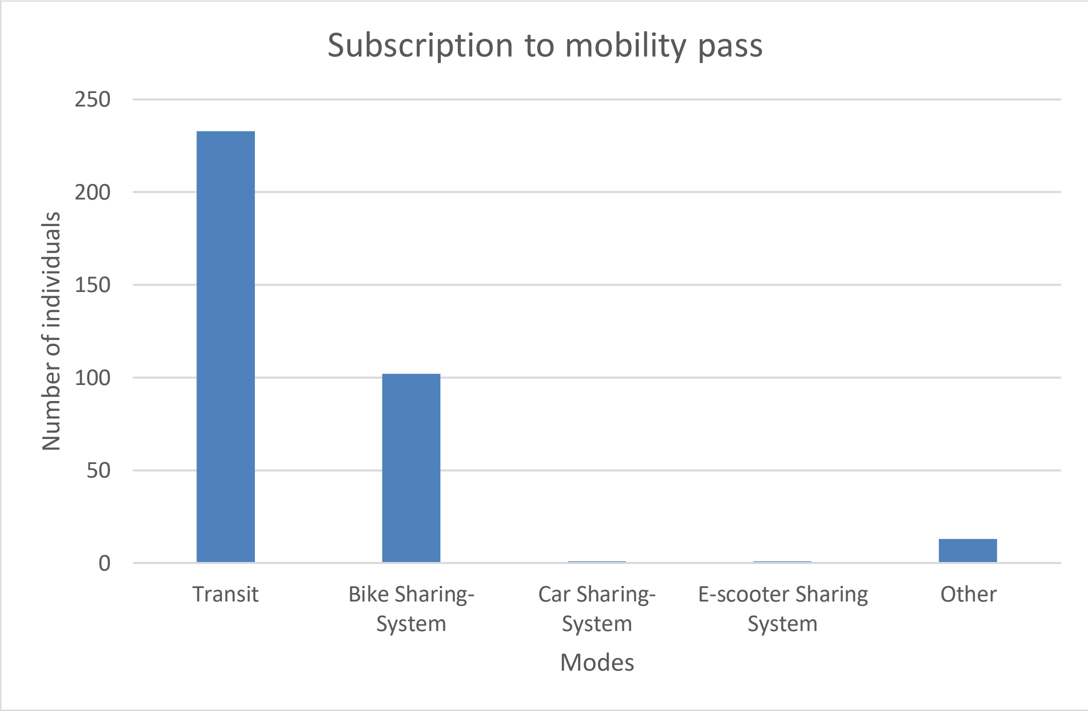


Fig 7. Number of subscriptions per mode

3.2. Choice in the SP-experiment

Now that we have seen how members of the ENTPE make their trips we can compare it to people's choice in the SP experiment (Fig 8).

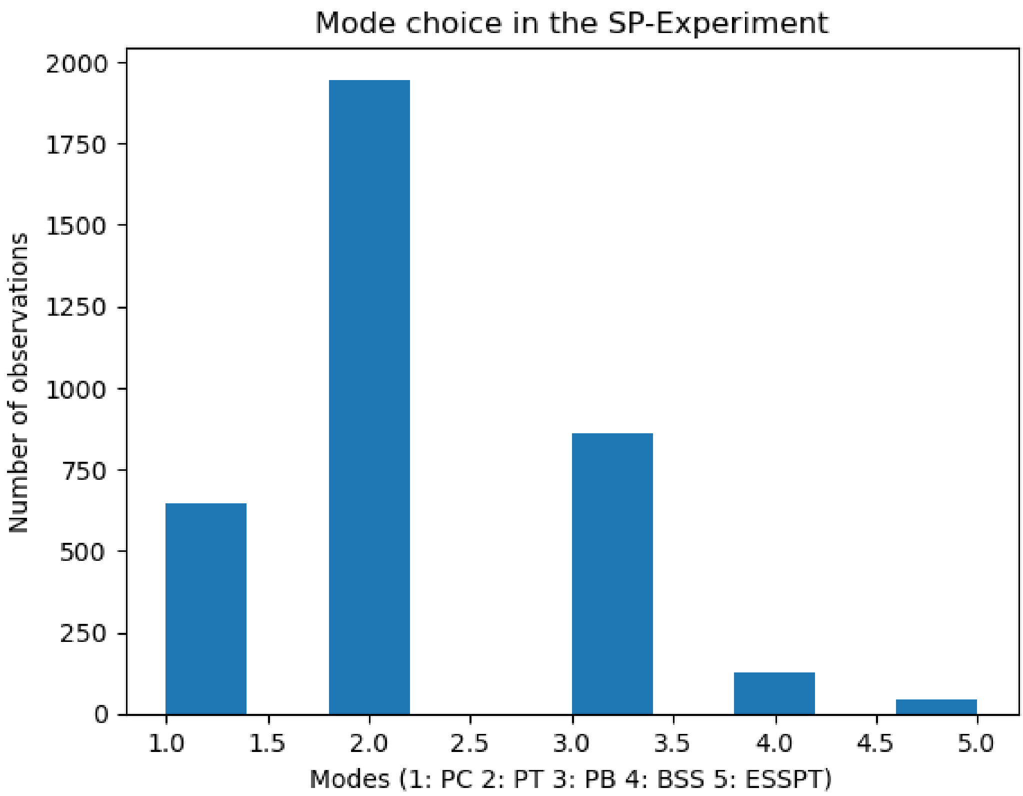


Fig 8. Number of observation per mode in the SP-dataset

When we look at this graph, we note that the hierarchy between chosen mode and main/secondary is the same modes which might means that people just chose what they were accustomed to and went with it during the whole experiment. In order to check if they did, we can check if their decision seems rational. For that I will display the distribution of each chosen variable and the mean of the unchosen ones (fig. 9 - 18) I also added the proportion of chosen values according to weather attributes.

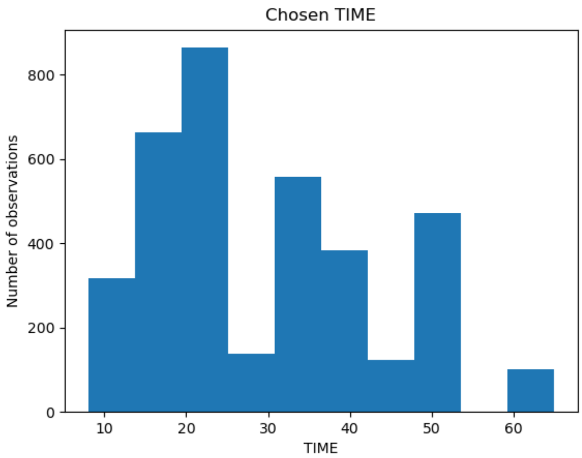


Fig 9. Number of observations for each chosen time of the experiment

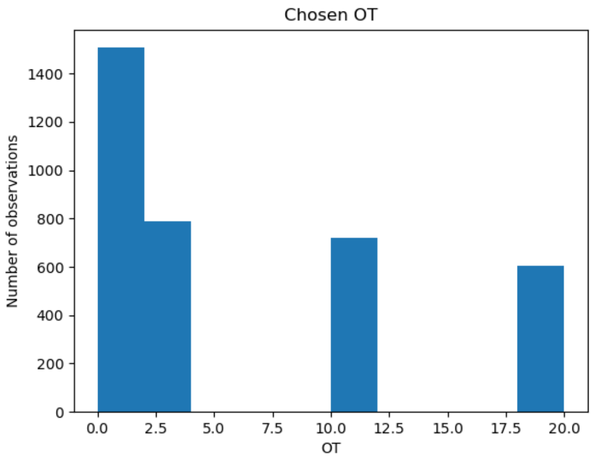


Fig 10. Number of observations for each chosen out-of-vehicle time of the experiment

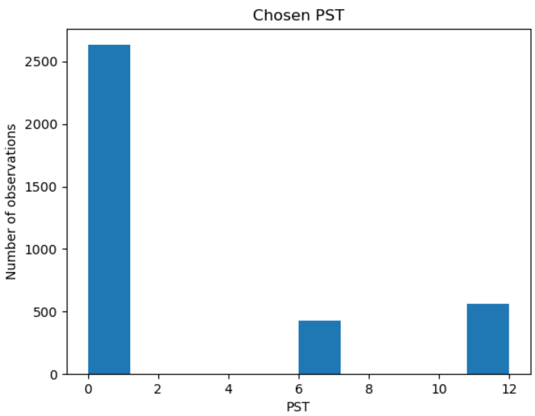


Fig 11. Number of observations for each chosen parking search time of the experiment

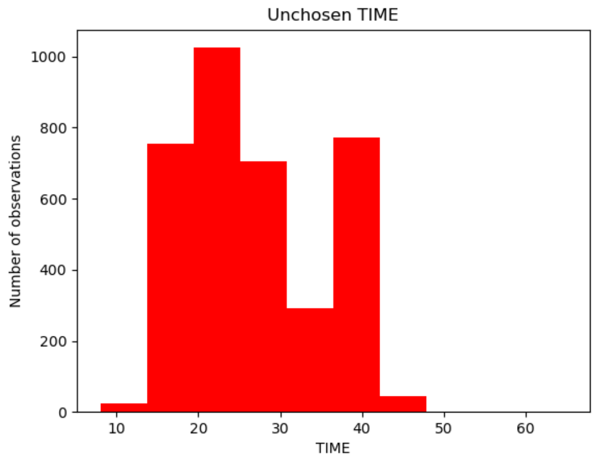


Fig 12. Number of observations for each mean of unchosen time of the experiment

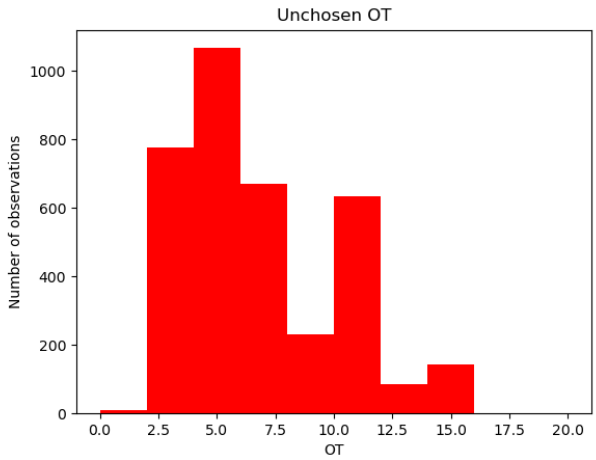


Fig 13. Number of observations for each mean of unchosen out-of-vehicle time of the experiment

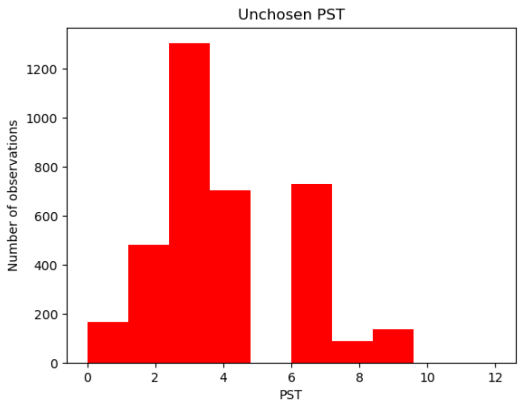


Fig 14. Number of observations for each mean of unchosen parking search time of the experiment

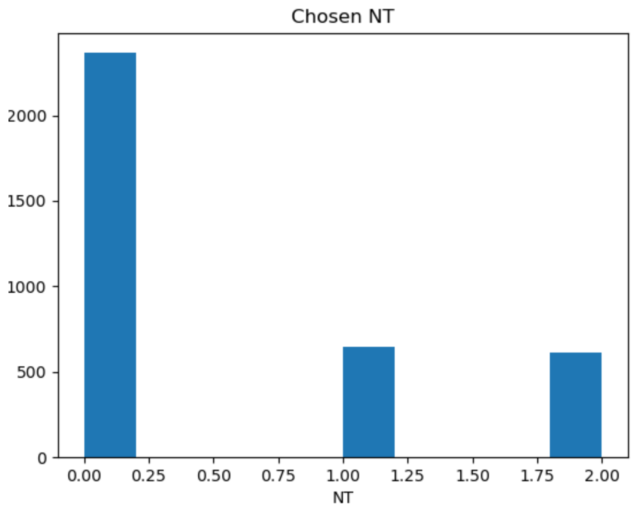


Fig 15. Number of observations for each mean of chosen number of transfer of the experiment

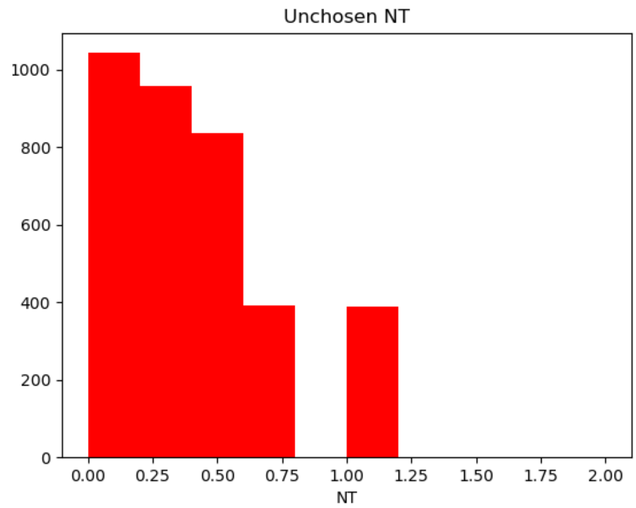


Fig 17. Number of observations for each mean of unchosen number of transfer of the experiment

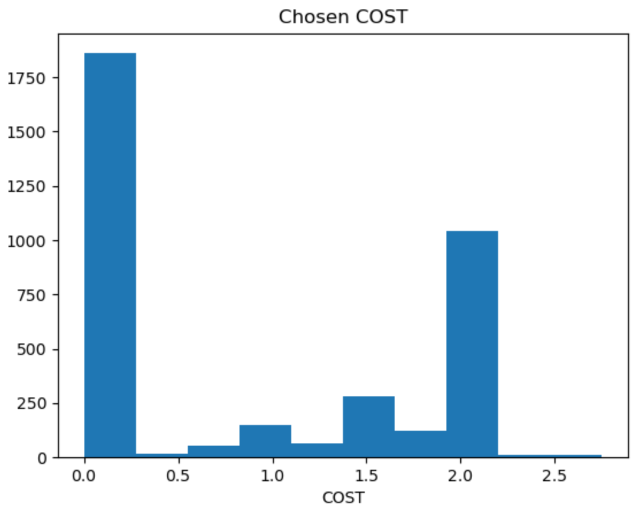


Fig 16. Number of observations for each mean of chosen cost of the experiment

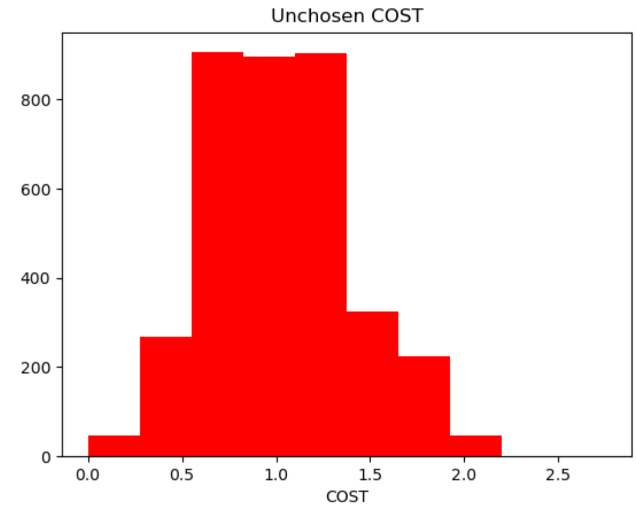


Fig 18. Number of observations for each mean of unchosen cost of the experiment

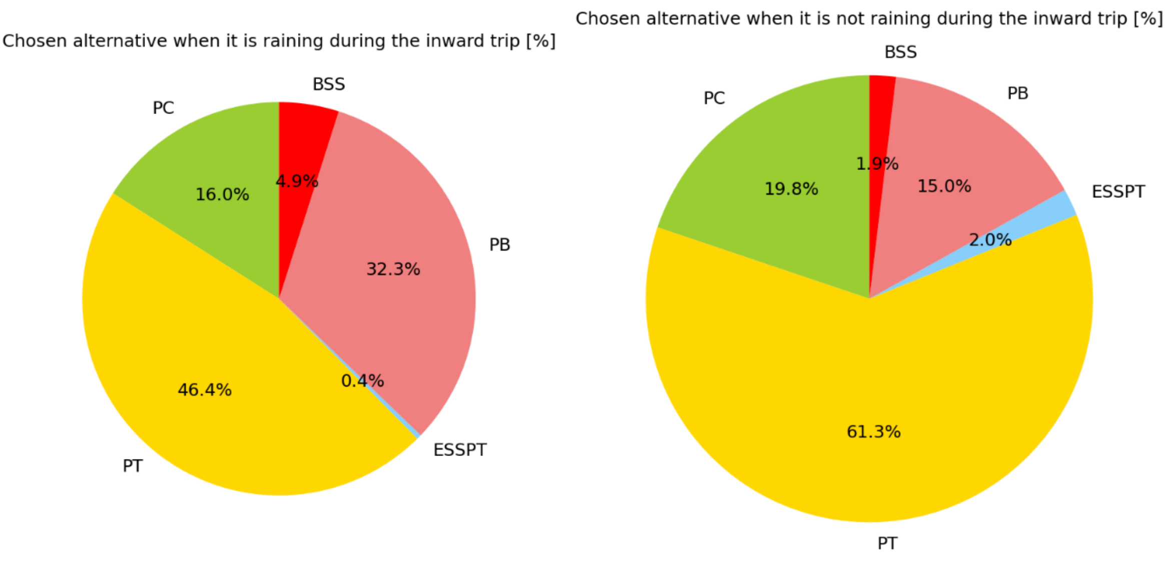


Fig 19. Proportion of chosen alternative when it is rain in the inward trip and when it is not

Fig 20. Proportion of chosen alternative according to the probability of rain on the way back

What shows up from those graphs (fig. 8 - 18) is that for all variables it seems like the peak of chosen values is always lower or similar to the peak of the mean of the unchosen values on the abscise, which means that even if people chose something close to what they actually do, there is still some rationality in their decision making. Of course, there are still peaks of chosen value that are higher than the mean of the unchosen one, those are due to trade-off between variables and will allow the estimation of our parameters when doing choice modeling. When looking at the chosen modes according to weather attributes' values we see that there is a change in shares depending on those variables, this change seems logic (less bike and less BSS when it's raining) meaning that individuals took those factors into account when making their decision revealing, again, a sense of rationality.

3.3. Descriptive statistics

I calculated some descriptive statistics in order to investigate the impact of the socio-economic variables on the mode choice. For the categorical variables, I did either a Chi² test, when all groups of that variable had at least 5 observations, or a Fisher test otherwise (Table 8). For the continuous variables I wanted to do a T-Test but unfortunately the variables did not follow a normal distribution, so I had to use a Shapiro-Wilk test (Table 9), which test if the median of a group is significantly different from the one of the other groups.

Variable	Categorie	Test	CAR	PT	BIKE	BSS	ESSPT
GENDER	Total	Chi2	0.764006	0.058796	**0.004906**	0.082464	1
EDUC_LEV	0	Fisher	**6.53E-11**	**9.12E-13**	**0.000239**	0.120778	**0.0425**
EDUC_LEV	1	Fisher	**1.63E-21**	**3.62E-06**	**8.06E-05**	0.118063	0.625574
EDUC_LEV	2	Fisher	**0.020608**	**0.038661**	0.955098	1	0.113043
EDUC_LEV	3	Fisher	**5.16E-08**	**1.16E-06**	0.116699	0.529181	**0.018933**
EDUC_LEV	4	Fisher	0.334135	**7.84E-05**	**7.57E-06**	0.837981	**0.006038**
EDUC_LEV	5	Fisher	**0.012142**	0.120894	**3.4E-05**	0.51789	0.572106
JOB	0	Fisher	**6.94E-19**	**6.63E-18**	**0.000446**	**4.05E-05**	**0.007766**
JOB	1	Fisher	**9.59E-06**	**0.001254**	1	0.457512	**0.038536**
JOB	2	Fisher	**1.91E-09**	**0.001046**	0.056989	0.727441	0.540438
JOB	3	Fisher	**1.17E-57**	**1.25E-17**	**2.16E-05**	**1.31E-05**	0.081068
JOB	4	Fisher	0.330453	**7.03E-14**	**5.52E-20**	0.339265	0.05765
DRIVER_LICENSE	Total	Chi2	**9.3E-08**	**0.001902**	0.112215	0.229848	0.514691
RESIDENCY_TYPE	1	Fisher	**2.54E-06**	**6.01E-08**	**0.001313**	0.08682	0.257673
RESIDENCY_TYPE	2	Fisher	**2.59E-12**	0.331791	**7.6E-06**	0.23331	0.876373
RESIDENCY_TYPE	3	Fisher	**2.45E-42**	**1.99E-15**	0.078183	**4.74E-05**	0.069359
MM_W	Total	Chi2	**4.52E-20**	**8E-13**	**1.67E-07**	0.057071	**0.034007**
MM_C	0	Fisher	**1.49E-28**	**8.59E-39**	**1.91E-11**	**0.043547**	**0.028157**
MM_C	1	Fisher	**1.49E-28**	**8.59E-39**	**1.91E-11**	**0.043547**	**0.028157**
MM_TC	Total	Chi2	**5.91E-12**	**1.27E-15**	**1.37E-17**	0.14112	0.970154
MM_PMM	Total	Chi2	**6.58E-07**	**3.17E-16**	**3.73E-60**	0.227847	0.257575
MM_SS	0	Fisher	**0.000712**	0.198827	0.365695	**9.56E-08**	0.518067
MM_SS	1	Fisher	**0.000712**	0.198827	0.365695	**9.56E-08**	0.518067
MM_O	0	Fisher	0.603051	0.304597	0.81369	1	**0.002484**
MM_O	1	Fisher	0.603051	0.304597	0.81369	1	**0.002484**
SM_W	Total	Chi2	**4.77E-19**	**1.15E-11**	**8.6E-07**	0.057112	**0.020023**
SM_C	0	Fisher	**1.1E-146**	**1.2E-49**	**1.21E-18**	**1.98E-05**	0.122698
SM_C	1	Fisher	**1.1E-146**	**1.2E-49**	**1.21E-18**	**1.98E-05**	0.122698
SM_TC	Total	Chi2	**4E-11**	**7.89E-16**	**1.38E-19**	0.117579	0.764545
SM_PMM	Total	Chi2	**4.51E-06**	**6.22E-17**	**5.35E-59**	0.184948	0.343834
SM_SS	0	Fisher	**1.1E-146**	**1.2E-49**	**1.21E-18**	**1.98E-05**	0.122698
SM_SS	1	Fisher	**1.1E-146**	**1.2E-49**	**1.21E-18**	**1.98E-05**	0.122698
SM_O	0	Fisher	0.603051	0.304597	0.81369	1	**0.002484**
SM_O	1	Fisher	0.603051	0.304597	0.81369	1	**0.002484**
TYPE	1	Fisher	0.093192	0.599682	0.287762	0.293729	0.181246
TYPE	2	Fisher	**4.46E-23**	**1.59E-06**	**0.020725**	**0.001232**	1
TYPE	3	Fisher	**1.51E-21**	**0.000212**	**1.41E-06**	**0.033581**	0.645748
SUB_TC	Total	Chi2	**7.76E-23**	**3.06E-25**	**1.59E-21**	**0.0411**	0.140087
SUB_BSS	Total	Chi2	**7.5E-14**	0.099529	0.742947	**4.09E-13**	0.062054
SUB_CSS	0	Fisher	0.141101	**0.001979**	**1.22E-06**	1	1
SUB_CSS	1	Fisher	0.141101	**0.001979**	**1.22E-06**	1	1
SUB_ESS	0	Fisher	0.141101	0.40328	0.315074	1	**0.000297**
SUB_ESS	1	Fisher	0.141101	0.40328	0.315074	1	**0.000297**
SUB_O	0	Fisher	0.904165	0.853354	0.191882	**0.021505**	0.401872
SUB_O	1	Fisher	0.904165	0.853354	0.191882	**0.021505**	0.401872

Table 8.p-values of the Chi^2 test and the Fisher test for the categorical socio-economic variables

Variable	CAR	PT	BIKE	BSS	ESSPT
AGE	**2.11E-25**	**4.38E-06**	0.812687	**0.00172**	0.144009
NB_CAR	**2.28E-48**	**1.69E-09**	0.137496	0.11428	**0.007531**
NB_BIKE	**2.87E-05**	**2.49E-14**	**9.55E-23**	**0.02791**	0.980536
NB_ADULTS	0.381964	0.90754	0.824709	0.966184	0.089276
NB_MINOR	**3.09E-22**	**0.002993**	**0.004114**	**0.001802**	0.375189
NB_2RM	**1.13E-06**	**0.028782**	0.244809	0.494347	0.680278
NB_SCOOTER	0.247547	0.660117	0.787168	0.850097	0.596512
NB_PROVEH	**0.097436**	0.748379	0.319877	0.373278	0.163493
DISTANCE	**0.078243**	**0.047817**	**0.049127**	0.723127	0.797972

Table 9. p-value for the Shapiro-Wilk test for the continuous socio-economic variables

In Table 8 and Table 9 I put significant variables in bold. We see that we have a lot of them meaning that the study captured some heterogeneity between groups which I will be able to model when doing the choice modeling.

3.4. Sample adjustments

To complete the analysis of the dataset I wanted to check if the sample was representative of the whole population of the ENTPE. To do so I went on the website of the school and look for reference values that could describe the population. Combining data from the social report (ENTPE, 2021), the description of the different labs that exist on campus (ENTPE, n.d.), and the school presentation booklet for potential new students (ENTPE, 2023), I found reference values to compare our sample with the whole population. To compute the adjustment factors I used Eq. 6 for the lines that have nothing in other count (see Table 10) for the others the issue is that I was missing data for one of the categories (for example: I don't know how many non-binary people there is in the school) so I used Eq. 7. Results are displayed in Table 10.

					Count Population	Weight population	Count Sample	Other Count	Weight Sample	Adj Factor
ADMIN and OTHER										
TOTAL	187	MEN	99	25-35	7	0.006	0		0.000	
				36-45	25	0.023	12		0.003	6.811
				46-55	34	0.031	60		0.017	1.853
				56-65	34	0.032	70		0.019	1.635
				-1	NA	0.090	154	12	0.043	2.123
		WOMEN	88	25-35	6	0.006	36		0.010	0.570
				36-45	22	0.020	135		0.037	0.543
				46-55	30	0.028	73		0.020	1.366
				56-65	31	0.028	81		0.022	1.268
RESEARCHER										
TOTAL	199					0.182	464		0.128	1.423
STUDENTS										
TOTAL	705	MEN	367	STUDENTS	191	0.175	601		0.166	1.052
				CIVIL SERVANTS	176	0.161	682		0.188	0.856
		WOMEN	338	STUDENTS	176	0.161	597		0.165	0.978
				CIVIL SERVANTS	162	0.149	736		0.203	0.732
		NB	NA	STUDENTS	367	0.336	1222	24	0.338	0.995
				CIVIL SERVANTS	296	0.271	1466	48	0.405	0.670

Table 10. Number of people in the population and in the sample and their adjustment factors

$$AdjFactor = \frac{Weight_{population}}{Weight_{sample}}$$

$$Weight_{population} = \frac{Count_{population}}{Total_{population}}$$

$$Weight_{sample} = \frac{Count_{sample}}{Total_{sample}}$$

Eq. 6. Computation of the adjustment factors

$$Weight_{population} = \frac{TotalCountCategory_{population}}{Total_{population}}$$

$$Weight_{sample} = \frac{TotalCountCategory_{sample}}{Total_{sample}}$$

Eq. 7. Adjustments factors for special case

Example for the category -1 in ADMIN and OTHER $Weight_{population} = \frac{99}{1091}$ et $Weight_{sample} = \frac{0+12+60+70+12}{3619}$.

We notice that the collected sample was not representative of the whole population. Specifically, women staff members that are between 25 and 45 years old were overrepresented such as civil-servant women and non-binary students (adjustment factors below 1). On the other hand, men staff members who are between 36 and 45 years old were highly underrepresented and those who are between 18 and 35 years old are not represented at all (adjustment factors above 1).

4. Model specifications

Once I had these data, I set out to find the best possible specifications for the mode choice model. In this section, I present 5 of them, representing the main stages of the specification finding. Section 4.1 describes the estimate for the Ngene specification. Section 4.2 presents a much simpler model that uses only trip attributes collected during the survey (ATA model). Section 4.3 shed light on the facts that there is heterogeneity between groups (SEV model). Section 4.4 introduce the impact of environmental characteristics (WA model). Finally, section 4.5 highlights heterogeneity between individual in the decision making (ECP model). The models were estimated using PythonBiogeme (Bierlaire, 2016). To compare those models, I used likelihood ratio tests that are summarized in Table 11, in an Excel I took the number of parameters and the maximum likelihood of each model and computed the test. In each cell is test for the column model and the line model. For example, line ATA model and column WA model is the likelihood ratio test of those models. As it is written WA model, it means that this model was better according to the likelihood ratio test.

Likelihood ratio Test	***NGENE MODEL***	***ATA MODEL***	***SEV MODEL***	***WA MODEL***	***ECP MODEL***
NGENE MODEL	-	NGENE MODEL	SEV MODEL	WA MODEL	ECP MODEL
ATA MODEL	NGENE MODEL	-	SEV MODEL	WA MODEL	ECP MODEL
SEV MODEL	SEV MODEL	SEV MODEL	-	WA MODEL	ECP MODEL
WA MODEL	WA MODEL	WA MODEL	WA MODEL	-	ECP MODEL
ECP MODEL	ECP MODEL	ECP MODEL	ECP MODEL	ECP MODEL	-

Table 11. Likelihood ratio tests for the 6 models

In order to compare the alternative specific constant, the continuous variables have been centered, and the alternative specific constant for the private car alternative has been normalized to 0.

4.1. Ngene model statistics

First, let's look at the Ngene Specification. For this first specification, I have adopted the utility functions (UF) used for the design of the SP survey (Eq. 5). The table below summarizes the estimation of the parameters and their p-value.

Parameters	*Estimate*	*Rob. p-value*	*Parameters*	*Estimate*	*Rob. p-value*
$\boldsymbol{ASC_{PB}}$	**0.198**	**0.00142**	$\beta_{RI\ PB}$	-0.00336	0.748
$\boldsymbol{ASC_{BSS}}$	**-1.84**	**0**	$\beta_{RI\ BSS}$	-0.0314	0.125
$\boldsymbol{ASC_{ESSPT}}$	**-3.15**	**0**	$\beta_{RI\ PC}$	0.00298	0.784
$\boldsymbol{ASC_{PT}}$	**1.03**	**0**	$\beta_{RI\ ESS}$	-0.0852	0.317
$\boldsymbol{\beta_{AVAIL\ BSS}}$	0.0109	0.333	$\boldsymbol{\beta_{RI\ PT}}$	**-0.0148**	**0.0101**
$\boldsymbol{\beta_{Cost}}$	**-0.123**	**0.00372**	$\boldsymbol{\beta_{RO\ PB}}$	**-0.0121**	**1.18E-13**
$\boldsymbol{\beta_{CT\ PC}}$	-0.0118	0.198	$\beta_{RO\ BSS}$	-0.00179	0.544
$\boldsymbol{\beta_{NbT\ ESSPT}}$	0.0182	0.925	$\beta_{RO\ ESSPT}$	0.00727	0.139
$\boldsymbol{\beta_{NbT\ PT}}$	-0.101	0.0908	$\beta_{RO\ PT}$	-0.00205	0.153
$\boldsymbol{\beta_{OT\ BSS}}$	-0.00598	0.686	$\boldsymbol{\beta_{TT\ PB}}$	**-0.0155**	**0.0483**
$\boldsymbol{\beta_{OT\ ESSPT}}$	**-0.0792**	**0.000651**	$\beta_{TT\ BSS}$	0.0196	0.262
$\boldsymbol{\beta_{OT\ PT}}$	**-0.051**	**2.19E-12**	$\boldsymbol{\beta_{TT\ PC}}$	**-0.0385**	**7.83E-06**
$\boldsymbol{\beta_{PST\ PB}}$	-0.0119	0.164	$\beta_{TT\ ESS}$	0.0797	0.163

$\boldsymbol{\beta_{PST\,BSS}}$	-0.03	0.116	$\beta_{TT\,ESSPT}$	0.0214	0.0604
$\boldsymbol{\beta_{PST\,PC}}$	-0.0135	0.157	$\boldsymbol{\beta_{TT\,PT}}$	**-0.0264**	**1.04E-12**

Table 12. Estimated parameter values of the model used for Ngene and corresponding robust p-value

Number of parameters: 30

Log likelihood: -3987.322

First let's look at β_{cost}, this parameter is significant, and its estimate is negative. This parameter is also generic which means that if any cost of any alternative increases while other parameters remain the same, then people are less likely to choose that alternative, which makes sense.

Regarding the private car utility function, we see that $\beta_{TT_{PC}}$ is negative which means that if the travel time of car increases while all other parameters remain the same, people are less likely to choose car, which seems rational. Unfortunately, $\beta_{PST\,PC}$, $\beta_{RI\,PC}$ and $\beta_{CT\,PC}$ are not significant which means that the specification should be change or that the variables should be dropped.

Concerning the transit alternative, we see that ASC_{PT} is significant and positive which means that PT is preferred to car when all variables are equal to 0. $\beta_{TT\,PT}$ and $\beta_{OT\,PT}$ are both significant and both of their estimate is negative which means that if the travel time or the out-of-vehicle time of PT should increase while all other parameters remain constant, the decision makers will be less attracted by that option. $\beta_{RI\,PT}$ is also significant and negative meaning that when it rains the longest the trip is, the less people are likely to choose that option which does not really makes sense since PT is safe from the rain in theory. $\beta_{NbT\,PT}$ and $\beta_{WR\,PT}$ are not significant.

Now for the private bike option, ASC_{PB} is significant and positive meaning that the bike alternative is preferred to the car alternative when all variables are null. $\beta_{WR\,PB}$ is significant and negative meaning that when the probability of rain in the outward trip, people will less likely choose the PB alternative. $\beta_{TT_{PB}}$, $\beta_{PST\,PB}$ and $\beta_{RI\,PB}$ are not significant.

Regarding the bike sharing system alternative, ASC_{BSS} is significant and negative meaning that when all variables are equal to zero car is preferred to BSS. Unfortunately, all other parameters involved in the utility function of bike sharing-system are not significant which is an issue especially regarding $\beta_{TT\,BLS}$ because it would mean that the travel time has not impact on the utility function is not rational.

Finally, for the ESSPT utility function, we have ASC_{ESSPT} which is significant and negative meaning that car is preferred to ESSPT when all variables are null. $\beta_{OT\,ESSPT}$ is also negative and significant meaning that when all variables are fixed, and out-of-vehicle time is increased the probability of choosing ESSPT will decrease. All other parameters are non-significant.

What we see here is that a lot of parameters are not significant. The plan is then to work on the specification to have more significant parameters. To do that I'll begin with a simpler model and augment progressively the number of variables. In the following sections, non-significant parameters will be excluded from the model specification.

4.2. A simpler model

For this model, I put all trip attributes, dropped all environmental variables and all non-significant attributes. There are the classic variables (TT, OT, COST, NbT) but also the new one (PST and CT). Utility functions are detailed in Eq. 8, parameters estimate, and p-values are available in Table 13

$$U_{ATA_{PC}} = ASC_{PC} + \beta_{TT\,PC} * TT_{PC} + \beta_{Cost} * COST_{PC} + \beta_{AdT} * (CT_{PC} + PST_{PC})$$

$$U_{ATA_{PT}} = ASC_{PT} + \beta_{TT\,PT} * (TT_{PT} + OT_{PT}) + \beta_{Cost} * COST_{PT} + \beta_{\mathrm{NbT}\,PT} * NbT_{PT}$$

$$U_{ATA_{PB}} = ASC_{PB} + \beta_{TT\,B} * TT_{PB} + \beta_{AdT} * PST_{PB}$$

$$U_{ATA_{BSS}} = ASC_{BSS} + \beta_{TT\,B} * TT_{BSS} + \beta_{Cost} * COST_{BSS} + \beta_{AdT} * (OT_{BSS} + PST_{BSS})$$

$$U_{ATA_{ESSPT}} = ASC_{ESSPT} + \beta_{TT\,ESSPT} * (TT_{ESS} + TT_{ESSPT}) + \beta_{OT\,ESSPT} * OT_{ESSPT} + \beta_{Cost} * (COST_{ESS} + COST_{PT})$$

Eq. 8 Utility functions of the ATA model

Name	Value	Rob. p-value
$\boldsymbol{ASC_{PB}}$	0.282	2.34E-07
$\boldsymbol{ASC_{BSS}}$	-1.73	0
$\boldsymbol{ASC_{ESSPT}}$	-2.89	0
$\boldsymbol{ASC_{PT}}$	1.08	0
$\boldsymbol{\beta_{AdT}}$	-0.0185	3.46E-05
$\boldsymbol{\beta_{Cost}}$	-0.155	1.02E-05
$\boldsymbol{\beta_{NbT\,PT}}$	-0.178	6.06E-05
$\boldsymbol{\beta_{OT\,ESSPT}}$	-0.0958	3.58E-05
$\boldsymbol{\beta_{TT\,B}}$	-0.0157	0.000104
$\boldsymbol{\beta_{TT\,PC}}$	-0.0306	8.86E-09
$\boldsymbol{\beta_{TT\,ESSPT}}$	-0.019	0.0354
$\boldsymbol{\beta_{TT\,PT}}$	-0.0351	0

Table 13. Estimated parameters of the model containing the new explanatory variables and their T-Test

Number of parameters: 12

Log likelihood: -4039.073

First, we see that ASC_{PT}, ASC_{PB}, ASC_{BSS} and ASC_{ESSPT} are still significant and goes in the same direction as in the Ngene model meaning that the interpretation still holds. Same goes for β_{cost}, $\beta_{TT\,PC}$, $\beta_{TT\,PT}$and $\beta_{OT\,ESSPT}$except that now $\beta_{TT\,PT}$ is the parameter of both TT_{PT} and OT_{PT} because when they were separated they were not significantly different.

Regarding the utility function of the car, we see that there is only one parameter for both additional return times which is β_{AdT}. Its value is negative meaning that an increase in CT_{PC} or PST_{PC} while keeping all other parameters constant will make people less likely to choose the car option.

Now for the PT alternative we observe that reducing the number of variables, and so potentially retrieving correlation between some of them, allowed $\beta_{NbT\,PT}$ to be significant. Its value is negative meaning that an augmentation of the number of transfers for the transit option will result in making that alternative less attractive.

On the private bike utility function, we now have $\beta_{TT\,B}$that is significant which is good. It is shared with the bike sharing-system alternative, meaning that an increase of travel time for PB (resp. BSS) while having all other parameters fixed will make the decision makers less likely to choose the bike (resp. BSS) option. β_{AdT} is also in both utility functions. In the bike it is the parameter for the parking search time, and in bike sharing-system it holds for out-of-vehicle time and parking search time. Again, its sign is negative meaning that an increase in one those variables will make its alternative less likely to be chosen. Now looking at magnitude we see that β_{AdT} is higher than $\beta_{TT\,B}$meaning that in order to make the BSS option more competitive there is a need to reduce the out-of-vehicle time and the parking search time meaning that people want to have a the insurance that if they choose the bike sharing-system option they will have one next to their origin and that they will have a free spot when coming back in order to optimize their walking time.

Concerning the e-scooter combined with transit alternative, $\beta_{TT\,ESSPT}$ is significant and shared with both travel times of the ESS leg and the PT leg. Its sign is negative meaning that if the travel time of one leg increase it will have the same negative impact on the attractivity of the ESSPT option. $\beta_{OT\,ESSPT}$ is also significant and significantly different than $\beta_{TT_{ESSPT}}$, meaning that for the same increase but in the out-of-vehicle time the effect while be worst on the attractivity of that alternative, probably because as PT is combined with ESS creating an extra fare, people expect that the ESS leg is enough to reduce significantly the out-of-vehicle time of that option, so that it can be competitive to the PT alternative. The interpretation can go further by acknowledging the fact that $\beta_{NbT_{ESSPT}}$ is not significant meaning that what bothers people is the walking time more than the waiting time of the out-of-vehicle time otherwise $\beta_{NbT_{ESSPT}}$ would have been significant and negative.

So, this specification gives us a model where all parameters are significant and in which all important variables (travel time and cost) are, meaning we can move further and try capturing more complex behavior such as heterogeneity between groups.

4.3. Capturing groups heterogeneity

The next step was to add socio-economic variables to our model. These variables allow the acknowledgment of differences in decision mechanics among different groups of people. Due to the large number of socio-economic variables and the large number of classes in some of these groups, I can't test every possibility. As a guide, I used the descriptive statistics (Table 8 and 9) to find out which are the most relevant and in which utility functions it is most suitable to include them. The final model contains a total of 24 socio-economic variables (counting each class of an included variable as a variable) that can be used in several utility functions (see Eq. 9) Parameters estimation and their p-values are available in Table 14.

$$U_{SEV_{PC}} = ASC_{PC} + \beta_{TT_{PC}} * TT_{PC} + \beta_{Cost} * COST_{PC} + \beta_{AdT} * (CT_{PC} + PST_{PC}) + \beta_{DL} * (DRIVER.LICENSE == 1) + \beta_{EDUC.LEV_0} * (EDUC.LEV == 0) + \beta_{EDUC.LEV_2} * (EDUC.LEV == 2) + \beta_{EDUC.LEV_{4PC}} * (EDUC.LEV == 4) + \beta_{JOB_{0\ OR\ 1}} * ((JOB == 0) + (JOB == 1)) + \beta_{TYPE_{2\ OR\ 3}} * ((TYPE == 2) + (TYPE == 3)) + \beta_{NB.MV} * (NB.CAR + NB.2RM) + \beta_{M.C} * (MM_C + SM_c) + \beta_{SUB_{TC}} * SUB_{TC} + \beta_{SUB_{BSS_{PC}}} * SUB_{BSS}$$

$$U_{SEV_{PT}} = ASC_{PT} + \beta_{TT_{PT}} * (TT_{PT} + OT_{PT}) + \beta_{Cost} * COST_{PT} + \beta_{NbT_{PT}} * NbT_{PT} + \beta_{EDUC.LEV_1} * (EDUC.LEV == 1) + \beta_{EDUC.LEV_4} * (EDUC.LEV == 4) + \beta_{RT_{0\ OR\ 1}} * ((RT == 0) + (RT == 1)) + \beta_{MM_{TC}} * MM_{TC} + \beta_{MM_W} * MM_W + \beta_{SM_{TC_{PT}}} * SM_{TC} + \beta_{SM_W} * SM_W + \beta_{SUB_{BSS_{PT}}} * SUB_{BSS}$$

$$U_{SEV_{PB}} = ASC_{PB} + \beta_{TT_{PB}} * TT_{PB} + \beta_{AdT} * PST_{PB} + \beta_{EDUC.LEV_0} * (EUC.LEV == 0) + \beta_{EDUC.LEV_4} * (EDUC.LEV == 4) + \beta_{JOB_{0\ OR\ 1}} * ((JOB == 0) + (JOB == 1)) + \beta_{GENDER_0} * (GENDER == 0) + \beta_{MINOR} * NB.MINOR + \beta_{NB.VELO} * NB.VELO + \beta_{MM_{PMM}} * MM_{PMM} + \beta_{SM_{TC_{PB}}} * SM_{TC} + \beta_{SUB_{TC}} * SUB_{TC}$$

$$U_{SEV_{BSS}} = ASC_{BSS} + \beta_{TT_{BSS}} * TT_{BSS} + \beta_{Cost} * COST_{BSS} + \beta_{AdT} * (OT_{BSS} + PST_{BSS}) + \beta_{TYPE_{1\ OR\ 2}} * ((TYPE == 1) + (TYPE == 2)) + \beta_{M.SS} * (MM_{SS} + SM_{SS}) + \beta_{SUB_{TC}} * SUB_{TC} + \beta_{SUB_{BSS_{BSS}}} * SUB_{BSS}$$

$$U_{SEV_{ESSPT}} = ASC_{ESSPT} + \beta_{TT_{ESSPT}} * (TT_{ESS} + TT_{ESSPT}) + \beta_{OT_{ESSPT}} * OT_{ESSPT} + \beta_{Cost} * (COST_{ESS} + COST_{PT}) + \beta_{SUB_{ESS}} * SUB_{ESS}$$

Eq. 9 Utility functions of the SEV model

Name	Value	Rob. p-value	Name	Value	Rob. p-value
ASC_{PB}	0.11	0.601	$\beta_{NB.MV}$	0.441	3.57E-10
ASC_{BSS}	-9.21	0	$\beta_{NB.VELO}$	0.234	2.91E-12
ASC_{ESSPT}	-4.06	0	$\beta_{NbT_{PT}}$	-0.217	2.25E-05
ASC_{PT}	0.669	0.00437	$\beta_{OT_{ESSPT}}$	-0.106	3.05E-06
β_{AdT}	-0.0251	1.07E-06	$\beta_{RT_{0\ OR\ 1}}{}^{*}$	-0.893	2.06E-10
β_{Cost}	-0.195	9.81E-07	$\beta_{SM_{TC_{PB}}}$	-0.458	0.000658
β_{DL}	0.463	0.00458	$\beta_{SM_{TC_{PT}}}$	1.1	0.00566
$\beta_{EDUC.LEV_0}$	-0.413	2.39E-05	β_{SM_W}	-1.67	0.000197
$\beta_{EDUC.LEV_1}$	0.673	0.0307	$\beta_{SUB_{BSS_{BSS}}}$	1.1	2.41E-07

$\beta_{EDUC.LEV_2}$	-0.607	0.00153	$\beta_{SUB\,BSS_{PC}}$	-0.583	0.000257
$\beta_{EDUC.LEV_4}$	-1.7	3.55E-07	$\beta_{SUB\,BSS_{PT}}$	-0.225	0.0408
$\beta_{EDUC.LEV_{4\,PC}}$	-3.44	2.22E-16	$\beta_{SUB\,ESS_{ESSPT}}$	2.93	0.000487
$\beta_{JOB_{0\,OR\,1}}$	-0.621	1.80E-06	$\beta_{SUB_{TC}}$	-0.792	0
β_{GENDER_0}	0.228	0.0164	$\beta_{TT_{PB}}$	-0.0259	1.17E-07
β_{MINOR}	-0.625	1.03E-07	$\beta_{TT_{BSS}}$	-0.0157	0.0767
$\beta_{MM_{PMM}}$	1.34	0	$\beta_{TT_{PC}}$	-0.0427	2.78E-11
$\beta_{MM_{TC}}$	-0.701	0.0725	$\beta_{TT_{ESSPT}}$	-0.0316	0.00145
β_{MM_W}	1.88	1.83E-05	$\beta_{TT_{PT}}$	-0.0472	0
β_{M_C}	1.19	0	$\beta_{TYPE_{1\,OR\,2}}$	6.4	0
$\beta_{M_{SS}}$	0.763	1.38E-05	$\beta_{TYPE_{2\,OR\,3}}$	-0.472	0.000343

Table 14. Estimated parameters of the model containing socio-economic variables and their T-Test

*RT = RESIDENCY_TYPE

Number of parameters: 40

Log likelihood: -3175.636

Except for ASC_{PB} and β_{TT_B} all variables from model ATA are still significant and goes in the same direction meaning that the interpretation still holds.

Regarding ASC_{PB} we see that it is not significant anymore meaning that the bike option is neither preferred nor overlooked compared to the private car alternative.

Concerning β_{TT_B} we see that is has been separated in two parameters $\beta_{TT_{PB}}$ and $\beta_{TT_{BSS}}$ that are both significant and negative meaning that the direction is the same only that the impact of travel time is higher in bike than in bike sharing-system.

Now for the socio-economic variables, we note that $\beta_{EDUC.LEV_0}$ is negative and in the PC and PB utility functions meaning that people who have the Bac as their last diploma are less likely to choose the private car or the private bike alternative compared to the other options when all other variables are null. $\beta_{EDUC.LEV_1}$ is positive and only in the PT UF meaning that people whose last degree is a DUT or a BTS are more likely to choose the transit alternative compared to the other ones. $\beta_{EDUC.LEV_2}$ is negative and in the utility function of the private car alternative meaning that people who have a master's degree, an engineer diploma or a maitrise as their last diploma are more likely to overlook the car alternative when all other variables are equal to 0. $EDUC.LEV_4$ is present in the PC, PT and PB alternatives. It has one parameter for both the bike and transit option and another one for the private car alternative. All parameters are negative and the magnitude of the car's UF is higher than the other which means that people who have a PhD or an aggregation are less likely to choose PT or PB compared to BSS and ESSPT and even less likely to choose the car. $\beta_{JOB_{0\,OR\,1}}$ is negative and in the utility functions of car and bike. The variable JOB is equal to 0 or 1 when the person is a student and student of the ENTPE have as a last diploma a bac and this variable is also present in the PC and PB option meaning that people who have the bac are even less likely to choose one of these alternatives if they also are a student.

β_{GENDER_0} is positive and in the bike UF meaning that men are more likely to choose this option than the others when all variables are null.

β_{MINOR} is negative in the bike UF meaning that people who have minors in their household are less likely to take the bike, which makes sense because they might have to bring their children to school which is harder with a bike than with a car or transit.

$\beta_{MM_{PMM}}$ is positive and in the bike alternative meaning that people who have a private non-motorized mean of transport are more likely to choose the private bike option. In the PT alternatives there is MM_W and MM_{TC}, those two variables have two different parameters the one of MM_W is positive and the one of MM_{TC} is negative which seems to not makes sense because people who use PT in their main modes of transportation should be more likely to choose the PT option. The issue is that the two variables are correlated because when you have PT in your main modes of transportation you need to have walk too. So actually what the value of these parameters tell us is that if you have walk in your main modes of transportation you are more likely to choose PT than any other alternative, and if you have PT in your main modes of transportation you also are more likely to choose PT (because 1.88-0.701=1.179 which is positive) but less than if you only have walk as main modes of transportation.

And after that there are SM_W and SM_{PT} which are opposite signs. As said in section 3.1. if you have one mode as main modes there is a high probability that you also have it in your secondary modes (Fig 3). So, if you only have them in main modes the interpretation still holds. Then if you have walk in your secondary modes, you still prefer the PT option (1.88-1.67 = 0.21 which is greater than 0) but less than if you only have it in main modes. And if you have PT in secondary modes as well then you have walk as secondary modes as well and you are still more likely to choose transit (1.88-0.701-1.67+1.1 = 0.609) less than if you only have it in main modes but more than if you only have walks in secondary modes.

SM_{PT} is also in the bike option and its parameter is negative which means that if you have transit in your secondary modes then you are less likely to choose the private bike option. Then there is β_{M_C} is the private car utility function and that is positive which means that if a respondent has car in their main modes or their secondary modes, they are more likely to choose the car alternative when all other parameters are null. $\beta_{M_{SS}}$ tells us that if you have sharing system as one of your main modes of transportation.

$\beta_{NB.MV}$ is in the car utility function and is positive which means that the more you motorized vehicles in your household the more likely you are to choose car alternative, which looks rational. $\beta_{NB.VELO}$ is the UF of bike and is positive which means than the more bike you access to, the more attractive will be the private bike alternative, which seems reasonable.

$\beta_{RT_{0\ OR\ 1}}$ tells us that if the respondent is an owner or the renter of social housing or a student’s residence, then you are less likely to use public transportation which is hard to interpret.

Regarding the subscription, if you have one for the Vélo’V then you are more likely to choose BSS compared to the other alternative, less likely to choose transit and even less likely to choose car compared to bike, BSS and ESSPT. If you have a pass for the e-scooter sharing system, then you are more likely to choose ESSPT. And if you have a transit pass then you are less likely to choose car, bike or bike sharing-system compared to transit and ESSPT which makes sense.

Finally, $\beta_{TYPE_{2\ OR\ 3}}$ tells that if you live in the city center or in the suburb you are less likely to choose the car alternative and $\beta_{TYPE_{1\ OR\ 2}}$ says that if you live in the suburbs or the countryside you are more likely to choose the bike sharing-system alternative compared to the other one when all parameters are null.

4.4. Impact of the weather

After, the socio-economic variables I added the variables related to the rain. Rain on the way in was not significant when interacted with the time of travel so I used it as a binary variable, rain on the way out was divided by a 100 in order to be able to compare the parameters of rain on the way in with the ones of the way out. As they are exterior factors, I normalized the parameter of the car alternative to 0. For the elucidation of the specifications (Eq. 10) I only showed where the rain factors were added to make it clearer (the rest of the utility functions was the same as in the SEV model).

$$U_{WA_{PC}} = U_{SEV_{PC}}$$

$$U_{WA_{PT}} = U_{SEV_{PT}}$$

$$U_{ATA_{PB}} = U_{SEV_{PB}} + \beta_{RI_B} * RI + \beta_{RO} * RO$$

$$U_{WA_{BSS}} = U_{SEV_{BSS}} + \beta_{RI_B} * RI$$

$$U_{WAA_{ESSPT}} = U_{SEV\,ESSPT} + \beta_{RI\,ESSPT} * RI$$

Eq. 10 Utility functions for the WA model

Name	Value	Rob. p-value	Name	Value	Rob. p-value
ASC_{PB}	0.8	0.000293	$\beta_{NbT\,PT}$	-0.216	2.49E-05
ASC_{BSS}	-8.39	0	$\beta_{OT\,ESSPT}$	-0.085	0.000124
ASC_{ESSPT}	-4.79	0	$\beta_{RI\,B}$	-1.84	0
ASC_{PT}	0.675	0.00463	$\beta_{RI\,ESSPT}$	0.884	0.0202
β_{AdT}	-0.018	0.00033	$\beta_{WR\,PB}$	-1.69	0
β_{Cost}	-0.0774	0.0512	$\beta_{RT\,0\,OR\,1}$*	-0.917	3.77E-10
β_{DL}	0.514	0.00214	$\beta_{SM\,TC\,PB}$	-0.525	0.000157
$\beta_{EDUC.LEV\,0}$	-0.457	7.44E-06	$\beta_{SM\,TC\,PT}$	1.21	0.00294
$\beta_{EDUC.LEV\,1}$	0.718	0.0288	$\beta_{SM\,W}$	-1.8	0.000104
$\beta_{EDUC.LEV\,2}$	-0.669	0.000664	$\beta_{SUB\,BSS\,BSS}$	1.06	9.32E-07
$\beta_{EDUC.LEV\,4}$	-1.75	2.35E-07	$\beta_{SUB\,BSS\,PC}$	-0.584	0.000372
$\beta_{EDUC.LEV\,4\,PC}$	-3.55	0	$\beta_{SUB\,BSS\,PT}$	-0.247	0.0311
$\beta_{JOB\,0\,OR\,1}$	-0.682	5.66E-07	$\beta_{SUB\,ESS\,ESSPT}$	2.76	0.00443
$\beta_{GENDER\,0}$	0.296	0.00299	$\beta_{SUB\,TC}$	-0.857	0
β_{MINOR}	-0.737	1.07E-09	$\beta_{TT\,PB}$	-0.0708	0
$\beta_{MM\,PMM}$	1.49	0	$\beta_{TT\,BSS}$	-0.0602	2.62E-11
$\beta_{MM\,TC}$	-0.851	0.0337	$\beta_{TT\,PC}$	-0.045	4.35E-13
$\beta_{MM\,W}$	1.95	1.66E-05	$\beta_{TT\,ESSPT}$	-0.0204	0.0775
$\beta_{M\,C}$	1.24	0	$\beta_{TT\,PT}$	-0.0504	0
$\beta_{M\,SS}$	0.873	7.48E-07	$\beta_{TYPE\,1\,OR\,2}$	6.4	0
$\beta_{NB.MV}$	0.442	6.68E-10	$\beta_{TYPE\,2\,OR\,3}$	-0.486	0.000334
$\beta_{NB.VELO}$	0.281	6.66E-16			

*RT = RESIDENCY_TYPE

Table 15. Estimation of model parameters and their T-Test

Number of parameters: 43

Log likelihood: -2972.941

We note that ASC_{PB} is significant again and that is still positive meaning that when all variables are equal to 0, private bike is preferred to car. All parameters that were in the SEV model are still significant and in the same directions meaning that the interpretation still holds except for the magnitude of $\beta_{TT\,BSS}$ and β_{AdT} because the weather attributes reverse their order of magnitude.

Regarding the weather attributes, rain on the inward trip is significant in the utility functions of private bike, bike sharing-system and e-scooter combined with transit, it was not in the transit UF, meaning that when it rains respondent are equally attracted to public transportation and car which seems correct as both options are protected from the rain. In the PB and BSS options they have a shared parameter $\beta_{RI\,B}$ which is negative, that means that when it rains people are less likely to choose private bike or bike sharing-system than car or PT which seems correct. On the other hand, $\beta_{RI_{ESSPT}}$ is positive which means that when it rains people are more likely to choose that alternative than the other ones. This might be due to the fact that in the PT option the first leg is walk while here it is e-scooter, meaning that this first leg should be shorter in the ESSPT option than the PT option and so that people will be safe from the rain sooner. For the rain in the outward trip only one parameter is significant in the bike utility functions, this makes sense cause if it rains on the way out the car alternative is not impacted, and for transit, bike sharing-system and ESSPT you're not constrained for your choice in the way out so you can choose an option that is protected from the rain. $\beta_{RO_{PB}}$ is negative meaning that when the probability of rain on way out increases the likelihood to choose the private bike alternative decreases which, again, makes a lot of sense.

4.5. Capturing heterogeneity between individuals

In order to fully capture heterogeneity between individuals I changed the choice model by adding individual specific random error components. Socio-economic variables aimed at capturing heterogeneity between groups and with the error components the model will be able to capture heterogeneity between individuals. The error components are here to capture what could not be captured with attributes or the variables that were used in the study, hence capturing heterogeneity between individuals without regarding the socio-economic group they're in. For the error components I used a normal distribution and to estimate their parameters I used five hundred draws. Utility functions can be found in Eq. 11.

$$U_{ECP_{PC}} = ASC_{PC} + \beta_{TT_{PC}} * TT_{PC} + \beta_{Cost} * COST_{PC} + \beta_{AdT} * (CT_{PC} + PST_{PC}) + \beta_{EDUC.LEV_{4_{PC}}} * (EDUC.LEV == 4) + \beta_{NB.MV} * (NB.CAR + NB.2RM) + \beta_{M.C} * (MM_C + SM_c) + \beta_{SUB_{TC}} * SUB_{TC} + \beta_{SUB_{BSS_{PC}}} * SUB_{BSS}$$

$$U_{ECP_{PT}} = ASC_{PT} + \beta_{TT_{PT}} * (TT_{PT} + OT_{PT}) + \beta_{Cost} * COST_{PT} + \beta_{NbT_{PT}} * NbT_{PT} + \beta_{MM_W} * MM_W + \beta_{SM_W} * SM_W + \gamma_{PT} * \sigma_{PT} + \gamma_{PTBSS} * \sigma_{BSS}$$

$$U_{ECP_{PB}} = ASC_{PB} + \beta_{TT_{PB}} * TT_{PB} + \beta_{AdT} * PST_{PB} + \beta_{MINOR} * NB.MINOR + \beta_{NB.VELO} * NB.VELO + \beta_{MM_{PMM}} * MM_{PMM} + \beta_{SM_{TC_{PB}}} * SM_{TC} + \beta_{SUB_{TC}} * SUB_{TC} + \beta_{RI_B} * RI + \beta_{RO} * RO + \gamma_{BIKE} * \sigma_{BIKE} + \gamma_{BIKEPT} * \sigma_{PT}$$

$$U_{ECP_{BSS}} = ASC_{BSS} + \beta_{TT_{BSS}} * TT_{BSS} + \beta_{Cost} * COST_{BSS} + \beta_{AdT} * (OT_{BSS} + PST_{BSS}) + \beta_{TYPE_{1\ OR\ 2}} * \big((TYPE == 1) + (TYPE == 2)\big) + \beta_{SUB_{TC}} * SUB_{TC} + \beta_{RI_B} * RI + \gamma_{BSS} * \sigma_{BSS} + \gamma_{BSSBIKE} * \sigma_{BIKE}$$

$$U_{ECP_{ESSPT}} = ASC_{ESSPT} + \beta_{TT_{ESSPT}} * (TT_{ESS} + TT_{ESSPT}) + \beta_{OT_{ESSPT}} * OT_{ESSPT} + \beta_{Cost} * (COST_{ESS} + COST_{PT}) + \beta_{SUB_{ESS}} * SUB_{ESS}$$

Eq. 11 Utility functions for the ECP model

Name	Value	Rob. p-value	Name	Value	Rob. p-value
ASC_{PB}	1.51	6.12E-09	$\beta_{NB.MV}$	0.762	5.50E-05
ASC_{BSS}	-12.4	0	$\beta_{NB.VELO}$	0.453	1.67E-07
ASC_{ESSPT}	-3.69	0	$\beta_{NbT_{PT}}$	-0.321	1.45E-06
ASC_{PT}	1.2	7.65E-05	$\beta_{OT_{ESSPT}}$	-0.113	6.69E-06
β_{AdT}	-0.0271	2.85E-05	β_{RI_B}	-2.42	0
β_{Cost}	-0.121	0.0296	$\beta_{WR_{PB}}$	-2.44	0
$\beta_{EDUC.LEV_{4_{PC}}}$	-2	0.00245	$\beta_{SM_{TC_{PB}}}$	-1.41	7.17E-06
γ_{PB}	-1.38	0	β_{SM_W}	-1.83	0.182
γ_{PBPT}	-0.763	8.46E-14	$\beta_{SUB_{BSS_{PC}}}$	-1.01	0.00327
γ_{BSS}	-1.47	0	$\beta_{SUB_{ESS_{ESSPT}}}$	4.52	0
γ_{BSSPB}	-0.893	0.000401	$\beta_{SUB_{TC}}$	-1.05	8.60E-05
γ_{PT}	-1.46	0	$\beta_{TT_{PB}}$	-0.0885	0

γ_{PTBSS}	-0.615	1.29E-06	$\beta_{TT_{BSS}}$	-0.0676	1.09E-07
β_{MINOR}	-0.977	0.00117	$\beta_{TT_{PC}}$	-0.0618	6.44E-15
$\beta_{MM_{PMM}}$	2.27	6.09E-12	$\beta_{TT_{ESSPT}}$	-0.0357	0.00205
β_{MM_W}	2.32	0.0853	$\beta_{TT_{PT}}$	-0.0754	0
β_{M_C}	2.07	2.27E-11	$\beta_{TYPE_{1\ OR\ 2}}$	10.7	0

Table 16. Parameters estimate and their robust p-values for the EP model

Number of parameters: 34

Likelihood: -2515.822

Regarding the alternative specific constants and the trip attributes, they are all still significant and in the same direction except for RI_{ESSPT} which is not significant anymore, but otherwise interpretation still holds.

Concerning the socio-economic variables, a lot of their parameters were not significant anymore (β_{DL}, $\beta_{EDUC.LEV_0}$, $\beta_{EDUC.LEV_1}$, $\beta_{EDUC.LEV_2}$, $\beta_{EDUC.LEV_4}$, $\beta_{JOB_{0\ OR\ 1}}$, β_{GENDER_0}, $\beta_{MM_{TC}}$, $\beta_{M_{SS}}$, $\beta_{RT_{0\ OR\ 1}}$,$\beta_{SM_{TC_{PT}}}$, $\beta_{SUB_{BSS_{BSS}}}$, $\beta_{SUB_{BSS_{PT}}}$ and $\beta_{TYPE_{2\ OR\ 3}}$) for some of them, this means that the differences between individuals are more important that differences between groups. The parameters related to socio-economic variables that are still significant are in the same directions than before meaning that their interpretation still hold.

Now for the parameters related to the error components. I made three different specifications before coming with this one. The first one was one error component in each alternative (except car for normalization issues) and one generic parameter. The parameter was not significant, meaning that the people who had a preference did not have the same level of preference for each alternative. So, then I tried specific parameters and the one of transit and bike were significant. Which means that one group of individuals like public transportation and bike more than the others. Finally, I implemented four different error components in each alternative and the goodness of fit was way better according to the likelihood ratio test. The one of ESSPT was not significant so I dropped it. The other ones were significant meaning that there are some groups of people that like PT, bike and bike-sharing system more than the others and that those groups are made of different people, otherwise the specification with one error component would have had a better goodness of fit and the parameter of BSS would have been significant. Then I added parameters to cross the error components with other alternatives to see if the group of people who had a preferred option had an opinion on another one. Three of those parameters were significant: γ_{PBPT}, γ_{BSSPB} and γ_{PTBSS}. γ_{PBPT} shares the sign of γ_{PT}, γ_{BSSPB} shares the sign of γ_{PB} and γ_{PTBSS} shares the sign of γ_{BSS} which means that the group of people that likes PT, also likes bike, which could make sense if we suppose that it is because of environmental consciousness; the group of people that prefers bike also likes BSS, if we suppose that there are people who just like biking; and the group of people that like BSS also like PT, which makes sense if we suppose that it's people who likes shared modes of transportation.

5. Forecasting

Now that I have those models, I will look at the values of time that they give. Then, I will compute the modal split of the WA model and do some forecasting. The objective of forecasting is to see which trip attributes' variation the model is sensible to. This will help to understand how the new mobility can emerge and become more attracting for a Home – Work trip. Section 5.1 presents the values of time for each model. Section 5.2 will exhibit cost. Section 5.3 will be about the time elasticities. Section 5.4 will introduce the modal split of the WA model. Finally, section 5.5 will show several scenarios that aims at increasing the modal split of the bike sharing-system alternative and the e-scooter sharing-system combined with transit options.

5.1. Values of Time

The value-of-time is an indicator which tells how much one is willing to pay to save one hour of travel time. In our case, as we have simple utility functions, the value-of-time can be computed using Eq. 12.

$$VOT_j = \frac{\beta_{TT\,j}}{\beta_{Cost}}$$

Eq. 12 Formula of the value-of-time

where j is an alternative, $\beta_{TT\,j}$ the parameter associated with travel time and β_{Cost} the parameter associated with cost.

In 2019, the reference value prescribed by the Ministry of Ecology Energy Territory for socio-economic calculations in urban areas for all modes is €$_{2015}$10.6/h for the home-work-study journey (MEET, 2019), which gives us a value of time of around €13.2/h. I will use this value as a reference to check the pertinence of each model.

In Table 17 are the value of time computed with the ATA model. In Table 18 are presented the value of time resulting from the SEV model. Table 19 displays the value of time from the WA model. Finally, Table 20 shows the values of time from the ECP model.

	AVERAGE	PC	PT	BSS	ESSPT
VALUES OF TIME	12.5 €/h	11.9 €/h	13.6 €/h	6.09 €/h	7.36 €/h

Table 17. Value of Time for the ATA model

	AVERAGE	PC	PT	BSS	ESSPT
VALUES OF TIME	13.4 €/h	13.1 €/h	14.5 €/h	4.83 €/h	9.71 €/h

Table 18. Value of Time for the SEV model

	AVERAGE	PC	PT	BSS	ESSPT
VALUES OF TIME	35.88 €/h	34.89 €/h	39.03 €/h	46.62 €/h	15.81 €/h

Table 19. Time values for the WA model

	MOYENNE	PC	PT	BSS	ESSPT
VALEUR DU TEMPS	17.81 €/h	15.32 €/h	17.70 €/h	33.52 €/h	17.70 €/h

Table 20. Value of Time for the ECP model

When looking at the four tables we see that all models differ on those values of time. First two seem rather reasonable, the second one is really close to what we expect, but for the WA model the values are too high to be consider consistent. Then for the ECP model values of time decrease to more acceptable values except for BSS. The only difference between the SEV model and the WA model is the weather attributes and in fact when computing $\frac{\beta_{RIB}}{\beta_{Cost}}$ we see that people are willing to pay 23.77€ to avoid the rain on the way in which is a lot. Unfortunately, we cannot do the same with rain on the way back because it is in the utility function of bike in which there is no cost. First what this tells is that having a good goodness of fit doesn't mean that the model makes sense. If the objective is to compute the value of time the SEV model seem way more pertinent. The second thing is that since the issue seems to come from the variables related to the weather there might an issue related to how I constructed them.

In the remaining of this section will keep the WA model as reference because due to time constraints I cannot forecast with the ECP model however I'll keep in mind that the values-of-time are higher than expectations and that it can have an impact on the forecast.

5.2. Cost Elasticities of the probabilities

Cost elasticities are derived from the cost of an alternative and aims at computing how the probability of an alternative will react to a variation of the cost of that alternative (direct point elasticity) or another (cross point elasticity). The cost elasticities on the alternatives' probability are available on Table 21 for the EC model. Lines correspond to the alternatives from which we took the cost and columns represent the alternative that we took the probability from, so line Car, column Bike is the cross elasticity of cost car on the probability of choosing bike.

		Alternative				
		Car	**PT**	**Bike**	**BSS**	**ESSPT**
COST_	Car	-0.050106	0.012498	0.011849	0.01059	0.015637
	PT	0.019575	-0.025768	0.031249	0.041772	0.030719
	Bike	NaN	NaN	NaN	NaN	NaN
	BSS	0.001657	0.002648	0.003497	-0.077163	0.002758
	ESSPT	0.001036	0.002147	0.00124	0.001625	-0.148979

Table 21. Cost elasticities for the EC model

Here we see that all direct point cost elasticities are negative meaning that an increase in the cost of an alternative will decrease the probability of that alternative which makes sense, the magnitude of those elasticities tells us that the bigger variation will happen in the e-scooter sharing-system combined with transit, with its cost, indeed if the cost of ESSPT increase by 10% then the probability will decrease by 1.5%. On the contrary, the cross-point elasticities are all positive meaning that an increase of the cost of an alternative will increase the probability of choosing any other alternative. In both model when ranking the elasticities, we clearly see that the cost of PT is the one that have the higher impact on the alternatives.

5.3. Travel Time Elasticities of the probabilities

Time elasticities works as cost elasticities unless instead of deriving probabilities according to cost, we derive probabilities according to travel time. Travel time elasticities from model EC are available in Table 22. The tables work the same way than the one for cost elasticities so line bike, column BSS correspond to the cross elasticity ofTT_{PB} on the probability of the BSS alternative.

		Alternative				
		Car	**PT**	**Bike**	**BSS**	**ESSPT**
TT_	Car	-0.618403	0.153262	0.14767	0.133548	0.198473
	PT	0.544208	-0.510045	0.478858	0.731748	1.195665
	Bike	0.396577	0.388598	-1.250774	0.740725	0.359661
	BSS	0.03893	0.064918	0.080946	-1.844201	0.06081
	ESSPT	0.006677	0.010316	0.004846	0.007311	-0.719823

Table 22. Time elasticities for the EC model

Again all direct point elasticities are negative and all cross elasticities are negative meaning that when the time of one alternative increases the probability of choosing that alternative decreases and the probability of choosing the other increases which makes sense, the direct elasticity that is the highest in absolute value is the BSS alternative with an elasticity of -1.84 meaning that an increase of 10% in TT_{BSS} will reduce the probability of choosing that alternative by 18.4%. For the cross elasticities Time PT is still the variables that has the most impact on the other alternatives.

Those elasticities help us to understand what the impact of the cost and the travel time of each alternative have on the probabilities of all alternatives however if we want to have the impact of other variables, we would need to compute other elasticities. Instead of doing that I focused on specific variables by creating scenarios that directly gives the change in the modal split that is induced by the variation of one variable. But first let's look at our modal split.

5.4. Modal split

Concerning the modal split, we can refer to the article (Manout et al., 2023). This article presents a modal split calculation for the city of Lyon. However, there are a number of differences between this study and theirs. Firstly, the data used includes information on the entire population of Lyon, not just ENTPE members. Secondly, the data are RP part of it is trip information for car, PT, bike and walk is from the EMD 2015 and the BSS and ESS trip records are from 2022 and this study uses SP data that is from 2023. Also, all trip types are included in the data of this article, not just home-work trips. Finally, the alternatives considered in their survey are the private car (driver or passenger), public transport, shared bicycle, shared electric scooter and walking. Thus, I merged the alternatives personal car driver and personal car and the alternative walking to have the most comparable modal split Fig 21.

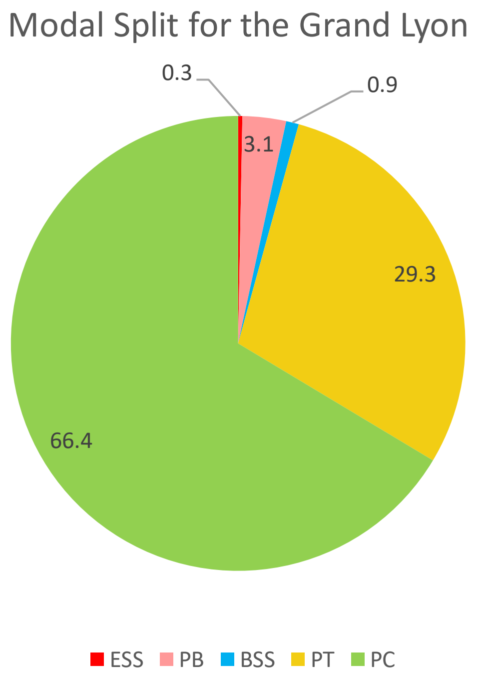


Fig 21. Modal split for the city of Lyon (Manout et al., 2023)

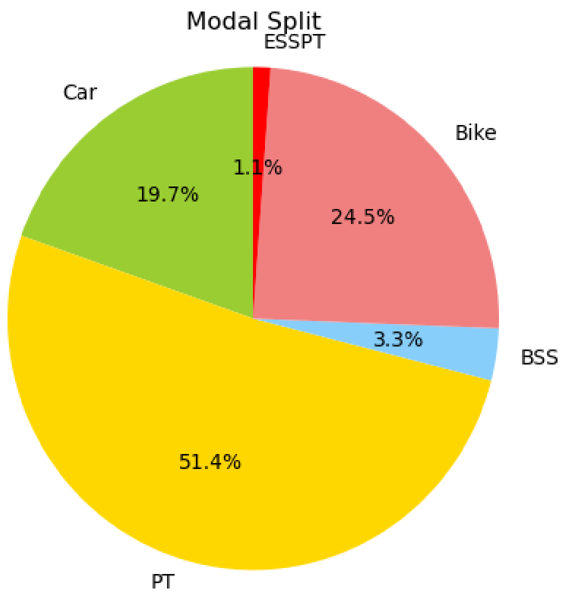


Fig 22. Modal split resulting from the WA model

We see that the modal split is very different from that of the city of Lyon. The hierarchy is completely different, and the orders of magnitude are not respected. This is due to the fact that the population surveyed in the RP data they used is not the same as that of this survey. First, students are overrepresented compared to the Lyon metropolitan area. Other factors could be put forth, the infrastructures available are not necessarily the same. For example, parking spaces are limited at ENTPE, which may explain the car's low share of the modal split. On the other hand, there are plenty of bicycle facilities on campus, which may encourage people to use them, especially as civil servants, who make up a large if not majority of ENTPE members, can get a bonus if they use this mode of transport for the majority of their home-to-work commute. The social aspect must also be taken into account, particularly the difference in social class, which is higher at the ENTPE than in the overall population of the Lyon metropolitan area. Finally, the ENTPE presents itself as a school involved in the ecological transition, and we can therefore imagine that its members share a certain number of common values, notably ecological values, which may explain the very high share of public transport and cycling compared to the shares of the Grand Lyon. This means that, as expected, this model cannot be used as such for the Lyon metropolitan area and further work would be needed to do so.

5.5. Revealing new modes' potential

In this section I present seven scenarios that aims at seeing which policy could be led to increase the modal shares of the new forms of mobility (BSS and ESSPT). Section 5.5.1 presents three scenarios regarding the bike sharing-system. Section 5.5.2 shed light on two scenarios about e-scooter combined with transit. Finally, section 5.5.3 highlights two scenarios for both alternatives.

5.5.1. Exploiting the bike sharing-system

In this section are presented 3 scenarios that could increase the share of the BSS alternative.

Scenario 1

In the first one, I wanted to investigate what would happen if we had more parking spot for the bike sharing-system. As the number of stops at arrival was not significant, I could not use that to do the simulation, so what I did instead is that I decreased out-of-vehicle time, which is basically the time from origin to first station, by 30% and the parking search time on the way back by 30%, considering that if you have more spots, you'll need less time to find one, and looked at how the modal split change. The new modal split is displayed in Fig 24.

Scenario 2

In this second scenario I wanted to see what would happen if Lyon Metropole made the bike sharing-system free for students, this was straight forward I made the cost of BSS equal to zero (-1.44 indeed cause the data is centered) for the students and see what happen. Results are shown with Fig 25.

Scenario 3

In this last scenario, I intended on simulating what would happen if the fleet of the bike sharing-system was replaced with e-bike. To do so I simulated a reduction of travel time of 15/19 (15km/h being the mean speed of a bike in a city and 19km/h being the speed of an e-bike in a city (GPS Zapp, n.d.)) while augmenting the price by 30% to compensate the fleet change. Results are displayed with Fig 26.

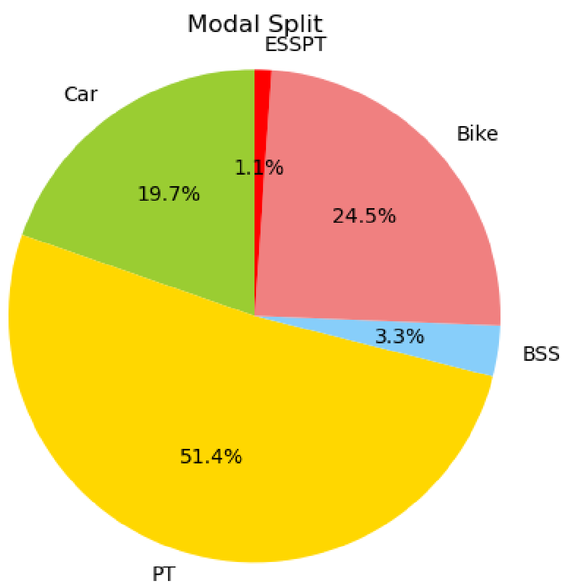


Fig 23. Reference modal split

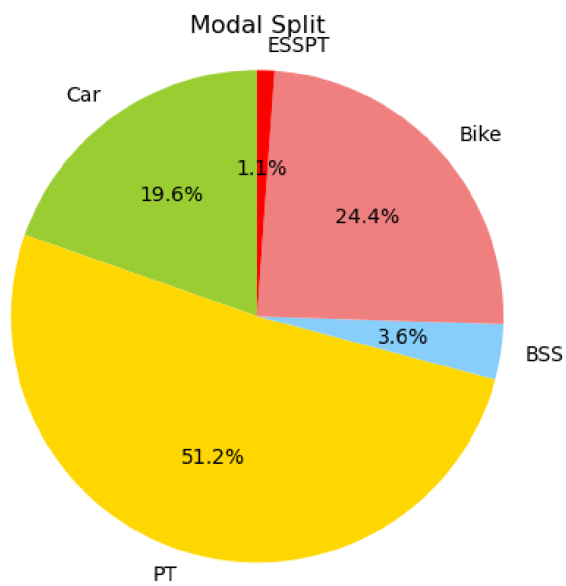


Fig 24. Modal split resulting from scenario 1

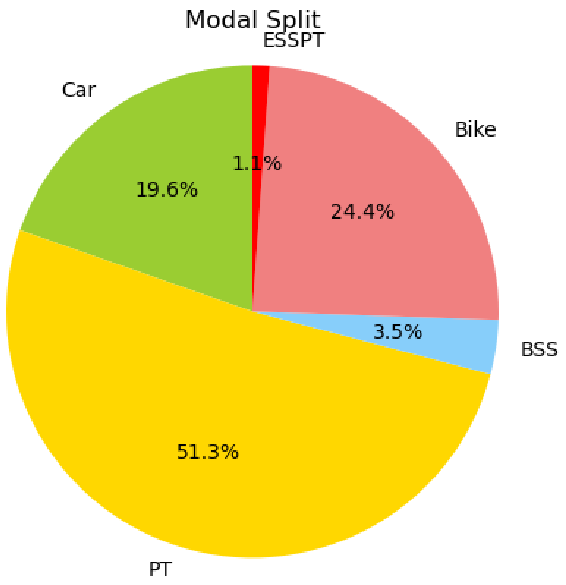


Fig 25. Modal split resulting from scenario 2

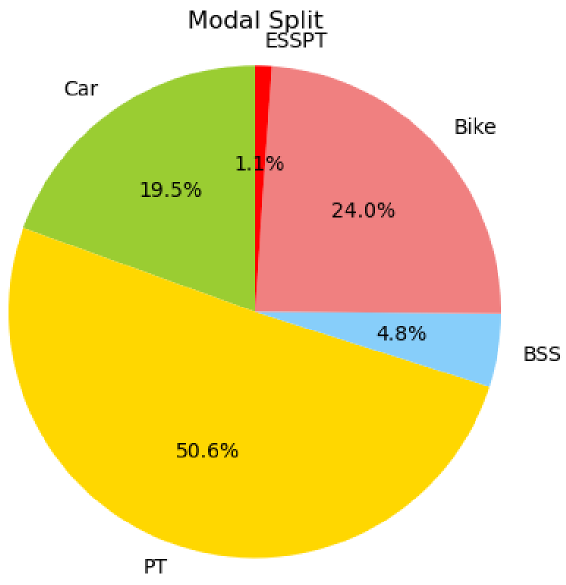


Fig 26. Modal split resulting from scenario 3

Just by looking at the shares we see that the first two scenarios are slightly increasing the share of bike sharing-system, and that the e-bike increases it significantly.

In fact, in the first scenario, the share of BSS increases by 8.00%, 12.05% of the new users come from the car alternative, 53.81% comes from transit, 33.10% from bike, and 1.04% from ESSPT. In the second scenario, the increase in the modal share is of 6.40%. 11.95% of the transfer comes from car, 56.53% comes from public transportation users, 30.47% are from the private bike and 1.05% are from e-scooter combined with transit. For the last scenario, the share of bike sharing-system is heightened by 45.50%. 12.68% comes from the car, 53.91% from the PT option, 32.28% are coming from the bike alternative and 1.12% from ESSPT.

In order to increase the bike sharing-system share in the modal split it seems that the best option is the e-bike scenario. However, this increase is really high and is due, in fact, to really high values of time. If we reject that scenario, then the first was better, not only because it increases more the share of BSS but also because it the one that impact the least ESSPT. Moreover, this policy is interesting because doing one thing you reduce two variables.

5.5.2. Revealing the capacity of ESSPT

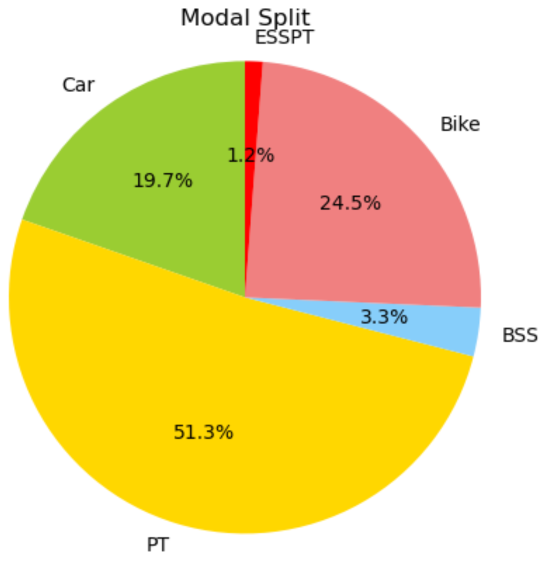
In this section I present two scenarios that aims at increasing the modal share of e-scooter combined with public transportation.

Scenario 1

The objective of this scenario is to see what would happen if we increased the availability of e-scooter. Again, there is no availability or density variables in the model so I will use the out-of-vehicle time instead and decreased by 30%. Results are shown in Fig 27

Scenario 2

This second scenario aims at seeing the impact of having a student free pass for the ESS leg will have on the share of ESSPT. Results are displayed in Fig 28.



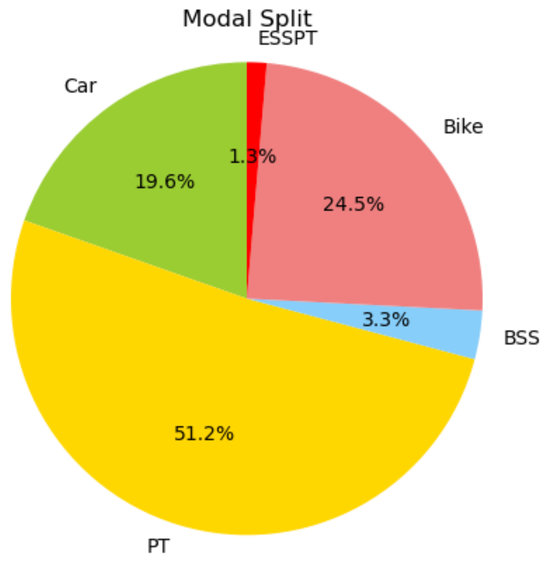


Fig 27. Modal split with the EC model when more e-scooter are available

Fig 28. Modal split with the EC model when e-scooter are free for students

	CAR	TRANSIT	BIKE	BSS	ESSPT
SHARES	19.7%	51.4%	24.5%	3.3%	1.1%

Table 23. Reference modal split

When looking at the resulting split we don't observe a huge change in the repartition of the shares for both scenarios but the second one seems to be better.

Indeed, in the first scenario, the share of ESSPT increases by 8.30%, 11.90% of the new users come from the car alternative, 69.79% comes from transit, 14.95% from bike, and 3.37% from BSS. In the second scenario, the increase in the modal share is of 20.73%. 14.70% of the transfer comes from car, 63.01% comes from public transportation users, 18.00% are from the private bike and 4.28% are from bike sharing-system.

Again, scenario 1 is best because it's the one that increases most the share of ESSPT and decrease less the shares of BSS.

5.5.3. Improving both shares'

In this last section I present two scenarios that aims at enhancing the shares of bike sharing-system and e-scooter sharing system combined with PT.

Scenario 1

In this first proposition I wanted to see what the effects of combining the first scenarios of the two other sections would be, meaning having more spots for the bike-sharing option and increase the number of e-scooters available. Modal split is available in Fig 29.

Scenario 2

In the last scenario I wanted to see if a negative policy on another alternative than BBS and ESSPT could have a positive and significant impact on the shares of those alternatives and thus reinforce the potential of those new mobility. The most effective way to that would be to simulate damage in the transit system because it the one that have the higher share, however, due to the cost of such system this doesn't seem realistic. The second share in ranking in the one of bike but due to environmental issues it seems unreasonable to damage the bike option nowadays. So, I turned to the car alternative and investigated what would happen if there were fewer parking slots available for the car option resulting in an increase of the parking search time of 30%. Resulting modal shares are displayed in Fig 30.

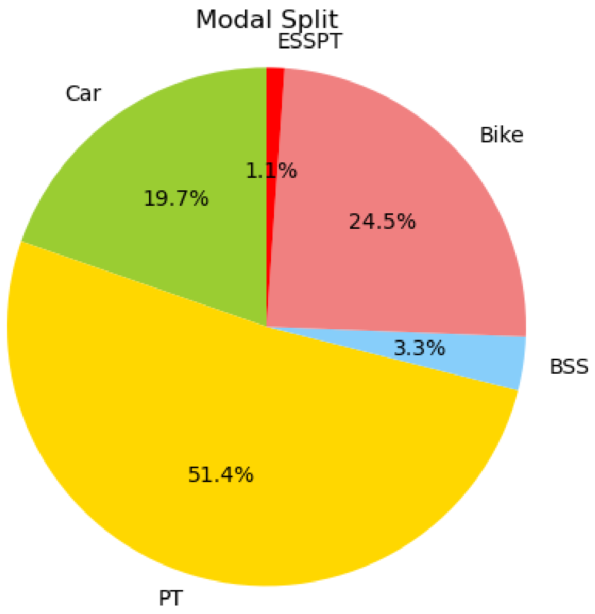


Fig 29. Reference modal split

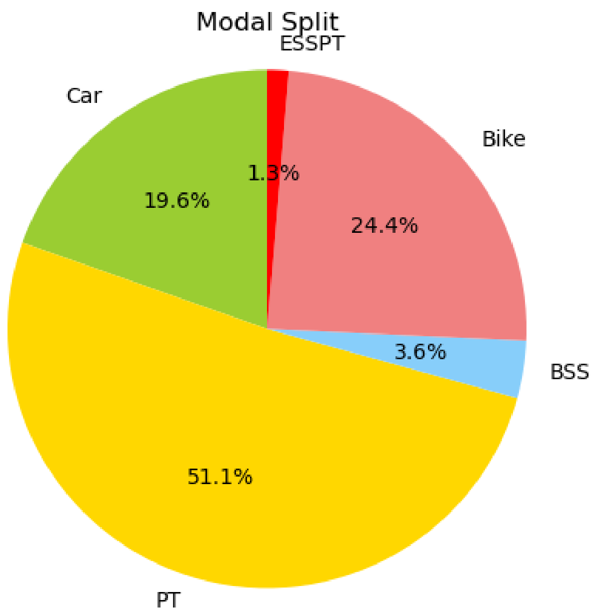


Fig 30. Modal split with EC model when the availability of e-scooters and bike sharing-system is increased

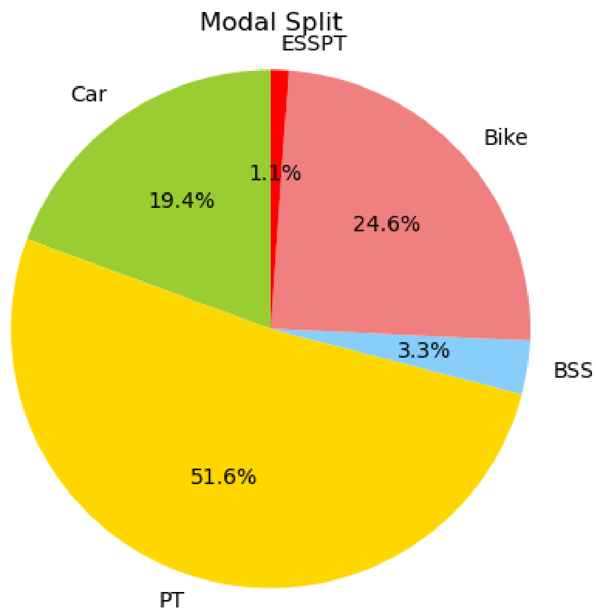


Fig 31. Modal split with EC model when there are less parking spot available for the car alternative

From what we see scenario 1 improved both the share of BSS and ESSPT while scenario two seems to make no difference compared to reference in those shares.

In numbers we have that for the first scenario, the increase in the modal share is of 7.68% for BSS and of 20.42% for ESSPT. 13.58% of the transfer comes from car, 59.35% comes from public transportation user and, 27.01% are from the private bike. For the other scenario, the shares of bike sharing-system is heightened by 0.37% and by 0.23% for ESSP. Of course, a 100% comes from the car option.

Here we can say with confidence that the first scenario is better meaning that in this case, improving a service is better than damaging one. However, even if the best model makes no doubt there is a second question which is where are this people coming from? In the first scenario only 13.58% come from the car against a 100% for the other and even if the main objective is two improves the shares of new mobilities decreasing the shares of the car should be an objective as well meaning that is could be interesting to go further in order to find a better trade-off.

So, what we see here is that there is some potential that can be exploited in order to enhance the shares of the new forms of mobility. However, this potential is limited, as mentioned in section 3.1. shared mobility might be not adapted for this kind of trip.

6. Discussion and conclusion

This section concludes this report. During this internship I designed an SP survey, analyzed the data obtained and created a mode choice model for which I tested numerous specifications with the aim of finding the most suitable one, starting with a logit model and ending with a mixed logit model incorporating error components panel. These different specifications enabled me to test the impact of new variables on demand estimation. Section 6.1. presents the main research findings. Section 6.2. recognizes the limitations of the current study. Finally, section 6.3. introduce some recommendation for future research.

6.1. Main research findings

In summary, the results show that adding additional information to the ones used in the classical models can contribute to explain mode choice. For example, the information that users have about the return trips, in particular the parking search time of the return trip, the extra time spent in traffic jams the day before and the probability of rain on the way back have an impact on the mode choice on the outward trip. Rain on the inward trip also has a significant impact on the utility functions of bike and bike sharing-system and with probability of rain on the return trip, has a large impact on the estimate of values of time. On the other hand, available spots at arrival are never significant.

I showed that the out-of-vehicle time of BSS and ESSPT and the parking search time on the way back of BSS were the most important time of those alternatives meaning that people need a high availability of these modes to consider them for a Work-Home trip which makes sense as being late could have bad consequences for them.

Now comparing our results with the priors, we see that the results are not completely as expected. First, the specification for rain on the way in was changed cause not significant, it was supposed to be interacted with travel time, but it didn’t work. Still with the rain, we expected it to be significant for all alternative whether it was for the inward or backward trip, however it was only significant for bike for the way out, and bike and BSS for the way in. Otherwise, the direction of every parameter is as expected but the magnitude of the ratios is different than expected. This can be explained by the fact that the parameters used come from a survey (Krauss et al., 2022) that has many differences from our own. First, the geographical area. I focused on a very small and specific part of the Lyon metropolitan area in France, whereas the other survey was carried out in several German cities. Secondly, the alternatives proposed were not the same: PC, PT, PB, BSS, ESSPT for this study and in the other: self-service electric scooter, self-service bike, walking and private car for short distances and carpooling, self-service car, public transport and private car for medium distances (Krauss et al., 2022). In addition, the alternatives offered differed according to the distance travelled and people only had short distance our medium distance in their survey.

Also, the use of advance choice models allows to reveal that there was a correlation between answers in the survey that could be shown using only explanatory variables. Adding random error components showed that there were several groups of respondents not belonging in a certain group that could have been represented using the socio-economic variables collected that preferred one option over the others.

Then when computing indicators such as the value of time. For some models, I obtain a value that is very close to the one recommended by the French Ministry of Ecology, Energy and Territory, which reassures me on the validity of those models and the quality of the data.

Finally, by computing the model split with the developed model I was able to show that today the forms of mobility have a low share in the modal split. However, using forecasting methods I was able to create scenarios that could support the emergence of new forms of mobility. More specifically, I was able to support the idea that it is important to have a strong availability for sharing-systems to insure reliability for the decision makers. Nevertheless, the increase of the modal split seemed limited.

6.2. Limitations

This study has several limitations. Firstly, the survey was carried out on a very specific and very small sample of the Lyon population and is, a priori, not representative of the latter, which prevents us from using this data alone to predict the overall behavior of the Lyon population. Specifically, the modal split, I obtain quite different results from the modal split I was able to find for the city of Lyon. This result can be explained by the fact that the populations surveyed are not the same and that ENTPE employees and students are not representative of the Lyon population.

Secondly, further analysis is needed to understand if the random error components showed a real heterogeneity among individual's preference or that people only considered part of the alternative available. Indeed, several things could support that assumption. Firstly, the instructions might haven't been clear enough about the fact that, in choice situations, respondents had access to all the proposed transport options, and not just those to which the respondent actually has access in everyday life (particularly for the private car). It is also possible that, in trying to represent as accurately as possible, through the choice situations, the transport options actually available for a home - ENTPE journey, respondents got confused between hypothetical choices and reports of the mode they actually use, placing the survey in an in-between RP and SP. Similarly, some choice situations may not make sense for the respondent. Indeed, if an individual is accustomed to using public transport and has the means to do so, he will choose his home in such a way as to optimize his public transport travel time, making choice situations presenting very long travel times for this alternative of little relevance. Finally, not all potential decision-making factors have been taken into account, particularly the environmental factor, which could have a major impact at the ENTPE.

In addition some results are hard to interpret, for example parking search time on the way back should have not be significant for the bike sharing-system because of the fact that people don't have a mode constraints for the way back when choosing that alternative for the inward trip.

There were also some simplifications made when creating the survey, which may have complicated the estimation of certain parameters. For example, in real life, in terms of vehicle travel time, the bike sharing-system and bicycle options are very similar. However, bike sharing system has a cost, whereas a bicycle alone does not, which puts the BSS option at a major disadvantage and complicates the estimation of the parameters of the utility function of this alternative. In view of the number of respondents obtained, I could have imagined two different surveys, one with the Car, Public Transport, Bicycle and Electric Scooter Self-Service and Public Transport modes, and another with the Car, Bike Sharing-system, Bicycle and Electric Scooter Self-Service and Public Transport options depending on whether or not the respondent owned a bicycle, so as not to have to confront these two options and thus facilitate the estimation of the parameters of the two utility functions. Also, the non-in-vehicle trip times could have been better constructed, for example by having different levels for each alternative, and in particular for the alternative of e-scooter sharing system and public transport, so as to make this option more competitive, given that it is very expensive compared with the others. I could have added more variability to our variables, and in particular our cost variable, even if this meant departing a little from reality, in order to make it easier to estimate the parameters of our model. Finally, the parameters used in the construction of the design with Ngene were very far from the estimates we obtained, even in terms of ratios, which certainly distorted the efficiency of our design.

Also, there is definitely something happening with the weather attributes. When I add these factors, I obtain values-of-time that are higher than the expectations and are not realistic. One hypothesis could be that the construction of those variables (binary variable for the rain in, and probability of having rain on the way out) was not optimal.

Finally, the implementation of random error components in the model is not optimal. Indeed, I tried to have only one with a generic parameter or with specific parameters, but the model had a higher goodness of fit with three random errors. The problem is that if we use each of them for a single alternative, we can't account for exchanges between the different alternatives, i.e. when the respondent prefer an alternative do they have another preference. To overcome this problem, I have added parameters to implement each random error in each alternative, keeping only the significant parameters. This solution is not optimal in terms of interpretability. For better interpretability, I should have performed a Cholesky decomposition.

6.3. Implications for practice and future research

When designing an SP experiment for mode choice modeling, information on the return trip should be included. It should also be used when implementing the models in simulations. The results of the present study should be used to design a new survey using the estimates of the model as priors in order to have a design that is efficient according to the subject.

If the present study should be conducted again, the subject of the weather attributes would need to be explored in greater depth in order to acknowledge its effect, because it has one, while having consistent results. An attention should be given to the additional return-trip times to make sure that each alternative is competitive.

The number of levels should be increased for certain variables whose parameters were more complicated to estimate, notably cost. Also, the out-of-vehicle time attributes should be separated in different times (e.g. access, waiting and egress time) in order to understand which one are the most impactful.

The survey should be made so that there are no options that are too close in terms of travel time (such as private bike, and bike sharing-system), for example by using socio-economic variables to determine which alternatives is given to the respondent.

The study could be lead on a larger scale to be able to model mode choice of a larger environment (e.g. Lyon Metropole) making sure the sample is representative of the population and/or use INSEE data for adjustments.

Now that it has been shown that the information of the return trip has a significant impact on the choice of the inward trip the subject can be investigated in a wider range to maybe found other factors of the return trip that might have an effect on the inward choice.

For further use of the collected data, first a Cholesky decomposition should be implemented to better understand unobserved factors influencing the choices of each individual. Forecasting should be done with the ECP model. Also, the most recent RP data regarding the surveyed population (e.g. EMD 2015) should be use to refine the parameters estimated and thus reduce bias. Finally, the model should be integrated with a traffic simulation.

References

Aubagna, M., 2021. Quel taux de réponses moyen à un questionnaire de satisfaction ? skeepers. URL https://skeepers.io/fr/blog/taux-de-reponses-moyen-questionnaire/ (accessed 8.11.23).

Baek, K., Lee, H., Chung, J.-H., Kim, J., 2021. Electric scooter sharing: How do people value it as a last-mile transportation mode? Transportation Research Part D: Transport and Environment 90, 102642. https://doi.org/10.1016/j.trd.2020.102642

Becarie, C., Tettarassar, C., Seppecher, M., Cortina, M., Leclercq, L., 2022. GitHub - licit-lab/MnMS: Multimodal simulator for DIT4TraM project [WWW Document]. URL https://github.com/licit-lab/MnMS (accessed 5.4.23).

Bekhor, S., Dobler, C., Axhausen, K.W., 2011. Integration of Activity-Based and Agent-Based Models: Case of Tel Aviv, Israel. Transportation Research Record 2255, 38–47. https://doi.org/10.31412255-05

Ben-Akiva, M.E., Lerman, S.R., 1985. Discrete choice analysis: theory and application to travel demand, MIT Press series in transportation studies. MIT Press, Cambridge, Mass.

Bierlaire, M., 2016. PythonBiogeme: a short introduction.

Cantelmo, G., Viti, F., 2019. Incorporating activity duration and scheduling utility into equilibrium-based Dynamic Traffic Assignment. Transportation Research Part B: Methodological 126, 365–390. https://doi.org/10.1016/j.trb.2018.08.006

Cantelmo, G., Viti, F., Cipriani, E., Nigro, M., 2018. A utility-based dynamic demand estimation model that explicitly accounts for activity scheduling and duration. Transportation Research Part A: Policy and Practice 114, 303–320. https://doi.org/10.1016/j.tra.2018.01.039

ChoiceMetrics, n.d. Ngene. https://choice-metrics.com/

Daganzo, C.F., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. Transportation Research Part B: Methodological 41, 49–62. https://doi.org/10.1016/j.trb.2006.03.001

De Clercq, G.K., Van Binsbergen, A., Van Arem, B., Snelder, M., 2022. Estimating the Potential Modal Split of Any Future Mode Using Revealed Preference Data. Journal of Advanced Transportation 2022, 1–11. https://doi.org/10.1155/2022/6816851

Dissanayake, D., Morikawa, T., 2010. Investigating household vehicle ownership, mode choice and trip sharing decisions using a combined revealed preference/stated preference Nested Logit model: case study in Bangkok Metropolitan Region. Journal of Transport Geography, Tourism and climate change 18, 402–410. https://doi.org/10.1016/j.jtrangeo.2009.07.003

Dosman, D., Adamowicz, W., 2006. Combining Stated and Revealed Preference Data to Construct an Empirical Examination of Intrahousehold Bargaining. Rev Econ Household 4, 15–34. https://doi.org/10.1007/s11150-005-6695-1

Dott, 2023. Application mobile Dott. https://play.google.com/store/search?q=dott&c=apps&hl=fr&gl=US

ENTPE, 2023. Plaquette de présentation du cursus ingénieur de l'ENTPE. https://www.calameo.com/read/0047010491a92f89f867b

ENTPE, 2021. Bilan Social 2019-2020. https://www.entpe.fr/sites/default/files/2023-05/Bilan%20Social-%202019-2020%20opt.pdf

ENTPE, n.d. Présentation des laboratoires de recherche de l'ENTPE. https://www.entpe.fr/5-laboratoires

Flötteröd, G., Chen, Y., Nagel, K., 2012. Behavioral Calibration and Analysis of a Large-Scale Travel Microsimulation. Netw Spat Econ 12, 481–502. https://doi.org/10.1007/s11067-011-9164-9

Fu, X., Lam, W.H.K., 2018. Modelling joint activity-travel pattern scheduling problem in multi-modal transit networks. Transportation 45, 23–49. https://doi.org/10.1007/s11116-016-9720-8

Fu, X., Lam, W.H.K., 2014. A network equilibrium approach for modelling activity-travel pattern scheduling problems in multi-modal transit networks with uncertainty. Transportation 41, 37–55. https://doi.org/10.1007/s11116-013-9470-9

Geroliminis, N., Zheng, N., Ampountolas, K., 2014. A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. Transportation Research Part C: Emerging Technologies 42, 168–181. https://doi.org/10.1016/j.trc.2014.03.004

Global Petrol Prices, 2023. Prix moyen du diesel pour la ville de Lyon.

GPS Zapp, n.d. Quelles différences entre vélo électrique ou vélo normal ?

Halat, H., Mahmassani, H.S., Zockaie, A., Vovsha, P., 2017. Activity Cancellation and Rescheduling by Stressed Households: Improving Convergence in Integrated Activity-Based and Dynamic Traffic Assignment Models. Transportation Research Record 2664, 100–109. https://doi.org/10.3141/2664-11

Halat, H., Zockaie, A., Mahmassani, H.S., Xu, X., Verbas, O., 2016. Dynamic network equilibrium for daily activity-trip chains of heterogeneous travelers: application to large-scale networks. Transportation 43, 1041–1059. https://doi.org/10.1007/s11116-016-9724-4

Hensher, D.A., Button, K.J. (Eds.), 2007. Handbook of Transport Modelling: 2nd Edition, Handbooks in Transport. Emerald Group Publishing Limited. https://doi.org/10.1108/9780857245670

Javanmardi, M., 2012. Integration of TRANSIMS with the ADAPTS Activity-based Model.

JCDecaux, Lyon Métropole, 2023. Grille tarifaire Vélo'V. https://play.google.com/store/search?q=v%C3%A9lo%27v+officiel&c=apps&hl=fr&gl=US

Johari, M., Keyvan-Ekbatani, M., Leclercq, L., Ngoduy, D., Mahmassani, H.S., 2021. Macroscopic network-level traffic models: Bridging fifty years of development toward the next era. Transportation Research Part C: Emerging Technologies 131, 103334. https://doi.org/10.1016/j.trc.2021.103334

Kim, S.H., Mokhtarian, P.L., Choo, S., Circella, G., 2022. Exploring Heterogeneous Structural Relationships Between E-Shopping, Local Accessibility, and Car-Based Travel: An Application of Enriched National Household Travel Survey Add-on Data. Transportation Research Record: Journal of the Transportation Research Board 036119812211328. https://doi.org/10.1177/03611981221132854

Krauss, K., Krail, M., Axhausen, K.W., 2022. What drives the utility of shared transport services for urban travellers? A stated preference survey in German cities. Travel Behaviour and Society 26, 206–220. https://doi.org/10.1016/j.tbs.2021.09.010

Li, Z.-C., Lam, W.H.K., Wong, S.C., 2014. Bottleneck model revisited: An activity-based perspective. Transportation Research Part B: Methodological 68, 262–287. https://doi.org/10.1016/j.trb.2014.06.013

Lin, D.-Y., Eluru, N., Waller, S.T., Bhat, C.R., 2008. Integration of Activity-Based Modeling and Dynamic Traffic Assignment. Transportation Research Record 2076, 52–61. https://doi.org/10.3141/2076-06

Manout, O., Diallo, A.-O., Gloriot, T., 2023. Implications of pricing and fleet size strategies on shared bikes and e-scooters: a case study from Lyon, France (preprint). In Review. https://doi.org/10.21203/rs.3.rs-2670644/v1

MEET, 2019. Valeurs de référence prescrites pour le calcul socio-économique. https://www.ecologie.gouv.fr/sites/default/files/V.2.pdf

Meng, L., Somenahalli, S., Berry, S., 2020. Policy implementation of multi-modal (shared) mobility: review of a supply-demand value proposition canvas. Transport Reviews 40, 670–684. https://doi.org/10.1080/01441647.2020.1758237

Nash, J.F., 1950. Equilibrium points in *n* -person games. Proc. Natl. Acad. Sci. U.S.A. 36, 48–49. https://doi.org/10.1073/pnas.36.1.48

Qualtrics, n.d. https://www.qualtrics.com/fr/

Ramadurai, G., Ukkusuri, S., 2010. Dynamic User Equilibrium Model for Combined Activity-Travel Choices Using Activity-Travel Supernetwork Representation. Netw Spat Econ 10, 273–292. https://doi.org/10.1007/s11067-008-9078-3

Reck, D.J., Haitao, H., Guidon, S., Axhausen, K.W., 2021. Explaining shared micromobility usage, competition and mode choice by modelling empirical data from Zurich, Switzerland. Transportation Research Part C: Emerging Technologies 124, 102947. https://doi.org/10.1016/j.trc.2020.102947

Reck, D.J., Martin, H., Axhausen, K.W., 2022. Mode choice, substitution patterns and environmental impacts of shared and personal micro-mobility. Transportation Research Part D: Transport and Environment 102, 103134. https://doi.org/10.1016/j.trd.2021.103134

Rose, J.M., Bliemer, M.C.J., 2013. Sample size requirements for stated choice experiments. Transportation 40, 1021–1041. https://doi.org/10.1007/s11116-013-9451-z

Rudloff, C., Straub, M., 2021. Mobility surveys beyond stated preference: introducing MyTrips, an SP-off-RP survey tool, and results of two case studies. Eur. Transp. Res. Rev. 13, 49. https://doi.org/10.1186/s12544-021-00510-5

Shui, C.S., Szeto, W.Y., 2020. A review of bicycle-sharing service planning problems. Transportation Research Part C: Emerging Technologies 117, 102648. https://doi.org/10.1016/j.trc.2020.102648

Statista Research Department, 2023. Consommation moyenne de carburant d'une voiture particulière en France de 2004 à 2020, selon le type de carburant.

SYTRAL, 2023. Grille tarifaire du réseeau TCL. www.TCL.fr

Tier, 2023. Application mobile Tier. https://play.google.com/store/search?q=tier&c=apps&hl=fr&gl=US

Train, K., 2009. Discrete choice methods with simulation, 2nd ed. ed. Cambridge University Press, Cambridge ; New York.

van Essen, M., Thomas, T., van Berkum, E., Chorus, C., 2020. Travelers' compliance with social routing advice: evidence from SP and RP experiments. Transportation 47, 1047–1070. https://doi.org/10.1007/s11116-018-9934-z

Verbas, İ.Ö., Mahmassani, H.S., Hyland, M.F., Halat, H., 2016. Integrated Mode Choice and Dynamic Traveler Assignment in Multimodal Transit Networks: Mathematical Formulation, Solution Procedure, and Large-Scale Application. Transportation Research Record 2564, 78–88. https://doi.org/10.3141/2564-09

Vo, K.D., Lam, W.H.K., Chen, A., Shao, H., 2020. A household optimum utility approach for modeling joint activity-travel choices in congested road networks. Transportation Research Part B: Methodological 134, 93–125. https://doi.org/10.1016/j.trb.2020.02.007

Vo, K.D., Lam, W.H.K., Li, Z.-C., 2021. A mixed-equilibrium model of individual and household activity–travel choices in multimodal transportation networks. Transportation Research Part C: Emerging Technologies 131, 103337. https://doi.org/10.1016/j.trc.2021.103337

Wardrop, J.G., Whitehead, J.I., 1952. CORRESPONDENCE. SOME THEORETICAL ASPECTS OF ROAD TRAFFIC RESEARCH. Proceedings of the Institution of Civil Engineers 1, 767–768. https://doi.org/10.1680/ipeds.1952.11362

Weis, C., Kowald, M., Danalet, A., Schmid, B., Vrtic, M., Axhausen, K.W., Mathys, N., 2021. Surveying and analysing mode and route choices in Switzerland 2010–2015. Travel Behaviour and Society 22, 10–21. https://doi.org/10.1016/j.tbs.2020.08.001

Xu, X. (Alex), Fakhrmoosavi, F., Zockaie, A., Mahmassani, H.S., 2017a. Estimating Path Travel Costs for Heterogeneous Users on Large-Scale Networks: Heuristic Approach to Integrated Activity-Based Model–Dynamic Traffic Assignment Models. Transportation Research Record 2667, 119–130. https://doi.org/10.3141/2667-12

Xu, X. (Alex), Zockaie, A., Mahmassani, H.S., Halat, H., Verbas, Ö., Hyland, M., Vovsha, P., Hicks, J., 2017b. Schedule Consistency for Daily Activity Chains in Integrated Activity-Based Dynamic Multimodal Network Assignment. Transportation Research Record 2664, 11–22. https://doi.org/10.3141/2664-02

Zargayouna, M., 2021. Modèles multi-agents de déplacement à base d'activités : Etat de l'art (report). Université gustave eiffel.

Zhang, L., Yang, D., Ghader, S., Carrion, C., Xiong, C., Rossi, T.F., Milkovits, M., Mahapatra, S., Barber, C., 2018. An Integrated, Validated, and Applied Activity-Based Dynamic Traffic Assignment Model for the Baltimore-Washington Region. Transportation Research Record 2672, 45–55. https://doi.org/10.1177/0361198118796397

Annexes

Appendix 1

Choice situation	RI	RO	TIME_CAR	COST_CAR	PST_CAR	DIS_CAR	TIME_PT	OT_PT	NT_PT	COST_PT	TIME_BIKE	PST_BIKE
1	0	10	29	1.42	12	14.5	52	2	0	2	48	0
2	0	0	40	1.66	6	0	52	20	2	0	51	0
3	0	90	18	1.66	6	0	52	2	0	0	51	6
4	1	0	40	1.46	6	4	25	11	2	2	36	12
5	1	90	40	1.66	0	4	65	2	0	0	51	6
6	1	50	16	0.86	6	1.6	13	20	1	0	22	6
7	1	90	35	0.74	0	3.5	22	20	2	0	20	6
8	0	90	40	1.46	0	20	19	11	2	2	36	12
9	1	0	26	0.86	12	13	13	11	0	2	22	12
10	0	10	30	1.46	6	3	19	11	0	2	34	12
11	1	0	26	0.86	6	7.8	22	2	1	2	22	0
12	1	10	16	0.86	6	0	13	11	1	0	24	6
13	0	50	20	1.25	0	0	19	2	0	0	38	6
14	1	10	35	0.86	0	10.5	22	2	1	2	24	0
15	0	0	18	1.42	12	1.8	39	11	2	0	45	0
16	1	10	40	1.46	6	20	25	20	1	2	34	0
17	0	50	30	1.25	12	15	19	20	2	2	36	12
18	1	50	35	0.86	0	3.5	22	20	2	0	22	6
19	0	10	20	1.25	0	0	19	20	1	0	38	0
20	0	50	30	1.25	12	9	25	11	1	2	34	12
21	0	0	29	1.42	6	8.7	65	20	1	0	51	12
22	0	90	35	0.74	12	17.5	13	11	1	2	20	0
23	0	50	35	0.86	12	0	16	20	2	0	24	12
24	0	10	20	1.25	0	6	31	2	0	2	34	12
25	1	10	18	1.42	0	0	52	2	0	0	48	6
26	1	0	20	1.46	12	10	31	11	2	2	36	12
27	1	90	40	1.42	12	0	39	2	0	0	51	6
28	1	90	20	1.46	6	2	31	20	2	0	38	6
29	1	0	16	0.86	12	4.8	16	2	0	2	20	0

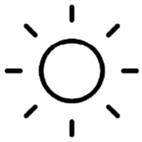
30	0	50	40	1.66	0	12	52	11	1	2	45	0

Choice situation	TIME_BSS	OT_BSS	AVAIL_BSS	COST_BSS	PST_BSS	TIME_ESS	TIME_ESSPT	OT_ESSPT	NT_ESSPT	COST_ESS	COST_ESSPT
1	43	11	12	0.65	0	2	65	11	0	0.5	2
2	46	11	3	2.6	6	4	41	20	2	1	0
3	49	20	0	2.75	6	4	41	20	0	2	0
4	38	20	0	2.2	6	1	35	11	2	1.25	2
5	49	11	0	2.75	6	4	39	2	1	1	0
6	20	2	30	0	0	2	13	20	2	1.5	0
7	20	11	30	0	12	12	10	11	0	3	0
8	34	2	30	2	0	1	35	20	2	1.25	2
9	22	20	12	0	0	2	22	2	1	1.5	2
10	30	11	0	0	12	8	23	11	1	3	2
11	18	11	3	0	12	12	10	2	0	3	2
12	18	2	30	0	12	12	12	11	0	4	0
13	34	20	3	2	6	1	30	2	0	0.25	0
14	22	11	0	0	0	2	16	11	1	0.5	2
15	43	11	12	2.45	6	2	61	2	2	1.5	0
16	30	2	30	1.8	12	1	40	20	1	1.25	2
17	38	2	3	0.4	0	1	40	20	1	1.25	2
18	18	2	12	0	0	12	12	11	0	3	0
19	38	20	3	2.2	6	8	27	2	2	3	0
20	38	2	12	0.4	12	8	19	11	1	2	2
21	49	20	3	0.95	12	2	65	20	2	1.5	0
22	22	20	0	0	12	12	8	11	1	4	2
23	18	20	30	1.8	0	12	12	20	2	4	0
24	38	20	0	2.2	6	8	19	2	0	3	2
25	43	11	0	2.45	6	4	44	2	0	1	0
26	34	2	12	0.2	12	1	30	20	1	0.25	2
27	49	20	3	0.95	6	2	61	2	0	0.5	0

28	34	2	30	0.2	0	1	30	20	2	0.25	0
29	22	11	12	0	0	2	22	2	1	0.5	2
30	46	2	3	0.8	12	2	63	11	2	0.5	2

Appendix 2

Considérez la situation suivante : vous devez vous rendre à l'ENTPE depuis votre domicile pour 9h. La météo pour la journée est la suivante :

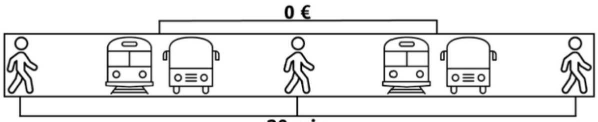
Météo à l'aller	Probabilité de pluie au retour
	 10%

Vous avez les options suivantes à disposition, quelle option choisissez-vous pour faire l'aller ?

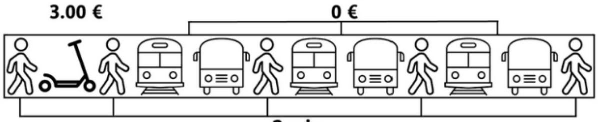
Vélo

Information sur le trajet aller :	Information sur le trajet retour :	Résumé :
38 min 	Temps de recherche d'une place de parking : **0 min**	Temps aller : **38 min** Temps additionnel retour : **+ 0 min**

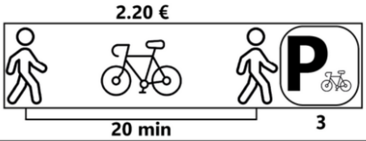
Transport en Commun

Information sur le trajet aller :	Résumé :
19 min **0 €**  **20 min**	Temps aller : **39 min** Coût : **0.00 €** Correspondance : **1**

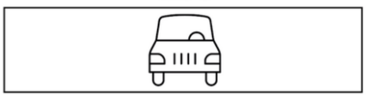
Trottinette électrique et Transport en Commun

Information sur le trajet aller :	Résumé :
8 min **3.00 €** **27 min** **0 €**  **2 min**	Temps aller : **37 min** Coût : **3.00 €** Correspondance : **2**

Vélo'V

Information sur le trajet aller :	Information sur le trajet retour :	Résumé :
38 min **2.20 €**  **20 min** **3**	Temps de recherche d'une place de parking : **6 min**	Temps aller : **58 min** Coût : **2.20 €** Place libre : **3** Temps additionnel retour : **+ 6 min**

Voiture

Information sur le trajet aller :	Information sur le trajet retour :	Résumé :
20 min **1.25 €** 	Temps de recherche d'une place de parking : **0 min** Temps perdu la veille dans les embouteillages : **0 min**	Temps aller : **20 min** Coût : **1.25 €** Temps additionnel retour : **+ 0 min**